UNITEXT 139

Alberto Rotondi · Paolo Pedroni
Antonio Pievatolo

Probability, Statistics and Simulation

With Application Programs Written in R

Springer

UNITEXT

La Matematica per il 3+2

Volume 139

The **UNITEXT - La Matematica per il 3+2** series is designed for undergraduate and graduate academic courses, and also includes advanced textbooks at a research level.

Originally released in Italian, the series now publishes textbooks in English addressed to students in mathematics worldwide.

Some of the most successful books in the series have evolved through several editions, adapting to the evolution of teaching curricula.

Submissions must include at least 3 sample chapters, a table of contents, and a preface outlining the aims and scope of the book, how the book fits in with the current literature, and which courses the book is suitable for.

For any further information, please contact the Editor at Springer: francesca.bonadei@springer.com

THE SERIES IS INDEXED IN SCOPUS

***

UNITEXT is glad to announce a new series of **free webinars and interviews** handled by the Board members, who will rotate in order to interview top experts in their field.

In the first session, going live on June 9, Alfio Quarteroni will interview Luigi Ambrosio. The speakers will dive into the subject of Optimal Transport, and will discuss the most challenging open problems and the future developments in the field.

Click here to subscribe to the event!

https://cassyni.com/events/TPQ2UgkCbJvvz5QbkcWXo3

Alberto Rotondi • Paolo Pedroni •
Antonio Pievatolo

Probability, Statistics and Simulation

With Application Programs Written in R

Springer

Alberto Rotondi
Dipartimento di Fisica
Università di Pavia
Pavia, Italy

Paolo Pedroni
Istituto Nazionale di Fisica Nucleare
Università di Pavia
Pavia, Italy

Antonio Pievatolo
Istituto di Matematica Applicata e
Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Milano, Italy

ISSN 2038-5714 ISSN 2532-3318 (electronic)
UNITEXT
ISSN 2038-5722 ISSN 2038-5757 (electronic)
La Matematica per il 3+2
ISBN 978-3-031-09428-6 ISBN 978-3-031-09429-3 (eBook)
https://doi.org/10.1007/978-3-031-09429-3

Cover illustration: “The face number three and two numbers three, one chance over 1326” (photo by the authors)

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book, based on the fourth Italian edition, comes from the collaboration between two experimental physicists and one statistician. Among non-statisticians, physicists are perhaps the ones who most appreciate and use probability and statistics, but most of the time in a pragmatic and manual way, having in mind the solution of specific problems or technical applications. On the other hand, in the crucial comparison between theory and experiment, it is sometimes necessary to use sophisticated methods which require knowledge of the fundamental logical and mathematical principles at the basis of the study of random phenomena. More generally, even those who are not statisticians have often to face, in any research field, problems that require particular attention and expertise for the treatment of random or aleatory aspects. These skills are naturally mastered by the statistician, whose research interests are the laws of chance.

This text has been prepared with the aim to seek a synthesis between these different approaches, to provide the reader not only with tools useful to address problems, but also with a guide to the correct methodologies needed to understand the complicated and fascinating world of random phenomena.

Such an objective obviously involved choices, sometimes even painful, both of content type and style. As for style, we tried not to give up the precision needed to properly teach the important concepts. When treating applications, we privileged the methods that do not require excessive preliminary conceptual elaborations.

As an example, we have tried to use, whenever possible, approximate methods for interval estimation, with Gaussian approximations for the estimator distributions. Similarly, in the case of least squares, we have extensively adopted the approximation based on the χ^2 distribution to verify the fitting of a model to the data.

We also avoided insisting on the formal treatment of complicated problems in cases where a solution using computer and simple simulation programs could be easily found.

In our book, simulation plays an important role in the presentation of many topics and in the verification of the accuracy of many techniques and approximations. This

feature, already present in the first Italian edition, is now common to many data science texts and, in our opinion, confirms the validity of our initial choice.

This book is aimed primarily at students of scientific undergraduate courses, such as engineering, computer science, and physics. However, we think that it can also be useful to all those scientific researchers who have to solve practical problems involving probabilistic, statistical, and simulation aspects. For this reason, we have given space to some topics, such as Monte Carlo methods and their applications, minimization techniques, and data analysis methods, which, usually, are only briefly mentioned in introductory texts.

The mathematical knowledge required by the reader is that which is normally given in the teaching of the basic calculus course in the scientific degrees, with the addition of minimum notions of linear algebra and advanced calculus, such as the elementary concepts of the derivation and integration of multidimensional functions.

The structure of the text allows different learning paths and reading levels. The first seven chapters deal with all the topics usually developed in a standard, basic statistical course. At the choice of the teacher, this program can be integrated with some more advanced topics from the other chapters. For example, Chap. 8 should certainly be included in a simulation-oriented course.

The notions of probability and statistics usually taught to physics students in undergraduate laboratory courses are enclosed in the first three chapters, in Chaps. 6 and 7 (basic statistics) and in Chap. 12, written explicitly for physicists and for all those who need to process data from laboratory experiments.

Many pages are devoted to the complete resolution of several exercises inserted directly inside the chapters to better explain the covered topics. We also recommend to the reader the problems (all with solutions) reported at the end of each chapter.

This book makes use of the statistical software R, which has now become the world standard for solving statistical problems. The 2019 ranking of the Institute of American Electrical and Electronic Engineers (IEEE) places R in fourth position among the most popular programming languages, after Python, Java, and C. Many R routines have been written by us, to guide the reader while going through the text. These routines can be easily downloaded from the link specified below. We therefore recommend an interactive reading, in which the study of a topic is followed by the use of R routines in the way showed both in the text and in the technical instructions included in the indicated Web pages.

We thank again the readers who reported errors or inaccuracies present in the previous Italian editions, and the publisher, Springer, for the continued trust in our work.

Pavia, Italy — Alberto Rotondi
Pavia, Italy — Paolo Pedroni
Milano, Italy — Antonio Pievatolo
March 2022

How to Use the Text

Figures, equations, definitions, theorems, Tables, and exercises are numbered progressively.

The abbreviations of quotations (e.g., [57]) refer to the bibliographic list at the end of the book.

Solutions of the problems are given in Appendix D. The table of symbols reported in Appendix A may also be useful.

Calculation codes as `hist` are marked with a different text style. Routines starting with a lowercase letter are (with some exceptions) the original R codes, which can be freely copied from the CRAN (Comprehensive R Archive Network) website, while those starting with an uppercase letter are written by the authors and are in:

`https://tinyurl.com/ProbStatSimul`

In this site, you will also find all the information for the installation and the use of R, a guide to the use of routines written by the authors and complementary teaching materials.

Contents

About the Authors

Alberto Rotondi is formerly Full Professor of Nuclear Physics at the University of Pavia (Italy), where he is now Adjunct Professor of Data Analysis in the Physics Department. He is the author of several hundred publications in the field of experimental nuclear physics. During his research activity, carried out mainly at the CERN laboratories of Geneva, he applied many statistical and simulation methods to the project of new experiments and to the analysis of experimental results.

Paolo Pedroni is a senior researcher at INFN (Italian National Institute of Nuclear Physics), Section of Pavia, and an adjunct professor in the Physics Department at the University of Pavia (Italy), where he teaches the course Statistical Methods in Physics. He has been carrying out experimental research on the interactions of quarks within protons and neutrons in the framework of several international collaborations.

Antonio Pievatolo is director of research at CNR IMATI in Milan, Italy, where he works in the applications of statistics in industry and technology. He has led his local unit in applied research projects commissioned by national and regional public bodies and companies. He has been president of the European Network for Business and Industrial Statistics (ENBIS) from 2017 to 2019, and Contract Professor of Statistics at the University.

Chapter 1
Probability

There seems to be no alternative to accepting some sort of incomprehensible quality to existence. Take your pick. We all fluctuate delicately between subjective view and objective view of the world, and this quandary is central to human nature.

Douglas R. Hofstadter, "THE MIND'S I".

1.1 Chance, Chaos and Determinism

In this introduction, before looking into the phenomena known as casual, stochastic or random, we will briefly analyse the importance and the role of these physical processes into our reality.

At the beginning of a scientific measurement or observation of a natural phenomenon, one usually tries to identify all the causes, conditions and external factors that determine its evolution. Subsequently, one operates in order to keep these external causes fixed or, as much as possible, under control, and then one proceeds to record the results of the observations.

When repeating the observations, two situations can occur:

- One always gets the same result. As an example, think of the measurement of a table with a commercial meter tape.
- One gets a different result each time. Think of a very simple natural phenomenon: the toss of a coin.

While, for the moment, in the first case there is not much to say, in the second case, we could ask ourselves what causes the observed variations of the results. Possible reasons are not having checked all the conditions that influence the phenomenon or having incorrectly defined the quantity to be observed. Once these corrections have been applied, the phenomenon can become stable or continue to show fluctuations.

Let's explain with an example: suppose we want to measure the time of sunrise on the horizon at a given location. We will observe that repeated measurements in successive days give different results. Obviously, in this case the variation of the results is due to a bad definition of the measure. The time of sunrise must be

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3_1

measured in a certain place and for a certain day of the year and must be repeated one year later in the same day and place. Redefining the observation in this way, the results of repeated measurements coincide. As it is well known, the laws of planet motion provide in this case a model that allows to predict (apart from small corrections that we don't want to discuss) the times of dawn for every day of the year.[1]

Let us now consider the measurement of another quantity, the temperature at a certain time of a day. In this case, even considering a certain day of the year and repeating the measurements from year to year, different results are observed. Unlike the time of sunrise, we are not in possession of a model that allows us to accurately predict the result. Why does the temperature, unlike dawn, seem to have an inherently random behaviour? The reason is that, while the time of dawn depends on a few body interactions (the sun and the planets), the temperature depends not only on astronomical conditions but also on the state of the atmosphere, which is the result of the interaction of countless factors, which *not even in principle* can be determined with absolute precision or, in any case, kept under control.

This distinction is crucial and is the key to establish the difference between quantities that fluctuate and those that appear to be fixed or are accurately predictable based on deterministic models.

Historically, deterministic systems have been considered, for a long time, free of fluctuations, and their study, in the context of classical physics, is continued in parallel to that of the systems called stochastic, casual or random, born with the study of gambling: toss of dices, card games, roulette, slot machines, lotto games and so on. The latter systems are specifically designed and built to ensure the randomness of the results. There were therefore two separate physics domains: the one of deterministic phenomena, without fluctuations, governed by the fundamental laws of classical physics, usually consisting of simple systems (generally few bodies systems), and the world of the random phenomena, subject to fluctuations, often related to complex systems (usually consisting of many bodies).

However, already at the beginning of the last century, the French mathematician-physicist H. Poincaré noticed that, in some cases, the knowledge of the deterministic laws was not enough to make exact predictions on the dynamics of some systems starting from known initial conditions. The problem, which today is called the study of chaos, was thoroughly investigated only much later, starting from the 1970s, thanks to the help of computers. Today, we know that, in macroscopic systems, the origin of the fluctuations can be twofold, that is, due to deterministic laws which present high sensitivity regarding the initial conditions (chaotic systems) and due to the impossibility of defining in a deterministic way all the variables of the system (stochastic systems). One of the best paradigms for explaining chaos is the logistic map, proposed since 1838 by the Belgian mathematician P.F. Verhulst and studied

[1] Actually we do not know exactly how stable the solar system is. Some models indicate that forecasts cannot be extended beyond a time interval of the order of one hundred million years [AAN+07].

in detail by the biologist R. May in 1976 and by the physicist M. Feigenbaum in 1978:

$$x(k+1) = \lambda\, x(k)\,[1 - x(k)]\,, \tag{1.1}$$

where k is the population growth cycle, λ is related to the growth rate and $0 \leq x(k) \leq 1$ is a state variable proportional to the number of individuals in the population. The condition $0 \leq \lambda \leq 4$ assures that x remains within the fixed limits. The logistic law well describes the dynamics of evolution of populations where there is an increase per cycle proportional to $\lambda\, x(k)$ with a negative feedback $-\lambda\, x^2(k)$ proportional to the square of the size already reached by the population.

Without going too far into the study of the logistic map, we notice that the behaviour of the population evolves with the number of cycles according to the following characteristics (also shown in Fig. 1.1):

- When $\lambda \leq 1$ the model always leads to the extinction of population.
- When $1 < \lambda \leq 3$ the system reaches a stable level, which depends on λ *but is independent of the initial condition* $x(0)$.
- When $3 < \lambda \leq 3.56994\ldots$ the system oscillates between some fixed values. For example, as shown in Fig. 1.1 for $\lambda = 3.5$, there are four possible values (in the

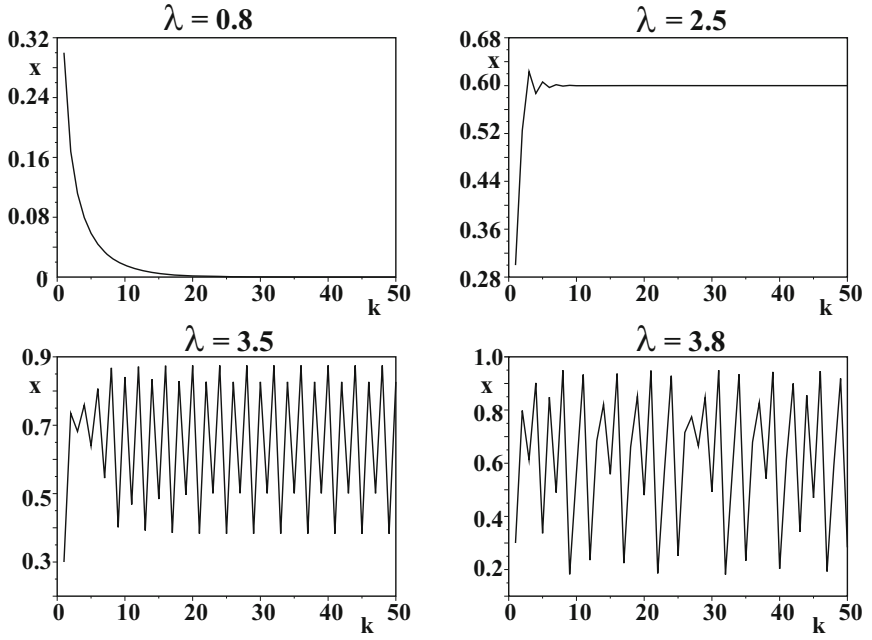


Fig. 1.1 x values from logistic equation (1.1) having as initial starting value $x = 0.3$ for different λ values

figure a continuous line joins the discrete x values). Also in this case the states reached by the system do not depend on the initial condition.

- When $\lambda > 3.56994\ldots$ the system is chaotic: the fluctuations seem regular, but, as can be seen by looking carefully at Fig. 1.1 for $\lambda = 3.8$, they are neither periodic nor do they seem entirely random. A thorough study also shows that the fluctuations are not even predictable precisely, because the initial condition values $x(0)$ very close to each other *lead to completely different evolutions*. This phenomenon, which is called *sensitive dependence on initial conditions or butterfly effect*,[2] is one of the main characteristics of chaos. Note that the fluctuations in chaotic systems are *objective, intrinsic or essential*, since the reproducibility of the results would require initial conditions at an accuracy level comparable to that of the atomic scale, which is not possible, *not even in principle*.

You can gain numerical experience with the logistic map and check the butterfly effect using our R `Logist` and `LogiPlot` routines[3] with which we produced (Fig. 1.1).

The methods for distinguishing the chaotic systems from the stochastic ones are based essentially on the study of dispersions, that is, the difference between the values of the same state variable in subsequent evolutions of the system, as a function of the whole number of the state variables.

In a chaotic system, once a certain number of variables have been identified, the deviations stabilize or tend to decrease. This behaviour indicates that a number of state variables adequate to describe the system have been reached and that the deterministic law that regulates its dynamics can be obtained. The fluctuations in the results of repeated experiments in this case are attributed, as we have seen, to the sensitivity of the system with respect to the initial conditions.

In a stochastic system, conversely, the number of state variables needed for the complete description of the system is never reached, and the sum of the deviations, or the quantities connected to them, continues to grow with the number of state variables considered [AAN+07]. The fluctuations of the system variables appear *random* and follow the distributions of probability theory.

The study of chaos and of the transitions from chaotic to stochastic states (and vice versa) is a very recent and still open research area, where many problems still remain unsolved. The interested reader can enter into this fascinating topic through the introductory readings [AAN+07, Rue96, Ste97].

In the remainder of the book, we will not deal with chaos, but we will instead devote ourselves to the study of random or stochastic systems, that is, of all the systems in which, as we have previously noted, there are variables following, in principle, the statement:[4]

[2] Referring to chaos in meteorological systems, it is often said: "a flap of butterfly wings in the tropics can trigger a hurricane in Florida".

[3] Most of the original R routines start with a lowercase letter, ours with a capital letter.

[4] In the following the non-mathematical operational definitions will be called "statements".

Statement 1.1 (Random Variable in a Broad Sense) *A stochastic, random or aleatory variable is the result of the interaction of many factors, each of which is not dominant over the others. These factors (and their dynamic laws) cannot be completely identified, fixed and in any case kept under control, not even in principle.*

In the present book, we will mainly use the term "random variable". Let us now try to identify some stochastic systems or processes which in nature produce random variables. All many-body systems have a very high degree of randomness: the dynamic observables of molecular systems, ideal gases and thermodynamic systems generally follow Statement 1.1 very well. These are systems studied by statistical physics.

At this point we can specify the meaning of the term *"factors and dynamic laws impossible to determine, not even in principle"* we used in Statement 1.1. Suppose we roll a dice 100 times. To build a deterministic model that can predict the outcome of the experiment, it would be necessary to introduce in the dice equations of motion all the initial conditions of the toss, the constraints given by the surfaces of the hands or the cup in which the dice is shaken before throwing, the constraints given by the table where the dice falls down and perhaps more. We would thus have a huge set of numbers, describing the initial conditions and constraints for each one of the hundred tosses, enormously larger than the one hundred numbers giving the final result of the experiment. Clearly, the predictive power of such a theory and its practical applicability are totally absent. A deterministic model, to be such, must be based on a compact set of equations and initial conditions and must be able to predict a vast set of phenomena.

For example, this is the case of the logistic law (1.1) or of the simple law of the fall of the bodies, which connects the path space s to the gravitational acceleration g and to the fall time t through the formula $s = gt^2/2$. This formula alone allows you to predict, assigning s or t as the initial conditions, the results of any experiment.

We can summarize the above considerations by saying that a deterministic model becomes meaningless when it generates algorithms requiring a numerical set of initial conditions, constraints and equations enormously larger than the set of results that one intends to predict. Alternatively, one should use the statistical approach which, based on the a posteriori study of the results obtained, try to quantify the extent of the fluctuations and extract global regularities that can be useful for the prediction of future results.

This line of thinking, developed during the last three centuries, arrived, by studying the pure stochastic systems, at identifying the fundamental mathematical laws for the description of random phenomena. The set of these laws is now known as the probability theory.

All the books dealing with probability theory, including the present one, make extensive use of examples taken from the games of chance, such as dice throwing. These examples well delineate the essence of the problem, because only by studying pure stochastic systems it is possible to discover the laws of chance. Great mathematicians and statisticians, like P. Fermat (1601–1665), P.S. Laplace (1749–

1827) and J. Bernoulli (1654–1705), often discuss experiments they performed with dices, cards or other devices taken from games. One of their goals was precisely to provide winning strategies for gambling games, which were already widespread at that time and that they played too. In this way they set the foundation of the probability calculus and statistics, based exclusively on experimental facts, as the scientific method requires.

In addition to traditional games, today there is another "artificial" laboratory, consisting of computer-generated random processes. As we will see, it is indeed possible to simulate pure stochastic systems of any kind and complexity using a uniform random number generator (a kind of electronic roulette): rolls of the dice, card games, many-body physical systems, and more.

These techniques, named Monte Carlo (recalling the homeland of the games of chance) or simulation methods, are very practical and effective, because they allow to obtain artificial datasets in a few seconds, whereas a real experiment in some cases would take years. However, it is important to note that, conceptually, these methods do not introduce new elements. The aim is always to obtain random variables from *models* consisting of stochastic systems also including, when necessary, deterministic components. These data are then used to develop and optimize the logical-mathematical tools to be applied to the study of real systems.

And now let's start to examine real systems in general. For example, consider Fig. 1.2, which represents the average temperature of the earth's surface over the past 142 years. As you can well imagine, our future depends on the trend of this curve in the next years. Comparing "by eye" this curve with that of Fig. 1.3, representing a pure stochastic process, it seems that, starting from the beginning of the last century, an increasing trend is superimposed to a random behaviour. We do

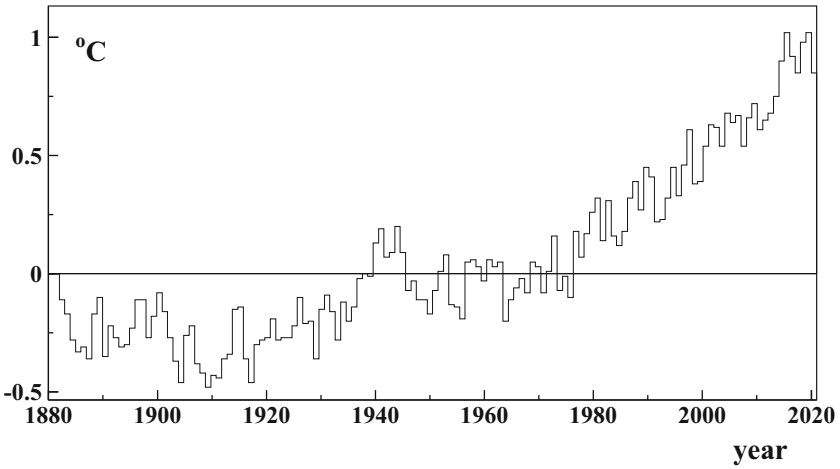


Fig. 1.2 Average global terrestrial surface temperature for the period 1880–2021. The line at zero represents the average of the years 1951–1980 [Tea22, LSH+90]

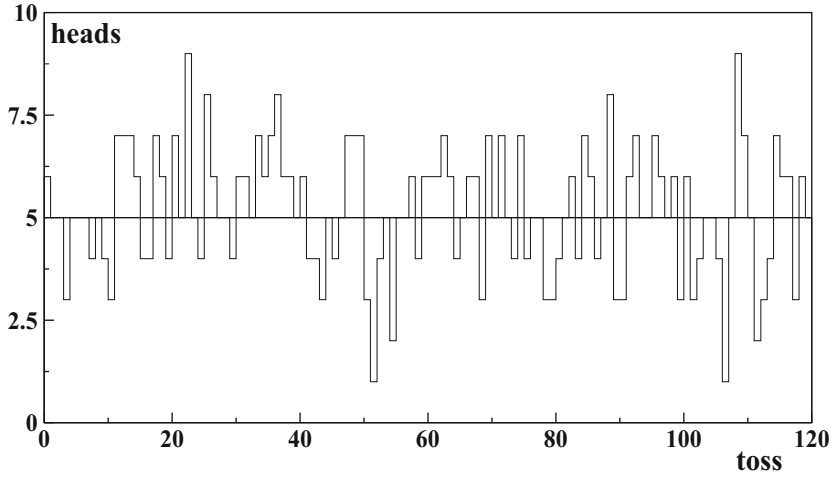


Fig. 1.3 Computer simulation of the number of heads obtained by throwing 10 coins in 120 tosses. The progressive number of tosses is reported on the abscissas, the number of heads in ordinates. The continuous line is the expected mean value (five heads). Compare this figure with Fig. 1.1 for $\lambda = 3.8$, which displays chaotic fluctuations

not go further into this rather alarming example that just served us to show that, in real cases, the simultaneous presence of both stochastic and deterministic effects is very common.

To account for these possible complications, the study of a real system is performed with a gradual approach, according to the following steps:

(a) To identify the purely stochastic processes of the system and deduce, based on the rules of probability and statistics, their evolution laws.
(b) To separate stochastic from non-stochastic components (sometimes called systematic), if any. This step is generally performed using statistical methods.
(c) If the problem is particularly difficult, to perform a computer simulation of the system on the computer and compare the simulated data with the real ones.

It is often necessary to repeat steps (a) to (c) until the simulated data are in a satisfactory agreement with the real ones. This recursive technique is a powerful method of analysis and is now applied in many fields of scientific research, from physics to economics.

Before closing this introduction, we would like to mention what happens in the microscopic world. Let us consider, for simplicity, a system consisting of a single subatomic particle as an electron. In this case the fundamental equations of physics provide a complex state function $\psi(r)$ whose square modulus gives the localization probability of a particle in space: $P(r) = |\psi(r)|^2$. The probability thus defined *obeys the general laws of probability which will be described in the following*.

Since the fundamental laws of the microscopic world contain a probability function and so far no one has been able to find more basic fundamental laws based on different quantities, one deduces that probability is a fundamental quantity of nature. Indeterminism, being present in the fundamental laws that govern the dynamics of the microscopic world, assumes in this case an objective character (called non-epistemic), not linked to ignorance or limited abilities of the observer.

1.2 Some Basic Terms

Here we informally introduce some fundamental definitions of current use in the study of stochastic phenomena. In the following, these terms will gradually be redefined in a mathematically rigorous way.

- *Sample space:* it is the set of all possible different values (cases) that a random variable can assume. For example, the random variable *card of a playing deck* gives rise to a sample space of 52 elements. The structure of the space depends on the way used to define the random variable. In fact the space relative to the random variable *card of a playing deck* is consisting of 52 cards, or 52 integer numbers if we create a correspondence between cards and numbers.
- *Event:* it is a particular combination or a particular subset of cases. For example, in the case of playing cards, if you define an event as an odd card, the set of cases obtained is 1, 3, 5, 7, and 9, for each of the four colours. This event gives rise to a subset of 20 elements selected among the 52 elements of the sample space (all the cards in the deck).
- *Spectrum:* it is the set of all the different elements of the subset of cases defining the single event. For odd playing cards, the spectrum is given by 1, 3, 5, 7, and 9. Obviously, the spectrum can coincide with the entire space of the random variable under study (if, e.g. the event is defined as any card of a deck).
- *Probability:* is the quantitative evaluation of the possibility of obtaining a certain event. It is evaluated based on experience, using mathematical models or even on a purely subjective basis. For example, the probability that, at this point, you continue reading our book is, in our opinion, 95% . . .
- *Trial:* it is the set of operations that realize the event.
- *Experiment, measurement, sampling:* it is a collection of trials. The term familiar to statisticians is sampling, whereas the physicists usually use the term experiment or measurement. In physics an experiment can be a sampling, but not necessarily.
- *Sample:* it is the result of an experiment or sampling.
- *Population:* it is the result of that number of trials, finite or infinite, which run through all the possible events. For example, in the lottery game, the population can be the finite set of all possible combinations of 5 numbers drawn from an urn of 90 numbers; in the case of the height of the Italians, we can imagine the set of measurements of the heights of each individual. When the population is

thought as a sample of an infinite number of elements, it should be considered as a mathematical abstraction not achievable in practice.

These ideas can be summarized as in Fig. 1.4. Once the elementary probabilities have been assigned to the elements of the sample space (inductive step), using probability theory one can calculate the probability of all events, thus deducing mathematical models for the population (deductive step). Instead, by running a series of measurements, one can get a sample of events (experimental spectrum) representative of the population under consideration. Then, using the statistical data analysis (inductive/deductive step), one tries to identify, from a detailed examination of the sample, the properties of the parent population. These techniques are called *statistical inference*. Once a model has been assumed, it is possible to verify its congruence with the collected data samples. This method is called *hypothesis testing*.

In this text, the fundamentals of probability calculus will be at first explained with particular regard to the assignment of elementary probabilities to the components of the sample space. Then, calculus and combinatorial analysis will be used to obtain some fundamental mathematical models of populations. Afterwards, the methods of statistical analysis will be explained. They allow to estimate, starting from measured

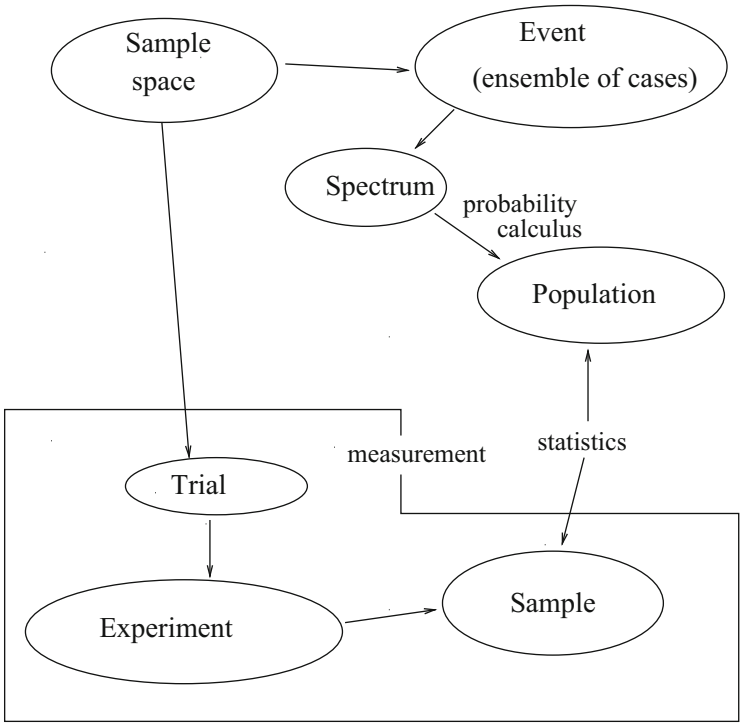


Fig. 1.4 The relationships between probability calculus, statistics and measurement

quantities, the "true" values of physical parameters or to verify the congruity of experimental samples with mathematical models of population. The elements of probability and statistics previously acquired will then be extensively applied to simulation techniques.

1.3 The Concept of Probability

Experience shows that, when a stochastic or random phenomenon is stable over time, some values of the spectrum occur more frequently than others. If we flip ten coins and count the number of heads, we see that the outcome of five heads occurs more frequently than eight, while ten heads is a really rare, almost impossible, event. If we consider an experiment consisting of 100 trials (where each trial is the toss of 10 coins), we observe that the number of times one gets 5, 8 and 10 heads is quite regular, even if with little variations from experiment to experiment, because the values 5, 8 and 10 always (or almost always) show up with decreasing frequency. If we imagine all the possible alignments of 10 coins, we can have an intuitive explanation of this fact: the event 10 heads (or 10 crosses) corresponds to only one alignment, while for the event 5 crosses (or 5 heads) many possible alignments are possible. (5 tails and then 5 heads, tails-to-heads alternately, and so on up to 252 different alignments). When tossing ten coins, we then choose at random, on the same footing, one of the possible alignments, and it is intuitive that almost always we will get balanced results (more or less five heads) and almost never the extreme cases. A reasoning of this type, common to everyone's daily experience, leads instinctively to think that this regularity of the stochastic phenomena is due to the existence of fixed quantities, called *probabilities*, that one can define, for example, as the ratio between favourable and possible cases (alignments). These considerations led J. Bernoulli to the formulation of the first mathematical law able to predict the trend of the results in experiments such as the coin toss, taking also into account the random fluctuations.

In the case of coins, the probability is introduced to account for the variability of experimental results; however, each of us uses probability also to manage the uncertainty of many non-repeatable situations that occur in real life, quantifying subjectively the realistic possibilities and choosing those with the highest probability, taking into account the resulting costs or benefits.

For example, when we are driving the car and we meet a red traffic light, we have two options: stop or continue. If, around noon, we are crossing in a high traffic road, we surely stop, because we know, based on our experience, that the collision probability with other vehicles is very high. Instead, if we are in a low traffic road in the middle of the night, we are tempted to continue, because we know that the probability of a collision is very low.

Another example of a subjective and discrete probability is given by the judgement of a defendant in a trial by a jury. In in this case, the probability can be expressed with two values, 0 or 1, i.e. guilty or innocent. In general, current jurispru-

dence formulates the final judgement combining subjective individual probabilities expressed by the individual jurors.

Given these observations, the approach currently considered more appropriate, effective and ultimately cheaper for the study of random phenomena is to consider the choice of probability as a subjective act, based on experience. A first possible effective definition of probability is:

Statement 1.2 (Subjective or Bayesian Probability) *The probability is the subjective degree of belief about the occurrence of an event.*

The subjective probability is free, but it is generally assumed that it must be *consistent*, that is, expressed as a real number $0 \leq p \leq 1$, $p = 1$ for a known event and $p = 0$ for an impossible event. Then, considering two or more *exclusive* events (like the faces 2 or 4 on a die roll), consistency requires their probabilities to be additive. These assumptions are sufficient for the axiomatization according to the Kolmogorov scheme, which will be presented shortly.

The subjective probability is widely used in soft sciences such as jurisprudence, economics, part of medicine, etc. In hard sciences as physics (we will specify better later, in Chap. 12, the meaning of the term "hard science"), the subjective probability is generally avoided and the definitions of a priori and frequentist probabilities are used (Laplace, 1749–1827) (Von Mises, 1883–1953).

Definition 1.3 (Classical or a Priori Probability) If N is the total number of cases of the sample space of a random variable and n is the number cases with outcome A, the classical or a priori probability of A is given by:

$$P(A) = \frac{n}{N} . \tag{1.2}$$

For example, the a priori probability of a given face when throwing a fair die is:

$$P(A) = \frac{n}{N} = \frac{\text{number of favorable cases}}{\text{number of possible cases}} = \frac{1}{6} ,$$

while the probability of drawing the ace of diamonds from a deck of cards is 1/52, the probability of extracting a suit of diamonds is 1/4 and so on.

Definition 1.4 (Frequentist Probability) If m is the number of occurrences of outcome A over a total of M trials, the probability of A is given by:

$$P(A) = \lim_{M \to \infty} \frac{m}{M} . \tag{1.3}$$

The limit appearing in this definition has an experimental meaning rather than a mathematical one, because the true probability should be found only by carrying out an infinite number of trials. In the following, we will call this operation, with the limit written in italics, as *frequentist limit*.

The choice of the elementary probabilities to be assigned to the different events is therefore inductive and arbitrary. The probability calculus applied to complex events starts from arbitrarily assigned elementary probabilities and then proceeds deductively, as we shall see, without departing from mathematical rigor. The use of subjective probabilities is also called *Bayesian approach*, because in this case the initial probabilities are often readjusted according to the results obtained using the famous Bayes' formula, which we will soon deduce in Sect. 1.7.

The frequentist approach is the one prevalent in physical and technical frameworks. Based on our experience, we believe that in experimental physics the frequentist approach is followed in 99% of cases, and this is a ... subjective evaluation! Within this framework, it is believed that Eq. (1.3) allows the "objective" evaluation of probability for those natural phenomena that can be easily sampled or easily repeated in the laboratory. In many cases, experience shows that the frequentist probability tends to coincide with the a priori one:

$$\lim_{M\to\infty} \frac{m}{M} \simeq \frac{n}{N} \qquad \text{(from the experience!)} \,. \tag{1.4}$$

When this condition holds, one says that the cases are equiprobable and mutually exclusive. Consider, for example, the roll of a dice: if you are sure that it is not rigged, it is intuitive to assume that the probability of getting a certain face (let's say, the number 3) in a throw is equal to 1/6. Experience shows that, after several throws, the frequentist probability (also called frequency limit, Eq. (1.3)) tends actually to 1/6, according to (1.4). If the die is not balanced, the probability of obtaining face number 3 can only be evaluated by running many trials. Since the limit for an infinite number of trials is not practically reachable, one usually stops to a high but finite number n of trials and the true probability is estimated by the confidence interval method (see Chap. 6).

The frequentist definition (1.3) would therefore seem the most general and reliable; however, this is not true:

- Since an experiment cannot be repeated an infinite number of times, the probability (1.3) will never be determined.
- The experiment must be repeatable, and the limit appearing in (1.3) does not have a precise mathematical sense. This leads to insurmountable mathematical inconsistencies in proving the validity of the empirical case law (1.4).

The statistician B. de Finetti, in one of his famous articles [DF33], comments on this last point as follows: *"... for a large category of the problems for probability theory (but not for all, as it is shown by the absurdities found and by the ones which could easily be found), by imagining an infinite sequence of similar experiences, one can build up an example of a possible course of results in a way as to obtain a limit frequency equal to probability, for each sequence of similar events."*

The decision on the best approach to use (subjective-Bayesian, a priori-classical or frequentist), based on the type of problem to be addressed (uncertainty in a broad sense or variability of the results of repeatable experiments), is still an open question and is a continuous source of disputes.

To definitively get out of this confused situation, the modern probability theory resorts to axiomatization. In the next paragraph, we will see in fact that, after defining the probability in an abstract mathematical way, it is possible to outline a consistent mathematical theory for the study of random phenomena. The inductive and arbitrary approach is limited to the initial decision about what probability to adopt: once the choice of a probability that obeys the required axioms is made, this theory can be applied correctly. Then, if the obtained results are in disagreement with the experimental outcomes, it will be necessary to change the type of probability to be used for that problem. For example, it is perfectly possible to invent a probability that, in a lottery, assigns a higher probability to the delayed numbers. If this probability obeys the axioms, the approach is mathematically correct. However, in this case you will always get results totally different from those observed. Therefore, in fair games, as well as in statistical physics, the assumed probabilities are classic and frequentist, which leads to results in accordance with experience.

This book, which is dedicated to students and researchers in technical-scientific fields, is based on the frequentist approach. However, we will mention in some cases even the Bayesian point of view, referring the reader to more specific texts, such as [Gre06].

1.4 Axiomatic Probability

To formalize in a mathematically correct way the concept of probability, it is necessary to apply the set theory to the fundamental notions introduced so far. If S is the sample space of a random variable, we consider the family $\mathcal{F}$ of subsets of S according to the

Definition 1.5 (σ-algebra) Any collection $\mathcal{F}$ of subsets of S having the properties:

(a) the empty subset belongs to $\mathcal{F}$: $\emptyset \in \mathcal{F}$;
(b) if a countable collection of subsets $A_1, A_2, \ldots \in \mathcal{F}$, then

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F} ;$$

(c) if $A \in \mathcal{F}$, the same holds for the complement: $\overline{A} \in \mathcal{F}$;

is named σ-algebra.

Using the well-known properties:

$$\overline{\overline{A} \cup \overline{B}} = A \cap B ,$$
$$A \cap \overline{B} = A - B ,$$

it is easy to show that also the intersection of a countable collection of sets belonging to $\mathcal{F}$ and the difference $A - B$ of two subsets of $\mathcal{F}$ are included in $\mathcal{F}$:

$$\bigcap_{i=n}^{\infty} A_i \in \mathcal{F}\,, \tag{1.5}$$

$$A - B \in \mathcal{F}\,. \tag{1.6}$$

The correspondence between probability and set theories is summarized in Table 1.1. If, to fix ideas, we consider a deck of cards and we define the draw of an ace as event A and the extraction of a diamonds suit (Fig. 1.5) as event B, we get the following correspondence between sets (events) and elements of S:

- S: all the 52 playing cards;
- a: 1 of the 52 playing cards;
- $A \cup B$: diamonds suit or heart, clubs, aces of spades;
- $A \cap B$: diamonds suit;
- $A - B$: hearts, clubs or aces of spades;
- $\overline{A}$: any card except aces;
- $\overline{B}$: a non-diamonds suit

Table 1.1 Correspondence between probability and set theories

Symbol	Set theory	Probability theory
S	Total set	Sample space
a	An element of S	Result of a trial
A	Subset of S	If $a \in A$ the event A occurs
$\emptyset$	Empty set	No events occur
$\overline{A}$	Collection of elements of S not belonging to A	The event A does not occur
$A \cup B$	Elements belonging either to A or to B	The events A or B occur
$A \cap B$	Elements that belong to both A and B	The events A and B occur
$A - B$	Elements of A not Belonging to B	The event A occurs, but The event B does not
$A \subseteq B$	The elements of A belong also to B	If A occurs B also occurs

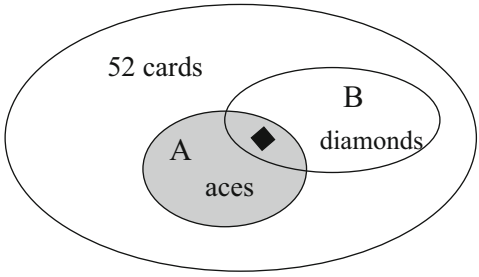


Fig. 1.5 The random variable "playing card" and the events "extraction of an ace" and "extraction of a diamonds suit" according to set theory

Let us now consider a function $P(A)$, for A belonging to a σ-algebra $\mathcal{F}$, that brings the set A to a real number in the range [0.1]. In symbols,

$$P : \mathcal{F} \rightarrow [0, 1] \, . \tag{1.7}$$

According to the Kolmogorov approach, the probability follows the

Definition 1.6 (Kolmogorov Axiomatic Probability) A function $P(A)$ satisfying (1.7) and the properties:

$$P(A) \geq 0 \, ; \tag{1.8}$$

$$P(S) = 1 \, ; \tag{1.9}$$

and,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \;\; \text{if} \;\; A_i \cap A_j = \emptyset \;\; \forall i \neq j \, , \tag{1.10}$$

for any countable collection $A_1, A_2, \ldots$ of mutually disjoint subsets included in $\mathcal{F}$, is called probability.

Definition 1.7 (Probability Space) The probability triplet:

$$\mathcal{E} \equiv (S, \mathcal{F}, P) \, , \tag{1.11}$$

composed by the sample space, a σ-algebra $\mathcal{F}$ and P is named probability space.

The Kolmogorov probability satisfies the following important properties:

$$P(\overline{A}) = 1 - P(A) \, , \tag{1.12}$$

$$P(\emptyset) = 0 \, , \tag{1.13}$$

$$P(A) \leq P(B) \quad \text{if} \quad A \subseteq B \, . \tag{1.14}$$

Equation (1.12) is valid since the complement $\overline{A}$ is such that by definition $\overline{A} \cup A = S$; therefore, $P(\overline{A}) + P(A) = P(S) = 1$ from (1.9, 1.10), since A and $\overline{A}$ are disjoint. Moreover:

$$P(S \cup \emptyset) = P(S) = 1 \qquad \text{from (1.9)} \, , \tag{1.15}$$

$$P(S \cup \emptyset) = P(S) + P(\emptyset) = 1 \qquad \text{from (1.10)} \, , \tag{1.16}$$

from which Eq. (1.13) follows: $P(\emptyset) = 1 - P(S) = 1 - 1 = 0$. Finally, when $A \subseteq B$ one can write $B = (B - A) \cup A$, where $B - A$ is the set of the elements of

B not in A. Then:

$$P(B) = P[(B - A) \cup A] = P(B - A) + P(A)$$

and, since $P(B - A) \geq 0$, the property (1.14) is also proved.

Another important proposition is:

Theorem 1.1 (of Addition) *The probability of the event given by the occurrence of the events A or B, when* $A \cap B \neq \emptyset$*, is given by:*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \, . \tag{1.17}$$

Proof It easy to show that (you can draw the sets):

$$A \cup B = A \cup [B - (A \cap B)] \, ,$$

$$B = [B - (A \cap B)] \cup (A \cap B) \, ;$$

since $A \cup B$ and B are disjoint sets, it is possible to apply Eq. (1.10) to obtain:

$$P(A \cup B) = P(A) + P[B - (A \cap B)] \, ,$$

$$P(B) = P[B - (A \cap B)] + P(A \cap B) \, .$$

Then, one gets, by subtraction:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \, .$$

□

Both classical and frequentist probabilities follow the axioms (1.8–1.10). In fact, for the classical probability, we have:

$$\begin{aligned} P(A) &= (n_A/N) \geq 0 \qquad \text{always, because } n, \ N \geq 0 \, , \\ P(S) &= N/N = 1 \, , \\ P(A \cup B) &= \frac{n_A + n_B}{N} = \frac{n_A}{N} + \frac{n_B}{N} = P(A) + P(B) \, . \end{aligned}$$

Similarly, the validity of the axioms can also be proved for the frequentist probability, since its limit can be considered as a linear operator.

The classical and frequentist probabilities previously defined satisfy therefore to the properties (1.8–1.17). For example, the classical probability to draw an ace or a red card from a deck of cards, based on (1.17), is given by:

A = ace of hearts, ace of diamonds, ace of clubs, ace of spades,
B = 13 diamonds cards, 13 hearts cards,

$P(A \cap B) =$ ace of hearts, ace of diamonds ,
$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 4/52 + 1/2 - 2/52 = 7/13$.

The probability associated with the set $A \cap B$ covers, as we will see, a particularly important role in the algebra of the probability. It is called compound probability:

Definition 1.8 (Compound Probability) The compound probability

$$P(A \cap B) \quad \text{or} \quad P(AB)$$

is the probability that events A and B both occur.

Now we introduce a new kind of probability. Suppose we are interested in the probability that, after extracting a suit of diamonds, the card is an ace or that, when an ace is drawn, the suit is diamonds. We denote by A the set of aces, with B that of the diamond cards and with $P(A|B)$ the probability of A occurring after B, that is, once a suit of diamonds is drawn, the card is an ace. Obviously, we have:

$$P(A|B) = \frac{\#(\text{outcomes of the diamonds ace })}{\#(\text{outcomes of the diamonds suit })}$$

$$= \frac{1}{13} = \frac{1}{52} \Big/ \frac{13}{52} = \frac{P(A \cap B)}{P(B)} \, . \tag{1.18}$$

Similarly, the probability of getting a suit of diamonds if an ace is drawn is given by:

$$P(B|A) = \frac{\#(\text{outcomes of the diamonds ace})}{\#(\text{outcomes of an ace})} = \frac{1}{4} = \frac{1}{52} \Big/ \frac{4}{52} = \frac{P(B \cap A)}{P(A)} \, .$$

In the example just seen, the conditional probability $P(A|B)$ to get an ace once a suit of diamonds is drawn is equal to the unconditional probability $P(A)$ of hitting an ace; indeed:

$$P(A|B) = \frac{1}{13} = P(A) = \frac{4}{52} \, .$$

In this case, we say that *the events A and B are independent*. However, if A is the set [ace of diamonds, aces of spades] and B is, as before, the set of diamonds cards, we have:

$$P(A|B) = \frac{1}{13} \neq P(A) = \frac{2}{52} = \frac{1}{26} \, .$$

We see that events A and B are now dependent, because, if one draws a diamonds suit, the probability of A is modified. However, Eq. (1.18) is also valid in this case:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1}{52}\frac{52}{13} = \frac{1}{13} .$$

These examples suggest the following.

Definition 1.9 (Conditional Probability) The conditional probability of B given A is the quotient of the probability of the occurrence of A and B and the probability of A:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{if} \quad P(A) > 0 . \tag{1.19}$$

It is easy to show (this is left as an exercise) that the definition of conditional probability (1.19) is in agreement with the general axioms of Kolmogorov (1.8–1.10). It is also important to note that

$$P(A|B) \neq P(B|A) , \tag{1.20}$$

a fact that appears obvious from the examples just made but that often does not appear obvious to our logical-intuitive abilities. Failure to comply with Eq. (1.20) is perhaps the source of most of the errors which are done by dealing with probabilities. The crucial point is that *the correct connection between the two probabilities is possible only through Bayes' theorem*, as we will see shortly. On this point we recommend Problems 1.16 and 1.17 and also to read about the so-called Sally Clark case (see, e.g. [Wik22]).

We also note that the conditional probability has been introduced as a *definition*. However, for the probabilities we are dealing with, the following property holds.

Theorem 1.2 (Product of Probabilities) *In the classical and frequentist frameworks, the probability of the event formed by the occurrence of both A and B is:*

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) . \tag{1.21}$$

Proof For the classical probability, if N is the total number of cases and n_{AB} that of the favourable ones to both A and B, we have:

$$P(A \cap B) = \frac{n_{AB}}{N} = \frac{n_{AB}}{n_B}\frac{n_B}{N} = P(A|B)P(B) ,$$

since, by definition, $n_{AB}/n_B \equiv P(A|B)$. This property obviously continues to hold by exchanging A and B, hence Eq. (1.21).
For the frequentist probability, the proof is analogous if one replaces the number of cases with that of trials. □

In the previous examples, we have introduced the notion of independent events; in a general way, we can adopt the

Definition 1.10 (Independent Events) Two events A and B are independent if

$$P(A \cap B) = P(A)P(B) \, .$$

More generally, the events of a family $(A_i, \; i = 1, 2 \ldots, n)$ are independent if

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i) \, , \tag{1.22}$$

for any subset J of different indices of the family.

From Eq. (1.19) it follows that for independent events $P(A|B) = P(A)$ and $P(B|A) = P(B)$. Another useful definition is:

Definition 1.11 (Incompatible Events) Two events are incompatible or disjoint when the condition

$$A \cap B = \emptyset$$

holds. From Eqs. (1.13) and (1.19) we then have:

$$P(A \cap B) = 0 \, , \qquad P(A|B) = P(B|A) = 0 \, .$$

For example, if A is the ace of spades and B the suit of diamonds, A and B are incompatible events. According to these definitions, the essence of the probability calculus can be summarized in the following formulae:

- For incompatible events:

$$P(A \text{ or } B) \equiv P(A \cup B) = P(A) + P(B) \, . \tag{1.23}$$

- For independent events:

$$P(A \text{ and } B) \equiv P(A \cap B) \equiv P(AB) = P(A) \cdot P(B) \, . \tag{1.24}$$

1.5 Repeated Trials

Up to now we have considered experiments performed with one single trial. However, often one has to deal with experiments consisting of many trials: two cards drawn from a deck, the score obtained rolling five dices and so on. We address

this problem by considering two repeated trials because the generalization to any finite number of trials is obvious, as we shall see later.

Two repeated trials can be considered as the realization of two events related to two experiments $(S_1, \mathcal{F}_1, P_1)$ and $(S_2, \mathcal{F}_2, P_2)$ which satisfy Definitions 1.6 and 1.7. It is therefore natural to define a new sample space $S = S_1 \times S_2$ as a Cartesian product of the two initial sample spaces, in which a single event is constituted by the ordered pair (x_1, x_2), where $x_1 \in S_1$ and $x_2 \in S_2$ and the new space S contains $n_1\, n_2$ elements, if n_1 and n_2 are the elements of S_1 and S_2, respectively. For example, [ace of hearts, queen of clubs] is an element of the set S of the probability space relative to the extraction of two cards from a deck. Note that the Cartesian product can also be defined when S_1 and S_2 are the same sample space.

Using definition of events, and since $A_1 \subseteq S_1$ and $A_2 \subseteq S_2$, it is easy to realize that:

$$A_1 \times A_2 = (A_1 \times S_2) \cap (S_1 \times A_2)\ . \tag{1.25}$$

The next step is now to define a probability P in $S_1 \times S_2$, which satisfies the axioms of Kolmogorov (1.8–1.10) and can be associated in a unique way with experiments consisting of repeated trials. Equation (1.24), which is valid for independent events, and Eq. (1.25) allow to write:

$$\begin{aligned} P(A_1 \times A_2) &= P[(A_1 \times S_2) \cap (S_1 \times A_2)] \\ &= P(A_1 \times S_2 | S_1 \times A_2)\, P(S_1 \times A_2) \\ &= P(A_1 \times S_2)\, P(S_1 \times A_2) \\ &= P(A_1) P(A_2) \quad A_1 \in \mathcal{F}_1\ , \quad A_2 \in \mathcal{F}_2\ , \end{aligned} \tag{1.26}$$

where the last equality is valid because the probability of the set of pairs $A_k \times S_j$ in the sample space $S_k \times S_j$ obviously has the same probability as the A_k event in the S_k sample space. The probabilities of the events $A_1 \in S_1$ and $A_2 \in S_2$ can therefore be computed in the space S using the equalities:

$$\begin{aligned} P(A_1 \times S_2) &= P_1(A_1)\, P_2(S_2) = P_1(A_1)\ , \\ P(S_1 \times A_2) &= P_1(S_1)\, P_2(A_2) = P_2(A_2)\ , \end{aligned} \tag{1.27}$$

which are obvious both for classical and frequentist probabilities. For example, in the drawing of two playing cards, the probabilities of the events $A_1 =$ [draw an ace the first time] and $A_1 \times S_2 =$[ace, any card] are equal, like those of the events $A_2 =$ [extraction of a diamonds suit the second time] and $S_1 \times A_2 =$[any card, suit of diamonds].

As we said, Eq. (1.26) is only considered valid for independent events, for which, based on (1.24), the occurrence of any event does not alter the probability of the others.

To better fix ideas with an example, suppose we pull out two cards from a playing deck (with replacement into the deck of the first card after the first draw) and let be A_1 the set of aces and A_2 the set of diamonds suits. Equation (1.27) becomes:

$$P_1(A_1) = \frac{4}{52} = P(A_1 \times S_2) = \frac{4}{52}\,\frac{52}{52}\,,$$
$$P_2(A_2) = \frac{13}{52} = P(S_1 \times A_2) = \frac{52}{52}\,\frac{13}{52}\,,$$

whereas Eq. (1.26) gives:

$$P(A_1 \times A_2) = \frac{4}{52}\,\frac{13}{52}\,,$$

according to the ratio between the number of favourable cases $(4 \cdot 13)$ and the possible ones $(52 \cdot 52)$ in the sample space $S_1 \times S_2$.

The family of sets $\mathcal{F}_1 \times \mathcal{F}_2 = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$ is not in general a σ-algebra, but it is possible to show that a single σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ of subsets of $S_1 \times S_2$ exists containing $\mathcal{F}_1 \times \mathcal{F}_2$ and that Eq. (1.26) allows the extension, in a unique way, of the probability of each event $A \subset S_1 \times S_2$ from the family set $\mathcal{F}_1 \times \mathcal{F}_2$ to the product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$ [GS92]. Therefore, we can write the product probability space as:

$$\mathcal{E} = \mathcal{E}_1 \otimes \mathcal{E}_2 \equiv (S_1 \times S_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P)\,.$$

An extension of (1.26) is used when the space S_2 cannot be defined in advance but depends from the results of the previous experiment $\mathcal{E}_1$. A good example is given by the Italian lottery, in which five numbers are drawn, without replacing them in the box.

In the case of two trials, we can imagine the extraction of two playing cards: if you reinsert the first card drawn into the deck, S_2 consists of 52 elements and 51 otherwise. Given these conditions, you need to define the space $S = S_1 \times S_2$ not as a Cartesian product but as the set of all possible ordered pairs of the two initial sets, as they result from each particular experiment. We can say that, in this situation, event A_2 depends on event A_1 and generalize Eq. (1.26) as:

$$P(A \times B) = P_2(B|A)\, P_1(A)\,, \tag{1.28}$$

resulting in an extension of the product Theorem 1.2 (see also Eq. 1.21). It is immediate to show that the a priori and frequentist probabilities match Eq. (1.28). The proof for the frequentist probability is identical to that of Theorem 1.2, whereas, for the classical probability, it is required to redefine N as the set of possible pairs, n_{AB} as the set of favourable pairs and n_A as the set of pairs in which, at the first extraction, event A occurred.

At this point, to avoid confusion, it is important to distinguish between *independent experiments* and *independent events*. The hypothesis of independent experiments, which we will use throughout the text, is completely general and implies that *the experimental procedures that lead to the occurrence of any event are independent of those which lead to the occurrence of all the other events*. This hypothesis has no connection with the number of elements of the sample space.

On the contrary, in the repeated trial scheme the events will be considered dependent when the size of the i-th space S_i depends on the $(i-1)$ trials carried out previously. This *is the only kind of dependency that one assumes, considering repeated trials, when writing the conditional probability* $P(A|B)$. Let us consider a simple example, where $(W_1 \times B_2)$ is the event which consists in the extraction, from an urn containing two white and three black marbles, of one white marble and one black marble in this order (without replacing the marble into the urn after the drawing). In this case we have (with obvious meaning of the symbols):

$$
\begin{aligned}
S_1 &= [W1, W2, B1, B2, B3] \,, \\
S_2 &= [\text{set formed by 4 marbles, 2 white and 2 black ones} \\
&\qquad \text{or 1 white and 3 black ones}] \,,
\end{aligned}
$$

$$
S_1 \times S_2 = \begin{matrix} W1W2, & W2W1, & B1W1, & B2W1, & B3W1 \\ W1B1, & W2B1, & B1W2, & B2W2, & B3W2 \\ W1B2, & W2B2, & B1N2, & B2B1, & B3B1 \\ W1B3, & W2B3, & B1N3, & B2B3, & B3B2 \end{matrix} = 5 \times 4 = 20 \text{ elements.}
$$

Since the marbles are not reinserted after the drawing, the events as $W1W1$, $B2B2$, etc. are excluded. Now we define:

$$
\begin{aligned}
W_1 \times S_2 &= [\text{white marble, any marble}] \,, \\
S_1 \times B_2 &= [\text{any marble, black marble}] \,, \\
W_1 \times B_2 &= [\text{white marble, black marble}] = \begin{pmatrix} W1B1, W2B1 \\ W1B2, W2B2 \\ W1B3, W2B3 \end{pmatrix} = 6 \text{ elements} \,.
\end{aligned}
$$

In the situation of equally probable cases, the classical probability gives:

$$P(W_1 \times B_2) = \frac{\text{six favorable cases}}{\text{twenty possible pairs}} = \frac{6}{20} = \frac{3}{10} .$$

So far we have used in a general way the definition of classical or a priori probability. Now we note that the probabilities (1.27) of the events W_1 and B_2 are given by:

$$P_1(W_1) = 2/5 , \qquad P_2(B_2|W_1) = 3/4 ,$$

because initially we have in the urn $W1$, $W2$, $B1$, $B2$, $B3$, (i.e. two white marbles over a total number of five) and, in the second drawing, we have three black marbles and one white marble in the urn if the first extracted marble was white, so there are three black marbles over four. Now we apply Eq. (1.28) and obtain again:

$$P(W_1 \times B_2) = P_2(B_2|W_1)P_1(W_1) = 3/4 \times 2/5 = 6/20 = 3/10 ,$$

according to the direct calculation of the favourable cases over the total ones.

If we neglect the order of extraction and define as event the extraction of a white and a black marble or vice versa, we must define the sets:

$$W_1 \times S_2 = [\text{white marble, any marble}] ,$$
$$B_1 \times S_2 = [\text{black marble, any marble}] ,$$
$$S_1 \times B_2 = [\text{any marble, black marble}] ,$$
$$S_1 \times W_2 = [\text{any marble, white marble}] ,$$

and apply Eq. (1.28):

$$P[(B_1 \times W_2) \cup (W_1 \times B_2)] = P_2(W_2|B_1)P_1(B_1) + P_2(B_2|W_1)P_1(W_1) = \frac{3}{5} .$$

The result agrees with the ratio between favourable (12) and possible pairs (20).

The generalization of this scheme for a higher number of repeated trials requires the natural extension of the equations discussed here for two trials only and does not present any relevant difficulty.

1.6 Elements of Combinatorial Analysis

Assuming you are already familiar with the topic, we briefly summarize here the basic formulae of combinatorial analysis, which are often helpful in calculating probabilities by counting the number of possible or favourable cases.

To count well, it must be kept in mind that the number of possible pairs (matches) $A \times B$ between two sets A of a elements and B of b elements is given by the product ab and that the number of possible permutations of n objects is given by the factorial $n!$. A selection or arrangement in which order is important is called a permutation; a selection in which order is neglected is called a combination.

Based on these properties, four fundamental formulae can be easily demonstrated, which refer to arrangements without repetition $D(n,k)$ of n objects in groups of k (using k of the objects at a time), to those with repetition $D^*(n,k)$ and to combinations without and with repetition $C(n,k)$ and $C^*(n,k)$, in which the order of the k elements does not matter.

The formulae, as perhaps you already know, are:

$$D(n,k) = n(n-1)\cdots(n-k+1)\,, \tag{1.29}$$

$$D^*(n,k) = n^k\,, \tag{1.30}$$

$$C(n,k) = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!} \equiv \binom{n}{k}\,, \tag{1.31}$$

$$\begin{aligned} C^*(n,k) &= \frac{(n+k-1)(n+k-2)\cdots n}{k!} \\ &= \frac{(n+k-1)!}{k!(n-1)!} \equiv \binom{n+k-1}{k}\,, \end{aligned} \tag{1.32}$$

where the binomial coefficient formula has been used.

To understand these formulae, just imagine the group of k objects such as the Cartesian product of k sets. In $D(n,k)$ the first set contains n elements; the second set contains $n-1$ elements because the first element is excluded, until you get, after k times, a set of $(n-k+1)$ elements. Instead, if the repetitions in the group of k objects are allowed, all the sets will contain n elements each; hence, we obtain Eq. (1.30). The base n number system is just a $D^*(n,k)$ arrangement: if, for instance, $n = 10$, $k = 6$, we have 10^6 numbers, from 000,000 to 999,999.

In Eq. (1.31), where $C(n,k) = D(n,k)/k!$, the number of groups containing the same k objects is not counted, because in this case the order does not matter.

Finally, to obtain Eq. (1.32) one has to imagine to write, for instance, a combination $C^*(n,5)$ as $a_1a_2a_2a_2a_7$ in a new way: $a_1 * a_2 * * * a_3a_4a_5a_6a_7*$, where any element is followed by a number of asterisks equal to the number of times of its occurrence; it is easy to verify that there is a one-to-one correspondence between the original combinations and all possible permutations in the alignment of letters and asterisks in the alternative representation. Since each alignment starts with a_1, it is possible to permute in total $n-1+k$ objects, that is, k asterisks and $n-1$ elements (a_i with $i = 2,\ldots,n$) equal to each other, obtaining Eq. (1.32).

In R, it is possible to calculate $n!$ with the routine `factorial(n)` and the binomial coefficients (1.31) with the routine `choose(n,k)`. Moreover, the routine `combn(n,k)` prints the combinations (1.31) by columns, but a routine for

the calculation of the permutations is not available. For this purpose our routines `Perm`, `Combn` and `Dispn` are available to print permutations and combinations by rows.

A particularly useful formula is the *hypergeometric law*, which allows the calculation of the probability to extract k marbles of type A having extracted $n \leq a + b$ marbles without replacement from an urn containing a marbles of type A and b marbles of type B. Assuming that all marbles have the same probability of being extracted and that extractions are independent, adopting the a priori definition (1.2) and using the binomial coefficients, we have:

$$P(k; a, b, n) = \frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}, \quad \max(0, n-b) \leq k \leq \min(n, a)\,. \tag{1.33}$$

In fact, the number of possible cases in the denominator is given by the binomial coefficient, while in the numerator we have the number of favourable cases, given by the number of elements of the Cartesian product of the two sets consisting of $C(a, k)$ and $C(b, n-k)$ elements, respectively.

In R, the hypergeometrical law probabilities are calculated by the routine `dhyper(k,a,b,n)`.

Exercise 1.1
Find the probability, in a lottery, of a combination of two (pair) or three (triplet) numbers out of five numbers between 1 and 90 drawn from an urn (Italian lottery).

Answer The solution, if the game is not rigged, is given by the hypergeometric law (1.33) with $a = k$ and $b = 90 - k$:

$$P(2; 2, 88, 5) = \frac{\binom{88}{3}}{\binom{90}{5}} = \frac{2}{800} \quad \text{(pair)}\,,$$

$$P(3; 3, 87, 5) = \frac{\binom{87}{2}}{\binom{90}{5}} = \frac{1}{11\,748} \quad \text{(triplet)}\,.$$

(continued)

Exercise 1.1 (continued)
The same results are obtained by calling `dhyper(2,2,88,5)` and `dyper(3,3, 87,5)`. The pair probability is about 1 over 400 and that of the triplet is about 1 over 12,000. A game is fair if the payout equals the inverse of the probability of the bet; in the Italian lottery, the pair is paid 250 times and the triplet 4250 times . . .

1.7 Bayes' Theorem

In principle, any problem involving the use of probability can be solved with the two fundamental laws of additivity and product. However, the algebra of probability leads quickly to complicated formulae, even in the case of relatively simple situations. In these cases two basic formulae are of great help, those of total probabilities and the Bayes' theorem, as we will show. If the sets B_i $(i = 1, 2, ..n)$ are pairwise disjoint and collectively exhaustive:

$$\bigcup_{i=1}^{n} B_i = S, \qquad B_i \cap B_k = \emptyset \qquad \forall i, k \ , \tag{1.34}$$

by means of Eq. (1.21), it is easy to show that, for every set A in S:

$$\begin{aligned} P(A) &= P[A \cap (B_1 \cup B_2 \cup \cdots \cup B_n)] \\ &= P[(A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)] \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n) \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i) \ . \end{aligned} \tag{1.35}$$

Equation (1.35) is called *partition theorem* or *law of total probability*. When $B_1 = B$ e $B_2 = \overline{B}$, the theorem gives:

$$P(A) = P(A|B)P(B) + P(A|\overline{B})P(\overline{B}) \ . \tag{1.36}$$

With these formulae, you can solve problems that happen frequently, such as those shown in the following two examples.

Exercise 1.2
A disease H affects 10% of men and 5% of women per year. Knowing that the population is composed by 45% men and 55% women, find the expected number N of sick persons in a population of 10,000 people.

Answer The probability of getting sick for each man or woman of the population is given by the probability that the individual is a woman times the probability that a woman has to get sick plus the probability that the individual is a man times of the probability a man has of getting sick. This situation is summarized into Eqs. (1.35, 1.36). Therefore, we have:

$$P(H) = 0.45 \cdot 0.10 + 0.55 \cdot 0.05 = 0.0725 \,.$$

The expected number of sick persons is obtained by multiplication of the number of trials (individuals) times the probability $P(H)$ we have just found. We then have:

$$N = 10{,}000 \cdot 0.0725 = 725 \,.$$

Exercise 1.3
A box contains six white and four black marbles. After two extractions without replacement, what is the probability to get a white marble at the second draw?

Answer By indicating with A and B the outcome of a white marble at the first and second extraction, respectively, from Eq. (1.36) one obtains, with obvious meaning of symbols:

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) = \frac{5}{9}\frac{6}{10} + \frac{6}{9}\frac{4}{10} = 0.60 \,.$$

If we now use Eq. (1.35) to express the probability $P(A)$ that appears in (1.21), we get the famous Bayes' theorem.

Theorem 1.3 (Bayes) *When the sets B_k follow Eq. (1.34), the conditional probability $P(B_k|A)$ can be written as:*

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)} , \quad P(A) > 0 . \tag{1.37}$$

This theorem is perhaps the most relevant result of the elementary algebra of probability, because it allows us to reverse the conditional probabilities, avoiding the errors resulting from the violation of Eq. (1.20). It is often used to "readjust", based on a real data set A_k, the probabilities $P(B_k)$ arbitrarily assigned a priori. The procedure to be used is shown in the following examples: we also recommend physics students to solve the Problem 1.8 at the end of the chapter.

Exercise 1.4
A test for the diagnosis of a disease is 100% sensitive for sick people but is also positive in 5% of the healthy people. Knowing that the illness is present on average in 1% of the population, what is the probability of being really sick if your test is positive?

Answer Since the diagnostic testing is an important medical problem, let's deal with the topic in a general way. We can define the following conditional probabilities:

$P(P|H) = 0.05$ False Positive (FP): probability to be positive when healthy,
$P(N|H) = 0.95$ True Negative (TN): probability to be negative when healthy,
$P(P|S) = 1.$ True Positive (TP): probability to be positive when sick,
$P(N|S) = 0.$ False Negative (FN): probability to be negative when sick.

$P(P|S)$ and $P(N|H)$ are known as sensitivity and specificity, respectively. From the probability laws one obviously has:

$$P(P|H) + P(N|H) = 1.$$

$$P(P|S) + P(N|S) = 1.$$

A test is ideal when the following conditions hold:

$$P(P|H) = 0 , \qquad P(N|H) = 1 ,$$

$$P(P|S) = 1 , \qquad P(N|S) = 0 .$$

(continued)

Exercise 1.4 (continued)
Now we have to find the probability $P(S|P)$ of being sick conditioned by the positivity of the test. Applying Bayes' theorem (1.37) and bearing in mind that from the data we know that the probabilities to be healthy or sick are, respectively, $P(H) = 0.99$ and $P(S) = 0.01$, we obtain:

$$P(S|P) = \frac{P(P|S)P(S)}{P(P|S)P(S) + P(P|H)P(H)} = \frac{1 \times 0.01}{1 \times 0.01 + 0.05 \times 0.99} = 0.168\,,$$

that is, a probability of about 17%.
The result (a low probability with the positive test) seems paradoxical at first sight. To help your intuition, we invite you to examine Fig. 1.6, which shows the graphical representation of Bayes' theorem. If 100 people are subjected to the test, on average 99 will be healthy and only 1 will be sick; the test, applied to the 99 healthy, will fail in 5% of cases, corresponding to $0.05 \times 99 = 4.95 \simeq 5$ positive cases; to these the correctly diagnosed case of disease must be added. Eventually, we will have only one really sick person of a total six positive tests:

$$\frac{1}{6} = 16.67\%$$

where the small difference with the exact calculation is due only to rounding effects.
The test is then repeated for the positive persons. If the result is negative, then the person is healthy, because the test here considered can never go wrong on sick people. If the test results were still positive, then it is necessary to calculate, based on Eq. (1.24), the probability of a doubly positive test on a healthy person:

$$P(PP|H) = P(P|H)\,P(P|H) = (0.05)^2 = 0.0025\,,$$

which is about 2.5 per thousand and that of a doubly positive test on a sick (which obviously gives again $P(PP|S) = 1$) and reapply the Bayes' theorem:

$$\begin{aligned} P(S|PP) &= \frac{P(PP|S)P(S)}{P(PP|S)P(S) + P(PP|H)P(H)} \\ &= \frac{1 \times 0.01}{1 \times 0.01 + 0.0025 \times 0.99} = 0.802 \simeq 80\%\,. \end{aligned}$$

(continued)

Exercise 1.4 (continued)
The same result is obtained if one uses the initial probabilities $P(P|S)$ and $P(P|H)$ to people who have already undergone a test, for which $P(S) = 0.168$ and $P(H) = 0.802$.
As you can see, not even two positive tests are enough for the certainty of the disease. You can calculate by yourself that, in these conditions, the certainty comes only after three consecutive tests (about 99%).
The example shows how careful you need to be with testing which may result positive even on healthy people. The opposite is true with the tests that are always negative on the healthy persons but not always positive on the sick ones. In this case a positive test assures the disease, whereas a negative test leaves some uncertainty. There are also cases where the tests have an efficiency limited to both the healthy and sick persons. In all these situations, Bayes' theorem allows you to exactly calculate the probabilities of interest.

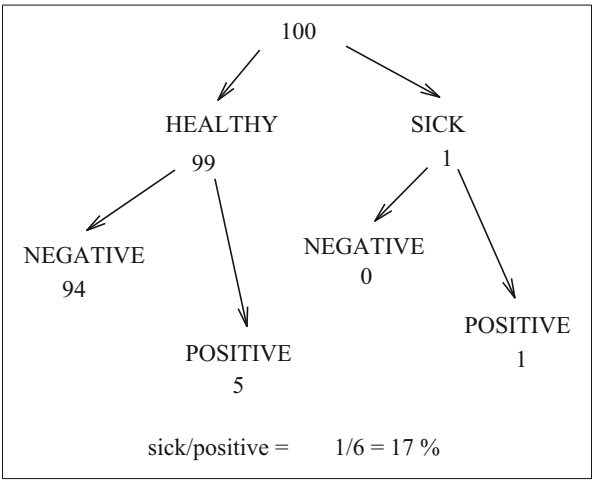


Fig. 1.6 Graphical illustration of Bayes' theorem for a test which gives 5% of false positives for a disease affecting 1% of the population

Exercise 1.5
A group of symptoms A_1, A_2, A_3, A_4 can be due to three diseases H_1, H_2, H_3, which, based on epidemiological data, have a relative frequency of 10%, 30% and 60%, respectively. The relative probabilities are therefore:

$$P(H_1) = 0.1, \qquad P(H_2) = 0.3, \qquad P(H_3) = 0.6\,. \tag{1.38}$$

According to epidemiological data, the occurrence of the symptoms above in the three diseases are as follows:

	A_1	A_2	A_3	A_4
H_1	.9	.8	.2	.5
H_2	.7	.5	.9	.99
H_3	.9	.9	.4	.7

from which it results, for example, that the symptom A_2 occurs in 80% of cases in the H_1 disease, the symptom A_4 occurs in 70% of cases in the H_3 disease and so on.
A patient presents only A_1 and A_2 symptoms. Which of the three considered diseases is the most likely?

Answer First of all, to apply Bayes' theorem, it is necessary to define the patient as an event A such that:

$$A = A_1 \cap A_2 \cap \overline{A}_3 \cap \overline{A}_4\,,$$

and to calculate the probabilities of this event, conditional on the three diseases (hypotheses) H_1, H_2, H_3:

$$P(A|H_i) = P(A_1|H_i)\; P(A_2|H_i)\; P(\overline{A}_3|H_i)\; P(\overline{A}_4|H_i) \qquad (i = 1, 2, 3)\,.$$

From the table, we also obtain:

$$P(A|H_1) = .9 \times .8 \times .8 \times .5 = 0.288\,,$$
$$P(A|H_2) = .7 \times .5 \times .1 \times .01 = 0.00035\,,$$
$$P(A|H_3) = .9 \times .9 \times .6 \times .3 = 0.1458\,.$$

The most likely disease seems to be H_1, but we have not yet taken into account the epidemiological frequencies (1.38); to deal with this crucial point, it is necessary to use Bayes' theorem!
We therefore apply Eq. (1.37) and finally get the probabilities for each of the three diseases (note that the sum gives 1, thanks to the denominator of Bayes'

(continued)

Exercise 1.5 (continued)
formula, which is just the normalization factor):

$$P(H_1|A) = \frac{0.288 \times 0.1}{0.288 \times 0.1 + 0.00035 \times 0.3 + 0.1458 \times 0.6} = 0.2455\,,$$
$$P(H_2|A) = \frac{0.00035 \times 0.3}{0.288 \times 0.1 + 0.00035 \times 0.3 + 0.1458 \times 0.6} = 0.0009\,,$$
$$P(H_3|A) = \frac{0.1458 \times 0.6}{0.288 \times 0.1 + 0.00035 \times 0.3 + 0.1458 \times 0.6} = 0.7456\,.$$

The final result shows that H_3 is the most likely disease, with a probability of about 75%.
The solution to the problem can also be found graphically, as shown in Fig. 1.7: since there are small probabilities, in the figure we consider 100,000 subjects, who are divided according to the three diseases weighted with the epidemiological frequencies 0.1, 0.3, 0.6; applying to these three groups the probabilities of the set of symptoms A (0.288, 0.00035, 0.1458), one gets the final numbers 2880, 10, 8748. Also in this way we obtain the results provided by Bayes' formula.

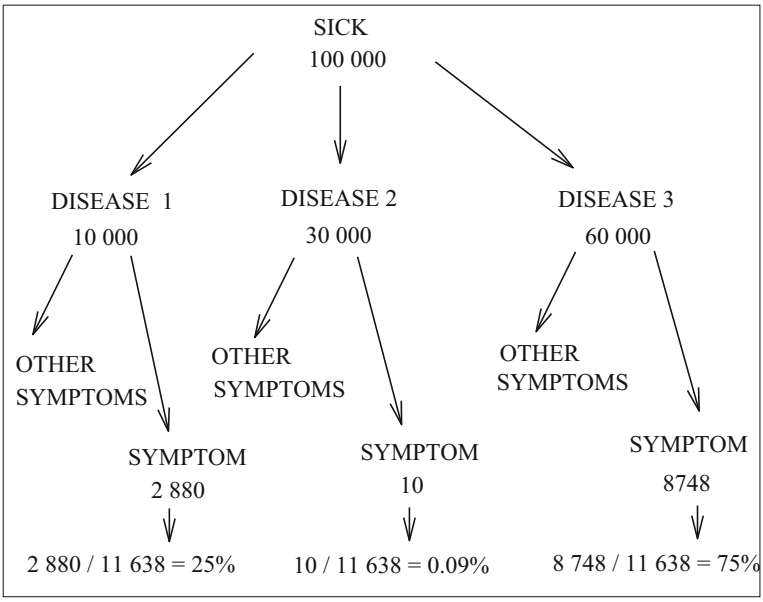


Fig. 1.7 Graphic illustration of Bayes' theorem in the case of three diseases with some symptoms in common

1.8 Learning Algorithms

Bayes' formula is the basis of many machine learning codes and artificial intelligence algorithms, from spam mail recognition to the proper function of electric appliances and to the learning of neural networks. The topic is very broad and we just want to give you a general idea with a simple example. Suppose Bob is attracted to Alice (the example also applies to parts inverted) and that he wants to test if the interest is mutual by inviting Alice to have a coffee. Having no information, Bob assumes the following probabilities:

$$
\begin{aligned}
P(OK) &= 0.5\,, \text{ attraction,}\\
P(\overline{OK}) &= 1 - P(OK) = 0.5\,, \text{ indifference,}\\
P(YES|OK) &= 0.9\,, \text{ positive answer with attraction,}\\
P(NO|OK) &= 1 - P(YES|OK) = 0.1\,, \text{ negative answer with attraction,}\\
P(YES|\overline{OK}) &= 0.5\,, \text{positive answer and indifference,}\\
P(NO|\overline{OK}) &= 1 - P(YES|\overline{OK}) = 0.5\,, \text{negative answer and indifference,}
\end{aligned}
$$

which give 50% probability to the possible existence of attraction by Alice. In the case of Alice's first affirmative answer, the probability of mutual attraction becomes, by using the initial data:

$$
\begin{aligned}
P(OK|YES) &= \frac{P(YES|OK)P(OK)}{P(YES|OK)P(OK) + P(YES|\overline{OK})P(\overline{OK})} \qquad (1.39)\\
&= \frac{0.9 \cdot 0.5}{0.9 \cdot 0.5 + 0.5 \cdot 0.5} = 0.643\,.
\end{aligned}
$$

Instead, in the case of a first negative answer, one has:

$$
\begin{aligned}
P(OK|NO) &= \frac{P(NO|OK)P(OK)}{P(NO|OK)P(OK) + P(NO|\overline{OK})P(\overline{OK})} \qquad (1.40)\\
&= \frac{0.1 \cdot 0.5}{0.1 \cdot 0.5 + 0.5 \cdot 0.5} = 0.167\,.
\end{aligned}
$$

Now the crucial step for learning takes place: in evaluating the probability after a second answer, the substitution $P(OK|YES) \rightarrow P(OK)$ is performed (and consequently ($P(\overline{OK}) = 1 - P(OK)$) in Eq. (1.39) in the case of affirmative answer or $P(OK|NO) \rightarrow P(OK)$ in Eq. (1.40) in case of a first negative answer. In this way, basic learning is achieved based on the data accumulation, so that the probability $P(OK)$, assumed initially to be 50% in lack of initial information, is continuously updated and made more reliable. With our routine `BayesBobAl`, you can interactively check how the probabilities evolve as a function of the answers.

It turns out, for example, that if there are three consecutive negative answers, the chance of Alice's attraction to Bob assumes gradually the decreasing values $0.167, 0.038, 0.008$, confirming the advice given by a friend: "Bob, after three refusals, it is better to give up..."

In the previous example, we showed how learning algorithms extend the application of Bayes' formula also to the cases where, at the beginning, the probabilities are not known with reasonable certainty. In these situations, Eq. (1.37) can be also used to modify, during the data collection, the initial probabilities of hypotheses $P(H_i)$ *subjectively evaluated according to statement 1.2*.

Following the method we have outlined above, if we indicate with the generic event "data" the result of one or more trials (experiments), in the Bayesian approach one applies Eq. (1.37) as follows:

$$P(H_k|\text{data}) = \frac{P(\text{data}|H_k)P(H_k)}{\sum_{i=1}^{n} P(\text{data}|H_i)P(H_i)} \,. \tag{1.41}$$

The probabilities $P(H_k|\text{data})$ thus obtained are then substituted for $P(H_k)$ in the term to the right of Eq. (1.41) and the calculation can be repeated iteratively:

$$P_n(H_k|E) = \frac{P(E_n|H_k)P_{n-1}(H_k)}{\sum_{i=1}^{n} P(E_n|H_i)P_{n-1}(H_i)} \,, \tag{1.42}$$

where E_n is the n-th event. In the following example, the probabilities $P(E_n|H_k)$ remain constant.

Exercise 1.6
An urn contains five black and white marbles in an unknown proportion. Assuming the same probability $P(H_i) = 1/6$ for the six possible starting hypotheses, written with obvious notation as:

$$H_i = (b, n) \equiv \begin{cases} H_1 = (5, 0), \; H_2 = (4, 1), \; H_3 = (3, 2), \\ H_4 = (2, 3), \; H_5 = (1, 4), \; H_6 = (0, 5) \end{cases}$$

calculate the six probabilities $P(H_i|\text{data})$ when, with replacement into the urn, $n = 1, 5, 10$ black marbles are extracted consecutively.

Answer The exercise is easily solved by defining the event "data"$= E = E_n$ as the extraction of a black marble, using, for the a priori probabilities $P(E|H_k) = P(E_n|H_k)$, the values:

(continued)

Exercise 1.6 (continued)

$$P(\text{black}|H_1) = 0, \quad P(\text{black}|H_2) = 1/5, \quad P(\text{black}|H_3) = 2/5,$$
$$P(\text{black}|H_4) = 3/5, \quad P(\text{black}|H_5) = 4/5, \quad P(\text{black}|H_6) = 1$$

and applying iteratively Eq. (1.42). One gets Table 1.2, from which we see that, when increasing the number of black marbles drawn consecutively, hypothesis H_6 (5 black marbles) becomes more and more probable.

It is important to note that this problem has not been solved with a pure frequentist approach, because for the initial hypotheses the a priori probabilities $P(H_i = 1/6$ $(i = 1, 2, \ldots, 6)$ were used, which are subjective and arbitrary. However, this increased flexibility is paid with a certain amount of ambiguity, because with different initial hypotheses, different results would have been obtained, as in Problems 1.12 and 1.13. The dilemma "greater flexibility of application in spite of some ambiguity of the results" often gives rise to heated debates, as in [JLPe00].

The example just seen is therefore fundamental to understand the difference between the frequentist and the Bayesian approach:

- Frequentist approach (followed in this book): no arbitrary subjective probabilities are assumed for the hypotheses. Therefore, probabilities of hypotheses of the type $P(H|\text{data}) \equiv P(\text{hypothesis}|\text{data})$ are never determined. For a frequentist solution of the exercise just seen, you can see later Exercise 6.1.
- Bayesian approach: probabilities as $P(\text{hypothesis}|\text{data})$ are determined. They depend, via Eq. (1.41), on the initial probabilities arbitrarily assumed for the hypotheses and from the data obtained during the trials.

Table 1.2 Calculation of the a posteriori probabilities $P_n(H|n\bullet)$ starting from equal a priori probabilities in the case of consecutive extractions with replacement of n black marbles from an urn containing five black and white marbles

Hypothesis	H_1	H_2	H_3	H_4	H_5	H_6
Urn content	○ ○ ○ ○ ○	○ ○ ○ ○ ●	○ ○ ○ ● ●	○ ○ ● ● ●	○ ● ● ● ●	● ● ● ● ●
$P(H_i)$ a priori	1/6	1/6	1/6	1/6	1/6	1/6
$P_n(H_i\|n = 1\bullet)$	0	0.07	0.13	0.20	0.27	0.33
$P_n(H_i\|n = 5\bullet)$	0.	0.00	0.01	0.05	0.23	0.71
$P_n(H_i\|n = 10\bullet)$	0.	0.00	0.00	0.00	0.11	0.89

1.9 Problems

1.1 Monty Hall's game is named after the host of a television game that in 1990 made a lot of Americans discuss about probabilities. The competitor is placed in front of three doors: behind one door there is a car, and behind the others, there are goats. He picks a door, say n. 1, and Monty, who knows what's behind the doors, opens another door, say n. 3, which has a goat. He then says to you, "Do you want to pick door number 2?" Is it better to change, not to change or the choice is indifferent?

1.2 In the game of bridge, a deck of 52 cards is divided into 4 groups of 13 cards and dealt to 4 players. Calculate the probability that 4 players who play 100 games a day for 15 billion years (the age of the universe) can repeat the same game.

1.3 A device is made up of three elements, which can fail independently of each other. The probabilities of operation of the three elements during a fixed time T are $p_1 = 0.8$, $p_2 = 0.9$, $p_3 = 0.7$. The machine stops due to a fault in the first element or for failure of the second and third elements. Calculate the probability P for the device to work within T.

1.4 A device is made up of four elements all having the same probability $p = 0.8$ of operation within the time T. The device stops for a simultaneous failure of the elements 1 and 2 or for a simultaneous failure of elements 3 and 4. (a) Draw the device operating flow and (b) calculate the probability P of working within T.

1.5 Calculate the probability of getting at least a face with 6 by rolling three dices.

1.6 A quality check of a batch containing ten pieces accepts the whole lot if all three pieces chosen at random are good. Calculate the probability P that the lot will be discarded in case of (a) one defective piece or (b) four defective pieces.

1.7 The famous "encounter problem": two friends X and Y decide to meet in a certain place at an hour between 12 and 13, randomly choosing the arrival time. X arrives, waits for 10 minutes, and then leaves. Y behaves like X but waits for 12 minutes. What is the probability P that X and Y meet each other?

1.8 The trigger problem, common in physics: a physical system randomly produces the events A and B with probability 90% and 10%, respectively. A device, designed to select the good B events, enables (triggers) the recording of the events A and B in 5% and 95% of cases, respectively. Calculate the percentage $P(T)$ of events accepted by the trigger and the percentage $P(B|T)$ of B type events among those accepted.

1.9 Evaluate the probability $P\{X \leq Y\}$ that in a test, in which two coordinates $0 \leq X, Y \leq 1$ are randomly extracted in a uniform way, one gets the values $x \leq y$.

1.10 Three electronic firms, A, B and C, supply identical components to a laboratory. The supply percentages are 20% for A, 30% for B and 50% for C. The percentage of defective components of the three suppliers is 10% for A, 15% for B and 20% for C. What is the probability that a component chosen at random will turn out to be defective?

1.11 A certain type of pillar has the breaking load R uniformly distributed between 150 and 170 kN. Knowing that it is subjected to a random load C evenly distributed between 140 and 155 kN, calculate the probability of the pillar failure.

1.12 Solve Exercise 1.6 in the case of consecutive extraction of $n = 5$ black marbles, assuming the following initial probabilities (of binomial type): $P(H_1) = 0.034$, $P(H_2) = 0.156$, $P(H_3) = 0.310$, $P(H_4) = 0.310$, $P(H_5) = 0.156$, $P(H_6) = 0.034$.

1.13 If you assume that your friend is 50% honest and 50% cheating, find the final probability that the friend is a cheater after $n = 5, 10, 15$ consecutive wins.

1.14 The probability of the three events A, B and C is different from zero. State whether the following statements are true or false:
1) $P(ABC) = P(A|BC)P(B|C)P(C)$; 2) $P(AB) = P(A)P(B)$; 3) $P(A) = P(AB) + P(A\bar{B})$; 4) $P(A|BC) = P(AB|C)P(B|C)$.

1.15 A randomly chosen thermometer from a sample marks 21° Celsius. From the production standards, you know that the probabilities that the thermometer shows the temperature decreased by one degree, the right one and that increased by one degree are 0.2, 0.6, 0.2, respectively. The subjective a priori probabilities about the temperature of the environment, according to a survey, are: $P(19°) = 0.1$, $P(20°) = 0.4$, $P(21°) = 0.4$, $P(22°) = 0.1$. Calculate the a posteriori probabilities of the measured temperature.
(*Hint: indicate the temperatures to be evaluated as* $P(true|measured) \equiv P(true|21°)$).

1.16 The probability of honestly winning a lottery is estimated at one over a million (10^{-6}). Prove that the probability for a winner to be honest is *not* 10^{-6}!

1.17 The likelihood of a DNA test making a wrong association is evaluated in one case over 10,000. In a town of 20,000 inhabitants, in which it is certain that the responsible for a serious crime is present, all the inhabitants are tested for DNA. What is the probability that a positive tested person is guilty?

1.18 Find the probability of the event depicted on the book cover.

Chapter 2
Representation of Random Phenomena

Science is predicated upon the belief that the Universe is algorithmically compressible and the modern search for a Theory of Everything is the ultimate expression of that belief, a belief that there is an abbreviated representation of the logic behind the Universe's properties that can be written down in finite form by human beings.

John D. Barrow, "THEORIES OF EVERYTHING: THE QUEST FOR ULTIMATE EXPLANATION".

2.1 Introduction

We will begin this chapter by better defining the formalism, notation and terminology that will accompany us throughout the rest of the book. Without this very important step, the reader would risk to misunderstand the meaning of most of the basic equations of probability and statistics.

We will continue by describing the representation of events in histograms, which is the most convenient for the correct development of both probabilistic and statistical theories and the one closest to applications. This choice will lead us to immediately define the first probability distribution, the binomial, while the other distributions will be studied later, in Chap. 3.

From now on we begin the systematic use of the R software to explain all the new concept and topics that will be presented. We therefore recommend the reader to install R on her/his computer before reading this chapter, and to get some practice through the online instructions and the good manuals that can be downloaded online. In addition, Appendix B should also be read in parallel with this chapter.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3_2

2.2 Random Variables

Given the Definitions 1.5, 1.6 and 1.7 from the previous chapter, we consider a probability space $\mathcal{E} = (S, \mathcal{F}, P)$: let $a \in A \subseteq S$ be the results of an experiment realizing the event A of the σ-algebra $\mathcal{F}$. We associate to each element a a real number using the function:

$$X : S \to (-\infty, +\infty) , \quad \text{that is,} \quad X(a) = x . \tag{2.1}$$

Then we have the following definition.

Definition 2.1 (Random Variable) A random variable $X(a)$ is a function having the space S as domain, the real axis as codomain and such that the set of elements a for which the relation

$$X(a) \leq x \tag{2.2}$$

holds is an event for any $x \in \mathbb{R}$.

Be careful because this is an important conceptual step: the random variable is defined as *a correspondence or function* leading from the sample space to the real axis. Obviously, it would be more appropriate to speak about a random function instead of a random variable, but this terminology is the standard one. If $a_1 \in A_1$ and $a_2 \in A_2$ are elements of two sets (events) of $\mathcal{F}$, based on Eq. (1.6), the element $a \in (A_2 - A_1)$ will also belong to a set (event) of the field. If now $X(a_1) \leq x_1$ and $X(a_2) \leq x_2$, $(A_2 - A_1)$ will be the event corresponding to the numerical set:

$$x_1 < X(a) \leq x_2 \quad \text{that is} \quad x_1 < X \leq x_2 . \tag{2.3}$$

Since a random variable establishes a correspondence between events and real numbers, from Definition 2.1 and from Eq. (1.5), it follows that also the set:

$$\bigcap_{n=1}^{\infty} \{a : x_0 - \frac{1}{n} < X(a) \leq x_0\} = \{a : X(a) = x_0\} \equiv \{a : X = x_0\} , \tag{2.4}$$

where x_0 is a real number, is an event. Let us take, as an example, the experiment consisting in the extraction of a numbered marble in the lottery. According to the formalism just introduced, Eq. (2.2) becomes:

$$\text{EXTRACT AND READ (marble)} \leq \text{integer number} .$$

The domain of this law is the set of marbles, the codomain is a subset of the real axis and the random variable is defined as "random extraction of a marble and read out of the number". In general, if $\mathbb{R}_0$ is any subset of the real axis, representable as

union (sum), intersection (product) or difference of intervals, it is easy to show that also the numerical set:

$$\{a : X(a) \in \mathbb{R}_0\}$$

can be referred to a σ-algebra $\mathcal{F}$ and defines an event.

To summarize, in this book we will use the notation:

$$\{X \geq x_0\}\,, \quad \{X \leq x_0\}\,, \quad \{x_1 \leq X \leq x_2\}\,, \quad \{X \in \mathbb{R}_0\}\,, \tag{2.5}$$

to indicate *a set $A \in \mathcal{F}$ of a elements obtained experimentally* (not a set of real numbers!) for which the function $X(a)$ satisfies the numerical condition within braces. The probability:

$$P(A) = P\{X(a) \in \mathbb{R}_0\} \tag{2.6}$$

is named *distribution* of X. It is an application $\mathbb{R}_0 \rightarrow P\{X(a) \in \mathbb{R}_0\}$ which associates the probability of X to assume values in $\mathbb{R}_0$ for any subset $\mathbb{R}_0 \subseteq \mathbb{R}$. The distribution is a function defined on a set. It can be put in correspondence with linear combinations of sums or integrals of functions, which are easier to handle. These are the cumulative and density functions, which we will define shortly. We consider probabilities for which the condition:

$$P\{X = \pm\infty\} = 0$$

holds. Note that this property does not exclude the variable from assuming the infinity as a value; however, the probability of these events is zero. In other books you can find the notation:

$$\boldsymbol{x}(a) \leq x$$

or other equivalent forms instead of Eq. (2.2).

We will use uppercase letters (usually the latter ones) to indicate the random variables, and we will reserve the lowercase letters to represent the occurrences (sometimes called random variates), that is, particular numerical values obtained in an experiment; when we have conflicts of notation, we use uppercase bold letters such as $\boldsymbol{X}$ for random variables.

A random variate is sometimes called deviate when it's different from the mean or other central parameters (often divided by the standard deviation of the distribution). Moreover, to understand what we indicate, you will have to pay attention to the brackets: the expression $\{\ldots\}$ will always represent a set of the sample space, corresponding to the set of real values indicated inside the curly brackets.

Table 2.1 Comparison between the present notation and other current notations

Meaning	Book notation	Other notations
Result of a trial	a	a
The result is an event	$a \in A$	$a \in A$
Random variable	$X(a) = x$	$\boldsymbol{x}(a) = x$
If $a \in A$, $X \in \mathbb{R}_0$	$\{X \in \mathbb{R}_0\}$	$\{\boldsymbol{x} \in \mathbb{R}_0\}$
Probability $P(A)$	$P\{X \in \mathbb{R}_0\}$	$P\{\boldsymbol{x} \in \mathbb{R}_0\}$
Real codomain of $X(a)$	Spectrum or support	Codomain, support
Operator on random variable	$O[X]$	$O[\boldsymbol{x}]$, $O(X)$
Expected value	$E[X]$, $\langle X \rangle$	$E(\boldsymbol{x})$, $E(X)$
Variance	$\mathrm{Var}[X]$	$\mathrm{Var}[\boldsymbol{x}]$, $\sigma^2(X)$

The notations we use are given in Table 2.1. We stress the difference between capital and lowercase notation:

- X is the random variable that can take on a finite or infinite set of possible numerical values, *before* one or more trials.
- x is a number, that is, a specific numerical value of X *after* a specific trial. The n-tuple of values $(x_1, x_2, \ldots, x_n)$ is the result of a sampling.

It is important to note that the random variable must be defined on all elements of the sample space S. As an example, consider a space $(S, \mathcal{F}, P)$ and an event A with probability p, so that $P(A) = p$. It is natural then to define the function X, called dummy function or variable, such that:

$$X(a) = 1 \ \text{ if } a \in A\,, \quad X(a) = 0 \ \text{ if } a \in \bar{A}\,. \tag{2.7}$$

It is easy to see that X is a random variable:

- If $x < 0$ then to $X \leq x$ corresponds the empty set $\emptyset$.
- If $0 \leq x < 1$ then to $X \leq x$ corresponds the set $\bar{A}$.
- If $x \geq 1$ then to $X \leq x$ corresponds the sample space $A \cup \bar{A} = S$.

From the properties of the σ-algebra $\mathcal{F}$, if A is an event also the sets $\emptyset$, $\bar{A}$ and S are events; hence X satisfies Definition 2.1 and is therefore a random variable. Several random variables can be defined on the same probability space. For example, if the sample space consists of a set of persons, the random variable $X(a)$ can be the weight of a person and the variable $Y(a)$ its height. We will also often have to deal with functions of one, two or more random variables, such as:

$$Z = f(X, Y)\,. \tag{2.8}$$

The Z domain is formed by the elements of the sample space S, since by definition $X \equiv X(a)$ and $Y \equiv Y(b)$ with $a, b \in S$, so that also $Z = f(X(a), Y(b))$ holds.

Since $X(a) = x$ and $Y(b) = y$, it is possible to associate to Z also a "traditional" function with real domain and codomain:

$$z = f(x, y) \, .$$

So far, the discussion shows that the random variable creates a correspondence between countable unions, intersections, differences and complements of sets of S and intervals of real numbers. It is then possible to prove (but it is quite intuitive) that, for every real z, if the inequality $f(x, y) \leq z$ is satisfied by unions, countable intersections or differences of real intervals of the variables x and y, then the variable Z of Eq. (2.8) is also a random variable obeying Definition 2.1 [Cra51, PUP02]. If X and Y are random variables defined on the same probability space, the same happens for all the functions $f(X, Y, \ldots)$ that will be considered later, so we will always assume, from now on, that the functions of random variables are also random variables.

Following our notation, the difference between $Z = f(X, Y)$ and $z = f(x, y)$ can be explained with the following example. Consider the sum:

$$Z = X + Y$$

and the experimental results:

$$z_1 = x_1 + y_1 \quad \text{and} \quad z_2 = x_2 + y_2 \, .$$

The random variable Z indicates a possible set of results $(z_1, z_2, \ldots)$, which are the outcomes of Z in an experiment consisting of a series of trials where X and Y occur and results are added together. Instead, z_1 represents the value of Z obtained in the first test, z_2 the one obtained in the second test, and similar for x_i and y_i.

The independence between events also defines the one between random variables:

Definition 2.2 (Independent Random Variables) If the random variables X_i, $(i = 1, 2, \ldots, n)$ are defined on the same probability space and the events $\{X_i \in A_i\}$, $(i = 1, 2, \ldots, n)$ are independent, according to Definition 1.10, for any possible choice of the intervals $A_i \in \mathbb{R}_x$, from Eq. (1.22), it results:

$$P\{X_1 \in A_1, X_2 \in A_2, \ldots, X_n \in A_n\} = \prod_i P\{X_i \in A_i\} \, . \tag{2.9}$$

In this case one says that variables X_i are *stochastically* independent.

In the following, we will simply state that variables are independent, but, when necessary, we will distinguish independence from a specific type of mathematical dependence, such as linear dependence (or independence), which implies the existence (or not) of linear equations between variables.

Finally, one last definition.

Definition 2.3 (Spectrum or Support) The spectrum or support is the real codomain of the random variable $X(a)$. The spectrum is called discrete when the codomain is a countable set and is called continuous when the possible values are in $\mathbb{R}$ or in subsets of $\mathbb{R}$.

A variable X with discrete spectrum is named *discrete random variable*, while a variable with a continuous spectrum is named *continuous random variable*. In mathematics the name spectrum is used in other contexts, as in the theory of transforms. For this reason, the term support is generally used in mathematical statistics. However, among physicists and engineers, during research or laboratory activities, it is quite usual to speak about "the spectrum" rather than about the range or support of the random variable being examined. There are also fields of physics, as atomic or nuclear spectroscopy, where the probabilities of discrete or continuous energy states assumed by a physical system are studied. However, since in the following the use of the term spectrum will not cause any conflict, we decided to maintain this term, beside to the support one. The law or distribution (2.6) therefore associates a probability to values (or sets of values) of the spectrum.

2.3 Cumulative or Distribution Function

The law or distribution of a random variable is usually expressed in terms of the *cumulative or distribution function*, which gives the probability that the random variable assumes values less than or equal to a certain assigned value x.

Definition 2.4 (Cumulative or Distribution Function) If X is a continuous or discrete random variable, the cumulative or distribution function:

$$F(x) = P\{X \leq x\} \tag{2.10}$$

represents the probability that X assumes a value not greater than an assigned value x. Cumulative functions will usually be indicated by uppercase letters.

Notice that x does not have to be part of the spectrum of X; for example, in the case of a die roll, where $X = 1, 2, 3, 4, 5, 6$:

$$F(3.4) = P\{X \leq 3.4\} = P\{X \leq 3\} = F(3) \, .$$

If $x_1 < x_2$, the events $\{X \leq x_1\}$ and $\{x_1 < X \leq x_2\}$ are incompatible and the total probability of the event $\{X \leq x_2\}$ is given by Eq. (1.10):

$$P\{X \leq x_2\} = P\{X \leq x_1\} + P\{x_1 < X \leq x_2\} \, ,$$

from which, according to Eq. (2.10):

$$P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1) \,. \tag{2.11}$$

Since the probability is non-negative, we have:

$$F(x_2) \geq F(x_1) \,.$$

If $x_{\max}$ and $x_{\min}$ are the maximal and minimal X values, respectively, from Eq. (1.13) and from Definition (1.9), it follows:

$$F(x) = 0 \quad \text{for } x < x_{\min} \,, \tag{2.12}$$

$$F(x) = \sum_{i=1}^{\infty} p(x_i) = 1 \quad \text{for } x \geq x_{\max} \,. \tag{2.13}$$

It also turns out, by construction, that $F(x)$ is continuous at each point x if approached from the right. This fact depends on the position of the equal sign appearing in Eq. (2.11):

$$\lim_{x_1 \to x_2} [F(x_2) - F(x_1)] = P\{X = x_2\} \,, \tag{2.14}$$

$$\lim_{x_2 \to x_1} [F(x_2) - F(x_1)] = 0 \quad \text{(continuity to the right)} \,. \tag{2.15}$$

It can be shown that any cumulative or distribution function fulfilling the properties:

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1 \,, \tag{2.16}$$

is non-decreasing and continuous to the right. Conversely, if a function satisfies these properties, it represents the cumulative function of a random variable. For mathematical details on the proof, you can see [Cra51].

We also note that from (2.10) it is easy to calculate the probability of obtaining values greater than an assigned limit x:

$$P\{X > x\} = 1 - F(x) \,. \tag{2.17}$$

The cumulative $F(x)$ allows the definition of a very useful quantity:

Definition 2.5 (Quantile) The α quantile is the smallest x_α value that obeys to the inequality:

$$P\{X \leq x_\alpha\} = F(x_\alpha) \geq \alpha \,. \tag{2.18}$$

The inequality $\geq$ in Eq. (2.18) takes into account that, for discrete variables, $F(x_\alpha)$ may not coincide with α when this value is assigned a priori. For continuous variables, $F(x_\alpha) = \alpha$ and the quantile is the value $x = F^{-1}(\alpha)$.

For example, if $P\{X \leq x_\alpha\} = F(x_\alpha) = 0.25$, it means that $x_{0.25}$ is the 0.25 quantile; if the α values are given as percentages, one says that x is the 25-th percentile or that x is between the second and third decile, or between 20% and 30%. Table B.2 of Appendix E indicates how to use R to get the quantiles of the main probability distributions. In R, the `quantile` routine estimates quantile values by interpolating from a set of random data. For example, if we generate a vector of 10 uniform random variates in [0, 1] with the instruction `x<-runif(10)`, we can estimate the 20% and 40% quantiles with the call:

```
quantile(x,c(0.20,0.40),names=FALSE) ,
```

where the variable `names` inhibits the complete output and produces a numerical vector containing the quantiles. In this case, we will see that the two output values, interpolated between the data of the ordered vector in ascending order, have 2 and 4 values to the left, respectively.

2.4 Data Representation

The most used representation to analyse data samples coming from an experiment is called *histogram*. In the case of a discrete random variable, the histogram is built as follows:

- The x axis represents the spectrum of X.
- On the y axis the *number of times* that each spectrum value appeared in the sample is recorded.

Consider an experiment formed by 100 trials, each trial consisting of tossing 10 coins. We define the event as the number X of heads in a toss; the spectrum of X is given by the 11 integers 0, 1, 2, ..., 10. Every value is labelled with the number of times that number of heads occurred in those 100 trials.

If we report in abscissa the spectrum of the event (shown also in the first column of Table 2.2), and in ordinate the results obtained in a real experiment (shown in the second column of Table 2.2), we obtain the histogram of Fig. 2.1. The number of events having x_i heads is denoted by $n(x_i)$, which is called the number n of events or trials fallen into the i-th bin of the histogram. Obviously, one has:

$$\sum_{i=1}^{C} n(x_i) = N \,, \tag{2.19}$$

where C is the number of bins of the discrete spectrum and N the total number of events of the trials made in the experiments. One can also say that the histogram represents a sample of N events. The histogram thus constructed has

Table 2.2 Results of a real experiment made of 100 observations, each consisting in the tossing of 10 coins. The second column reports the number of observations in which the number of heads reported in the first column was obtained. The third column reports the empirical frequencies obtained simply by dividing the values of the second column by 100, the fourth column reports the cumulative frequencies

Spectrum (number of heads)	Number of trials	Frequency	Cumulative
0	0	0.00	0.00
1	0	0.00	0.00
2	5	0.05	0.05
3	13	0.13	0.18
4	12	0.12	0.30
5	25	0.25	0.55
6	24	0.24	0.79
7	14	0.14	0.93
8	6	0.06	0.99
9	1	0.01	1.00
10	0	0.00	1.00

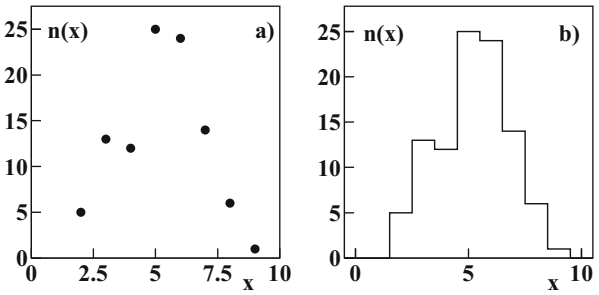


Fig. 2.1 Two ways to build a histogram of an experiment where x heads are counted after the toss of 10 coins: the values $0 \leq x \leq 10$ of the spectrum are reported in abscissa; the number of events $n(x)$, for each value x of the spectrum, is reported on the ordinate, as a point with abscissa x_i (**a**) or as a bar as wide as the distance between two spectrum values (**b**). These data, reported also in Table 2.2, refer to an experiment of 100 trials

the disadvantage of having the ordinates diverging with the number of trials, since $n(x_i) \to \infty$ for $N \to \infty$. This drawback is corrected by representing the sample as a *normalized histogram* where the number of events $n(x_i)$ is replaced by the frequency:

$$f(x_i) = \frac{n(x_i)}{N} \,. \tag{2.20}$$

The frequencies of the example of Fig. 2.1 are shown in the third column of Table 2.2. In this way, when $N \to \infty$, the contents of the bins remain finite and, based on Eq. (1.3), $f(x_i)$ tends to the probability $p(x_i)$ to fall into the i-th bin of the spectrum. Equation (2.19) then becomes:

$$\sum_{i=1}^{C} f(x_i) = 1 \,, \tag{2.21}$$

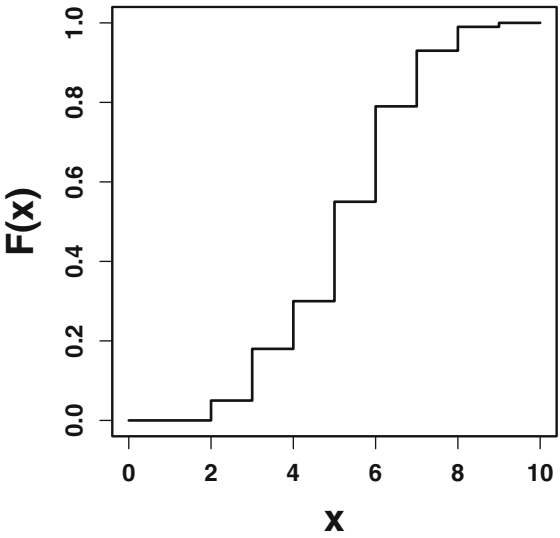


Fig. 2.2 Histogram of the cumulative frequencies of the data shown in Fig. 2.1

which is the normalization condition.

The frequencies of the histogram can also be represented as cumulative frequencies, usually denoted by capital letters: for the k-th bin they are obtained, as in Fig. 2.2, by adding its content to those of all the bins to the left and dividing by the total number of events:

$$F_k = \frac{\sum_{i=1}^{k} n_i}{N} = \sum_{i=1}^{k} f(x_i) \; . \tag{2.22}$$

The cumulative frequency $F_k \equiv F(x_k)$ gives the percentage of sample values $x \leq x_k$. It is the "experimental" estimate of the cumulative function (2.10).

Now we show how it is possible to extend the representation by histograms also to the case of continuous variables that can assume values over any range $[a, b]$ of the real number field $\mathbb{R}$. Let x be the values of the continuous spectrum under consideration, belonging to the interval $[a, b]$. We divide this interval into equal parts $[x_1, x_2)$, $[x_2, x_3)$, $\ldots$, $[x_{m-1}, x_m]$ of width Δx, and assign to each new interval (channel or bin) a value given by the the number of events having the value of the spectrum contained in that interval. If we divide for the total number of events, we obtain the analogue of Eq. (2.20):

$$f(\Delta x_k) = \frac{n(\Delta x_k)}{N} \; . \tag{2.23}$$

For $N \to \infty$, based on Eq. (1.3), $f(\Delta x_k) \to p(\Delta x_k)$, which is the probability to obtain spectrum values in $[x_{k-1}, x_k]$ that is in the k-th bin. The graphical representation of continuous variables is typically that of the histogram of Fig. 2.1b, which indicates that values are distributed within the whole bin; if you want to use the representation of Fig. 2.1a, the abscissa of the point is the central value of the bin. Histograms can be obtained using the R routine `hist(x)`, which plots the histogram of a vector `x` of raw data. The graphical style can be changed with the options listed in the online R manual. Alternatively, you can use our `HistoBar(x,fre)` routine which, in addition to the raw data, allows you also to draw histogram of data collected in two vectors: `x` containing the bin values and `fre` containing the frequencies $n(x)/N$ or the number of events $n(x)$. If `fre` and `x` have the same size, `x` is interpreted as the average value of the bin or as the spectrum of a discrete variable; if `x` has one position more than `fre`, it is interpreted as the vector of the bin breakpoints of a continuous variable.

For example, Fig. 2.2 has been obtained with the following lines:

```
fre <- c(0,0,0.05,0.18,0.30,0.55,0.79,0.93,0.99,1.)
x <- c(0,1,2,3,4,5,6,7,8,9,10)
HistoBar(x,fre,xex='x',yex='F(x)')
```

where the calling sequence is explained in the comments of `HistoBar`.

2.5 Discrete Random Variables

As we saw in Definition 2.3, a discrete random variable X takes at most a countable infinity of values $(x_1, x_2, \ldots, x_n)$. We also know, from Eq. (2.4), that the sets $\{X = x_i\}$ are events. We can then define the probabilistic analogue of frequency histograms, as follows:

Definition 2.6 (Probability Density Function for Discrete Variables) Given a discrete variable X, the function $p(x)$, given by:

$$p(x_i) = P\{X = x_i\} \tag{2.24}$$

for the discrete values of X and $p(x) = 0$ outside, is called probability density function (sometimes abbreviated as p.d.f.) or, more simply, density.

Physicists usually call a p.d.f. a distribution. Statisticians reserve the name distribution for the cumulative function. In the following we will use mainly the term cumulative function.

This density satisfies the important normalization condition:

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} P\{X = x_i\} = P\left(\bigcup_{i=1}^{\infty} \{X = x_i\}\right) = P(S) = 1\,. \tag{2.25}$$

The knowledge of the density allows the calculation of laws or statistical distributions as in Eq. (2.6):

$$P(A) = P\{X(a) \in \mathbb{R}_0\} = \sum_{x_i \in \mathbb{R}_0} P\{X = x_i\} = \sum_{x_i \in \mathbb{R}_0} p(x_i) \, . \tag{2.26}$$

The cumulative or distribution function (2.10) can then be expressed as:

$$P\{X \leq x_k\} = F(x_k) = \sum_{i=1}^{k} p(x_i) \, . \tag{2.27}$$

This function can be seen as the probabilistic correspondent of the cumulative frequency histogram (2.22). An example of density and cumulative functions is shown in Fig. 2.3. As we know, the cumulative function is defined for every x, not just for discrete values assumed by the variable X. Hence, in the case of Fig. 2.3 and Table 2.3 we have, for example:

$$F(6.4) = P\{X \leq 6.4\} = P\{X = 0, 1, 2, 3, 4, 5 \text{ or } 6\} = 0.625 \, .$$

$F(x)$ shows jumps of heights $P\{X = x_i\}$ for discrete X values and remains constant within $[x_k, x_{k+1})$. Continuity on the right is shown graphically in Fig. 2.3 with dots in bold to the left of the constant values in the ordinate. The important Eq. (2.11)

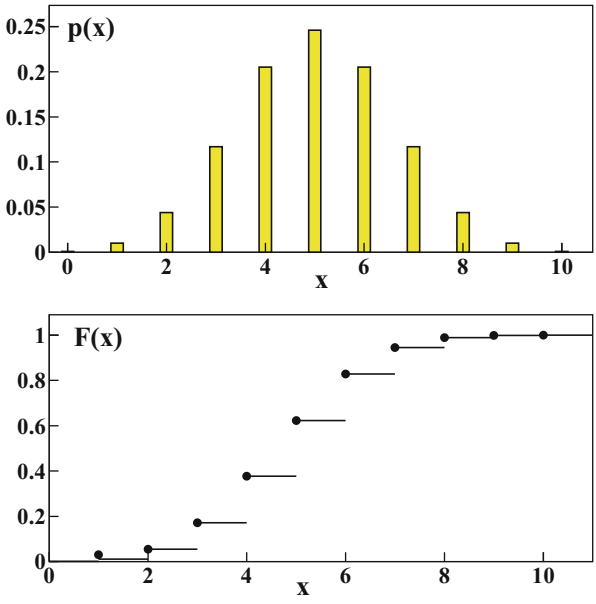


Fig. 2.3 Bar representation of the binomial probability density $p(x) = b(x; 10, 1/2)$ and corresponding cumulative function $F(x)$. This distribution is the population model for the 10 coin experiment of Table 2.2. The data are also reported in Table 2.3

can be written using the density function as follows:

$$P\{X = x_k\} = P\{x_{k-1} < X \leq x_k\} = F(x_k) - F(x_{k-1}) = p(x_k)\ , \tag{2.28}$$

which allows one to perform calculations in terms of density function or cumulative function, as it turns out easier.

There is a distribution, called binomial or Bernoulli, able to predict the results of experiments like those of Fig. 2.1 and Table 2.2. This is one of the most important results of probability theory.

2.6 Binomial Distribution

Consider an experiment consisting of n attempts and let p be the a priori probability to obtain the aimed event (success) in each attempt. We want to find the probability of the event consisting of x successes and $n - x$ failures in the considered experiment. The problem therefore requires the determination of a probability function $b(x; n, p)$ where (be careful!) n and p are *assigned parameters* and $b(x; n, p)$ is the probability of the event consisting of x successes.

Consider now a series of results consisting of x successes and $n - x$ failures, denoted by the symbols X and O, respectively:

$$\begin{array}{l} X\ X\ O\ O\ O\ X\ \ldots\ X\ O \\ X\ O\ O\ X\ O\ O\ \ldots\ O\ X \\ \ldots\ldots\ldots\ldots\ldots \\ X\ X\ O\ O\ O\ X\ \ldots\ X\ X \end{array}$$

According to the law of compound probabilities (1.24), if the events are *independent*, the probability of each configuration (row) is the same and is given by the product of the probabilities of obtaining x successes *and* $(n - x)$ failures, that is:

$$p \cdot p \cdot \ldots \cdot (1 - p) \cdot (1 - p) \cdot (1 - p) = p^x\ (1 - p)^{n-x}.$$

The possible alignments are as many as the combinations of n elements of which x and $n - x$ equal to each other. We know from combinatorial analysis that this number is given by the binomial coefficient (1.31):

$$\frac{n!}{x!(n-x)!} \equiv \binom{n}{x}\ .$$

Since an attempt realizing the requested event gives any one of the previous lines, according to the law (1.17), the probabilities of the rows must all be added up to

obtain the final probability of the event. We therefore have the final expression of the binomial density function:

$$P\{X = x \text{ favorable outcomes}\} = b(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$
$$= \binom{n}{x} p^x (1-p)^{n-x}. \tag{2.29}$$

It is important to always remember that this distribution is valid *if and only if* the n attempts are *independent* and the probability of success in an attempt is always *constant* and equal to p. The binomial distribution, when $n = 1$, is also called Bernoulli distribution, in memory of the Swiss mathematician Jacques Bernoulli, who first introduced it in the end of the seventeenth century. It has numerous applications, as the following examples show.

Exercise 2.1
Calculate the probability to obtain five successes in 10 attempts having each a 20% success probability.

Answer From Eq. (2.29) one immediately has:

$$b(5; 10, 0.2) = \frac{10!}{5!5!}(0.2)^5(0.8)^5 = 0.0264 \simeq 2.6\% \,.$$

Repeating this calculation for $0 \leq x \leq 10$ and plotting the corresponding probabilities, the representation of the binomial distribution of Fig. 2.4 is obtained. The binomial probabilities can be also obtained with the R routine `dbinom(x,n,p)` (see also Table B.2). For example, to obtain Fig. 2.4, one can write:

```
x <- c(0,1,2,3,4,5,6,7,8,9,10)
y <-  dbinom(x,10,0.2)
plot(x,y,type='p',pch='+')  # points are drawn as small +
```

Exercise 2.2
10% of the parts produced on an assembly line are defective. Calculate the probability to have 2 defective pieces over a total of 40.

Answer Since $n = 40$, $p = 0.1$ and $x = 2$:

$$b(2; 10, 0.1) = \binom{40}{2} (0.1)^2(0.9)^{38} = 0.142 \,.$$

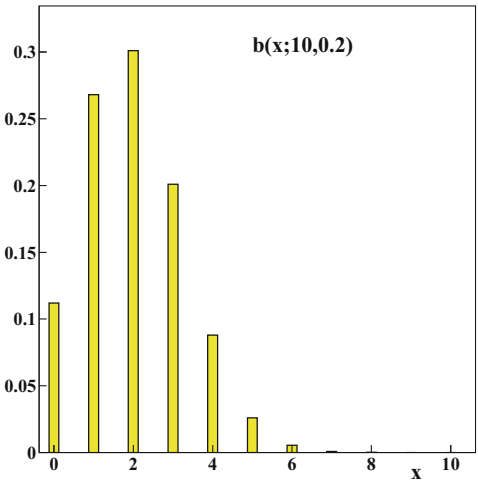


Fig. 2.4 Bar plot of the binomial distribution $b(x; 10, 0.2)$

Exercise 2.3
Assuming that the probability of having both male and female children is the same, calculate the probability of having five daughters.

Answer Since $p = (1 - p) = 1/2, n = 5, x = 5$, one has:

$$b(5; 5, 0.5) = \binom{5}{5} 2^{-5} = 3.12 \cdot 10^{-2} .$$

Exercise 2.4
The likelihood that a child will contract scarlet fever in preschool age is 35%. Calculate the probability that, in a classroom of 25 pupils, 10 pupils already had scarlet fever.

Answer Since $p = 0.35, n = 25, x = 10$, one has:

$$b(10; 25, 0.35) = \binom{25}{10} (0.35)^{10}(0.65)^{15} = 0.141 .$$

We can now compare Table 2.2 and Fig. 2.1 with the predictions of the probability theory. To do this, we just calculate the probabilities $b(x; 10, 0.5)$ for $0 \leq x \leq 10$.

Table 2.3 Results of the experiment of Table 2.2 compared with the predictions of the binomial law. The theoretical values, obtained by inserting $n = 10$ and $p = 1/2$ in Eq. (2.29), are reported in the third column, whereas the fourth and the fifth columns contain the cumulative frequencies and the probabilities calculated by inserting the probabilities of the third column in Eq. (2.27)

Spectrum (heads)	Frequency	Cumulative	Cumulative probability	Probability frequency
0	0.00	0.001	0.00	0.001
1	0.00	0.010	0.00	0.011
2	0.05	0.044	0.05	0.055
3	0.13	0.117	0.18	0.172
4	0.12	0.205	0.30	0.377
5	0.25	0.246	0.55	0.623
6	0.24	0.205	0.79	0.828
7	0.14	0.117	0.93	0.945
8	0.06	0.044	0.99	0.989
9	0.01	0.010	1.00	0.999
10	0.00	0.001	1.00	1.000

Obviously, we assume that coins are not rigged and that the probability of having head (or tail) is constant and equal to $1/2$. Results are shown in Fig. 2.3 and in Table 2.3. Notice that there are significant differences between the experimental frequencies and the theoretical probabilities. If they were only due to the limited size of the sample (100 tosses), we could be confident that, in the limit of an infinite number of tosses, we would obtain the values given by the binomial distribution. In such a situation, one says that theory and experiment are in agreement each other within the *statistical fluctuations*. If the differences were instead due to a wrong probabilistic model (which would happen in the case of correlated coin tosses, rigged coins or coins with memory, etc.), we should talk about *systematic* or non-statistical differences between theory and model. Only statistics can tell us whether the fluctuations under discussion are statistical or systematic. Without statistical notions, which are not intuitive at all, *even a truly trivial experiment like coin tossing cannot be correctly interpreted!* If you take our word for it, we can tell you that the differences between the binomial distribution predictions and the results of our 10 coin experiment are only due to statistical fluctuations and that the agreement between theory and experiment is good.

2.7 Continuous Random Variables

When a random variable can vary continuously inside a real interval, finite or infinite, the definition of the density function must be done with caution and by successive steps. The procedure is similar to that often done in physics, when, for example, one moves from a set of point electric charges within a volume to a

charge density, which gives the effective charge in a region of space only when it is integrated on the corresponding volume.

Functions or distributions of continuous random variables with points of discontinuity can be sometimes encountered. However, in practice, continuous distributions are much more frequent and we will then only analyse this type of function. Therefore, if we want to calculate the probability of a certain value of X, since $\{X = x\} \subset \{x - \varepsilon < X \leq x\}$ for all $\varepsilon > 0$, from (1.14, 2.11) we get:

$$P\{X = x\} \leq P\{x - \varepsilon < X \leq x\} = F(x) - F(x - \varepsilon)$$

for all $\varepsilon > 0$. Hence:

$$\lim_{\varepsilon \to 0} [F(x) - F(x - \varepsilon)] = 0 \,,$$

from the continuity of $F(x)$. As a consequence, *the probability of an assigned x value is always zero*. This result is intuitively compatible with the concept of classical and frequentist probability: the continuous spectrum includes an uncountable infinity of values and the probability to get in one trial *exactly* that x value on an infinite set of possible cases (a priori probability) or over an infinite number of occurrences (frequentist probability) must be zero. Therefore, for a continuous variable X, only the probabilities to fall within a finite interval are meaningful. The following equalities then hold:

$$P\{a < X < b\} = P\{a \leq X < b\} = P\{a < X \leq b\} = P\{a \leq X \leq b\} \,,$$

which also show that the inclusion of the extremes of the interval is insignificant. Consider now a continuous variable X assuming values in $[a, b]$. Let us divide this interval into sub-intervals $[a = x_1, x_2]$, $[x_2, x_3]$, ..., $[x_{n-1}, x_n = b]$ of amplitude Δx_k and define a discrete random variable X' which takes values only in the mean points x_k of these intervals and let

$$p_{X'}(\Delta x_k) \equiv p_{X'}(x'_k) = P\{x_k \leq X \leq x_{k+1}\}$$

be the density function of X', giving the probability of X to fall into the k-th interval. Since the density $p_{X'}(\Delta x_k)$ is a function *of the bin amplitude*, it depends on the arbitrary choice of Δx, because the spectrum is continuous. However, it is possible to define a function $p(x)$ as:

$$p_{X'}(\Delta x_k) = p(x_k) \Delta x_k \,, \tag{2.30}$$

where x_k is a point internal to the k-th bin (for instance, the middle point). According to the Riemann integral, we can write Eq. (2.25) as:

$$\lim_{\substack{\Delta x \to 0 \\ k \to \infty}} \sum_k p_{X'}(\Delta x_k) = \lim_{\substack{\Delta x \to 0 \\ k \to \infty}} \sum_k p(x_k) \Delta x_k = \int p(x)\, \mathrm{d}x = 1. \tag{2.31}$$

The *function* $p(x)$ so defined is the probability density function of the continuous random variable X.

Definition 2.7 (Probability Density Function (p.d.f.) for Continuous Variables) The probability density of a continuous variable $X \in \mathbb{R}$ is a function $p(x) \geq 0$ satisfying, for any x, the equation:

$$F(x) = \int_{-\infty}^{x} p(t)\, \mathrm{d}t \ . \tag{2.32}$$

The probability to obtain values into the interval $[x_k, x_{k+1}]$ of width Δx is given by (see also Fig. 2.5):

$$P\{x_k \leq X \leq x_{k+1}\} = F(x_{k+1}) - F(x_k) = \int_{x_k}^{x_{k+1}} p(x)\, \mathrm{d}x \ , \tag{2.33}$$

and the normalization property (2.25) in this case is written as:

$$\int_{-\infty}^{+\infty} p(x)\, \mathrm{d}x = 1 \ . \tag{2.34}$$

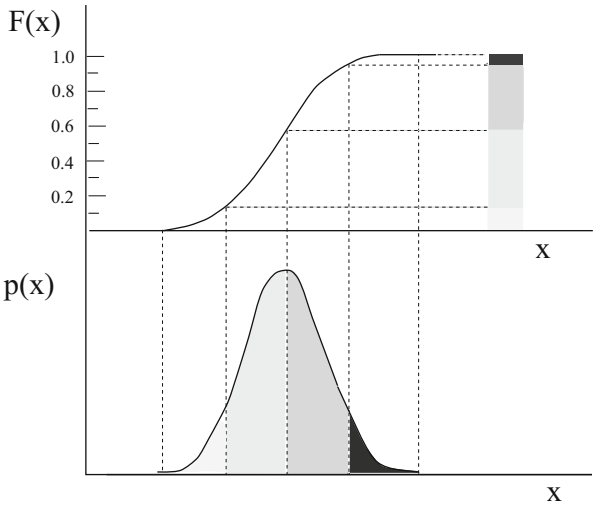


Fig. 2.5 Probability density function (lower plot) with its cumulative distribution function (upper plot) for continuous random variables. The shading shows that the relation between the areas of the density function and the increments of the ordinates of the cumulative function

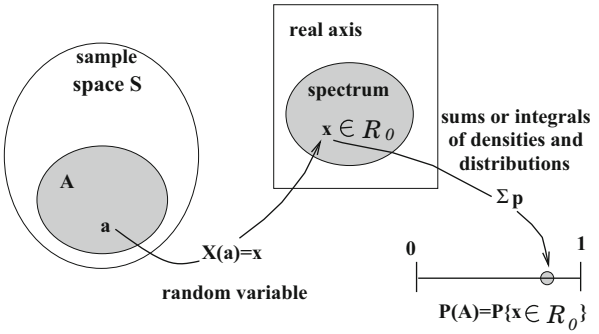


Fig. 2.6 Assignment of a probability $P(A)$ to an event A of the sample space S. The random variable $X(a) = x$ transforms the experimental result $a \in A$ to a codomain of the real axis $x \in \mathbb{R}_0$, called spectrum or support. The density and cumulative functions allow the calculation of $P(A)$ starting from the spectrum values

If the interval is small compared to the range of the function variations, we can approximate $p(x)$ with a straight line within each Δx (linear approximation) and Eq. (2.33) is replaced by Eq. (2.30), where x_k is the bin midpoint.

We can therefore symbolically indicate the transition from the discrete spectrum to the continuous one as:

$$\sum_k p(x_k) \;\rightarrow\; \int p(x)\,\mathrm{d}x\;, \qquad p(x_k) \;\rightarrow\; p(x_k)\,\mathrm{d}x\;,$$

from which we see that we move from a function with discrete values to the product of a continuous-valued function and a differential. The differential quantity $p(x)\,\mathrm{d}x$ gives the probability to obtain X values within $[x, x + dx]$.

In the points where $F(x)$ is differentiable, from Eq. (2.32) one also obtains the important relation:

$$p(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}\;. \tag{2.35}$$

Density and cumulative functions have as domain the spectrum of the random variable and as codomain a set of real values. They can be put in correspondence with laws or distributions of Eq. (2.6), as shown in Fig. 2.6.

2.8 Mean, Sum of Squares, Variance, Standard Deviation and Quantiles

The description of a random phenomenon in terms of mean and variance is less complete than the characterization of its density function, but it has the advantage of being simpler and often adequate enough for many practical applications. Basically, the mean identifies where the centre of gravity of values of the X variable is localized, while the standard deviation, which is the square root of the variance, gives an estimate of the dispersion (spread) of values around the mean.

To help your intuition, we first consider mean and variance of a set of data (although this topic will be explored further on, in statistics) and then mean and variance of distributions. We start by introducing a notation that will accompany us throughout the text.

Definition 2.8 (True and Sample Parameters) Mean, variance and standard deviation of a random variable X will be indicated in Greek letters; sample and experimental parameters, coming from finite experimental samples, will be indicated in Latin letters. The probability is an exception, because its true value will be always indicated with p, whereas for the relative frequency we will use the f symbol, which is used by most statistics book and avoids conflicts of notation.

We now define mean and variance of a set of experimental data:

Definition 2.9 (Mean, Sum of Squares and Variance of a Data Sample) If x_i $(i = 1, 2 \dots N)$ are N occurrences of a random variable, mean, sum of squares (SS) and variance are defined as:

$$m = \frac{1}{N} \sum_{i=1}^{N} x_i \ , \tag{2.36}$$

$$SS = \sum_{i=1}^{N} (x_i - \mu)^2 \ , \tag{2.37}$$

$$s_\mu^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2 \ , \tag{2.38}$$

where μ is the true mean assigned *a priori*. When the sample mean m of Eq. (2.36) is considered, the variance is given by:

$$s_m^2 = \frac{\sum_{i=1}^{N} (x_i - m)^2}{(N-1)} = \frac{N}{N-1} \frac{\sum_{i=1}^{N} (x_i - m)^2}{N} \ . \tag{2.39}$$

You may have noticed that in the denominator of the variance about the sample mean the term $N - 1$ appears instead of N: the reason, conceptually non-trivial, is statistical by nature and will be explained later on, in Chap. 6. From a practical point of view, the $N/(N-1)$ factor is relevant for very small samples only. When this difference is neglected, we will write $s^2 \equiv s_m^2 \simeq s_\mu^2$.

Later on, we will use the following property of the sum of squares:

$$\begin{aligned} SS &= \sum_i (x_i - \mu)^2 = \sum_i (x_i - m + m - \mu)^2 \\ &= \sum_i (x_i - m)^2 + N(m - \mu)^2 + 2\sum_i (x_i - m)(m - \mu) \\ &= \sum_i (x_i - m)^2 + N(m - \mu)^2 \equiv SS_r + SS_s \ , \end{aligned} \tag{2.40}$$

since: $\sum_i (x_i - m) = \sum_i x_i - m \sum_i = \sum_i x_i - Nm = 0$. Therefore, the total sum of squares is the sum over the sample SS_r (sometimes called residual sum of squares SS_r or RSS) and of N times the squared deviation of the sample mean from the true one (sometimes called explained sum of squares SS_s or ESS).

The square root $s = \sqrt{s^2}$ is the sample standard deviation or root mean square. This operation is needed to measure dispersion with the same units of the mean. Also variance can be considered as the average of the squared deviations $(x_i - \mu)^2$.

When the values of a discrete random variable X are collected in histograms, mean and variance can be calculated as:

$$m = \frac{\sum_{k=1}^{C} n_k x_k}{N} = \sum_{k=1}^{C} x_k f_k \ , \tag{2.41}$$

$$s^2 = \frac{\sum_{k=1}^{C} n_k (x_k - \mu)^2}{N} = \sum_{k=1}^{C} (x_k - \mu)^2 f_k \ . \tag{2.42}$$

where C is the number of histogram bins and n_k, f_k and e x_k are the bin content, relative bin frequency and bin midpoint, respectively.

In Eqs. (2.36–2.38) the sum is the overall raw data, for example, $(3 + 3 + 4 + 2 + 4 + \dots)$, whereas in Eqs. (2.41, 2.42) the same sum is evaluated in the compact way $(2 + 2 \cdot 3 + 2 \cdot 4 + \dots)$. Therefore, the two estimates give exactly the same result for a discrete spectrum. Instead, when spectrum is continuous, Eqs. (2.41–2.42) can still be used by assigning the bin midpoints to x_k; in this case the values of the continuous distribution are approximated in a discrete way.

We now define mean, variance and standard deviation when we know a priori the probability density of a variable X. The formulae, for a discrete random variable, are nothing more than the generalization of (2.41, 2.42), where the measured frequencies f_k are replaced by the a priori probabilities p_k of the density function. For a continuous random variable, the formulae are derived with the same limit operation of Eq. (2.31). We therefore obtain the following:

Definition 2.10 (Mean and Variance of a Random Variable) The mean (or expected value) μ and the variance σ^2 of a random variable X are given by:

$$\begin{aligned}\mu &= \sum_{k=1}^{\infty} x_k p_k \quad \text{(for discrete variables)}\,, \\ &= \int_{-\infty}^{+\infty} x p(x)\,\mathrm{d}x \quad \text{(for continuous variables)}\,, \end{aligned} \tag{2.43}$$

$$\begin{aligned}\sigma^2 &= \sum_{k=1}^{\infty} (x_k - \mu)^2 p_k \quad \text{(for discrete variables)}\,, \\ &= \int_{-\infty}^{+\infty} (x-\mu)^2 p(x)\,\mathrm{d}x \quad \text{(for continuous variables)}\,. \end{aligned} \tag{2.44}$$

As in the sample case, the standard deviation is given by $\sigma = \sqrt{\sigma^2}$.

The definition (2.43) is assumed to be valid only if there exists the limit of the series or integral of absolute values. Since $p_k,\ p(x) \geq 0$, this condition is equivalent to write:

$$\sum_{k=1}^{\infty} |x_k|\, p_k < \infty\,, \qquad \int_{-\infty}^{+\infty} |x|\, p(x)\,\mathrm{d}x < \infty\,. \tag{2.45}$$

Equations (2.45) obviously imply the convergence of Eqs. (2.43). Here, also the inverse property is requested because in probability theory one always assumes that the order of summation of terms in infinite series is indifferent and equalities as $\sum_k x_k f(x_k) \sum_i y_i g(y_i) = \sum_{ik} x_k y_i f(x_k) g(y_i)$ are verified. These properties then require the absolute convergence of series, and the same holds also for integrals. In the following, the existence of mean values will always imply also the existence of absolute mean values.

As for the variance, the definition is obviously valid if the series or the integrals (2.44) (which have always positive values) are not divergent. There are, however, some cases of random variables with an undefined variance; in this situation, variance cannot characterize the data dispersion and the density or cumulative functions must be used.

It can be shown that the sequence of moments (see Appendix C):

$$\sum_i (x_i)^k p_i\,, \quad \int x^k p(x)\,\mathrm{d}x \qquad (k = 1, 2 \ldots \infty)$$

allows the unique determination of the distribution function of a generic random variable X [Cra51]. The $k = 1$ moment is the mean, and hence, once the other

moments are known, also the *central* moments about the mean can be found:

$$\Delta_k = \sum_i (x_i - \mu)^k p_i \ , \quad \int (x - \mu)^k p(x)\, \mathrm{d}x \qquad (k = 1, 2 \ldots \infty) \ . \tag{2.46}$$

Notice that $\Delta_2 \equiv \sigma^2$ and that the first moment Δ_1 is always zero:

$$\Delta_1 = \sum_i (x_i - \mu) p_i = \sum_i x_i p_i - \mu = 0 \ .$$

The first two more significant moments are just the mean and the variance. As we will see, they are sufficient to study univocally or, at least, in an acceptable way the statistical distributions describing almost all of the practical applications commonly considered. In the following, we will rarely use moments beyond the second order.

The mean is perhaps the most effective parameter for evaluating the centre of a distribution, since *the second order moment, if calculated about the mean, is minimal*. Indeed:

$$\frac{\mathrm{d}\Delta_2}{\mathrm{d}\mu} = \frac{\mathrm{d}}{\mathrm{d}\mu} \sum_k (x_k - \mu)^2 p_k = -2 \left[\sum_k x_k p_k - \mu \right] \ , \tag{2.47}$$

and this derivative is zero only when μ is the mean (2.43), so that $\Delta_2 \equiv \sigma^2$.

If one calculates the squares in Eqs. (2.39, 2.44), the variance can be expressed as *the difference between the mean of squares and the square of the mean*:

$$s_\mu^2 = \sum_{k=1}^{C} (f_k x_k^2) - \mu^2 \ , \tag{2.48}$$

$$s_m^2 = \frac{N}{N-1} \left[\frac{\sum_{i=1}^{N} x_i^2}{N} - m^2 \right] \tag{2.49}$$

$$= \frac{N}{N-1} \left[\sum_{k=1}^{C} (f_k x_k^2) - m^2 \right] , \tag{2.50}$$

$$\sigma^2 = \sum_{k=1}^{C} (p_k x_k^2) - \mu^2 \to \int_{-\infty}^{+\infty} x^2 p(x)\, \mathrm{d}x - \mu^2 \ . \tag{2.51}$$

These equations are sometimes useful in practical calculations, as shown in Exercise 2.6.

The moments of a density function, including variance, are independent of the position of the mean, that is, invariants under translations along the x axis. This property is obvious (but important), since the intrinsic width of a function cannot depend on its position along the axis of abscissas. This can be demonstrated in a

formal way, by defining a generic translation:

$$x' = x + a \ , \quad \mu' = \mu + a \ ,$$

$$\mathrm{d}x' = \mathrm{d}x \ , \quad p(x) \to p(x' - a) \equiv p'(x') \ ,$$

and by verifying that:

$$\begin{aligned} \Delta_n(x) &= \int (x - \mu)^n \, p(x) \, \mathrm{d}x = \int (x' - \mu')^n \, p(x' - a) \, \mathrm{d}x' \\ &= \int (x' - \mu')^n \, p'(x') \, \mathrm{d}x' \equiv \Delta_n(x') \ . \end{aligned} \tag{2.52}$$

In summary, mean, variance and moments of Eq. (2.46) for histograms of experimental samples or for random variables are given by:

$$m = \sum_{k=1}^{C} f_k x_k \ , \tag{2.53}$$

$$\mu = \sum_{k=1}^{\infty} p_k x_k \to \int_{-\infty}^{+\infty} x p(x) \, \mathrm{d}x \ , \tag{2.54}$$

$$s_\mu^2 = \sum_{k=1}^{C} f_k (x_k - \mu)^2 \ , \tag{2.55}$$

$$s_m^2 = \frac{N}{N-1} \sum_{k=1}^{C} f_k (x_k - m)^2 \ , \tag{2.56}$$

$$\sigma^2 = \sum_{k=1}^{\infty} p_k (x_k - \mu)^2 \to \int_{-\infty}^{+\infty} (x - \mu)^2 p(x) \, \mathrm{d}x \ , \tag{2.57}$$

$$D_n = \sum_{k=1}^{C} f_k (x_k - \mu)^n \ , \tag{2.58}$$

$$\Delta_n = \sum_{k=1}^{\infty} p_k (x_k - \mu)^n \to \int_{-\infty}^{+\infty} (x - \mu)^n p(x) \, \mathrm{d}x \ , \tag{2.59}$$

where arrows denote the transition from a discrete to a continuous variable. In R, the calculation of mean and variance from a set `x` of raw data according to Eqs. (2.36, 2.39) can be done with the functions `mean(x)` and `var(x)`. If, instead, data are in the form of histograms, the calculation can be performed with Eqs. (2.54, 2.57) and our routines `MeanHisto(x,fre)` and

`VarHisto(x,fre)`, where `x` is the spectrum of the considered random variable and `fre` is the corresponding vector of frequencies or number of occurrences.

In data analysis, the so-called Q-Q, quantile-quantile, plot (see Definition 2.5) is often used to compare two probability distributions. If x_i and y_i are the elements of two vectors containing the values of two random variables X and Y sorted in ascending order, the position of these elements, divided by the length of the vector, is close to the theoretical quantile of the parent distribution function. For example, in a vector of 100 sorted variates, $x(40)$ will be the quantile value close to 0.4, because it has 40 out of 100 values to its left, and so on. Using R, we can generate two random vectors from the uniform distribution (defined later; see Eq. 3.79), which returns values between 0 and 1 with constant probability, through the commands: `x<-sort(runif(100))` and `y<-sort(runif(100))`, which sort the values of the two vectors in ascending order. If we now represent, on the two cartesian axes x and y, the points corresponding to the pairs of the values of these two vectors *having the same position*, i.e. $(x_1, y_1), (x_2, y_2), \ldots$, we obtain the Q-Q plot for the two samples shown in Fig. 2.7. This plot has been generated with the simple R command `qqplot(x,y)`, followed by the inline commands `grid()` and `abline(0,1)`. As clearly shown by this figure, data tend to cluster in the plane around the diagonal $y = x$. Indeed the Q-Q plot is used to check whether data from two samples come from the same parent population: the more the homogeneity hypothesis is true, the closer the quantiles of the two variables are and, hence, the more they cluster around the diagonal. It is also possible to compare data sample with a theoretical population: just place these data, sorted in ascending order on an axis, calculate the sample quantile probabilities and place the corresponding theoretical quantile values on the other axis. In our example with uniform variates, this can be performed by sorting `x` with the command `x<-sort(x)`, by calculating the probabilities and the theoretical quantiles of the experimental sample with `pth<-seq(0,1,by=1/length(x))` and `qth<-qunif(pth)`

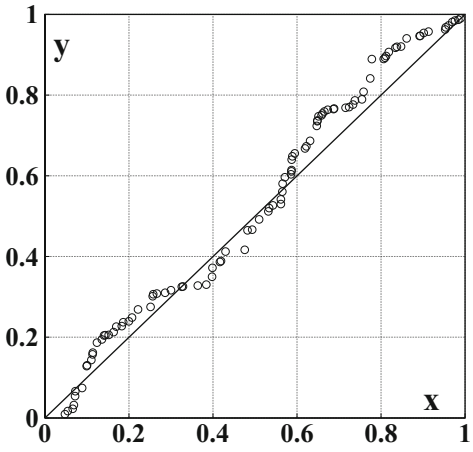


Fig. 2.7 Q-Q plot of quantiles between two uniform variates `x` and `y` generated with R. The continuous line is the straight line $y = x$

(note that `pth=qth` for uniform variates), and, finally, by plotting the result with `qqplot(qth,x)`. You will see that these pairs of data cluster around the diagonal.

In R, the `qqnorm` routine generates the Q-Q plot between a sample and the normal distribution, as we will see shortly. If random variables are discrete, the ordering no longer ensures the correct determination of the quantile, due to repeated values. This situation is discussed in Problem 2.11

2.9 Operators

As we have seen, the formulae for the calculation of mean and variance assume different forms, depending whether we have finite or infinite datasets, discrete or continuous random variables, raw data or histograms.

For this reason, it is convenient to consider mean and variance not just as numbers but also as operators on random variables or sets of data, in which the type of operation that is being carried out is given regardless of the particular representation used for data or variables.

In the following, we will indicate in italics and lowercased letters:

$$m_x, \quad m(x), \quad \langle x \rangle, \quad s_x^2, \quad s^2(x), \quad s_x, \quad s(x),$$

means, variances and standard deviations of variates of X obtained in a particular trial or sampling. Notice the symbol $\langle x \rangle$ for the mean, a very common notation among physicists and engineers.

The operators $\langle \ldots \rangle$ or $E[\ldots]$ refer to the mean, whereas Var[...] indicates the variance. The random variables on which the operators act will be always indicated in capital letters:

$$\langle X \rangle, \quad E[X], \quad \text{Var}[X]. \tag{2.60}$$

We will use mainly the notation $\langle X \rangle$, Var$[X]$. The standard deviation in operator form will be indicated as $\sqrt{\text{Var}[X]} \equiv \sigma[X]$.

At this point, we see from Table 2.1 that the functions that operate on random variables are probabilities and operators: *the random variable is therefore always enclosed in curly or square brackets*. We will use this convention throughout the rest of the book.

To ease the notation, we will sometimes indicate mean, variance and standard deviation as:

$$\langle X \rangle = \mu_x = \mu, \quad \text{Var}[X] = \sigma_x^2 = \sigma^2, \quad \sigma[X] = \sigma_x = \sigma \tag{2.61}$$

(note the subscripts written in lowercase letters). The writing μ and σ^2 (or μ_x and σ_x^2 if you need to specify the variable type) is intended as the numerical result of the correspondent statistical operator. Anyhow, the notation with Greek letters

(true values) uniquely defines the type of operation performed, avoiding possible confusions.

It is easy to verify, from Eqs. (2.53–2.57), that mean and variance operators have the following properties:

$$\mathrm{Var}[X] = \left\langle (X - \langle X \rangle)^2 \right\rangle , \tag{2.62}$$

$$\langle \alpha X \rangle = \alpha \langle X \rangle , \tag{2.63}$$

$$\mathrm{Var}[\alpha X] = \alpha^2 \mathrm{Var}[X] , \tag{2.64}$$

$$\langle X + \alpha \rangle = \langle X \rangle + \alpha , \tag{2.65}$$

$$\begin{aligned} \mathrm{Var}[X + \alpha] &= \left\langle (X + \alpha - \langle X + \alpha \rangle)^2 \right\rangle \\ &= \left\langle (X + \alpha - \langle X \rangle - \alpha)^2 \right\rangle = \mathrm{Var}[X] , \end{aligned} \tag{2.66}$$

where α is a constant. The last two equations show, of course, that the average of one constant coincides with the constant itself and that there is no dispersion for a constant. Using Eqs. (2.62, 2.63), we can rewrite in operatorial notation Eq. (2.51), which defines variance as the mean of squares minus the square of the mean:

$$\begin{aligned} \mathrm{Var}[X] &= \left\langle (X - \mu)^2 \right\rangle \\ &= \left\langle (X - \langle X \rangle)^2 \right\rangle \\ &= \left\langle X^2 \right\rangle - 2 \langle X \rangle \langle X \rangle + \langle X \rangle^2 \\ &= \left\langle X^2 \right\rangle - \langle X \rangle^2 . \end{aligned} \tag{2.67}$$

The mean operator allows the definition of the true mean value of any function of random variable (2.8):

Definition 2.11 (Expected Value) If $f(X)$ is a function of a random variable X with p.d.f. $p(x)$, the expected value of $f(X)$ is the quantity:

$$\langle f(X) \rangle \equiv E[f(X)] = \sum_k f(x_k)\, p(x_k) \to \int f(x)\, p(x)\, \mathrm{d}x , \tag{2.68}$$

where the arrow indicates the transition from a discrete to a continuous variable.

Functions of random variables will be treated in detail later, in Chap. 5. According to this definition, the true mean μ can be considered as the expected value of X. The expected value is also known as the expectation, mathematical expectation, mean, average or first moment.

As in the case of Definition 2.10, the existence of the sum $\sum_k |f(x_k)|p(x_k)$ or of the integral $\int |f(x)|p(x)\,dx$ is implied in the definition of the expected value of $f(X)$. Therefore, one always assumes that a variable or function of variable, for which one defines the expected value (2.68), is *absolutely summable or integrable* on the corresponding p.d.f. This implies, for a continuous variable, that the probability density tends to zero for $x \to \pm\infty$ at least like $1/|x|^\alpha$ with $\alpha > 2$. All densities which we will consider later have this property.

Exercise 2.5
Consider the space $\mathcal{E}$ where the event (A_1) can occur as x_1 with probability p_1 and the event (A_2), independent of the previous one, can occur as x_2 with probability p_2. Find the mean of the sum $(X_1 + X_2)$ where X_1 and X_2 are dummy variables defined as $X_1(A_1) = x_1$, $X_2(A_2) = x_2$ and $X_1(\bar{A}_1) = 0$, $X_2(\bar{A}_2) = 0$.

Answer The spectrum of the sum is given by the four values $(0+0)$, $(0+x_2)$, (x_1+0), (x_1+x_2), having, from Theorem (1.21), probabilities: $(1-p_1)(1-p_2)$, $(1-p_1)p_2$, $p_1(1-p_2)$ and p_1p_2, respectively. From Eq. (2.54) one has:

$$\begin{aligned}\mu &= (1-p_1)(1-p_2)(0+0) + (1-p_1)p_2(0+x_2) \\ &\quad + p_1(1-p_2)(x_1+0) + p_1p_2(x_1+x_2) \\ &= p_1x_1 + p_2x_2\ .\end{aligned}$$

Since $\langle X_1\rangle = (1-p_1)\cdot 0 + p_1x_1 = p_1x_1$, and the same holds for X_2, one can write:

$$\langle X_1 + X_2\rangle = \langle X_1\rangle + \langle X_2\rangle\ ,$$

that is, the mean of a sum is equal to the sum of the means.

Exercise 2.6
Find mean and standard deviation of the data shown in Table 2.2, and compare them with the values given by the binomial distribution.

Answer By applying Eq. (2.53) to the data of the first and third columns of the table, we obtain:

$$\langle x\rangle = m = (2\cdot 0.05 + 3\cdot 0.13 + \cdots + 9\cdot 0.01) = 5.21\ ,$$

(continued)

Exercise 2.6 (continued)
for a total of 521 heads over 1000 tosses. The mean of squares is given by:

$$\left\langle x^2 \right\rangle = (4 \cdot 0.05 + 9 \cdot 0.13 + \cdots + 81 \cdot 0.01) = 29.7 \, .$$

From Eq. (2.50) we then have

$$s^2 = \frac{N}{N-1} \left(\left\langle x^2 \right\rangle - \langle x \rangle^2 \right) = \frac{100}{99} (29.7 - 5.21^2) = 2.48 \, ,$$
$$s = \sqrt{2.48} = 1.57 \, .$$

By applying the same procedure to the true probabilities of the fourth column, we get:

$$\langle X \rangle = 5.00, \quad \left\langle X^2 \right\rangle = 27.5.$$
$$\mathrm{Var}[X] = \sigma^2 = \left\langle X^2 \right\rangle - \langle X \rangle^2 = (27.5 - 25) = 2.5 \, ,$$
$$\sigma = \sqrt{2.5} = 1.58 \, .$$

Notice the absence of the 100/99 factor in the calculation, since here the variance is evaluated with respect to the true mean. Again, the difference between the experimental values $m = 5.21, s = 1.57$ and the theoretical ones $\mu = 5.00, \sigma = 1.58$, will be explained in Chap. 6.

2.10 Simple Random Sample

Consider an experiment involving a random variable X and N independent observations. The result of this operation is an N-tuple of independent variables $(X_1, X_2, \ldots, X_N)$, for which condition (2.9) applies. We then arrive at the following definition.

Definition 2.12 (Simple Random Sample) The set of N independent variables $(X_1, X_2, \ldots, X_N)$ coming from the same probability density function $p(x)$ is called simple random sample of size N, extracted from the parent population of density $p(x)$. The variables are called independent and identically distributed and sometimes designated with the *iid* acronym.

The correct, but somewhat long, definition of "population of density $p(x)$" is sometimes abbreviated to "population $p(x)$". The concept of population, introduced on an intuitive basis in Sect. 1.2, here assumes a precise meaning. For a discrete random variable, the probability to obtain a certain set of random variates is,

from Eq. (2.9):

$$P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_N\} = \prod_{i=1}^{N} P\{X_i = x_i\} = \prod_{i=1}^{N} p(x_i)\,. \tag{2.69}$$

To better understand the concept of random sample, we must refer to a situation where a random variable X is sampled repeatedly, according to the following scheme:

$$\begin{array}{lll} \text{first sample} & x_1', x_2', \ldots, x_N' & \to m_1, s_1^2, \\ \text{second sample} & x_1'', x_2'', \ldots, x_N'' & \to m_2, s_2^2, \\ \text{third sample} & x_1''', x_2''', \ldots, x_N''' & \to m_3, s_3^2, \\ \ldots & \ldots & \\ \text{all samples} & X_1, X_2, \ldots, X_N & \to M, S^2; \end{array} \tag{2.70}$$

X_1 is the random variable "occurrence of the first trial" or "first element", whereas x_1'' is the random variate resulting from the first trial in the second sample. The values m_i ed s_i^2 are the mean and variance of the i-th sample:

$$m = \sum_k \frac{x_k}{N}\,, \qquad s^2 = \sum_k \frac{(x_k - \mu)^2}{N}\,.$$

These values, which estimate the corresponding true quantities μ and σ^2 from a sample of finite size, are called *estimates* of mean and variance. If we repeat the experiment or sampling, we will get different means and variances: the sample values m and s therefore have to be considered as realizations or variates of the random variables M and S, indicated in the last line of Eq. (2.70), which are functions, in the sense of (2.8), of the variables X_i:

$$M = \sum_i \frac{X_i}{N}\,, \qquad S^2 = \sum_i \frac{(X_i - \mu)^2}{N}\,. \tag{2.71}$$

The variables M and S^2 are sample functions. In general, any of these quantities is called a *statistic*.

Definition 2.13 (Statistic) For a given sample $(X_1, X_2, \ldots, X_N)$ from a density $p(x)$, any function

$$T = t(X_1, X_2, \ldots, X_N) \tag{2.72}$$

which does not contain any unknown parameter, is a random variable called statistic (singular).

Sample mean and variance are two examples of statistic. As we will see in the next section, they are also *estimators* of the mean and of the variance.

2.11 Convergence Criteria

At this point, it is necessary to well specify the meaning of the limits and of the convergence criteria used in the study of random variables.

As we have mentioned before, the frequentist limit of Eq. (1.3) is applied to the realizations of the random variable: it does not have a precise mathematical meaning and indicates *that trials must be repeated an infinite or finite number of times, until the population is used up*. This is the meaning to be attributed to a limit whenever it is applied to a sequence or sum of values of a variable (lowercase notation). We called this operation *frequentist limit*. However, expressions such as:

$$\lim_{N\to\infty} \frac{1}{N} \sum_k x_k n_k = \lim_{N\to\infty} \sum_k n_k f_k = \sum_k x_k p_k = \mu \quad \text{(wrong!)}\,,$$

should be avoided, because precise mathematical quantities are present on the right side, whereas we have an undefined limit on the left. We point out, however, that this limit is justified in the frequentist interpretation (1.3) and that it reproduces *qualitatively* what actually happens in many observations. For example, if we consider the toss of a dice where $\{X = 1, 2, 3, 4, 5, 6\}$ and we assign to each face a probability equal to $1/6$, according to Eq. (2.43) $\mu = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$. Experience shows that, when rolling a die and averaging progressively the scores according to Eq. (2.36), the result tends to 3.5 as long as the number of rolls is increased. For instance, after having arbitrarily assigned a probability of $1/6$ to each face, we simulated a dice roll using a computer and obtained, for $N = 100,\ 1000,\ 100{,}000,\ 1{,}000{,}000$, the sample means $m = 3.46,\ 3.505,\ 3.50327,\ 3.500184$, respectively.

For a mathematically rigorous study of the random phenomena, it is however necessary to establish, as the sample size increases, the type of convergence that might occur in sequences of random variables, and the extent of the deviations from the limit value.

A first rigorous definition states that a succession of variables X_N converges to a variable X if, given any $\epsilon > 0$ however small, the probability that X_N differs from X by a quantity $> \epsilon$ tends to zero for $N \to \infty$:

$$\lim_{N\to\infty} P\left\{|X_N(a) - X(a)| \geq \epsilon\right\} = 0\,, \quad \text{or} \quad \lim_{N\to\infty} P\left\{|X_N(a) - X(a)| \leq \epsilon\right\} = 1 \tag{2.73}$$

This limit, which fulfils the usual properties of the limits of real number sequences, is called *limit in probability or weak convergence*.

At this point, we draw your attention to a subtle distinction: the limit of Eq. (2.73) does not ensure that all the values $|X_N(a) - X(a)|$, for each element a of the sample space, will be less than ϵ above a certain N, but only that the set of values exceeding ϵ has a vanishing probability to exist.

If one requires that for the most part of the *sequences* the condition $|X_N(a) - X(a)| \leq \epsilon$ holds for any a, the *almost sure or strong convergence* on the set of elements a of the probability space $(S, \mathcal{F}, P)$ must be introduced:

$$P\{\lim_{N\to\infty} |X_N(a) - X(a)| \leq \epsilon\} = 1 \,. \tag{2.74}$$

In this case, we are sure that there is a set of elements a, converging to the sample space S for $N \to \infty$, such that the sequence of real numbers $X_N(a)$ tends to the standard mathematically defined limit. The convergence in probability and the almost sure convergence are graphically represented in Fig. 2.8.

When X_N is the sample mean estimator and $X(a) = \mu$, Eqs. (2.73) and (2.74) are called weak and strong law of large numbers, respectively.

The last type of convergence we consider is the convergence in law or distribution. A sequence of random variables X_N converges in distribution to a variable X if:

$$\lim_{N\to\infty} F_{X_N}(x) = F_X(x) \,, \tag{2.75}$$

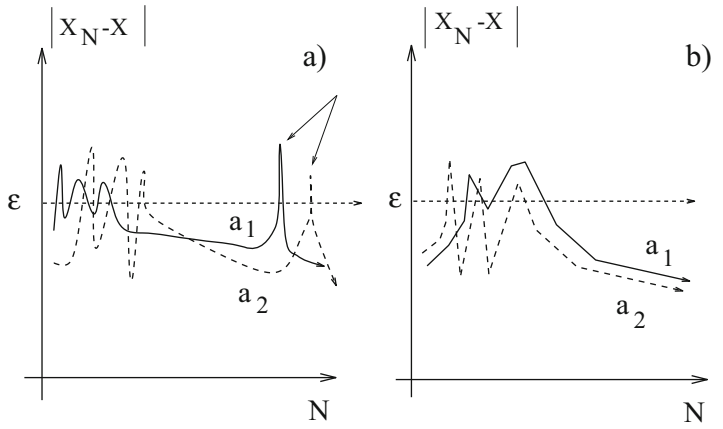


Fig. 2.8 (**a**) Convergence in probability: the set of values $|X_N(a) - X(a)|$ higher than an assigned value ϵ tends to have vanishing probability for $N \to \infty$. However, this does not prevent to have some points outside the limit (denoted by arrows). (**b**) Almost sure or strong convergence: most of sequences satisfies the inequality $|X_N(a) - X(a)| \leq \epsilon$ except for a set of elements $a \in S$ having null probability

for any point x where $F_X(x)$ is continuous; here F_{X_N} and F_X are the corresponding distribution or cumulative functions. This limit is widely used in statistics when studying the type of distribution followed by statistical estimators. It can be demonstrated (as, indeed, it is quite intuitive) that the almost sure convergence implies the convergence both in probability and in distribution and that convergence in probability implies the one in law, while the opposite is not true. Furthermore, a fundamental theorem of Kolmogorov, whose proof can be found in [Fel47], guarantees the almost sure convergence of sequences of independent random variables if the condition:

$$\sum_N \frac{\mathrm{Var}[X_N]}{N^2} < +\infty \tag{2.76}$$

holds. All statistical estimators considered in this book satisfy this property, so that, when dealing with statistics, we will not have convergence problems: *all the considered variables will converge almost surely (and, therefore, also in probability) to the requested limit values*. The convergence criteria are sometimes important in the theory of stochastic functions [PUP02], which will not be considered here.

In summary, we have described four different limits: the frequentist limit (acting on the "lowercase" variables), the one in probability, and the almost sure one (both acting on "uppercase" variables) and the one in law, which involves distribution functions.

The random variables considered as limits in Eqs. (2.73, 2.74) can also simply be constants, as often happens for the limits of statistical estimators. In fact, a statistic $T_N(X)$, function of a random sample of size N according to Eq. (2.72), and converging in probability to a constant:

$$\lim_{N\to\infty} P\left\{|T_N(X) - \mu| > \epsilon\right\} = 0\,, \tag{2.77}$$

is defined as a consistent estimator of μ.

The theory of statistical estimators, such as those defined in Eq. (2.71), will be described in detail later, in Chaps. 6 and 10. However, here we want to explain the meaning of mean and variance of an estimator. Since from (2.67) it results that the variance can be written in terms of mean operators, it will be enough to discuss only the mean of an estimator. For example, let us examine the meaning of the mean of the variance:

$$\langle T_N(X)\rangle \equiv \left\langle \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2 \right\rangle\,.$$

What is in $\langle\,\rangle$ brackets is a random variable composed by N observations x_i of X, combined to calculate their variance. The $\langle\,\rangle$ parenthesis indicates that one has to repeat the procedure an infinite amount of times and to take the average of the infinite variances thus obtained. Therefore, in the frequentist view, *at first* a sample

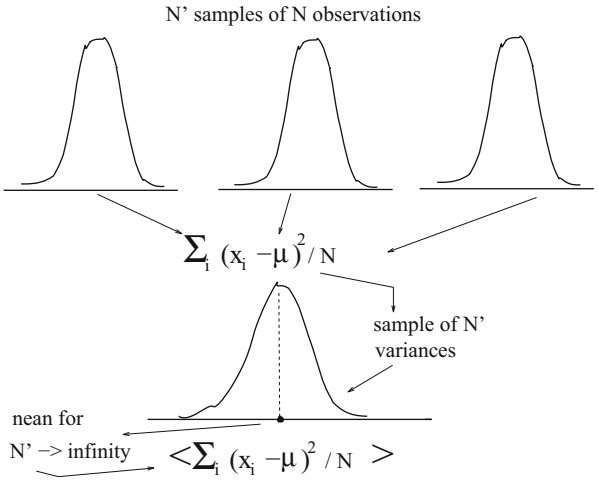


Fig. 2.9 The mean of a variance

of the variable $\sum_i (X_i - \mu)^2/N$ is obtained from a series of N' samples from N observations of the same variable X, and *then* this last sample is averaged by $N' \to \infty$. The procedure is shown in Fig. 2.9. From the properties (2.67) of the mean operator, one obtains:

$$\left\langle \sum_{i=1}^{N} (X_i - \mu)^2 \right\rangle = \sum_{i=1}^{N} \left\langle (X_i - \mu)^2 \right\rangle = N\sigma^2 \,, \tag{2.78}$$

that is:

$$\left\langle \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2 \right\rangle = \sigma^2 \,.$$

This last equation can be also written as:

$$\langle T_N(X) \rangle = \sigma^2 \,. \tag{2.79}$$

The consistent estimators which satisfy this property are unbiased. Basically, for this class of estimators, the true mean of a population consisting of T_N elements obtained from samples of size N coincides with the limit to infinity (true value) of the estimator.

As it will be later shown in Sects. 6.10 and 6.11, the sample mean (2.53) satisfies the properties (2.77) and (2.79), while the variance about the sample mean (2.56), without the factor $N/(N-1)$, does not satisfy Eq. (2.79).

Generally, the study of the asymptotic properties of random variables and their estimators is simpler if we apply the mean operator to a set of estimators $T_N(X)$, instead of studying directly the estimator limit (2.73) for $N \to \infty$.

2.12 Problems

2.1 Calculate the probability to obtain 2 times the face 6 by tossing three dices.

2.2 If one assigns a probability equal to $1/2$ to the event head, calculate the probability to obtain 3 heads in in 10 tosses.

2.3 A manufacturer knows that the percentage of defective pieces offered for sale is 10%. In a contract, he agrees to pay a penalty if, in one box of ten pieces, more than two pieces are defective. Find the probability to pay the penalty.

2.4 One player wagers 60 euros on a roulette by betting 50 euros on red (X strategy) and 10 euros on the black number 22 (Y strategy). Keeping in mind that the numbers range from 0 to 36 and that the dealer wins if the 0 hits and that the payout is equal to the bet if the red hits and is 36 times the post if 22 is the winning number, calculate the average capital value after a bet. What does this result mean?

2.5 A variable X can take the values $\{X = 2, 4, 6, 8, 10\}$ with equal probabilities. Find mean and standard deviation of the variables X and $Y = 2X + 1$.

2.6 Three marbles are extracted, without replacement, from an urn containing three red and seven black marbles. Determine the discrete p.d.f., mean and standard deviation of the number R of black marbles after three draws.

2.7 Find probability density $p(x)$, cumulative function $F(x)$, mean and standard deviation of a continuous variable X, defined in $[0, 1]$, and having a linearly increasing density such that $p(0) = 0$.

2.8 Find the 25-th percentile of the distribution of the previous problem.

2.9 To cover the round-trip distance between two points A and B, with a distance of 100 km from each other, a car travels at a speed of 25 km/h one way and 50 km/h on the way back. Calculate the average speed of the trip.

2.10 Find the total number of heads over a total of 1000 tosses from the data of Table 2.2.

2.11 Generate the Q-Q plot between two vectors of size 100 extracted from the binomial distribution $b(x; n = 20, p = 0.3)$. Then, produce the Q-Q plot of these generated random numbers versus the expected distribution.

Chapter 3
Basic Probability Theory

The question is not so much whether God plays dice, but how God plays dice

Ian Stewart, "DOES GOD PLAY DICE?".

3.1 Introduction

In this chapter we start by analysing the properties of the binomial distribution, and, then, we will gradually derive, using probability theory, all other fundamental statistical distributions. We will not avoid important mathematical steps, since we believe that this helps to have a general and consistent vision of the described topics. This will allow you, while analysing any scientific problem, to immediately understand its statistical and probabilistic aspects and to find the solution using the more appropriate statistical distributions.

We will end the chapter with some hints about the use of probability theory in hypothesis testing. This topic, which will be fully developed later on, in statistics, will allow you to appreciate better what you have learned and to fully understand many natural phenomena.

3.2 Properties of the Binomial Distribution

The binomial density, introduced in Sect. 2.6, is the p.d.f. of a random variable X which represents the number x of successes obtained in n independent trials with constant success probability. The properties of this density will allow us to develop, by successive stages, the basic scheme of density functions for one-dimensional random variables.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3_3

The binomial density is normalized, since from the formula of the Newton binomial coefficient and from Eq. (2.29), we have:

$$\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1. \tag{3.1}$$

The mean of the binomial distribution is given by Eq. (2.43), where the sum must be extended to all values $0 \leq x \leq n$, and the probability p_k is given by Eq. (2.29):

$$\mu = \sum_{x=0}^{n} x\, b(x; n, p) = \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n(n-1)!}{x(x-1)!(n-x)!} p\, p^{x-1} (1-p)^{n-x}\,, \tag{3.2}$$

where in the last row the change of the sum over x has to be noted since, when $x = 0$, the term of the sum is zero.

If we set in Eq. (3.2):

$$x' = x - 1, \qquad n' = n - 1, \qquad n' - x' = n - x\,, \tag{3.3}$$

we obtain, using Eq. (3.1):

$$\langle X \rangle = np \sum_{x'=0}^{n'} \frac{n'!}{x'!(n'-x')!} p^{x'} (1-p)^{n'-x'} = n\, p\,, \tag{3.4}$$

a result in agreement with intuition, which considers mean as an expected value, i.e. the product between the number of attempts (or trials) and the probability of each attempt. On the contrary, the value of the variance is not at all intuitive and, as we will see later, extremely important. It can be easily obtained by calculating the average of the squares, with the same procedure as for (3.2), and again using Eq. (3.3):

$$\left\langle X^2 \right\rangle = \sum_{x=1}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} = np \sum_{x'=0}^{n'} (x'+1) \binom{n'}{x'} p^{x'} (1-p)^{n'-x'}$$

$$= n\, p\, (n' p + 1) = n\, p\, [(n-1)p + 1]\,.$$

Then, from Eqs. (2.67) and (3.4), one can write:

$$\mathrm{Var}[X] = n\, p\, [(n-1)p + 1] - n^2 p^2 = n\, p\, (1-p)\,. \tag{3.5}$$

In conclusion, we have obtained the fundamental equations:

$$\mu = n\,p\;,\quad \sigma^2 = n\,p\,(1-p)\;,\quad \sigma = \sqrt{n\,p\,(1-p)}\;, \tag{3.6}$$

which are one of the milestones of probability calculus. The value of the mean is intuitive, while those of variance and standard deviation are not, and it helps if you memorize them right now.

Exercise 3.1
Find the true mean, variance and standard deviation for the 10 coin experiment of Table 2.2.

Answer Since the experiment is described by the binomial distribution with $n = 10$ and $p = 0.5$, from Eqs. (3.6) one has:

$$\mu = 10 \cdot 0.5 = 5$$
$$\sigma^2 = 10 \cdot 0.5\,(1 - 0.5) = 2.5$$
$$\sigma = \sqrt{2.5} = 1.58$$

The results are identical to those obtained in Exercise 2.6, where the basic formulae (2.54, 2.51) have been applied to the binomial probabilities reported in the fourth column of Table 2.2.

Exercise 3.2
The probability of hitting a target is equal to 80%. Find the p.d.f. of the number n of trials needed to be succesful. Find also the mean and standard deviation of n.

Answer We have to find the probability distribution of the number n of independent trials needed to get one success, when the probability of a single success is p.
In this case the random variable to be considered is no longer consisting of the number X of successes in n fixed attempts (leading to the binomial distribution) but is the number of attempts n, when the number of successes is fixed and equal to $x = 1$. For the n-th attempt to be successful, we must have at first $x = 0$ successes in $(n - 1)$ trials; the probability of this event is given by the binomial density (2.29):

$$b(x = 0; n - 1, p) = \frac{(n-1)!}{(n-1)!}\,(1-p)^{n-1} = (1-p)^{n-1}\;.$$

(continued)

Exercise 3.2 (continued)
Therefore, the probability to have a success in the n-th trial will be given by the compound probability:

$$g(n) = p(1-p)^{n-1} \,, \tag{3.7}$$

which is named *geometric density*. It can therefore be seen as a simple application of the law of compound probabilities (1.24), where the two probabilities refer to $(n-1)$ consecutive failures followed by a success at the n-th attempt.
The density is normalized, because from the theory of series we have:

$$\sum_{n=0}^{\infty} p^n = \frac{1}{1-p} \,, \quad \sum_{n=1}^{\infty} p^n = \frac{p}{1-p} \,, \quad 0 \le p < 1 \,, \tag{3.8}$$

so one gets:

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = p \sum_{k=1}^{\infty} (1-p)^{k-1} = p\,\frac{1}{p} = 1 \,, \qquad 0 < p \le 1 \,.$$

Mean and variance can be evaluated from Eqs. (2.43, 2.67) and from the properties of the geometric series. By differentiating Eq. (3.8) twice, one easily obtains:

$$\sum_{k=1}^{\infty} k\, p^{k-1} = \frac{1}{(1-p)^2} \,, \quad \sum_{k=1}^{\infty} k^2\, p^{k-1} = \frac{1+p}{(1-p)^3} \,. \tag{3.9}$$

From these equations one has:

$$\langle n \rangle = \sum_{k=1}^{\infty} kp\,(1-p)^{k-1} = p \sum_{k=1}^{\infty} k(1-p)^{k-1} = \frac{1}{p} \,, \tag{3.10}$$

$$\left\langle n^2 \right\rangle = \sum_{k=1}^{\infty} k^2 p\,(1-p)^{k-1} = \frac{2-p}{p^2} \,. \tag{3.11}$$

The variance is given by:

$$\sigma_n^2 = \left\langle n^2 \right\rangle - \langle n \rangle^2 = \frac{1-p}{p^2} \,. \tag{3.12}$$

(continued)

Exercise 3.2 (continued)
In this specific exercise, the probability, mean and standard deviation are then:

$$g(n) = 0.8 \cdot (0.2)^{n-1} ,$$
$$\mu_n = 1.25 ,$$
$$\sigma_n = \sqrt{(1-p)/p^2} = 0.31 .$$

A mean value $\mu = 1.25$ means that, in 12–13 runs, one will have, on average, 10 runs in which the first attempt is successful.

3.3 Poisson Distribution

The calculation of the binomial density is not easy, for large n, due to the factorials appearing in Eq. (2.29). If, in addition to the condition $n \gg 1$, the probability of the event is small ($p \ll 1$), then one will have few successes ($x \ll n$) and the approximation:

$$\frac{n!}{(n-x)!} = n(n-1)(n-2)\dots(n-x+1) \overset{x \ll n}{\simeq} n^x$$

holds. By writing $y = (1-p)^{n-x}$ and using logarithms, one has:

$$\ln y = (n-x)\ln(1-p) \overset{p \ll 1}{\longrightarrow} -p(n-x) \overset{x \ll n}{\longrightarrow} -np$$
$$y = (1-p)^{n-x} = e^{\ln y} \to e^{-np} .$$

After these transformations, the binomial density, for $n \gg 1$ and $p \ll 1$, assumes the form:

$$\lim_{\substack{n\to\infty \\ p\to 0}} b(x; n, p) = \lim_{\substack{n\to\infty \\ p\to 0}} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$
$$= \frac{n^x}{x!} p^x e^{-np} = \frac{(np)^x}{x!} e^{-np} , \tag{3.13}$$

from which, on the basis of Eq. (3.6) and of definition $\mu = np$, the Poisson density is obtained:

$$p(x; \mu) = \frac{\mu^x}{x!} e^{-\mu} . \tag{3.14}$$

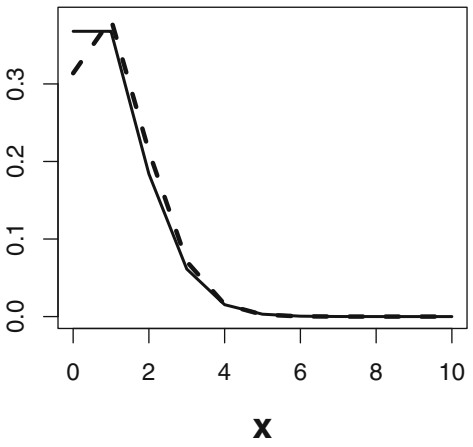


Fig. 3.1 Poissonian (full line) and binomial (dashed line) densities for $n = 10$, $p = 0.1$, $\mu = np = 1$. For a better comparison, the discrete values are joined with straight lines

It represents the probability to *obtain a value x when the mean value is* μ. The Poissonian is practically considered an acceptable approximation of the binomial already starting from $\mu > 10$ and $p < 0.1$, as shown in Fig. 3.1. The R code which reproduces this figure is:

```
BinPoisTest<- function(n=11){
 x <- seq(0,10,length=n)
 y <- dbinom(x,n,0.1)
 z <- dpois(x,lambda=1)
 plot(x,y,type='l',lwd=3,xlab='x',ylab=' ',
      lty='dashed',font.lab=2,cex.lab=1.5)
 lines(x,z,type='l',col='black',lwd=2) # lines adds z curve to plot
}
```

Exercise 3.3
In a city of 50,000 inhabitants, an average of five suicides occurs per year. Calculate the probability of ten suicides.

Answer From the binomial distribution (2.29), we have:

$$b(10;\,50000,\,\frac{5}{50{,}000}) = \binom{50{,}000}{10}\frac{5^{10}}{50{,}000^{10}}\left(1-\frac{5}{50{,}000}\right)^{49{,}990} = \,??\,.$$

Alternatively, from the Poissonian we obtain:

$$p(10;\,5) = \frac{5^{10}\,e^{-5}}{10!} \simeq 0.018 = 1.8\%\,.$$

The Poisson density is normalized, since:

$$\sum_{x=0}^{\infty} \frac{\mu^x}{x!} e^{-\mu} = e^{-\mu} \sum_{x=0}^{\infty} \frac{\mu^x}{x!} = e^{-\mu} e^{\mu} = 1 \,. \tag{3.15}$$

The mean and variance of the Poissonian can be found with the same method used in Eqs. (3.2–3.5) for the binomial distribution. We leave the explicit calculation of the mean as an exercise, which gives, as expected, the μ parameter of Eq. (3.14). For the variance, from Eqs. (2.67, 3.3), one has:

$$\begin{aligned} \mathrm{Var}[X] &= \sum_{x=0}^{\infty} \left[x^2 \frac{\mu^x e^{-\mu}}{x!} \right] - \mu^2 \\ &= \mu \sum_{x'=0}^{\infty} \left[(x'+1) \frac{\mu^{x'} e^{-\mu}}{x'!} \right] - \mu^2 = \mu^2 + \mu - \mu^2 = \mu \,. \end{aligned}$$

Therefore, we obtain the result:

$$\mu \equiv np \,, \quad \sigma^2 = \mu \,, \quad \sigma = \sqrt{\mu} \,, \tag{3.16}$$

which shows that, for a Poissonian, the mean and variance are equal.

3.4 Normal or Gaussian Density

Another important limiting case of the binomial density is the normal or Gauss density. This approximation is possible when the number of trials is huge: in this case only the values of the spectrum in the vicinity of the maximum of the density are important. Think, for example, about 1000 coin flips: the values of the spectrum for the event *head* range from 0 to 1000, but the important values will be concentrated around the expected value 500, which presumably is the value of higher probability. You will hardly ever get values such as 5, 995 and so on. Therefore, let us consider cases in which:

$$n \gg 1 \,, \qquad 0 < p < 1 \,, \qquad x \gg 1 \,, \tag{3.17}$$

and approximate factorials with the Stirling formula, valid for $x \gg 1$:

$$x! = \sqrt{2\pi x} x^x e^{-x} \, e^{-\varepsilon_x} \simeq \sqrt{2\pi x} \, x^x e^{-x} \,, \tag{3.18}$$

where $|\varepsilon_x| \leq 1/(12x)$. This is already a very good approximation for $x \geq 10$, since $\exp[-1/120] = 0.992$, corresponding to a relative error of 8 per thousand, as you can easily verify. Notice also that now x is, as requested, a *continuous variable*, which approximates the factorial for integer values.

Making use of the Stirling formula, the binomial density (2.29) assumes the form:

$$b(x; n, p) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$
$$\simeq \frac{1}{\sqrt{2\pi}} \frac{\sqrt{n} n^n}{\sqrt{x(n-x)} x^x (n-x)^{n-x}} p^x (1-p)^{n-x} . \quad (3.19)$$

Let us now expand the binomial density in the vicinity of its maximum. Turning to logarithms, we get:

$$\ln b(x; n, p) = \ln n! - \ln x! - \ln(n-x)! + x \ln p + (n-x) \ln(1-p) .$$

The point of maximum is obtained by setting the derivative of $\ln b$ to zero, using the Stirling formula and then considering x as a continuous variable:

$$\frac{\mathrm{d}}{\mathrm{d}x}(\ln b) = -\frac{\mathrm{d}}{\mathrm{d}x}(\ln x!) - \frac{\mathrm{d}}{\mathrm{d}x}[\ln(n-x)!] + \ln p - \ln(1-p) = 0 . \quad (3.20)$$

With these approximations, the derivative of $\ln x!$, reads:

$$\frac{\mathrm{d}}{\mathrm{d}x} \ln x! \simeq \frac{\mathrm{d}}{\mathrm{d}x} \ln \left[(2\pi x)^{1/2} x^x e^{-x} \right] = \frac{\mathrm{d}}{\mathrm{d}x} \left[\frac{1}{2} \ln 2\pi + \frac{1}{2} \ln x + x \ln x - x \right]$$
$$= \frac{1}{2x} + 1 + \ln x - 1 = \frac{1}{2x} + \ln x \xrightarrow{x \gg 1} \ln x . \quad (3.21)$$

Since $\mathrm{d}(\ln x!)/\mathrm{d}x \simeq \ln x$, Eq. (3.20) becomes:

$$\frac{\mathrm{d}}{\mathrm{d}x} \ln b \simeq -\ln x + \ln(n-x) + \ln p - \ln(1-p) = 0 , \quad (3.22)$$

and hence:

$$\ln \frac{p(n-x)}{x(1-p)} = 0 \Longrightarrow \frac{p(n-x)}{x(1-p)} = 1 \Longrightarrow x = \mu = np .$$

We have obtained a first result: for $\mu \gg 1$, that is, $x \gg 1$, the mean value also becomes the maximum. Let us now perform the Taylor expansion up to second order around this value. By writing $y(x) = \ln b$, we obtain:

$$y(x) = y(np) + \frac{1}{2} \left[\frac{d^2 y(x)}{dx^2} \right]_{x=np} (x - np)^2 + O\left(\frac{1}{n^2} \right)$$
$$\simeq y(np) + \frac{1}{2} \left[-\frac{1}{x} - \frac{1}{n-x} \right]_{x=np} (x - np)^2 , \quad (3.23)$$

where the second derivative is obtained by deriving Eq. (3.22) and the term in $(1/n^2)$ deriving again the second derivative. After some easy calculations, we get the equations:

$$y(x) \simeq y(np) - \frac{1}{2}\frac{(x-np)^2}{np(1-p)} ,$$

$$b(x; n, p) = e^{y(x)} \simeq e^{y(np)} \exp\left[-\frac{1}{2}\frac{(x-np)^2}{np(1-p)}\right] . \tag{3.24}$$

Since $e^{y(np)}$ is simply the binomial density at the maximum $x = np$ calculated with the Stirling formula, from Eq. (3.19) we easily obtain:

$$b(x = np; n, p) \simeq \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{np(1-p)}} , \tag{3.25}$$

and hence:

$$b(x; n, p) \simeq g(x; n, p) = \frac{1}{\sqrt{2\pi}\sqrt{np(1-p)}} \exp\left[-\frac{1}{2}\frac{(x-np)^2}{np(1-p)}\right] . \tag{3.26}$$

This is an approximate form of the binomial that holds for values of x around $\mu = np$, as long as the terms of order higher than $1/n$ in the Taylor expansion (3.23) are neglected. The density function $g(x; n, p)$, approximating the binomial for integer x values, *can be also viewed as a continuous function of* x. Indeed, the factor:

$$\Delta t \equiv \frac{1}{\sqrt{np(1-p)}}$$

of Eq. (3.26) is independent of x, $\to 0$ for $n \to \infty$ and has the meaning of differential of the variable t given by:

$$t = \frac{x - np}{\sqrt{np(1-p)}} .$$

By going to the Riemann limit, it is then possible, for large n values, to calculate the sum of the probabilities of a set of X values as:

$$\begin{aligned} \lim_{n\to\infty} P\{x_1 \le X \le x_2\} &= \lim_{n\to\infty} \sum_{x_1}^{x_2} b(x; n, p) \\ &= \lim_{\Delta t\to 0} \sum_{x_1}^{x_2} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x-np)^2}{np(1-p)}\right] \Delta t \\ &= \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{1}{2}t^2} \, \mathrm{d}t . \end{aligned} \tag{3.27}$$

This fundamental result is known as de Moivre-Laplace theorem. By using μ and σ from Eqs. (3.6), we then obtain:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\exp\left[-\frac{1}{2}t^2\right]\mathrm{d}t=\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{+\infty}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]\mathrm{d}x=1\,,$$

where the result $\int \exp(-z^2)\,\mathrm{d}z=\sqrt{\pi}$ has been used (see Exercise 3.4).

Therefore, the normal or Gaussian density:

$$g(x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right], \tag{3.28}$$

is normalized and, when it is integrated, allows the calculation of the probability that a value of x be in an interval about the mean μ when the variance is σ^2. This is by far the most important p.d.f. of probability theory and should therefore be remembered by heart.

Exercise 3.4
Prove that

$$\int_{-\infty}^{+\infty}\exp(-x^2)\,\mathrm{d}x=\sqrt{\pi}\,. \tag{3.29}$$

Answer By defining

$$I=\int_{-\infty}^{+\infty}e^{-x^2}\,\mathrm{d}x=\int_{-\infty}^{+\infty}e^{-y^2}\,\mathrm{d}y\,,$$

one has:

$$I^2=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-(x^2+y^2)}\,\mathrm{d}x\,\mathrm{d}y\,.$$

By converting to polar coordinates, we get:

$$I^2=\int_0^{+\infty}\int_{-0}^{2\pi}e^{-\rho^2}\rho\,\mathrm{d}\rho\,\mathrm{d}\theta=2\pi\int_0^{+\infty}e^{-\rho^2}\rho\,\mathrm{d}\rho=-\pi\left[e^{-\rho^2}\right]_0^{\infty}=\pi\,,$$

and hence $I=\sqrt{\pi}$. Similarly, it can be shown that

$$\int_{-\infty}^{+\infty}x^2\exp(-x^2)\,\mathrm{d}x=\sqrt{\frac{\pi}{2}}\,. \tag{3.30}$$

The mean and variance of the Gaussian distribution are given by the parameters μ and σ, which explicitly appear in Eq. (3.28). Indeed, from the integrals (3.29, 3.30),

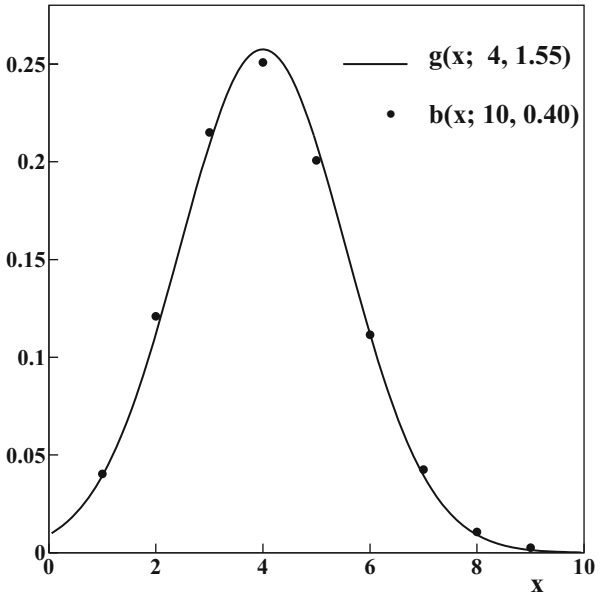


Fig. 3.2 Gaussian (full line) and binomial (dotted line) densities for $n = 10$, $p = 0.4$

it is possible to show that:

$$\langle X \rangle = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x \, \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \mathrm{d}x = \mu \,, \tag{3.31}$$

$$\mathrm{Var}[X] = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x-\mu)^2 \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \mathrm{d}x = \sigma^2 \,. \tag{3.32}$$

In statistics, we will have to use the 4-th order moment of the Gaussian; through integrals of the type (3.29, 3.30), one can demonstrate that:

$$\Delta_4 = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} (x-\mu)^4 \exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right] \mathrm{d}x = 3\,\sigma^4 \,. \tag{3.33}$$

The Gaussian density approximates the binomial for $\mu \gg 1$. Thanks to the rapid convergence of Stirling approximation (3.18), in many cases an error of less than a few percent is made when $\mu \geq 10$. As a rule of thumb, the Gaussian function can replace the binomial when the conditions:

$$\mu = np \geq 10 \,, \quad n(1-p) \geq 10 \quad \text{(in practice!)} \,, \tag{3.34}$$

hold, so that the approximation (3.23) is valid. If p and $1-p$ are not too close to the extremes of the interval [0, 1], the condition $\mu \geq 10$ is adequate. Otherwise, both conditions of Eq. (3.34) must be verified. This property is exemplified in Fig. 3.2,

where the Gaussian approximation of the binomial distribution $b(x; 10, 0.4)$ is reported. In this case, the mean and variance are given by $\mu = np = 10 \cdot 0.4 = 4$ and $\sigma^2 = np(1-p) = 2.4$. Even with the values $\mu = np = 4,\ n(1-p) = 6 < 10$, the approximation is already acceptable.

Exercise 3.5
Calculate the probability to obtain a given number of heads in the 10 coin experiment of Table 2.2 using the Gaussian distribution.

Answer In the ten-coin experiment, the mean, variance and standard deviation are given by:

$$\mu = np = 10 \cdot 0.5 = 5\ ,$$
$$\sigma^2 = np(1-p) = 10 \cdot 0.5(1-0.5) = 2.50\ ,$$
$$\sigma = \sqrt{2.5} = 1.58\ .$$

If we insert these values in Eq. (3.28), the requested Gaussian probabilities are obtained:

$$p(x) = 0.252\ \exp\left(-\frac{1}{2}\,\frac{(x-5)^2}{2.5}\right)\ , \quad (x = 0, 1, \ldots, 10)\ .$$

If we report the results in a new column of the table, we get the result:

Spectrum (number of heads)	Number of trials	Frequency	Binomial probability	Gaussian prob
0	0	0.00	0.001	0.002
1	0	0.00	0.010	0.010
2	5	0.05	0.044	0.042
3	13	0.13	0.117	0.113
4	12	0.12	0.205	0.206
5	25	0.25	0.246	0.252
6	24	0.24	0.205	0.206
7	14	0.14	0.117	0.113
8	6	0.06	0.044	0.042
9	1	0.01	0.010	0.010
10	0	0.00	0.001	0.002

(continued)

Exercise 3.5 (continued)
which shows that, even with a value $np = n(1-p) = 5 < 10$, the Gaussian density gives results already in good agreement with the binomial one. For this reason, the condition (3.34) is often further extended to:

$$\mu \geq 5\ ,\ n(1-p) \geq 5\ .$$

You can also study the Gaussian and all the other distribution functions by plotting them with R. For this, you should read Appendix B, where you can find some suggestions on the use of the R software.

3.5 The Three-Sigma Law and the Standard Gaussian Density

As we will see in the next section, the Gaussian density, besides being an excellent approximation of both the binomial and Poisson distributions, is also the limiting density of linear combinations of several independent random variables. For these reasons, it certainly represents the most important distribution of probability theory. Here we want to discuss a fundamental mathematical property of this function, known as the *three-sigma (or 3σ) law*:

$$P\{|X-\mu| \leq k\sigma\} \equiv \frac{1}{\sqrt{2\pi}\sigma} \int_{\mu-k\sigma}^{\mu+k\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \mathrm{d}x$$

$$= \begin{cases} 0.683 \text{ for } k=1 \\ 0.954 \text{ for } k=2 \\ 0.997 \text{ for } k=3 \end{cases} \tag{3.35}$$

which, in words, means: *if X is a Gaussian random variable with mean $\langle X \rangle = \mu$ and standard deviation $\sigma[X] = \sigma$, the probability to obtain an x value within an interval centred on μ and of width $\pm\sigma$ is about* 68%, *whereas if the interval width is $\pm 2\sigma$, it is about* 95%. *Moreover, the x values can occur outside an interval of width $\pm 3\sigma$ with a probability of* 3 *per thousand.*

In many practical cases, it is assumed that the spectrum values are all included in an interval centred on μ and $\pm 3\sigma$ wide (hence the name of the law (3.35)). The probabilities defined by Eq. (3.35) are also called Gaussian *levels or values*

of probability. In practice, the evaluation of the integral (3.35) is difficult because the primitive $\mathrm{E}(x)$ of the Gaussian:

$$\mathrm{E}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_0^x \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right] \mathrm{d}t \tag{3.36}$$

is not known analytically (strange, but true!). The integral is usually evaluated with numerical methods, by series expansion of the exponential and by numerically calculating the limit of the integrated series. In this way the probability levels of Eq. (3.35) can then be obtained. At this point, it would seem necessary to numerically calculate the cumulative probabilities in a given interval, for each Gaussian having a given mean and standard deviation. However, this complication can be avoided by resorting to a universal or standard Gaussian, which is obtained by defining a fundamental variable in statistics, the so-called standard variable:

$$T = \frac{X-\mu}{\sigma} \quad , \text{ which takes the values } \quad t = \frac{x-\mu}{\sigma} \, , \tag{3.37}$$

and measures *the deviation of a value x from its mean in units of standard deviation*. From Eqs. (2.63, 2.65, 2.66) it is immediate to notice that the standard variable has zero mean and unit variance.

The occurrences t of the random variable T are sometimes called deviates.

If we now insert the variable (3.37) in the integral (3.36), we get:

$$\mathrm{E}(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t \, . \tag{3.38}$$

It is easy to check that this primitive is related to the well-known error function $\mathrm{Erf}(x)$, since:

$$\mathrm{Erf}(x) \equiv \frac{1}{\sqrt{\pi}} \int_{-x}^{+x} \exp(-t^2)\, \mathrm{d}t = 2\, \frac{1}{\sqrt{2\pi}} \int_0^{\sqrt{2}x} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t \equiv 2\, \mathrm{E}(\sqrt{2}\, x) \, . \tag{3.39}$$

This universal function can be calculated once and for all, because it is independent of μ and σ. Since the Gaussian is symmetric, the conditions:

$$\mathrm{E}(x) = -\,\mathrm{E}(-x) \, . \tag{3.40}$$

holds. If we now consider the change of variable $t = \sqrt{2}\,z$ and integrate the exponential series, we obtain:

$$\begin{aligned} \mathrm{E}(x) &= \frac{1}{\sqrt{\pi}} \int_0^{x/\sqrt{2}} (1 - z^2 + \frac{z^4}{2!} - \frac{z^6}{3!} + \dots)\,\mathrm{d}z \\ &= \frac{1}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k\,(x/\sqrt{2})^{2k+1}}{k!\,(2k+1)}\,. \end{aligned} \tag{3.41}$$

The sum of this series is tabulated in all statistical books (in this one, it can be found in Table E.1 of Appendix E). The primitive of Eq. (3.38) corresponds to a standard Gaussian density:

$$g(x;\,0,\,1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{3.42}$$

with zero mean and unit standard deviation. A distribution having Gaussian density (3.28) is called *normal* and often denoted by $N(\mu, \sigma^2)$; the corresponding random variable sometimes is denoted by $X \sim N(\mu, \sigma^2)$, or $T \sim N(0, 1)$ when it is standard. The symbol $\sim$ means "distributed as". In R this function is called `dnorm(x)`. The cumulative function of the standard Gaussian of Fig. 3.3 is usually indicated as $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-x^2/2}\,\mathrm{d}x\,. \tag{3.43}$$

The link with the error function (3.38) is given by:

$$\Phi(x) = 0.5 + \mathrm{E}(x)\,, \tag{3.44}$$

which is valid also for $x < 0$, thanks to Eq. (3.40). $\Phi(x)$ is present in R with the function `pnorm(x)`, which can be used as an alternative to Table E.1 using the call `pnorm(x)-0.5`, which gives a result accurate up to seven significant digits. In summary, as shown by the following examples, the calculation of Gaussian probabilities can always be traced back to the case of the standard Gaussian of zero mean and unit variance, for which there is a universal table.

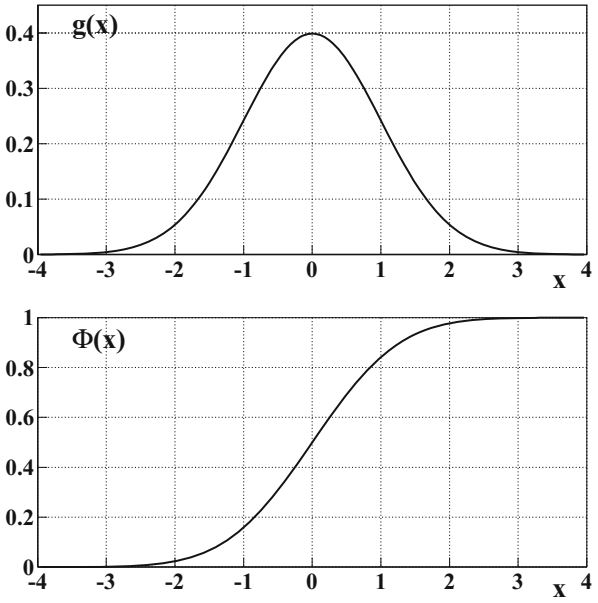


Fig. 3.3 Standard Gaussian $g(x; 0, 1)$ (3.42) and corresponding cumulative function $\Phi(x)$ (3.43)

Exercise 3.6
Find the probability $P\{\mu - \sigma \leq X \leq \mu + \sigma\}$ with 4 significant digits.

Answer In the two extremes of the interval to be considered, the standard variable (3.37) is equal to ± 1; from Table E.1, given in Appendix E, for $t = 1.00$ (1.0 reads in the row and the last zero in the column heading) we have $P = 0.3413$. Since the table gives the integral between zero and t, we have to double this value. We get therefore $P\{\mu - \sigma \leq X \leq \mu + \sigma\} = 2 \times 0.3413 = 0.6826$, according to Eq. (3.35). Note that the table only provides values for $0 \leq t \leq 0.5$, since the standard Gaussian is symmetric about the origin.

Exercise 3.7
Find the probability to obtain values $X > 12$ or $-2 \leq X \leq 12$ from a Gaussian $N(5, 9)$, that is, with mean $\mu = 5$ and standard deviation $\sigma = 3$.

Answer In this case the standard variable (3.37) is:

$$t = \frac{12 - 5}{3} = 2.33\,, \quad t = \frac{-2 - 5}{3} = -2.33\,.$$

(continued)

Exercise 3.7 (continued)
From the cumulative $\Phi(x)$ of Fig. 3.3, we see that the probability of occurrences $x > 2.33$ is very small. We can obtain a precise value from Table E.1, where, for $x = 2.33$, we find the number 0.4901, which is the probability to obtain normal deviates between 0 and 2.33. Therefore, the requested probabilities are given by:

$$P\{X > 12\} = 0.5000 - 0.4901 = 0.0099 \simeq 1\% \,,$$
$$P\{-2 \leq X \leq 12\} = 1 - 2 \times 0.0099 = 0.9802 \simeq 98\% \,.$$

Exercise 3.8
Find the extremes of the Gaussian probability interval of 95% centred around the mean.

Answer We have to evaluate a real number t such that $P\{\mu - t\,\sigma \leq X \leq \mu + t\,\sigma\} = 0.95$. As an alternative to Table E.1, we use, in this case, the R routine `qnorm`, which gives the Gaussian quantile values. To obtain the solution, we must enter the command `qnorm(0.5+0.5*0.95)`, which returns the value 1.9599. In conclusion, the answer is: a 95% probability is obtained by integration of the Gaussian in an interval of width $\pm 1.96\,\sigma$ centred around the mean. Indeed:

$$t = 1.96 = \left|\frac{x-\mu}{\sigma}\right| \qquad \text{implies} \qquad x = \mu \pm 1.96\,\sigma \,,$$

and hence:

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{\mu-1.96\sigma}^{\mu+1.96\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x = 0.95 \,.$$

3.6 Central Limit Theorem and Universality of the Gaussian Curve

The Gaussian density appears, so far, to be only a limiting distribution of the binomial density for high number of successes. However, a fundamental theorem of probability theory, proposed by Gauss and Laplace in 1809–1812, and demonstrated

in a general way in the early 1900s, assigns a crucial role to it, both for continuous and discrete variables, as follows.

Theorem 3.1 (Central Limit) *Consider a random variable Y which is a linear combination of N random variables X_i:*

$$Y_N = \sum_{i=1}^{N} a_i X_i \ , \tag{3.45}$$

where a_i are constant coefficients. If:

(a) X_i are mutually independent (see Eq. (2.9));
(b) X_i have finite variance;
(c) all variances (or standard deviations) have the same order of magnitude:

$$\frac{\sigma[X_i]}{\sigma[X_j]} = O(1) \quad \textit{for all} \quad i, j \ ; \tag{3.46}$$

then, for $N \to \infty$, the random variable Y_N converges in law, according to Eq. (2.75), towards the Gaussian distribution.
Therefore, we can write, using cumulative functions:

$$P\left\{ \frac{Y_N - \langle Y_N \rangle}{\sigma[Y_N]} \leq x \right\} \overset{N \gg 1}{\longrightarrow} \Phi(x) \ ,$$

where $\Phi(x)$ is the cumulative function of the standard Gaussian (3.43).

Proof The proof of the theorem, for variables having the same distribution, is based on generating functions and is reported in Appendix C. Also the de Moivre-Laplace formula (3.27) could be seen as a special case of the theorem for the sum of Bernoulli variables.
More generally, it can also be shown that this theorem holds for sums of variables, both discrete and continuous, each having a different distribution [Cra51]. However, it is essential for the variables X_i to be mutually independent (condition (a)) and, if taken individually, to have a weak influence on the final result (conditions (b) and (c)). Moreover, a very important and rather astonishing fact, which occurs in practice, should be immediately noted: the condition $N \to \infty$ can be replaced, in most cases, with a good approximation by the condition $N \geq 10$. □

In short, the theorem states that a random variable tends to have Gauss density if it is the linear superposition of several independent variables which, if taken individually, have weak influence on the final result.

Speaking somewhat freely, we can say that the theorem assigns to the Gaussian the role of a universal density function for variables of systems in "high statistical equilibrium".

Even more interesting is the fact that practice and computer simulations show that N does not have to be very big. As already highlighted in the proof of the theorem, it is often legitimate to use the condition $N \geq 10$.

It is often quoted: "experimentalists use the Gaussian because they think that mathematicians have proved that it is the universal curve; mathematicians use it because they think experimenters do have in practice demonstrated its universality . . .". This somewhat simplistic statement should be replaced with the following one: random variables often follow the Gaussian distribution because the conditions of the central limit theorem are often quite well verified. However, there are several and important exceptions to this general rule, so that the theorem should be considered as a fundamental reference point, not to be blindly applied.

Here are some examples of random variables which, according to the theorem, are Gaussian distributed:

- The height or weight of a population of ethnically homogeneous individuals.
- The weight of the beans contained in a standard can.
- The values of the intelligence quotient (IQ) of a group of people.
- The mean of a sample with a number of events greater than ten: in this case conditions (a) to (c) of the theorem are certainly satisfied, because these variables are independent and have the same variance.
- The velocity components of the molecules of an ideal gas.

On the contrary, the following variables are *not* Gaussian:

- The energy of the molecules of an ideal gas: this quantity is proportional to the square modulus of the velocity (which is a vector with Gaussian components) and therefore the linearity condition (3.45) is no longer valid. We will soon show that the square modulus of a vector of independent Gaussian components follows a particular distribution, called χ^2 or chi-square. In three dimensions, this distribution is the famous Maxwell distribution, well known to physicists.
- As we will see shortly, the arrival times of Poissonian events do not follow the Gaussian distribution.

The central limit theorem can be verified with simulated data with a few lines of R code:

```
N=10
for(j in 1:1000) y[j]=sum(runif(N))
hist(y)
plot(density(y,adj=0.01))
```

where a vector of 1000 values constructed as the sum of `N` uniform variables $U(0, 1)$ is generated. The raw data are then histogrammed with `hist` and also displayed in a different way with the `density` routine, which is described in Appendix B. By varying `N` from 1 to larger and larger values, you can check the speed of convergence of `y` towards the Gaussian distribution.

3.7 Poisson Stochastic Processes

In stochastic processes, random variables depend on continuous or discrete parameters. In this section, we will deal with a very frequent case, when time t is the continuous parameter.

Let's then consider a stochastic process in which a source generates events:

(a) Of discrete type, in such a way that *at most one event* can be emitted in an infinitesimal time interval $\mathrm{d}t$.
(b) With constant probability λ per unit of time, equal for all the events. This parameter, of dimension t^{-1}, represents the emission frequency per unit of time (e.g. the number of events per second). This property implies that the average number of events emitted in a given time interval depends only on the width of the interval, not on its position along the time axis.
(c) Mutually independent. In other words, the numbers of events occurring in disjoint time intervals are independent random variables.

A stochastic process following conditions (a) to (c) is called *stationary Poisson process*.

What is the statistical law of the number of events generated within a measurable time interval Δt? The answer becomes easy when the following quantities are defined:

- the number of attempts $N = \Delta t / \mathrm{d}t$. If Δt is a measurable macroscopic time interval and $\mathrm{d}t$ is a differential, we always have $N \gg 1$. Note that the duration of a useful attempt to generate an event is assumed to be a very small quantity, that is, the differential $\mathrm{d}t$. This is the crucial hypothesis of all the arguments we are developing;
- the probability to emit a single event in an attempt of duration $\mathrm{d}t$ is $p = \lambda\,\mathrm{d}t$. Then, we always have $p \ll 1$, if the emission process is discrete and $\mathrm{d}t$ is a differential;
- the average number of events generated within Δt is $\mu = \lambda \Delta t$.

The event emission process can then be considered as the binomial probability of having $\{X = n\}$ hits in $N = (\Delta t / \mathrm{d}t) \gg 1$ trials when the elementary probability of a success is $p = \lambda\,\mathrm{d}t \ll 1$. Keeping in mind the results of Sect. 3.3, it is immediate to infer that, in a discrete process, the event counting $X(\Delta t)$ in an interval Δt is a Poissonian random variable, with $\mu = \lambda \Delta t$. From Eq. (3.14) we obtain:

$$P\{X(\Delta t) = n\} = p_n(\Delta t) = \frac{(\lambda \Delta t)^n}{n!} e^{-\lambda \Delta t} \,. \tag{3.47}$$

It is easy to show that, if Eq. (3.47) holds, counts in disjoint time intervals are independent random variables (see Problem 3.13).

Poisson's law is followed by a truly vast class of phenomena: the number of fishes caught in 1 h in stable environmental conditions, the number of cars that stop

at a traffic light on the same day and the same time in, let's say, 1 month (if there are no exceptional events), the number of photons emitted by excited atoms, the number of nuclear particles emitted by a radioactive source, the number of shooting stars observed in 10 minutes on an August night, the number of traffic accidents in 1 year, if no new security measures are taken on and so on, in short, all those cases in which a *stable* source (λ constant) emits *discrete* ($\lambda\,\mathrm{d}t \ll 1$) and *independent* events.

So far we have considered the properties of the source of events; what is the role of the observer or of the counting apparatus? The observation of the process does not alter the Poissonian statistics if the detection or counting apparatus (the eye, an instrument or a more complicated device) *has both a short dead time and a short resolution time, compared to the average arrival time of the events*. The dead time is the time during which the instrument remains inactive after the detection of an event (e.g. in a Geiger counter, this time is of the order of a millisecond); the resolution time is the time below which the apparatus no longer records all emitted event (this quantity is slightly less than 1 second for the human eye, whereas it ranges from a fraction to a thousandth of a second in mechanical instruments, and it can reach a billionth of a second or even less in electronic instruments). Basically, an event arriving within the dead time is not recorded, while two or more events arriving within the resolution time are recorded as a single event. It is clear then that the original Poissonian statistics is not altered if the probability of an event arriving within the dead time and the probability of arrival of more than one event within the resolution time are both negligible.

The number of events emitted and/or counted in a finite time Δt is not the only important random variable of a stochastic process; the time T between events is also a random variable. This leads us to discuss a further fundamental statistical law of nature: the negative exponential law of arrival times. Imagine to reset the clock at an arbitrary time $t = 0$ or on the last recorded event: the time T of the next arrival is a random variable which is determined by the compound probability to have nothing until t *and* to have the occurrence of an event within $(t, t + \mathrm{d}t]$. Setting $n = 0$ in Eq. (3.47), we have $p_0(t) = e^{-\lambda t}$, while, according to the definition of λ, the probability of arrival in $(t, t + \mathrm{d}t]$ is simply given by $\lambda\,\mathrm{d}t$. The p.d.f. of the random variable T is therefore a negative exponential:

$$e(t)\,\mathrm{d}t = p_0(t)\lambda\,\mathrm{d}t = \lambda e^{-\lambda t}\,\mathrm{d}t\,, \quad t \geq 0\,. \tag{3.48}$$

It is straightforward to prove that this distribution is normalized and that its mean and variance are given by:

$$\mu = \int t\,e(t)\,\mathrm{d}t = \frac{1}{\lambda}\,, \quad \sigma^2 = \int \left(t - \frac{1}{\lambda}\right)^2 e(t)\,\mathrm{d}t = \frac{1}{\lambda^2}\,. \tag{3.49}$$

The discrete equivalent of the exponential density is the geometric density (3.7). It is indeed easy to see that, for $p \ll 1$, the logarithm of this equation becomes:

$$\ln g(n) = \ln[p(1-p)^{n-1}] = \ln p + (n-1)\ln(1-p) \to \ln p - (n-1)p \,,$$

and hence:

$$g(n) \overset{p \ll 1}{\to} pe^{-(n-1)p} \,. \tag{3.50}$$

The result obtained so far suggest some interesting considerations. The intuitive representation that we often have of the arrival times of an event flow (e.g. 10 events per second) is of Gaussian type: events should arrive at intervals of about $1/10$ of a second, with small symmetrical fluctuations around this value. This intuitive representation *is completely wrong*: the arrival of events is governed by the exponential law, according to which *clusters of events are much more likely than events separated by long waiting intervals* (see also Fig. 3.4). This effect is clearly visible with the particle counters that are often used in a laboratory: events seem to arrive "in clusters" separated by long waiting intervals, giving, to non-experts, the impression of instrumental malfunctions. Instead, this apparent temporal correlation is precisely due to the absolute lack of correlation between events!

Rare events, such as accidents or natural disasters, often occur within short time periods, generating the (false) belief of mysterious correlations between disasters. In fact, very often these are just effects due to chance, which have suggested some popular sayings such as " good things come in threes, bad things come in threes . . ." or "it never rains but it pours . . .". Maybe these are among the few proverbs that have some scientific foundation . . . Since the exponential density does not have a Gaussian-like "bell shape", the mean does not give much information about the data

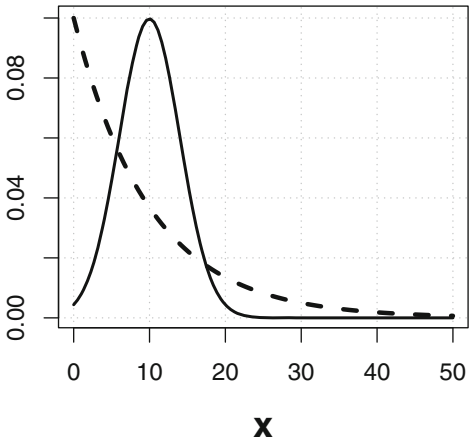


Fig. 3.4 In a Poisson process, the density of arrival times is a negative exponential, as shown by the dashed curve, which refers to an average flow of 0.1 event(s) ($\lambda = 0.1\,s^{-1}$, i.e. 1 event every 10 s). The solid line is a Gaussian with a mean of 10 s and arbitrary standard deviation of 4 s, which is an example of the intuitive (but incorrect!) idea that we often have of stochastic phenomena

localization in this case and is more useful to use the cumulative function of the density (3.48) that is easily obtained from Eq. (2.33):

$$P\{0 \leq T \leq t\} \equiv F(t) = \int_0^t \lambda e^{-\lambda\tau}\, \mathrm{d}\tau = 1 - e^{-\lambda t}\ . \tag{3.51}$$

This function gives the arrival probability of *at least one event* into $[0, t]$. The probability of not observing events up to a time t is given by:

$$P\{T > t\} = 1 - F(t) = 1 - P\{0 \leq T \leq t\} = e^{-\lambda t} = p_0(t)\ , \tag{3.52}$$

which is the Poisson density (3.47) when $n = 0$.

Since, from Eq. (3.49), we know that the time mean is $1/\lambda$, we can calculate the percentage of arrivals before and after this value:

$$P\{0 \leq T \leq 1/\lambda\} = 1 - e^{-1} = 0.63\ , \tag{3.53}$$

$$P\{1/\lambda \leq T\} = e^{-1} = 0.37\ . \tag{3.54}$$

Exercise 3.9
On average, 40 vehicles passed through a certain stretch of road between 23:00 and 23:30. What is the probability of observing a time interval less than 10 s and one longer than 5 minutes between one vehicle and the next one?

Answer The mean time between two consecutive vehicles is given by:

$$\frac{60 \times 30}{40} = \frac{1800}{40} = 45\, s\ ,$$

from which a mean arrival frequency of

$$\lambda = \frac{1}{45} = 0.022\, s^{-1}$$

is obtained. By applying Eq. (3.51) with $t = 10$ s, one obtains:

$$P\{0 \leq T \leq 10\ \mathrm{s}\} = 1 - e^{-0.022 \cdot 10} = 0.199 \simeq 20\%\ .$$

Then, the probability to observe time intervals longer than $t = 5$ min $= 300$ s is given by:

$$P\{T > 300\ \mathrm{s}\} = 1 - P\{0 \leq T \leq t\} = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}$$
$$= e^{-0.022 \cdot 300} = 1.27\, 10^{-3} \simeq 0.13\%\ .$$

We now demonstrate that the negative exponential is the only functional form ensuring the temporal independence of events. Suppose that $A = \{T > t + \Delta t\}$ and $B = \{T > t\}$ are the events "there are no arrivals up to $t + \Delta t$" and "up to t", respectively. Obviously one has $\{T > t + \Delta t\} \subset \{T > t\}$, because if A is verified, the same holds for B, and $P(A \cap B) = P(A)$ (the intersection of the events *does not* correspond to that of the time intervals!). The conditional probability (1.19) is in this case:

$$P\{T > t + \Delta t | T > t\} = \frac{P\{\{T > t + \Delta t\} \cap \{T > t\}\}}{P\{T > t\}} = \frac{P\{T > t + \Delta t\}}{P\{T > t\}}$$
$$= \frac{e^{-\lambda(t+\Delta t)}}{e^{-\lambda(t)}} = e^{-\lambda \Delta t} = P\{T > \Delta t\} \,. \tag{3.55}$$

This result shows that the information on the absence of events up to t has no influence on the arrival probability in subsequent times. The exponential law remains unchanged, independently of the arbitrary moment t in which the clock to reset; it therefore describes systems *without memory*.[1] The independence among counts in separate time intervals also implies the independence of events "there is an arrival at time t" and "there are no arrivals in $(t, t + \Delta t)$". Therefore, the relation:

$$P\{T > t + \Delta t | T = t\} = P\{T > \Delta t\} = e^{-\lambda \, \Delta t}$$

holds, which shows that the times between arrivals (interarrival times) *are independent random variables of negative exponential p.d.f.*

To describe biological systems or devices with memory, which tend to "age", a variant of the exponential function, called Weibull density, is sometimes used:

$$w(t) = a \, t^{b-1} \, e^{-(a/b) \cdot t^b} \,, \quad t \geq 0 \,, \quad a, \, b > 0 \,, \tag{3.56}$$

where the parameters a and b are often empirically determined from data analysis. The Weibull curve, for $a = b = 1$, is the negative exponential; instead it tends to assume a bell shape for $b > 1$. The corresponding cumulative function is $F(t) = \int_0^t w(t) \, \mathrm{d}t = 1 - \exp(-at^b/b)$. If, for example, $a = 2$, $b = 2$, then $w(t) = 2t \exp(-t^2)$ and from Eq. (3.55) one has $P\{T > t + \Delta t | t\} = \exp[-(\Delta t)^2 - 2t\Delta t]$. The probability of having no events up to $t + \Delta t$ (i.e. the survival probability, if the event is the system breakdown) decreases when t increases.

We now generalize the exponential law to non-contiguous Poissonian events and determine the probability distribution of the the waiting time until the arrival of the k-th event after the clock start (see Fig. 3.5).

[1] If we combine several systems without memory, we can have systems with memory, whose time probability densities do not satisfy Eq. (3.55); see, for instance, in the next two chapters, Exercise 4.1 and Problem 5.8.

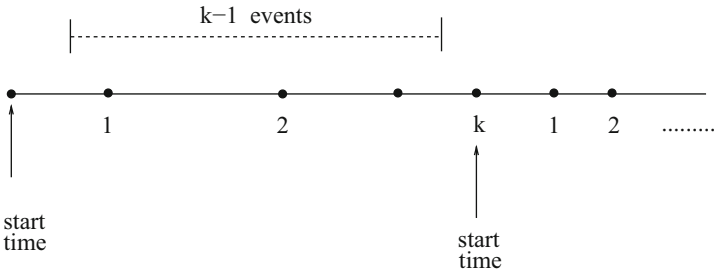


Fig. 3.5 Definition of k-th event of the gamma p.d.f., for the arrival times of non-contiguous events

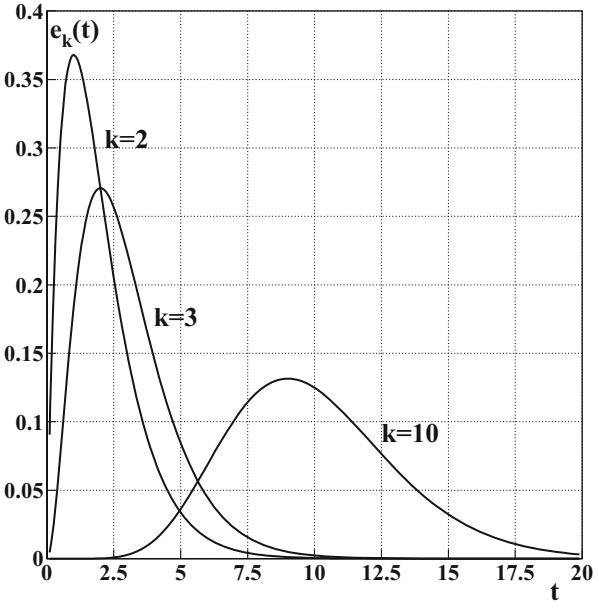


Fig. 3.6 Gamma density for $\lambda = 1$ and different k values

After inserting $p_{k-1}(t)$ in Eq. (3.48) instead of $p_0(t)$, we easily get the Erlangian density of order k or *gamma density* of Fig. 3.6:

$$e_k(t) = \lambda \frac{(\lambda t)^{k-1}}{(k-1)!} e^{-\lambda t} = \frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t} , \quad t \geq 0 , \ (k-1) \geq 0 , \tag{3.57}$$

where $\Gamma(k) = (k-1)!$ for integer k values (see, further on, Eq. (3.65)). The mean and variance of this density are given by:

$$\mu = \frac{k}{\lambda}\,, \qquad \sigma^2 = \frac{k}{\lambda^2}\,. \tag{3.58}$$

For contiguous events, $k - 1 = 0$ and we again obtain Eqs. (3.48–3.49). When k increases, the mean and variance increase and the flux density $\lambda_k = 1/\mu$ decreases. The gamma density can also be considered as the distribution of the sum of k independent negative exponential random variables. This property is evident from Fig. 3.6 where, thanks to the central limit Theorem 3.1, we see that the curve, for large k, tends to the Gaussian form. These results can also be derived from the theory of functions of random variables, which we will develop in Chap. 5 (see Problem 5.5).

In R, the Erlangian family is given by the function `dgamma(t, shape, rate)`, where `shape` $= k$ and `rate` $= \lambda$; if $k = 1$ one gets the negative exponential. As usual, by changing the prefix `d` to `p`, `q` or `r`, one obtains, the cumulative, quantile and simulated values of the distributions, respectively.

It is also easy to demonstrate that, between the Poisson and Erlang densities, the useful relation:

$$\frac{\mathrm{d}p_k(t)}{\mathrm{d}t} = e_k(t) - e_{k+1}(t) \tag{3.59}$$

holds.

Up to now, we have demonstrated that conditions (a) to (c), introduced at the beginning of the paragraph, which define the stationary Poisson process, imply Eq. (3.47), that is, the negative exponential density of arrival times. We will now show that the inverse is also true: Eq. (3.47) holds if interarrival times are independent and follow the negative exponential law. In fact, from the series expansion of Eqs. (3.51, 3.52), it follows that the probability to observe an event in an infinitesimal interval is $\lambda\,\mathrm{d}t$ and that to observe more events is negligible (an infinitesimal of higher order). Moreover, if k events occur within t, it means that t is within T_k and T_{k+1} arrival times of the k-th and $(k+1)$-th events, respectively. Since $P\{T_k \le t\} = P\{T_k \le t < T_{k+1}\} + P\{T_{k+1} \le t\}$, we can write:

$$\begin{aligned} P\{X(t) = k\} &= P\{T_k \le t < T_{k+1}\} \\ &= P\{T_k \le t\} - P\{T_{k+1} \le t\} \\ &= \int_0^t [e_k(t) - e_{k+1}(t)]\,\mathrm{d}t = p_k(t)\,, \end{aligned}$$

where, in the last step, Eq. (3.59) has been used. In this proof we have used the property that the Erlang distribution $e_k(t)$ can be also obtained as the distribution of the sum of k exponentially distributed random times, without explicitly invoking the Poisson distribution (see Problem 5.5 and Appendix C).

Having thus rediscovered Poisson's law, which implies the independence of counts in disjoint time intervals, ultimately we arrive at the following:

Theorem 3.2 (Stochastic independence) *A necessary and sufficient condition for a stationary Poisson process is to have independent interarrival times and a negative exponential p.d.f.*

3.8 χ^2 Density

What is the p.d.f. of the square modulus of a vector with random normal components? This problem requires the determination of the p.d.f. of a random variable which is a function of other random variables, a topic that will be analysed in detail in Chap. 5. However, here we anticipate a specific result, connected to the χ^2 density, which plays an important role within the one-dimensional density functions, which we are describing here.

We then want to determine the p.d.f. of the variable:

$$Q = \sum_{i=1}^{n} X_i^2 \,, \tag{3.60}$$

which is the sum of squares of n mutually independent standard normal variables. We immediately note that this discussion is valid, in general, for any independent Gaussian variables, because a normal variable can always be standardized through the transformation (3.37).

We indicate with $F_Q(q)$ the cumulative of the p.d.f. we are searching for: it gives the probability $P\{Q \leq q\}$ that the variable Q is within a hypersphere of radius $\sqrt{q}$. Since the variables X_i come from the standard Gaussian density (3.42) and are independent, we deduce, from Eqs. (1.23, 1.24, 2.9), that $P\{Q \leq q\}$ will be given by the product of the compound probability to obtain a set of values $(x_1, x_2, \ldots, x_n)$ summed (or integrated) over the set obeying the rule $\sum_i x_i^2 \leq q$. Therefore, we have:

$$\begin{aligned} P\left\{\sum X_i^2 \leq q\right\} \equiv F_Q(q) &= \int_{\sum x_i^2 \leq q} \cdots \int \left(\frac{1}{\sqrt{2\pi}}\right)^n \prod_i e^{-\frac{x_i^2}{2}} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \\ &= \int_{\sum x_i^2 \leq q} \cdots \int \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\sum_i \frac{x_i^2}{2}} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \,, \end{aligned} \tag{3.61}$$

where the product in the integrand follows from the compound probabilities theorem and from Eqs. (1.23, 1.24). We also note that the link between Q and the variables X_i shows up only in the definition of the integration domain. If we then pass to spherical coordinates by setting $\sqrt{\sum_i x_i^2} = r$, the integrand is angle independent

and the functional link (3.60) gives rise to an integration over the radius of the hypersphere. By a known result of differential geometry, we know that the element of a volume integrated over the angles becomes, from a three-dimensional sphere to a n-dimensional hypersphere:

$$\int_{\mathrm{d}\Omega} \mathrm{d}V = \int_{\mathrm{d}\Omega} r^2 \sin\theta \, \mathrm{d}r \, \mathrm{d}\theta \, \mathrm{d}\phi = 4\pi r^2 \, \mathrm{d}r \rightarrow D \, r^{n-1} \, \mathrm{d}r \, ,$$

where D is a constant that can be obtained by integration over the angular variables. The integral of Eq. (3.61) then becomes:

$$F_Q(q) = F_0 \int_0^{\sqrt{q}} e^{-\frac{r^2}{2}} \, r^{n-1} \, \mathrm{d}r \, ,$$

where F_0 includes all the constant factors present in the calculation. If we now operate the change of variable:

$$r = \sqrt{q} \, , \qquad \mathrm{d}r = \frac{1}{2} \frac{1}{\sqrt{q}} \, \mathrm{d}q \, ,$$

and collect again all constant factors in F_0, we can write $F_Q(q)$ as:

$$F_Q(q) = F_0 \int_0^{q} e^{-\frac{q}{2}} \, q^{\frac{n}{2}-1} \, \mathrm{d}q \, , \tag{3.62}$$

which is the primitive of the p.d.f. we are searching for. Therefore, we have, from Eq. (2.35):

$$p(q) = \frac{\mathrm{d}F_Q(q)}{\mathrm{d}q} = F_0 \, e^{-\frac{q}{2}} \, q^{\frac{n}{2}-1} \, . \tag{3.63}$$

Obviously, this function is defined only for $q \geq 0$. The constant F_0 is evaluated from the normalization condition:

$$F_0 \int_0^{\infty} e^{-\frac{q}{2}} \, q^{\frac{n}{2}-1} \, \mathrm{d}q = 1 \, .$$

The integral can be calculated from the gamma function definition:

$$\Gamma(p) = \int_0^{\infty} x^{p-1} \, e^{-x} \, \mathrm{d}x \, , \tag{3.64}$$

where $\Gamma(p)$, called gamma function, can be obtained via integration by parts from Eq. (3.64):

$$\begin{aligned}
\Gamma(1) &= 1 \\
\Gamma(1/2) &= \sqrt{\pi} \\
\Gamma(p+1) &= p\,\Gamma(p) \\
\Gamma(p+1) &= p! \qquad \text{for integer } p \\
\Gamma(p+1) &= p(p-1)(p-2)\ldots\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi} \quad \text{for half-integer } p,
\end{aligned} \tag{3.65}$$

and in R is calculated by the routine `gamma(x)` with $x > 0$. From Eqs. (3.63–3.64), we finally obtain the function:

$$p_n(q)\,\mathrm{d}q = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}\, e^{-\frac{1}{2}q}\, q^{\frac{1}{2}(n-2)}\,\mathrm{d}q\,, \quad q \geq 0\,, \tag{3.66}$$

which is the p.d.f. of the *square modulus of a vector with n independent Gaussian components*. If the linearly independent components were equal to $\nu < n$, then in Eq. (3.60) the sum can be transformed into a linear combination of the squares of ν independent Gaussian variables and the density function is always given by (3.66), after performing the $n \to \nu$ substitution. The linearly independent variables of a χ^2 distribution are called *degrees of freedom*.

In the international physics literature, the notation χ^2 is almost always used to indicate the distribution, the random variable and its numerical realizations (!). Since, for us, this seems to be an excessive ease, to remain (partially) consistent with our own notation, we will indicate with Q a random variable having χ^2 density (3.66) and with χ^2 (and not with q) the numerical values of Q obtained experimentally. We will then write that $Q(\nu) \sim \chi^2(\nu)$ takes values χ^2.

With this notation, the density (3.66) of the variable Q with ν degrees of freedom is written as:

$$p_\nu(\chi^2)\,\mathrm{d}\chi^2 \equiv p(\chi^2;\nu)\,\mathrm{d}\chi^2 = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\,(\chi^2)^{\frac{\nu}{2}-1}\, e^{-\frac{\chi^2}{2}}\,\mathrm{d}\chi^2\,, \tag{3.67}$$

which, based on Eqs. (3.64, 3.65), has mean and variance given by:

$$\langle Q\rangle = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int_0^\infty x(x)^{\frac{\nu}{2}-1}\, e^{-\frac{x}{2}}\,\mathrm{d}x = \nu\,, \tag{3.68}$$

$$\mathrm{Var}[Q] = \frac{1}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}\int_0^\infty (x-\mu)^2\,(x)^{\frac{\nu}{2}-1}\, e^{-\frac{x}{2}}\,\mathrm{d}x = 2\nu\,. \tag{3.69}$$

Sometimes, the reduced χ^2 distribution of the variable:

$$Q_R(\nu) = \frac{Q(\nu)}{\nu} \,, \tag{3.70}$$

is used, which, from Eqs. (2.63, 2.64, 3.68, 3.69), has mean and variance given by:

$$\langle Q_R(\nu) \rangle = 1 \,, \qquad \mathrm{Var}[Q_R(\nu)] = \frac{2}{\nu} \,. \tag{3.71}$$

In Table E.3 of Appendix E, the values of the integral of the reduced χ^2 density:

$$P\{Q_R(\nu) \geq \chi^2_R(\nu)\} = \int_{\chi^2_R(\nu)}^{\infty} \frac{\nu^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu x}{2}\right) \mathrm{d}x \tag{3.72}$$

are reported (see Fig. 3.7); they are obtained by applying Eq. (3.70) to Eq. (3.67). This table provides the probability of exceeding an assigned value of χ^2_R, a parameter which will be extensively used later in statistics.

We can summarize these results, found by Helmert in 1876 and generalized by Pearson in 1900, with the

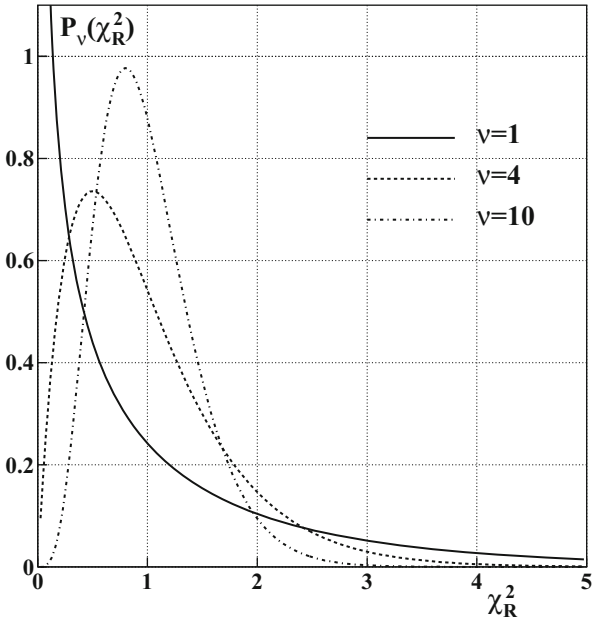


Fig. 3.7 Reduced χ^2 distribution for some degrees of freedom ν

Theorem 3.3 (of Pearson) *The sum of squares of ν independent Gaussian variables is a random variable with density (3.67) with ν degrees of freedom, called $\chi^2(\nu)$.*

In the following, also this theorem will be useful:

Theorem 3.4 (Additivity of the Variable χ^2) *If Q_1 and Q_2 are two independent random variables, having χ^2 density with ν_1 and ν_2 degrees of freedom, respectively, the variable*

$$Q = Q_1 + Q_2 \tag{3.73}$$

has $\chi^2(\nu)$ density with $\nu = \nu_1 + \nu_2$ degrees of freedom: $Q \sim \chi^2(\nu_1 + \nu_2)$. Moreover, if $Q \sim \chi^2(\nu)$ and $Q_1 \sim \chi^2(\nu_1)$, then $Q_2 \sim \chi^2(\nu - \nu_1)$.

Proof The proof of the first part of the theorem is immediate and can be seen as a lemma of Pearson's Theorem 3.3: since the sum of squares of ν_1 independent standard variables is distributed as χ^2, if other independent ν_2 variables are added to them, using Eq. (3.73), the result is a sum of squares of $\nu_1 + \nu_2$ independent standard Gaussian variables, hence the statement. The proof of the second part of theorem is also easy if one uses the generating functions of Appendix C and Eq. (C.12).
It is important to remember that the theorem applies to variables $Q \sim \chi^2$, not to the reduced ones $Q_R \sim \chi^2_R$ of Eq. (3.70). □

In R, the probabilities of the χ^2 distribution with `df` degrees of freedom for a value `x` are calculated by the `dchisq(x, df)` function, whereas the cumulative, quantile and simulated values are obtained from `pchisq`, `qchisq` and `rchisq`. In the next exercises, we will realize that the χ^2 density is of fundamental importance both in statistics and in physics.

Exercise 3.10
Find the p.d.f. of the modulus R of a three-dimensional vector having independent Gaussian components (X,Y,Z) of zero mean and variance σ^2.

Answer Let us assume to have Gaussian standard components with unit variance. The p.d.f. of the square modulus of this vector is then given by Eq. (3.66) with $n = 3$:

$$p_3(q)\,\mathrm{d}q = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{q}{2}}\, q^{\frac{1}{2}}\,\mathrm{d}q\ ,$$

since $\Gamma(3/2) = \Gamma(1/2 + 1) = \sqrt{\pi}/2$, from Eqs. (3.65). This density gives the probability that the square modulus is within $(q, q + dq)$. To have the

(continued)

Exercise 3.10 (continued)
density of the modulus (not of its square), we must use the transformation:

$$q = (x^2 + y^2 + z^2) = r^2 \,, \quad r = \sqrt{(x^2 + y^2 + z^2)} = \sqrt{q} \,, \quad 2r\,\mathrm{d}r = \mathrm{d}q$$

to obtain:

$$m(r)\,\mathrm{d}r = \sqrt{\frac{2}{\pi}}\, r^2 e^{-\frac{r^2}{2}}\,\mathrm{d}r \,. \tag{3.74}$$

So far we have considered the variables (X, Y, Z) as standard Gaussians. However, we know that R is the modulus of a vector with non-standard Gaussian components having a finite variance σ^2. To take into account the non-unit variance, we operate the transformation:

$$x \to \frac{x}{\sigma} \,, \quad y \to \frac{y}{\sigma} \,, \quad z \to \frac{z}{\sigma} \,,$$

which redefines r as:

$$r \to \frac{r}{\sigma} \quad \mathrm{d}r \to \frac{\mathrm{d}r}{\sigma} \,.$$

By inserting this transformation in Eq. (3.74), we obtain:

$$m(r)\,\mathrm{d}r = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma^3}\, r^2 e^{-\frac{r^2}{2\sigma^2}}\,\mathrm{d}r \,, \tag{3.75}$$

which is the well-known Maxwell density function.
Since R^2/σ^2 is a χ^2 variable with 3 degrees of freedom (i.e. $\chi^2(3)$), from Eqs. (2.64, 3.68, 3.69), the mean and variance are given by:

$$\left\langle \frac{R^2}{\sigma^2} \right\rangle = 3 \,; \quad \mathrm{Var}\left[\frac{R^2}{\sigma^2}\right] = \frac{1}{\sigma^4} \,; \quad \mathrm{Var}[R^2] = 6$$

so that:

$$\left\langle R^2 \right\rangle = 3\sigma^2 \,; \quad \mathrm{Var}[R^2] = 6\sigma^4 \,. \tag{3.76}$$

Exercise 3.11
Find the energy density of the molecules of an ideal gas at the absolute temperature T.

Answer In an ideal gas, the velocity components (V_x, V_y, V_z) of the molecules are random variables that satisfy the conditions of the central limit Theorem 3.1. Therefore, the velocity modulus follows the Maxwellian distribution (3.75). If we consider the relation between kinetic energy and velocity of a molecule of mass m:

$$E = \frac{1}{2}mv^2 \,, \quad v = \sqrt{\frac{2E}{m}} \,, \quad mv\,\mathrm{d}v = \mathrm{d}E \,,$$

we can write:

$$m(E)\,\mathrm{d}E = \frac{2}{\sqrt{\pi}}\frac{1}{\sigma^3}\frac{1}{m}\sqrt{\frac{E}{m}}e^{-\frac{E}{m\sigma^2}}\,\mathrm{d}E \,.$$

If we now use the known thermodynamics equation that links variance and temperature:

$$m\sigma^2 = KT \,, \quad \sigma = \sqrt{\frac{KT}{m}} \,, \tag{3.77}$$

where K is the Boltzmann constant, we finally obtain:

$$m(E)\,\mathrm{d}E = \frac{2}{\sqrt{\pi}}\frac{1}{KT}\sqrt{\frac{E}{KT}}\,e^{-\frac{E}{KT}}\,\mathrm{d}E \,, \tag{3.78}$$

which is the famous Boltzmann distribution.
From Eqs. (3.76, 3.77), another thermodynamics fundamental result is obtained, that is, the link between temperature and the molecule mean velocity $\langle U \rangle$ at the absolute temperature T:

$$\langle U \rangle = \frac{1}{2}m\left\langle V^2 \right\rangle = \frac{3}{2}m\sigma^2 = \frac{3}{2}KT \,.$$

At this point we would like to notice that we have obtained, both in this exercise and in the previous one, some fundamental results of statistical physics only using, as physics hypotheses, the central limit theorem and Eqs. (3.77).

3.9 Uniform Density

A continuous random variable X, assuming values in the finite interval $[a, b]$, is defined as uniform in $[a, b]$ and is indicated as $X \sim U(a, b)$ when it has a constant p.d.f., which is called uniform or flat density (see Fig. 3.8), given by:

$$u(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{for } x < a \ , x > b \end{cases} \tag{3.79}$$

The normalization condition:

$$P\{a \le X \le b\} = \int_a^b u(x)\, \mathrm{d}x = 1$$

is satisfied, and the mean and variance are given by:

$$\mu = \frac{1}{b-a} \int_a^b x \ \mathrm{d}x = \frac{b+a}{2} \ , \tag{3.80}$$

$$\sigma^2 = \frac{1}{b-a} \int_a^b \left(x - \frac{b+a}{2} \right)^2 \mathrm{d}x = \frac{(b-a)^2}{12} \ , \tag{3.81}$$

where the last integral is easily calculated, via the substitution $y = x - (b+a)/2$, $\mathrm{d}y = \mathrm{d}x$, between the limits $-(b-a)/2$ and $(b-a)/2$.

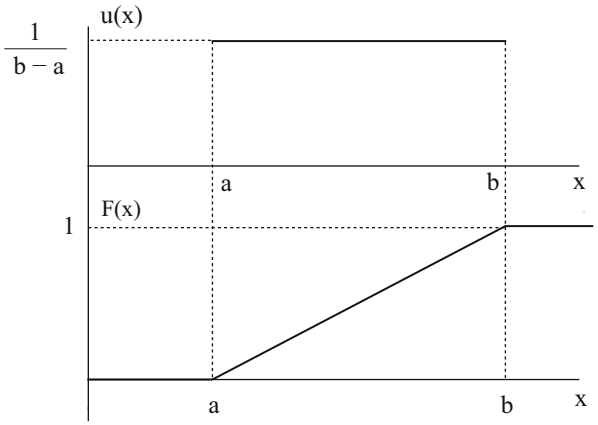


Fig. 3.8 Uniform density $u(x)$ and corresponding cumulative function $F(x)$ for a random variable assuming values in $[a, b]$

The density is often considered uniform within the interval $a = 0$ and $b = \Delta$; in this case the mean and variance are given by:

$$\mu = \frac{\Delta}{2}\,, \quad \sigma^2 = \frac{\Delta^2}{12}\,. \tag{3.82}$$

The support of the density is given by $\mu - \Delta/2 \leq x \leq \mu + \Delta/2$ and its standard deviation is $\sigma = \Delta/\sqrt{12}$. For a random variable $a \leq X \leq b$, having uniform density, the localization probability in (x_1, x_2) is proportional to the width of the interval:

$$P\{x_1 \leq X \leq x_2\} = \frac{1}{b-a} \int_{x_1}^{x_2} \mathrm{d}x = \frac{x_2 - x_1}{b-a}\,. \tag{3.83}$$

Conversely, if a continuous random variable satisfies Eq. (3.83), then it follows the uniform density.

We now present a very simple, but extremely important and general theorem related to this distribution.

Theorem 3.5 (Cumulative Random Variables) *If X is a random variable with continuous density $p(x)$, the cumulative random variable C:*

$$C(X) = \int_{-\infty}^{X} p(x)\,\mathrm{d}x \tag{3.84}$$

is uniform in $[0, 1]$*, that is,* $C \sim U(0, 1)$.

Proof The probability for X to be within $[x_1, x_2]$ coincides with the probability for the cumulative random variable C to be in the range $[c_1 \equiv C(x_1), c_2 \equiv C(x_2)]$. From Eq. (2.33), we then have:

$$\begin{aligned} P\{c_1 \leq C \leq c_2\} = P\{x_1 \leq X \leq x_2\} &= \int_{x_1}^{x_2} p(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{x_2} p(x)\,\mathrm{d}x - \int_{-\infty}^{x_1} p(x)\,\mathrm{d}x = c_2 - c_1\,, \end{aligned} \tag{3.85}$$

which implies that the cumulative variable $C \sim U(0, 1)$, since it satisfies Eq. (3.83) with $(b - a) = 1$. □

You should fully realize the conceptual and practical importance of the theorem: the cumulative variable *is always uniform, whatever the origin distribution is*.

If the integral (3.84) is known analytically, then the values of the cumulative variable C can be written as $c = F(x)$. If this function is also invertible, then the variable:

$$X = F^{-1}(C) \tag{3.86}$$

has density $p(x)$. If you have a uniform variable generator in [0, 1], a roulette or a computer random number generator, *continuous variables with any density can be generated* using the equation:

$$X = F^{-1}(\texttt{random}) \,. \tag{3.87}$$

The extension of Theorem 3.5 to discrete variables requires a minimum of attention and the use of Eq. (2.28). If C is a uniform variable, keeping in mind that $F(x)$ is defined in [0, 1], from (3.83), we obtain:

$$\begin{aligned} P\{X = x_k\} &= P\{x_{k-1} < X \leq x_k\} \\ &= F(x_k) - F(x_{k-1}) = P\{F(x_{k-1}) < C \leq F(x_k)\} \,. \end{aligned} \tag{3.88}$$

The discrete equivalent of Eq. (3.87), if one has a `random` generator, then becomes:

$$\{X = x_k\} \text{ if } \quad \{F(x_{k-1}) < \texttt{random} \leq F(x_k)\} \quad (\text{if } k = 1,\ F(x_0) = 0) \,. \tag{3.89}$$

All Monte Carlo simulation methods that will be examined in detail in Chap. 8 are based on Eqs. (3.87, 3.89). Equation (3.86) has a convincing graphical interpretation, reported in Fig. 3.9. Consider the probabilities $P\{x_1 \leq X \leq x_2\}$ and $P\{x_3 \leq X \leq x_4\}$ defined within the values $[x_1, x_2]$ and $[x_3, x_4]$: they are represented by the areas subtended by the density function $p(x)$ and displayed by the shaded zones of Fig. 3.9. The area between $[x_1, x_2]$ (corresponding to less probable X values) is smaller than the area between $[x_3, x_4]$ (more probable X values). By construction, these two areas are identical to the length of the intervals $[c_1, c_2]$ and $[c_3, c_4]$, obtained from the cumulative function $F(x)$. Let's now consider a variable $C \sim U(0, 1)$ on the cumulative ordinate axis: it will fall more frequently in $[c_3, c_4]$ rather than in $[c_1, c_2]$, with probabilities *exactly coinciding with the width of these intervals*. Therefore, if, given a value c_0 assumed by C, we find (graphically or analytically) the corresponding value x_0 and repeat this procedure several times, we will obtain a sample of X values from $p(x)$.

We can easily verify the theorem by generating cumulative variables with R. For example, we can consider the χ^2 density and write:

```
y <- pchisq(rchisq(1000,df=10),df=10)
hist(y)
```

to check that the histogram of `y` does follow the uniform distribution. It is also interesting to see that if one writes:

```
y <- pchisq(rchisq(1000,df=9),df=10)
hist(y)
```

a non-uniform distribution is obtained for `y`, because 1000 variables have been generated from a χ^2 distribution with 9 degrees of freedom (the value of the upper

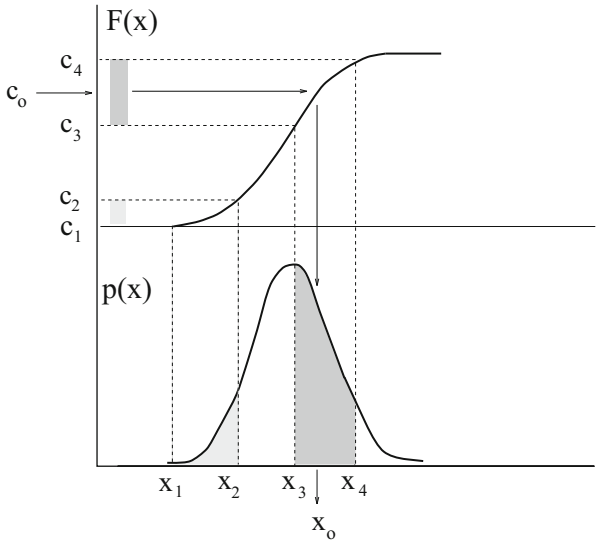


Fig. 3.9 Graphical representation of the cumulative variable theorem. The functions $p(x)$ and $F(x)$ are a generic p.d.f. and the corresponding cumulative, respectively

limit X in Eq. (3.84)), whereas the cumulative is calculated from a different p.d.f. $p(x)$ having `df=10`. Indeed, the theorem is valid *if and only if* in Eq. (3.84) X is sampled from $p(x)$. In the following, we often will use this property to check the parent distribution of some variables.

Exercise 3.12
Assuming to have a computer-generated `random` value uniformly distributed in [0, 1], randomly sample arrival times of stochastically independent events.

Answer The arrival time distribution of stochastically independent events is a negative exponential with cumulative function of an arrival in $[0, t]$ given by Eq. (3.51). Given Theorem 3.5, this function is a uniform random variable. We then have:

$$\texttt{random} = 1 - e^{-\lambda t} \ .$$

By inverting this equation, we get:

$$t = -\frac{1}{\lambda} \ln(1 - \texttt{random}) \ . \tag{3.90}$$

(continued)

Exercise 3.12 (continued)
Since, if `random` $\sim U(0,1)$, the same also holds for (1-`random`), this equation is equivalent to:

$$t = -\frac{1}{\lambda} \ln(\texttt{random}) \,. \tag{3.91}$$

Furthermore, it is easy to see that a vector of 1000 values
`y <- rgamma(1000,shape=1)`
obtained with the R routine for the generation of exponential variates with $\lambda = 1$ has a distribution identical to that of a vector `z <- -log(runif(1000))`, according to Eq. (3.91). Equations (3.90, 3.91) are an example of the general Eqs. (3.86, 3.87).

3.10 Chebyshev's Inequality

The standard deviation, which is an index of the dispersion of a variable around its mean value, satisfies an important and general property. Let us consider a p.d.f. of mean μ, *finite variance* σ^2, and the interval $[\mu - K\sigma, \mu + K\sigma]$, where K is a positive real number. Obviously, the points outside this range are defined by the condition $|x - \mu| > K\sigma$. Considering the expression (2.57), of the variance for continuous variables, we can write:

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{+\infty} (x-\mu)^2 p(x)\, \mathrm{d}x \\
&= \int_{\mu-K\sigma}^{\mu+K\sigma} (x-\mu)^2 p(x)\, \mathrm{d}x + \int_{|x-\mu|>K\sigma} (x-\mu)^2 p(x)\, \mathrm{d}x \\
&\geq \int_{|x-\mu|>K\sigma} (x-\mu)^2 p(x)\, \mathrm{d}x \geq K^2\sigma^2 \int_{|x-\mu|>K\sigma} p(x)\, \mathrm{d}x \\
&= K^2\sigma^2 \left(1 - \int_{\mu-K\sigma}^{\mu+K\sigma} p(x)\, \mathrm{d}x\right) .
\end{aligned}$$

From the last equality, we obtain:

$$\int_{\mu-K\sigma}^{\mu+K\sigma} p(x)\, \mathrm{d}x \geq 1 - \frac{1}{K^2} \,, \tag{3.92}$$

which is known as Chebyshev's inequality. This relation can be easily proved even for discrete variables and is quite general, because the only condition that has been imposed on $p(x)$ is to have a finite variance. Let us now see what information is present in this general law. Similar to Eq. (3.35), this inequality assumes the form:

$$P\{|X-\mu| \le K\,\sigma\} = \int_{\mu-K\sigma}^{\mu+K\sigma} p(x)\ \mathrm{d}x \ge 1 - \frac{1}{K^2} = \begin{cases} 0 & \text{for } K=1 \\ 0.75 & \text{for } K=2 \\ 0.89 & \text{for } K=3\,, \end{cases} \tag{3.93}$$

which shows that *intervals around the mean of width* 2σ *and* 3σ *cover at least* 75% *and* 90% *of the total occurrence probability*.

Equation (3.93) sometimes justifies the approximated 3σ law, which consists in considering probabilities outside $[\mu-3\sigma,\ \mu+3\sigma]$ to be negligible *for any statistical distribution*. Generally this is a good approximation, because Chebyshev's inequality, which predicts no more than a 10% probability outside of $\pm 3\sigma$, is almost always a significant overestimate of the actual values. For example, the Gauss density has only 0.3% of the values "outside 3σ", while, in the case of uniform density, all values are included within $\pm 2\sigma$ (check as exercise).

Considering only values within 3σ is very common, and, generally, this leads, as previously mentioned, to acceptable results. However, in special cases, if one is dealing with "very broad" densities, it is good to remember that with this method a significant error, up to 10%, can occur, as shown by Eq. (3.93).

3.11 How to Use Probability Calculus

Figure 3.10 and Table 3.1 report the fundamental probability distributions that have been so far obtained. The starting point is the binomial distribution of Eq. (2.29), for discrete and independent random events generated with constant probability.

The limit for $np,\ n(1-p) \gg 1$ (in practice $np,\ n(1-p) > 10$), where n is the number of attempts, leads to the Gaussian density, whereas the limit $n \gg 1,\ p \ll 1$ (in practice $n > 10,\ p < 0.1$) leads to the Poissonian density. This latter density, when $np > 10$ (and therefore also $n(1-p) > 10$, since $p \ll 1$), also evolves towards the Gaussian density while maintaining always the relation (3.16) $\mu = \sigma^2$, typical of Poissonian processes.

We have also seen that the Gaussian and Poissonian densities, far from being only limiting cases of the binomial density, are the reference distributions of many important natural phenomena. The Gaussian distribution is the limiting density of the linear superposition of independent random variables, none of which prevails over the others. This result comes from the central limit Theorem 3.1. The statistical distribution of the square modulus of a vector of Gaussian components is the χ^2 density, which in three dimensions is called Maxwell's density or Maxwellian.

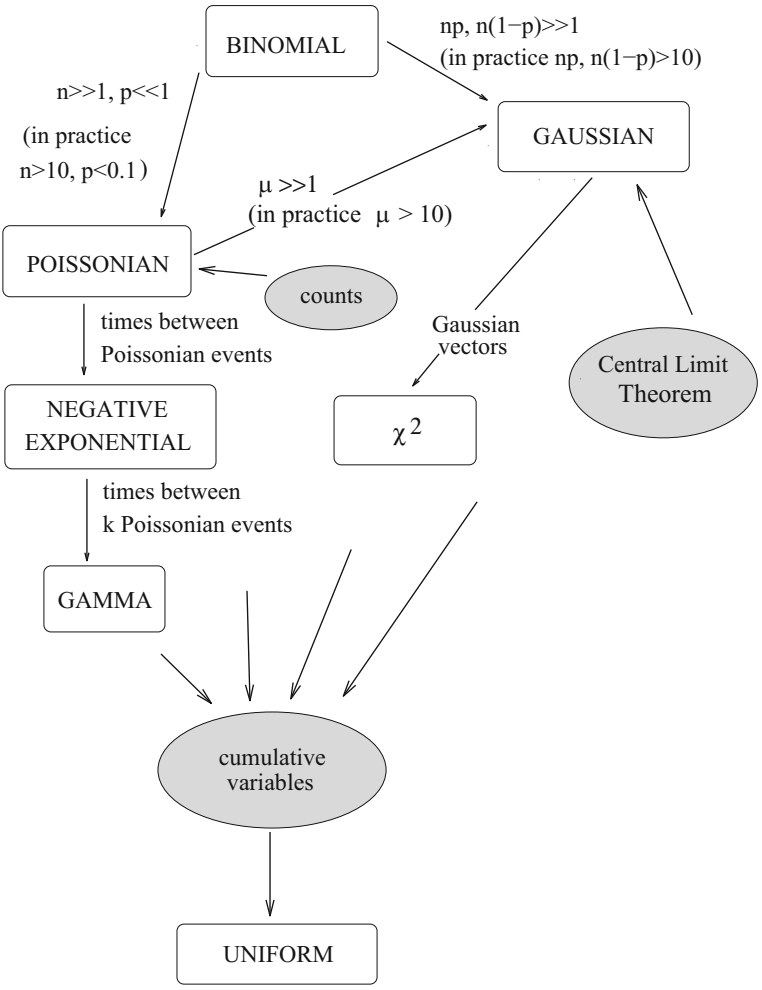


Fig. 3.10 The fundamental distributions of probability theory and the connections among them

The Poisson distribution is instead the universal distribution of the number of independent events generated discretely with constant probability over time (stochastic generation). The arrival times between these events also follow universal distributions, the negative exponential and the gamma or Erlangian ones.

Finally, the uniform density is also of general importance, because, as shown in Theorem 3.5, it is the distribution of all the cumulative random variables. As we will see, this principle is the basis of Monte Carlo simulations.

It should be already clear to the reader, and it will be anyway more and more so in the following, that the statistical distributions deduced in probability theory provide a specific and coherent scheme for the interpretation of a remarkable collection of

Table 3.1 The fundamental p.d.f. of probability theory

Name	Density	Mean	Standard deviation	Comment
Binomial	$\frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$	np	$\sqrt{np(1-p)}$	Number of successes in independent trials with constant probability
Gaussian	$\frac{\exp[-(x-\mu)^2/2\sigma^2]}{\sqrt{2\pi}\sigma}$	μ	σ	Linear combination of independent variables
Chi-square	$\frac{(\chi^2)^{\frac{\nu}{2}-1} \exp(-\frac{\chi^2}{2})}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})}$	ν	$\sqrt{2\nu}$	Modulus of a Gaussian vector
Poissonian	$\frac{\mu^x}{x!} e^{-\mu}$	μ	$\sqrt{\mu}$	Counts
Exponential	$\lambda\, e^{-\lambda t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda}$	Arrival times between Poissonian events
Gamma	$\frac{\lambda^k}{\Gamma(k)} t^{k-1} e^{-\lambda t}$	$\frac{k}{\lambda}$	$\frac{\sqrt{k}}{\lambda}$	Sum of k negative exponential variables
Uniform	$\frac{1}{\Delta}$ $(0 \le x \le \Delta)$	$\frac{\Delta}{2}$	$\frac{\Delta}{\sqrt{12}}$	Cumulative variables

natural phenomena. We will now try to better explain this point, with a series of examples and some insights.

Intuition alone is not enough to interpret random phenomena: if we toss a coin 1000 times, we guess that on average we will have 500 heads (or tails), but if we get 450 heads and 550 tails how do we know if the result is compatible with chance or if instead the coin is rigged or the flips were not regular? Using probability calculus, we can approach these problems in a quantitative way, using continuous or discrete p.d.f. and solving sums (for discrete variables) or integrals (for continuous variables) of the type:

$$P\{a \le X \le b\} = \sum_{x=a}^{b} p_x \to \int_a^b p(x)\,\mathrm{d}x\ , \tag{3.94}$$

which gives the probability level of the result, that is, the probability to obtain the observed values within the interval $[a, b]$, if the model given by $p(x)$ is valid.

It is therefore possible to judge whether a certain deviation from the expected value is due to chance or not. As we have already mentioned in Sect. 2.6, in the first case it is said that there is a normal *statistical fluctuation* and that the model

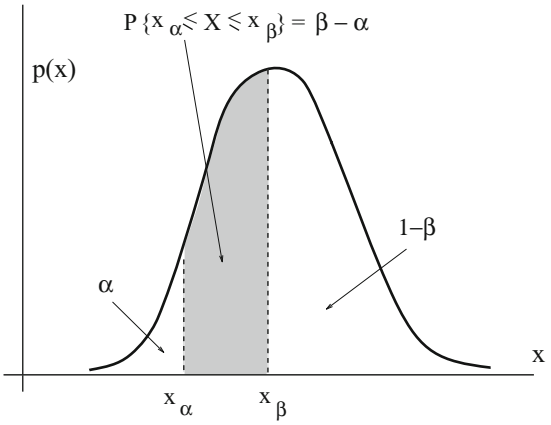


Fig. 3.11 Graphical representation of the probability level $\beta - \alpha$

represented by $p(x)$ is valid (or better, that is not *falsified* by observation); in the second case, the result is interpreted as a *significant deviation* and the model is rejected. In this type of studies, the quantile values (2.18) or the mean and standard deviation of the model distribution can be used. In this way, analysis is faster and allows us to acquire useful mental automatisms, easy to remember by heart, which also permits to evaluate, quickly and correctly, the experimental results. If the quantile values of the distribution are known, by setting $a = x_\alpha$ and $b = x_\beta$, the probability interval, as also shown in Fig. 3.11, becomes:

$$P\{x_\alpha \leq X \leq x_\beta\} = \beta - \alpha \; . \tag{3.95}$$

If X has a symmetric p.d.f., the following notation:

$$P\left\{-t_{1-\alpha/2} \leq \frac{X-\mu}{\sigma} \leq t_{1-\alpha/2}\right\} = 1 - \alpha \; . \tag{3.96}$$

is often used. In many cases, t_α is the standard normal quantile (easily obtained from Table E.1), or evaluated from Student's density (which we will discuss in the following). In several text books, Gaussian quantiles are indicated as z_α.

In R, it is easy to calculate Eq. (3.95) using cumulative functions (see Table B.2). For example, if you want to evaluate the area between the values $x_\alpha = -1.5$ and $x_\alpha = 2$ of the standard Gaussian, just write `pnorm(2)-pnorm(-1.5)`, obtaining the value 0.91044. Levels commonly used in statistics are $1 - \alpha = 0.90, 0.95, 0.99, 0.999$, which correspond to the Gaussian quantile values $t_{1-\alpha/2} = 1.64, 1.96, 2.58, 3.29$. Alternatively, as we will often do in this text, one can adopt the convention, common among physicists, that parameterizes probability intervals according to the 3σ law (3.35). *Standard deviation can be considered as the universal unit of measurement of statistical fluctuations.* To calculate its value, one should not proceed intuitively but use probability calculus. So, does a result of 450

hits in 1000 flips only represent a statistical fluctuation or a significant deviation from the expected value of 500? The answer is in the example below.

Exercise 3.13
A coin was flipped 1000 times and 450 heads were obtained. Is this result compatible with the hypothesis of random flipping of a non-rigged coin?

Answer The model taken as a reference, which in statistics will be called *null hypothesis*, predicts that the a priori probability of obtaining head in a single roll is $p = 0.50$ and that the probabilities of the possible values of a thousand flips, ranging from 0 to 1000, can be calculated from the binomial distribution with mean and standard deviation given by Eq. (3.6):

$$\mu = np = 500\,,$$
$$\sigma = \sqrt{np(1-p)} = \sqrt{500 \cdot 0.5} = \sqrt{250} = 15.8\,.$$

The observed frequency, that is, the experimental result, is $f = 0.45$, which corresponds to a standard value of:

$$t = \frac{450 - 500}{15.8} = -3.16\,.$$

The results differ 3.16 standard deviations from the expected value. Since $np = n(1-p) = 1000 \cdot 0.5 = 500 \gg 10$, we can assume the Gaussian approximation for the standard variable, and use Table E.1 of Appendix E. From this table we read, in correspondence of $t = 3.16$, the value 0.4992. The area of the tail to the left of t is given by:

$$P\{T < t = -3.16\} = 0.5000 - 0.4992 = 8 \cdot 10^{-4}\,,$$

which is the probability to obtain *by chance, if the model holds,* values ≤ 450. With R, we obtain, using the command `pnorm(-3.16)`, a value of 0.000788.
Now pay attention to this crucial step: *if we reject the model when it is true, the probability to be wrong is not greater than* 8 *over* 10 000. It is also said that the data agrees with the model with a *significance level* of $8 \cdot 10^{-4}$.
In conclusion, since the significance level is very small, we can say, with a small chance of being wrong, that 450 successes on a thousand tosses represent an event in disagreement with the binomial model with $p = 1/2$, which assumes independent flips of a non-rigged coin.

The hypothesis is generally considered to be falsified when the observed significance level is below 1–5%. However, this value depends on the type of problem being considered, and it is (at last partially) subjective, as we will discuss in more detail in Sect. 7.1. For instance, let's assume that instead of the fairness of a coin, we are considering the safety of an airplane. If a certain experiment results in a significance level of one per thousand with respect to the null hypothesis of a design flaw, we probably wouldn't feel like rejecting the hypothesis and concluding that the aircraft is well designed. In fact, the test indicates that, in this case, one in a thousand aircraft could crash. Probability theory thus allows us to quantify the possibilities that are taken into consideration in the study of a problem, but often in the final decision, it is necessary to make a cost/benefit analysis to take also into account factors that are not strictly mathematical or statistical by nature. However, there are cases in which the decision is easy, because a too small significance level is reached to practically coincide with certainty. For example, if 420 heads are obtained in a thousand flips of a coin, the standard variable would be 5.1 and the probability of being wrong by rejecting the true hypothesis of "fair coin" would be basically zero.

In the real experiment as shown in Table 2.2, and by Problem 2.10, 479 heads per thousand tosses have been obtained. This result corresponds to a value of the standard variable $t = |479 - 500|/15.8 = 1.39$, to which Table E.1 assigns a value of 0.4177, corresponding to a significance level of 8.2%. Here we have a first defined point in the analysis of the experiment of Table 2.2 (that we will reconsider many times in the following): the global number of heads obtained is in reasonable agreement with the binomial model and a priori probability of $1/2$.

In the previous exercise, the so-called one-tailed test was used. In the following example, the two-tailed test will be used, in which values to the left and to the right of a fixed interval are discarded.

Exercise 3.14
What is the probability to be wrong by adopting as a decision rule to define a fair coin (i.e. with $1/2$ probability) a number of hits in a thousand flips between 450 and 550?

Answer Using the data from the previous exercise, we immediately obtain the result:

$$P\{|T| > t = 3.16\} = 2\,(0.5000 - 0.4992) = 1.6 \cdot 10^{-3}\,.$$

This means that about 2 fair coins are discarded out of 1000 assuming that *all tested coins are fair*.

Referring to this last exercise, we could wonder how many bad coins are accepted. In other words, what is the probability to accept a false hypothesis as true? Generally,

finding this probability is not easy, because you should know the a priori probability of all the variables related to the problem. In the previous example, the true probabilities of all coins used in the test should be known, as explained in the following problem.

Exercise 3.15
In the coin stock of the previous exercise, there is one coin with a priori probability $p = 0.6$ for the face of interest. Which is the probability of accepting it as fair while still adopting $450 \leq x \leq 550$ as the decision rule?

Answer We have to calculate $P\{450 \leq X \leq 550\}$ for a Gaussian with mean and variance given by:

$$\mu = 0.6 \cdot 1000 = 600$$
$$\sigma = \sqrt{1000 \cdot 0.6 \cdot (1 - 0.6)} = 15.5 \, .$$

By considering the two standard values:

$$t_{450} = \frac{450 - 600}{15.5} = -9.68$$

$$t_{550} = \frac{550 - 600}{15.5} = -3.22 \, ,$$

from routine `pnorm`, we obtain the probability:

$$P\{450 \leq X \leq 550\} = \texttt{pnorm}(-3.22) - \texttt{pnorm}(-9.68) = 0.00064$$

which corresponds to a wrong probability of acceptance of about 6 over 10,000.

In the examples discussed so far, the Gaussian approximation of the binomial distribution has always been used. Here is a different situation.

Exercise 3.16
In a population of 10,000 inhabitants, historical data on a rare disease give four cases per year. If there are ten cases in a year, is there an increase in the disease or is it simply a statistical fluctuation?

(continued)

Exercise 3.16 (continued)
Answer The most reasonable working hypothesis is to assume a Poisson distribution with $\mu = 4$ as null hypothesis. In this case, we cannot use the Gaussian approximation, which requires $\mu > 10$, and we are forced to use Eq. (3.94). The probability of observing, by pure chance, at least ten cases of illness, when the annual average is four, is then:

$$\begin{aligned} P\{X \geq 10\} &= \sum_{x=10}^{\infty} \frac{(4)^x}{x!} e^{-4} = 1 - e^{-4} \sum_{x=0}^{9} \frac{(4)^x}{x!} \\ &= 1 - (0.0183 + 0.0733 + 0.1465 + 0.1954 + 0.1954 \\ &\quad + 0.1563 + 0.1042 + 0.0595 + 0.0298 + 0.0132) \\ &= 1 - 0.9919 \simeq 8 \cdot 10^{-3} \,. \end{aligned}$$

Alternatively, one can use the R command: `x=1-ppois(9,lambda=4)` which gives a result of `x=0.0081`.
This value represents the observed significance level of the hypothesis, that is, the probability to discard a true hypothesis. We can therefore reasonably say that we are in the presence of an increase of the disease, with a probability of being wrong, if the null hypothesis were correct, of about 8 per thousand.

Finally, let us consider a very instructive example, which can be considered as the paradigm of the way science operates as regards the possible rejection (falsification) of a theory.

Exercise 3.17
A committee of astrologers interviews a person, without knowing its personal details, to try to identify its zodiac sign. At the end of the examination, three zodiac signs, one correct and the other two incorrect, are submitted to the committee. If the commission got at least 50 successes on 100 tests, could astrology be reasonably considered a science?

Answer Let us assume pure chance as the null hypothesis. In this case, the probability for the committee to give the correct answer just by guessing it is equal to $1/3$, for each examined person. The terms of the problem are then:

- p.d.f.: binomial
- Number of attempts: $n = 100$
- Success probability at each attempt: $p = 1/3$

(continued)

Exercise 3.17 (continued)
- Number of successes: $x = 50$
- Expected mean of the distribution: $\mu = np = 33.3$
- Standard deviation: $\sigma = \sqrt{np(1-p)} = 4.7$

Since the conditions $np > 10$, $n(1-p) > 10$ are valid, we can use Gaussian probabilities to evaluate the probability level of the standard value:

$$t = \frac{50 - 33.3}{4.7} = 3.55\,.$$

The probability to get at least 50 hits is computable from Table E.1 of Appendix E:

$$P\{X \geq 50\} = P\{T \geq 3.55\} = 0.5000 - 0.4998 \simeq 2 \cdot 10^{-4}\,,$$

or with the R command: `1-pnorm(3.55)=0.0001926`. We can therefore reject the hypothesis of randomness, with a probability to be wrong of around 2 over 10,000.
If a series of results of this kind would be achieved, astrology would acquire scientific dignity and could well be taught in schools. However, reality is quite different: in 1985, the *Nature* magazine [Car85] reported the results of this test, conducted by a mixed committee of scientists and astrologers. The success rate, in 120 trials, was 34%. To date, no significant deviations from the laws of chance have been published, both for astrology and many other pseudo-sciences such as telepathy and clairvoyance, in any journal accredited by the international scientific community. In other words, we can state that, in this kind of experiments, the hypothesis of pure chance has never been falsified, that is, that astrology, telepathy and clairvoyance have no scientific validity.

In these last exercises, we have discussed whether or not to accept an experimental value, having previously adopted a certain probabilistic model *assumed as true*. These topics, which are the subject of statistics, will be discussed in detail later, starting from Chap. 6. However, we think it was helpful to have a preliminary look at these interesting problems.

3.12 Problems

3.1 Solve Problem 1.9 (i.e. to find probability $P\{X \leq Y\}$ for two uniform random variables $0 \leq X, Y \leq 1$) without using geometrical arguments.

3.2 Random walk: a particle moves in steps and, at each step Δx, it may remain at rest ($\Delta x = 0$) or deviate by $\Delta x = +1$ with the same probability. Calculate, after 500 steps, mean value and standard deviation of the path X.

3.3 Unlike the problem above, this time the particle chooses at each step, with equal probability $p = 1/2$, the deviations $\Delta x = -1$ and $\Delta x = +1$.

3.4 The probability to transmit a wrong bit is 10^{-3}. Calculate the probability that (a) a wrong bit is present in a 16-bit number and (b) the mean number of wrong bits.

3.5 The probability that a person has the flu virus in a certain season is 20%. Find the probability that, in a room with 200 people, the carriers of the flu are between 30 and 50.

3.6 Find, for $n \to \infty$, the density function of the variable $Y = X_1 X_2 \ldots X_n$, where X_i are positive random variables satisfying the central limit theorem conditions.

3.7 Sometimes, to describe the width of a density function, the "full width at half maximum" is used, which is defined as FWHM$= |x_2 - x_1|$, where $p(x_1) = p(x_2) = p_{\max}/2$. Find the relation between FWHM and σ for a Gaussian.

3.8 In a hospital there is, on average, one twin birth every 3 months. Assuming a negative exponential distribution, determine (a) the probability of not having twin births for at least 8 months and (b) the probability of not having twin births within a month if 8 months have passed without twin births.

3.9 The sum of the squares of 10 standard variables is 7. Find the probability to be wrong by stating that the variables are not Gaussian and independent.

3.10 Historical data shows 8500 deaths per year from traffic accidents. During the year following the introduction of seat belts, the deaths drop to 8100. Considering the historical data as the true average value and the annual data as a random variable, evaluate whether seat belts have significantly decreased the number of deaths at the 1% level.

3.11 If the average frequency of a Poissonian process is 100 events per second, calculate the fraction of time intervals between two events less than 1 millisecond.

3.12 A certain amount of radioactive substance emits a particle every 2 s. During a test performed on a sample, there were no counts for 10 s and it is concluded that this substance is absent. What is the probability that the conclusion is wrong?

3.13 Prove that, if Poisson's law (3.47) holds, the counts in disjoint time intervals are independent random variables.

3.14 Find the interval, centred on the mean, that contains with 50% probability the values of a standard Gaussian variable.

3.15 Find the mean and standard deviation of a Gaussian, knowing that the probability of obtaining values greater than 4.41 is 21% and that of obtaining values greater than 6.66 is 6%.

3.16 100 events were recorded in 5 days from a Poissonian process. Calculate how many days, on average, must pass before recording four events in 1 h.

3.17 Using Theorem 3.5, find an algorithm to generate random variables from the density $p(x) = 2x - 2$, $1 \leq x \leq 2$.

3.18 A sample of Gaussian-distributed electrical resistances has a mean value of 100 Ω, and a standard deviation of 5 Ω.
(a) What is the probability that a resistance value deviates by more than 10% from the expected value? (b) What is the probability that 10 resistances in series have a value $\geq 1050\ \Omega$? (c) What upper limit can be derived, for case (a), abandoning the Gaussian assumption?

3.19 In a rare decay process, a counter can record from 0 to 3 counts with the following probabilities:

X	0	1	2	3
probability	0.1	0.4	0.4	0.1

Considering this to be the true distribution, and using the Gaussian approximation, calculate the probability that in a month (30 days) a total number of counts greater than 80 is recorded, assuming independent daily counts.

3.20 An electronic company manufactures a particular component with a 5% percentage of defective parts. After having sold several batches of 200 items, the company declares a maximum of 15 defective pieces. Find the percentage of batches that do not meet the sales specification.

Chapter 4
Multivariate Probability Theory

The non-mathematician is seized by a mysterious shuddering when he hears of four-dimensional things, by a feeling not unlike that awakened by thoughts of the occult. And yet there is no more common-place statement than that the world in which we live is a four-dimensional space-time continuum.

Albert Einstein, "RELATIVITY, THE SPECIAL AND THE GENERAL THEORY; A POPULAR EXPOSITION".

4.1 Introduction

We will now address topics a little more complex than those of the previous chapter. However, this effort will be repaid by the results we will obtain, which are necessary for understanding and dealing with problems involving several random variables. In order not to unnecessarily complicate the mathematical formalism, the problems will be discussed initially for the case of two random variables. Then, since the hypothesis of only two variables will never enter into the proofs of theorems and into the discussion, the obtained results will easily be extended to any number of variables. They will also be mainly presented in integral form, considering continuous variables. The transition to the case of discrete variables is immediately obtained with transformations like:

$$\int_a^b (\dots)\, p(x)\, \mathrm{d}x \quad \rightarrow \quad \sum_{x=a}^{b} (\dots)\, p(x)\ , \tag{4.1}$$

where $(\dots)$ denotes any expression containing random variables and parameters. Equation (4.1) is immediately extendable also to the case of several variables.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3_4

4.2 Multivariate Statistical Distributions

If X and Y are two random variables defined on the same probability space, we define as A the event where x is within two values x_1 and x_2, $\{x_1 \leq X \leq x_2\}$ and similarly the event B as $\{y_1 \leq Y \leq y_2\}$. We write the compound probability $P(AB)$ as:

$$P(AB) \equiv P\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\} , \tag{4.2}$$

which is the probability that the value (x, y) falls into $[x_1, x_2] \times [y_1, y_2]$. A function $p(x, y) \geq 0$ such as:

$$P\{x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2\} = \int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y) \, \mathrm{d}x \, \mathrm{d}y \tag{4.3}$$

is called *joint probability density* of the two variables. It is the bidimensional extension of the p.d.f. defined in Sect. 2.7.

More generally, if $A \in \mathbb{R}^2$, it can be shown that if $p(x, y)$ satisfies Eq. (4.3), the probability that $(X, Y) \in A$ (e.g. $x + y \leq a$, with a constant) is given by [PUP02]:

$$P\{(X, Y) \in A\} = \int_A p(x, y) \, \mathrm{d}x \, \mathrm{d}y , \tag{4.4}$$

which becomes equivalent to the integral (3.94) and to the cumulative or distribution function (2.33) when:

$$P\{(X, Y) \in A\} = P\{-\infty \leq X \leq a, -\infty \leq Y \leq b\} .$$

Both the normalization condition and the variable means and variances are obtained as the immediate generalization of the corresponding one-dimensional formulae:

$$\begin{aligned} \int \int p(x, y) \, \mathrm{d}x \, \mathrm{d}y &= 1 , \\ \langle X \rangle = \int \int x \, p(x, y) \, \mathrm{d}x \, \mathrm{d}y &= \mu_x , \\ \langle Y \rangle = \int \int y \, p(x, y) \, \mathrm{d}x \, \mathrm{d}y &= \mu_y , \\ \mathrm{Var}[X] = \int \int (x - \mu_x)^2 \, p(x, y) \, \mathrm{d}x \, \mathrm{d}y &= \sigma_x^2 , \\ \mathrm{Var}[Y] = \int \int (y - \mu_y)^2 \, p(x, y) \, \mathrm{d}x \, \mathrm{d}y &= \sigma_y^2 , \end{aligned} \tag{4.5}$$

where the integration is extended to $(-\infty, +\infty)$, that is, to the whole range of existence of the density function. As in the one-dimensional case (see Definition 2.10), the absolute integrability or summability is required for the existence of mean values.

In the following, multiple integration will be indicated simply with the single integral symbol. According to Eq. (2.9), if X and Y are stochastically independent, the compound probability theorem and Eqs. (1.21, 1.24, 2.9) allow us to write, for each pair (x_1, x_2) and (y_1, y_2):

$$P\{x_1 \le X \le x_2, y_1 \le Y \le y_2\} = P\{x_1 \le X \le x_2\} P\{y_1 \le Y \le y_2\} ,$$

that is:

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} p(x, y) \, \mathrm{d}x \, \mathrm{d}y = \int_{x_1}^{x_2} p_X(x) \, \mathrm{d}x \int_{y_1}^{y_2} p_Y(y) \, \mathrm{d}y .$$

Therefore, the joint density can be defined as:

$$p(x, y) = p_X(x) \, p_Y(y) \quad \text{(if } X \text{ and } Y \text{ are independent)} , \tag{4.6}$$

where $p_X(x)$ and $p_Y(y)$ are the p.d.f. of the variables X and Y.

It is also possible to define means and variances of combinations of variables. For example:

$$\begin{aligned} \langle XY \rangle &= \int xy \, p(x, y) \, \mathrm{d}x \, \mathrm{d}y = \mu_{xy} , \\ \langle X + Y \rangle &= \int (x + y) \, p(x, y) \, \mathrm{d}x \, \mathrm{d}y = \mu_{x+y} , \\ \mathrm{Var}[XY] &= \int (xy - \mu_{x \cdot y})^2 \, p(x, y) \, \mathrm{d}x \, \mathrm{d}y , \\ \mathrm{Var}[X + Y] &= \int (x + y - \mu_{x+y})^2 \, p(x, y) \, \mathrm{d}x \, \mathrm{d}y . \end{aligned} \tag{4.7}$$

It is easy to show, from the second and third of Eqs. (4.5) and from the second of Eqs. (4.7), that:

$$\langle X + Y \rangle = \langle X \rangle + \langle Y \rangle , \tag{4.8}$$

and that, if X and Y are stochastically independent, from the first of Eqs.(4.7) and from Eq. (4.6) it follows:

$$\langle XY \rangle = \langle X \rangle \, \langle Y \rangle . \tag{4.9}$$

So far, there does not seem to be anything new, other than the obvious generalization of one-dimensional formulae. However, multidimensional distributions immediately reserve surprises, because definitions of at least three important new quantities between variables, i.e. *marginal density*, *conditional density* and *covariance*, can be defined.

Definition 4.1 (Marginal Density) If X and Y are two random variables with density $p(x, y)$ and A is an interval on the real axis, the marginal density $p_X(x)$

is defined as:

$$P\{X \in A\} \equiv \int_A p_X(x)\,\mathrm{d}x = \int_A \mathrm{d}x \int_{-\infty}^{+\infty} p(x, y)\ \mathrm{d}y\ , \tag{4.10}$$

from which:

$$p_X(x) = \int_{-\infty}^{+\infty} p(x, y)\ \mathrm{d}y\ . \tag{4.11}$$

The marginal density $p_Y(y)$ of y is obtained from the previous equation by substituting x with y.

It is easy to check that marginal densities are normalized:

$$\int p_X(x)\,\mathrm{d}x = \int p(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1\ .$$

These marginal densities give the probability of events of the type $\{X \in A\}$ for any value of Y (and vice versa); they therefore represent the one-dimensional probability densities of the variables X and Y. The following theorem is useful to establish a very important property, i.e. whether two or more variables are independent.

Theorem 4.1 (Independence of Variables) *Two random variables (X, Y), of joint density $p(x, y)$, are stochastically independent if and only if there are two functions $g(x)$ and $h(y)$ such that, for each $x, y \in \mathbb{R}$, we have:*

$$p(x, y) = g(x)\,h(y)\ . \tag{4.12}$$

Proof If (X, Y) are independent, Eq. (4.6) proves the first part of theorem with $g(x) = p_X(x)$ and $h(y) = p_Y(y)$. If, instead, Eq. (4.12) holds, one has in general:

$$\int_{-\infty}^{+\infty} g(x)\ \mathrm{d}x = G\ , \qquad \int_{-\infty}^{+\infty} h(y)\ \mathrm{d}y = H\ .$$

Since $p(x, y)$ is a normalized p.d.f., from the first of Eqs. (4.5) it results (the integration limits are implicit):

$$H\,G = \int g(x)\ \mathrm{d}x \int h(y)\ \mathrm{d}y = \int p(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1\ .$$

It is then possible to define two normalized marginal densities:

$$p_X(x) = \int g(x)h(y)\,\mathrm{d}y = Hg(x)\ , \qquad p_Y(x) = \int g(x)h(y)\,\mathrm{d}x = Gh(y)\ ,$$

from which , since $HG = 1$, Eq. (4.6) is obtained:

$$p(x, y) = g(x)\, h(y) = g(x)\, h(y) H\, G = p_X(x)\, p_Y(y)\ .$$

□

At a fixed value x_0, the function $p(x_0, y)$ should represent a one-dimensional p.d.f. of the variable Y. However, since $p(x_0, y)$ is not normalized, keeping in mind Eq. (4.11), we arrive to the following definition.

Definition 4.2 (Conditional Density) If X and Y are two random variables with joint density $p(x, y)$, the conditional density $p(y|x_0)$ of y for every fixed $x = x_0$ such that $p_X(x_0) > 0$, is given by:

$$p(y|x_0) = \frac{p(x_0, y)}{p_X(x_0)} = \frac{p(x_0, y)}{\int_{-\infty}^{+\infty} p(x_0, y)\ \mathrm{d}y}\ . \tag{4.13}$$

Also in this case, the conditional density $p(x|y)$ of x with respect to y is obtained by exchanging the two variables in the previous formula.

The conditional densities just defined are normalized. Indeed:

$$\int p(y|x_0)\, \mathrm{d}y = \frac{\int p(x_0, y)\, \mathrm{d}y}{\int p(x_0, y)\ \mathrm{d}y} = 1\ .$$

It is important to note that *the conditional density $p(y|x)$ is only a function of y*, since x is a fixed value (parameter). The same goes for $p(x|y)$ by swapping the two variables. Therefore, it is wrong to write:

$$P\{Y \in B|X \in A\} = \int_A \int_B p(y|x)\, \mathrm{d}x\, \mathrm{d}y = \int_A \int_B \frac{p(x, y)}{p_X(x)}\, \mathrm{d}x\, \mathrm{d}y \quad \text{(wrong!)}\ ,$$

since $p(y|x)$ is a function of y only. To find the correct formula, it is necessary to refer to the definition of conditional probability given by Eq. (1.19):

$$\begin{aligned} P\{Y \in B|X \in A\} &= \frac{P\{X \in A, Y \in B\}}{P\{X \in A\}} \\ &= \frac{\int_A \int_B p(x, y)\, \mathrm{d}x\, \mathrm{d}y}{\int_A \mathrm{d}x \int_{-\infty}^{+\infty} p(x, y)\, \mathrm{d}y}\ . \end{aligned}$$

The conditional mean and variance operators can be defined for a given fixed value. For example, the expected value of Y conditional on a x value is given by:

$$\langle Y|x\rangle = \int_{-\infty}^{+\infty} y\, p(y|x)\, \mathrm{d}y = \frac{\int_{-\infty}^{+\infty} y\, p(x, y)\, \mathrm{d}y}{p_X(x)}\ , \tag{4.14}$$

where $p_X(x) = \int p(x, y)\, \mathrm{d}y$ is constant because x is fixed.

The marginal and conditional densities descend from the compound probability Theorem 1.2. In fact, by inverting Eq. (4.13), we can write:

$$p(x, y) = p_Y(y)\, p(x|y) = p_X(x)\, p(y|x)\,, \tag{4.15}$$

which corresponds to the compound probability formula of Eq. (1.21) for continuous variables, i.e. the density of (X, Y) is given by the density of Y times the density of X for each fixed Y (or vice versa). When variables are independent, from Eqs. (4.6 and 4.13), one obtains:

$$p(x|y) = p_X(x)\,, \quad p(y|x) = p_Y(y)\,, \quad \text{(independent } X \text{ and } Y\text{)}. \tag{4.16}$$

The difference between marginal and conditional densities can be well understood with the help of Fig. 4.1; the marginal density is obtained by simply projecting the function in one dimension, while the conditional density is the projection of the function on the y axis (i.e. the y p.d.f.) for a given x value (or vice versa). We also note that means and variances defined in (4.5) can be expressed using the marginal densities:

$$\mu_x = \int x p_X(x)\, \mathrm{d}x\,, \quad \mu_y = \int y p_Y(y)\, \mathrm{d}y\,. \tag{4.17}$$

$$\sigma_x^2 = \int (x - \mu_x)^2 p_X(x)\, \mathrm{d}x\,, \quad \sigma_y^2 = \int (y - \mu_y)^2\, p_Y(y)\, \mathrm{d}y\,. \tag{4.18}$$

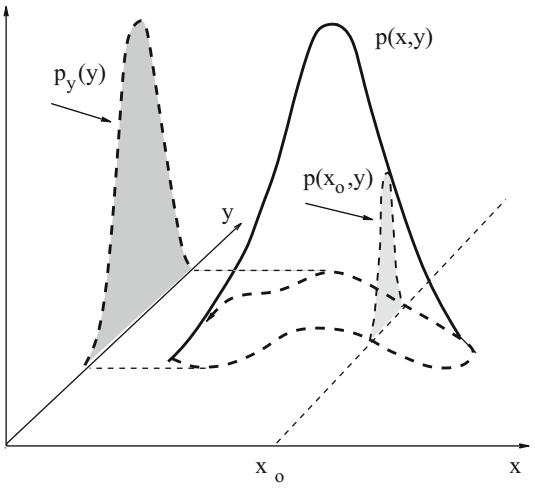


Fig. 4.1 The marginal density $p_Y(y)$ of y is the projection of the density of the (x, y) points on the y axis for any x value. Instead, the function $p(x_0, y)$ is the projection of the density on the y axis for a fixed value x_0 (dashed line). This function, after normalization with Eq. (4.13), is the conditional density of Y for a selected value x_0

Let us examine now the fourth of Eq. (4.7). Since, from Eq. (4.8), it results that the mean of a sum is equal to the sum of the means, we obtain:

$$\begin{aligned}\mathrm{Var}[X+Y] &= \int [(x-\mu_x)+(y-\mu_y)]^2 p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\ \mathrm{Cov}[X,Y]\,,\end{aligned} \tag{4.19}$$

where:

$$\mathrm{Cov}[X,Y] \equiv \int (x-\mu_x)(y-\mu_y)p(x,y)\,\mathrm{d}x\,\mathrm{d}y\ .$$

Therefore, we introduce the following definition.

Definition 4.3 (Covariance of Two Variables) The covariance of two random variables X and Y is defined as:

$$\mathrm{Cov}[X,Y] = \int\int (x-\mu_x)(y-\mu_y)\ p(x,y)\,\mathrm{d}x\,\mathrm{d}y \equiv \sigma_{xy}\ . \tag{4.20}$$

This parameter is a sort of "crossed" definition between the two possible variances σ_x^2 and σ_y^2 of Eqs. (4.5). It is a quadratic quantity, having as dimension the product of the X and Y dimensions. As an exercise, let us transcribe the general Definition (4.3) for the various possible cases, similarly to what has been done for the variance in Eqs. (2.38, 2.42, 2.67). In the case of a discrete density p_{XY}, covariance becomes:

$$\mathrm{Cov}[X,Y] = \sum_x \sum_y (x-\mu_x)(y-\mu_y)p_{XY}(x,y)\ , \tag{4.21}$$

where the sum is extended to the full support of the two variables. Instead, the direct computation from a dataset must be performed on the sum of all the numerical realizations of the variables:

$$s_{xy} = \frac{1}{N}\sum_i (x_i-\mu_x)(y_i-\mu_y)\ , \tag{4.22}$$

$$s_{xy} = \frac{1}{N-1}\sum_i (x_i-m_x)(y_i-m_y)\ . \tag{4.23}$$

Here the sum must be made only over one index, that is on all the observed (x_i, y_i) pairs. Notice that the sum of the deviations from the sample means must be divided by $N-1$. The set of these pairs exhausts, for $N \to \infty$, the set given by the pairs (x_i, y_k), where the indexes i and k select the values of the support of the two variables. This is a formal but important point that must be kept in mind in order to perform correct calculations: *when covariance is computed through the*

density function, the summation is double and must be done on the probabilities of all possible values of the two paired variables; when covariance is calculated from an experimental set of data, the sum is single and must be done on all the pairs obtained in the sampling. The R software provides the function `cov(x,y)` which calculates the covariance from Eq. (4.23), where `x` and `y` are two vectors with the raw experimental data.

Our routine `CovarHisto(x, y, matfre)` can be used to calculate covariance, using Eqs. (4.21), also from histograms. Here `x` and `y` are the vectors of the measured values and `matfre` is the matrix containing the frequencies or the number of events. Finally, we note that we have, in operator notation:

$$\mathrm{Cov}[X, Y] = \left\langle (X - \mu_x)(Y - \mu_y) \right\rangle \ . \tag{4.24}$$

Also the equation:

$$\begin{aligned} \mathrm{Cov}[X, Y] &= \left\langle (X - \mu_x)(Y - \mu_y) \right\rangle = \left\langle XY - \mu_x Y - \mu_y X + \mu_x \mu_y \right\rangle \\ &= \langle XY \rangle - \mu_x \langle Y \rangle - \mu_y \langle X \rangle + \mu_x \mu_y \\ &= \langle XY \rangle - \mu_x \mu_y \ , \end{aligned} \tag{4.25}$$

which corresponds to Eq. (2.51), is valid. Here we see that covariance is given by the mean of the product minus the product of the means.

Covariance has particularly interesting properties that will be discussed in the next section.

Exercise 4.1
An electronic device has an exponential lifetime of mean:

$$\mu = 1/\lambda = 1000 \text{ hours}$$

(see Eq. 3.49). An instrument composed by two devices in parallel works when at least one of them is operational. How much longer is the average lifetime of the instrument than that of the single device? What is the probability that the instrument and the single device are still working after 2000 h of functioning?

Answer If t_1 and t_2 are the failure times of the two devices, the instrument will stop working at a time t such that $t = \max(t_1, t_2)$. The probability of operation up to t, that is, of a failure at time t, will be given by the compound probability that both the first *and* the second device stop working. Since the two devices are independent, it is sufficient to integrate over $[0, t]$ the joint

(continued)

Exercise 4.1 (continued)
lifetime density (4.6), according to Eqs. (4.4, 4.12):

$$P\{\max(T_1, T_2) \le t\} = P\{T_1 \le t, T_2 \le t\} = \int_0^t e_1(t)\,\mathrm{d}t \int_0^t e_2(t)\,\mathrm{d}t \ .$$

Since the two exponential distributions e_1 ed e_2 are identical, from Eq. (3.51) one immediately has the cumulative function of the failure times of the instrument:

$$P\{\max(T_1, T_2) \le t\} \equiv P(t) = (1 - e^{-\lambda t})^2 \ .$$

By differentiating this distribution function, we will obtain the corresponding p.d.f. $e_p(t)$. We therefore have:

$$e_p(t) = 2\lambda e^{-\lambda t}\,(1 - e^{-\lambda t}) \ ,$$

which *is not* a simple exponential. The average of the failure times:

$$\langle T \rangle = 2\lambda \int_0^\infty t\,(e^{-\lambda t} - e^{-2\lambda t})\,\mathrm{d}t = \frac{3}{2}\frac{1}{\lambda} = \frac{3}{2}\,\mu = 1500 \ \text{hours} \ ,$$

correspond to a 50% increase in the mean life.

Now we consider the second part of the problem. According to Eqs. (3.51, 3.52), the probability for the single device to be working after $t_0 = 2000$ hours is given by:

$$P\{T \ge t_0\} = 1 - \int_0^{t_0} \lambda e^{-\lambda \tau}\,\mathrm{d}\tau = e^{-\lambda t_0} = e^{-2} = 0.135 = 13.5\% \ .$$

For the instrument with the parallel devices, the same probability is given by:

$$\begin{aligned} P_p\{T \ge t_0\} &= 1 - \int_0^{t_0} e_p(t)\,\mathrm{d}t = 1 - 2\lambda \int_0^{t_0} (e^{-\lambda t} - e^{-2\lambda t})\,\mathrm{d}t \\ &= 2e^{-\lambda t_0} - e^{-2\lambda t_0} = 2e^{-2} - e^{-4} = 0.252 \simeq 25\% \ . \end{aligned}$$

Then, after 2000 h, the improvement in reliability is about 100%.

4.3 Covariance and Correlation

The covariance of two random variables has the important property of vanishing when the variables are independent. In fact, in such condition, the density $p(x, y)$ appearing in Eq. (4.20) can be written, according to Eq. (4.6), as the product $p_X(x)p_Y(y)$. Since $p_X(x)$ and $p_Y(y)$ are normalized, one easily obtains:

$$\begin{aligned}\mathrm{Cov}[X, Y] &= \int\int (x-\mu_x)(y-\mu_y)\, p(x, y)\, \mathrm{d}x\, \mathrm{d}y \\ &= \int (x-\mu_x)\, p_X(x)\, \mathrm{d}x \int (y-\mu_y)\, p_Y(y)\, \mathrm{d}y \\ &= \left[\int x\, p_X(x)\, \mathrm{d}x - \mu_x \int p_X(x)\, \mathrm{d}x\right]\left[\int y\, p_Y(y)\, \mathrm{d}y - \mu_y \int p_Y(y)\, \mathrm{d}y\right] \\ &= \mu_x - \mu_x + \mu_y - \mu_y = 0\,. \end{aligned} \tag{4.26}$$

From Eq. (4.19) it follows that, for mutually independent X_i, the variance of a sum is equal to the sum of variances:

$$\mathrm{Var}\left[\sum_i X_i\right] = \sum_i \mathrm{Var}[X_i]\,. \tag{4.27}$$

Covariance therefore is a statistical indicator of correlation: the more it is different from zero, the larger is the correlation between variables. But how to evaluate the maximum degree of correlation? How to make it independent of the dataset you are examining? This difficulty can be overcome thanks to the following theorem:

Theorem 4.2 (of Cauchy-Schwarz) *If the variables X and Y have finite variances, then the following inequality holds:*

$$|\,\mathrm{Cov}[X, Y]| \le \sigma[X]\,\sigma[Y]\,. \tag{4.28}$$

Proof Equation (4.28) can be written as:

$$|\,\langle (X-\mu_x)(Y-\mu_y)\rangle\,| \le \sqrt{\langle (X-\mu_x)^2\rangle\, \langle (Y-\mu_y)^2\rangle}\,. \tag{4.29}$$

If the centred variables $X_0 = X - \mu_x$ and $Y_0 = Y - \mu_y$ are considered, Eq. (4.29) becomes:

$$\langle X_0 Y_0\rangle^2 \le \left\langle X_0^2\right\rangle \left\langle Y_0^2\right\rangle\,. \tag{4.30}$$

Let t be any real number and consider the variable $(tX_0 - Y_0)$. By exploiting the linearity properties of the mean, we easily obtain:

$$0 \leq \left\langle (tX_0 - Y_0)^2 \right\rangle = t^2 \left\langle X_0^2 \right\rangle - 2t \left\langle X_0 Y_0 \right\rangle + \left\langle Y_0^2 \right\rangle .$$

This second degree polynomial in t is always non-negative if and only if the discriminant remains ≤ 0. We therefore have:

$$4 \left\langle X_0 Y_0 \right\rangle^2 - 4 \left\langle X_0^2 \right\rangle \left\langle Y_0^2 \right\rangle \leq 0 ,$$

which is just Eq. (4.30). Therefore, Eq. (4.28) is verified.

In this equation the equality holds when

$$\left\langle (tX_0 - Y_0)^2 \right\rangle = 0 \quad \Longrightarrow \quad Y_0 = tX_0 \quad \Longrightarrow \quad Y = tX - t\mu_x + \mu_y \equiv aX + b ,$$

that is, when there is a linear dependence between X and Y. □

The definition of covariance and the Cauchy-Schwarz theorem leads intuitively to define the correlation coefficient between two variables as:

$$\rho_{xy} \equiv \rho[X, Y] = \frac{\mathrm{Cov}[X, Y]}{\sigma[X]\,\sigma[Y]} . \tag{4.31}$$

This coefficient lies between the limits:

$$-1 \leq \rho_{xy} \leq 1 ; \tag{4.32}$$

its values are null if variables are independent—are positive if they are correlated, i.e. when one of them increases (decreases) as the other variable increases (decreases), or are negative (anticorrelation) if an increase of one variable tends to be associated with a decrease of the other one. Finally, the $\rho = \pm 1$ limits occur when a linear relation of the type $Y = aX + b$ exists between the two variables. In this case we basically have a single random variable, written in two mathematically different ways.

In R, the correlation coefficient between two variables can be calculated with the simple command `cov(x,y)/sqrt(var(x)*var(y))`, or more briefly with `cor(x,y)`, where the vectors `x` and `y` contain the values of X and Y.

We must now discuss a very delicate point: the connection between statistical independence, correlation and causality. According to Definition 1.10, statistical independence occurs when the probability of the event can be factored into the product of single probabilities of each variable, as in Eq. (2.9) for discrete variables or in Eq. (4.6) for continuous ones. For correlation, we use the following definition:

Definition 4.4 (Correlation Between Variables) Two random variables are said to be uncorrelated when their correlation coefficient is zero; otherwise, they are said to be correlated.

Obviously, this definition depends on the correlation coefficient that is used in data analysis. For now, let us use the correlation coefficient defined in Eq. (4.31); in the following (see Eq. (11.10)) we will give the more general definition of this coefficient. We also note that here the meaning of correlation is very technical and does not necessarily coincide with the meaning that common sense assigns to this parameter. Let us analyse this problem in detail. We immediately notice that, if variables are correlated (i.e. if the coefficient ρ_{xy} of Eq. (4.31) is different from zero), then there exists some degree of dependence between them. However, the inverse is not true, and a simple counterexample is enough to prove this fact. Let us consider $X = U + V$ and $Y = U - V$, where U and V are uniform variables in $[0, 1]$; X and Y are dependent because Eq. (4.12) does not hold, but their correlation coefficient is zero because of a vanishing covariance. Indeed, since $\langle U + V \rangle = 1$ and $\langle U - V \rangle = 0$:

$$\begin{aligned}
\mathrm{Cov}[X, Y] &= \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle \\
&= \langle (U + V - \langle U + V \rangle)(U - V - \langle U - V \rangle) \rangle \\
&= \langle (U + V - 1)(U - V) \rangle = \left\langle U^2 - V^2 - U + V \right\rangle = \\
&= \left\langle U^2 \right\rangle - \left\langle V^2 \right\rangle - \langle U \rangle + \langle V \rangle = 0 \,,
\end{aligned}$$

because U and V have the same distribution. Another key point, as Eq. (4.26) shows, is that two statistically independent variables are also uncorrelated. We can then summarize our discussion as in the following:

- A sufficient condition for statistical dependence is the presence of a correlation.
- A necessary condition for statistical independence is the absence of correlation ($\rho_{xy} = 0$).

These properties are also recapped in Fig. 4.2.

A final important point relates to the connection between statistical correlation and causality. The important fact is that *a statistical correlation between two phenomena does not imply a causation relationship between them*. In fact, it is extremely easy to find artifacts and so-called spurious correlations: if you explore the web, you will easily find statistical correlations between the number of movies played by a well-known actor and the suicide trend, or between the increase in the pet sale and that of coffee pods and other similar amenities. Unfortunately, correlation and causation are often confused, and mistaking a statistical correlation for a cause-effect relationship is one of the most frequent and dangerous errors made by analysts.

A cause-effect link *can never be obtained with statistical analysis*, but only with the use of specific models.

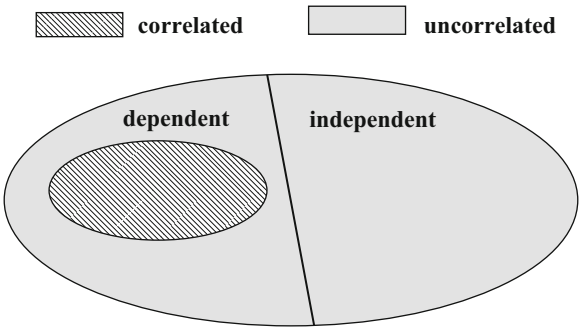


Fig. 4.2 Link between correlation and statistical dependence. For Gaussian variables, the set of (un)correlated variables coincide with the set of (in)dependent variables

Conversely, the study of statistical dependencies and correlations can instead be useful to verify models and theories. The approach must therefore be exactly the opposite: if a model or laboratory experiments suggest a cause-effect relationship producing a correlation (e.g. between the concentration of carbon dioxide in the atmosphere and the global Earth's temperature), from the statistical data analysis it is possible to compare the expected correlation coefficients with the experimental ones and falsify or not this model or theory. We will discuss these techniques in Chap. 7.

Exercise 4.2
If T, U and $V \sim U(0, 1)$, find the linear correlation coefficient between the variable:

$$X = T$$

and the variables:

$$Y = U\,, \quad Y_1 = X + V\,, \quad Y_2 = -X + V\,.$$

Also check the result with computer-simulated data.

Answer On the basis of Eq. (3.82), for a uniform variable U the following properties hold:

$$\langle U \rangle = 1/2\,, \quad \left\langle U^2 \right\rangle = 1/3\,, \quad \left\langle (U - 1/2)^2 \right\rangle = 1/12\,,$$

whereas for a pair of two uniform independent variables, one has:

$$\langle UV \rangle = \int r\, u(r)\, \mathrm{d}r \int s\, v(s)\, \mathrm{d}s = 1/4\,.$$

(continued)

Exercise 4.2 (continued)
From Eqs. (4.24, 4.31) one easily obtains:

$$\rho[X,Y] = \frac{\langle (T-1/2)(U-1/2)\rangle}{\left[\left\langle (T-1/2)^2\right\rangle \left\langle (U-1/2)^2\right\rangle\right]^{1/2}} = 0\,,$$

$$\rho[X,Y_1] = \frac{\langle (T-1/2)(T+V-1)\rangle}{\left[\left\langle (T-1/2)^2\right\rangle \left\langle (T+V-1)^2\right\rangle\right]^{1/2}} = \frac{1}{\sqrt{2}} \simeq 0.7071\,,$$

$$\rho[X,Y_2] = \frac{\langle (T-1/2)(-T+V)\rangle}{\left[\left\langle (T-1/2)^2\right\rangle \left\langle (-T+V)^2\right\rangle\right]^{1/2}} = -\frac{1}{\sqrt{2}} \simeq -0.7071\,.$$

The 2D plots of the three variable pairs are shown in Fig. 4.3, where it is possible to see clearly the difference between two uncorrelated variables (X, Y), two positively correlated variables (X, Y_1) and two negatively correlated variables (X, Y_2).

It is possible to test these results with a simulation directly using the `runif`, `cov` and `var` routines given by R. To generate $N = 20{,}000$ uniform variables `X, Y, Y1, Y2` and obtain their correlation coefficients, just use R in interactive way:

```
> T= <- runif(20000); U <- runif(2000); V <- runif(20000)
> X=T; Y=U; Y1=X+V; Y2=-X+V
> cov(X,Y)/sqrt(var(X)*var(Y))
>  ..... # the result is printed...
> cov(X,Y1)/sqrt(var(X)*var(Y1))
> cov(X,Y2)/sqrt(var(X)*var(Y2))
```

The correlation coefficient $\rho[X, Y]$ (and also the coefficients $\rho[X, Y_1]$ and $\rho[X, Y_2]$) are evaluated with Eq. (4.31).

From this simulation, the following values have been obtained:

$$r(x,y) = 0.007\,, \quad r(x,y_1) = 0.712\,, \quad r(x,y_2) = -0.705\,.$$

To estimate whether the difference between these "experimental" values and the theoretical ones (0, 0.7071, −0.7071) is significant or not requires the concepts that will be developed in Chap. 6.

A more complete calculation is performed in our routine `CorrelEst(X,Y)`, which gives also the errors on covariances and correlation coefficients. These last parameters will be evaluated later, in Chap. 6.

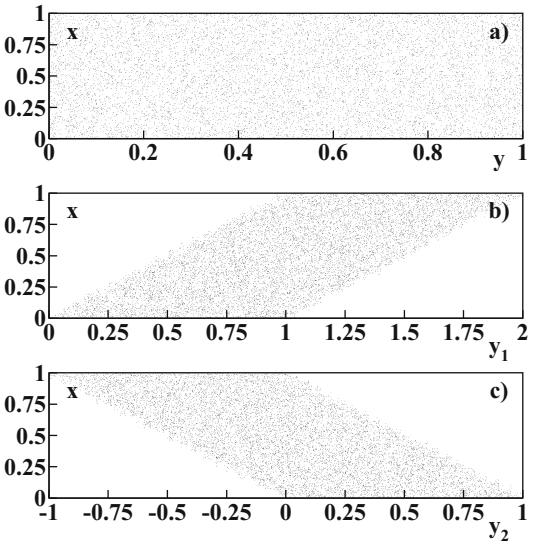


Fig. 4.3 2D plots of the pairs of variables considered in Exercise 4.2: no correlation (**a**), positive correlation (**b**) and negative correlation (**c**)

4.4 Two-Dimensional Gaussian Distribution

Even in the multidimensional case, many problems can be studied, in an exact or approximate way, assuming normal or Gaussian variables. It is therefore essential, at this point, to study the two-dimensional (named also bivariate) Gaussian density. Firstly, we immediately notice that, based on Eqs. (3.28, 4.6), the density $g(x, y)$ of two independent Gaussians variables is given by:

$$g(x, y) = g_X(x)\, g_Y(y) = \frac{1}{2\pi\sigma_x\,\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right]. \tag{4.33}$$

With the substitution:

$$U = \frac{X-\mu_x}{\sigma_x}\,, \quad V = \frac{Y-\mu_y}{\sigma_y}\,, \tag{4.34}$$

the two-dimensional analogous of the standard Gaussian density (3.42) will then be given by:

$$g(u, v; 0, 1) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(u^2+v^2)\right]. \tag{4.35}$$

However, the density we want to derive must refer to two normal variables that, in general, are mutually dependent and, therefore, with a linear correlation coefficient

$\rho_{uv} \equiv \rho \neq 0$. The corresponding standard Gaussian density must therefore satisfy the properties:

- Not to be factorizable into two separate terms depending on u and v only
- To have standard Gaussian marginal distributions
- To satisfy the equation $\sigma_{uv} = \rho$
- To satisfy the normalization requirement

Let us check if there is a functional form of the type:

$$g(u, v; 0, 1) = \frac{1}{d}\, e^{-(au^2+buv+cv^2)} , \tag{4.36}$$

which satisfies all these requirements. They correspond to the system of equations:

$$\begin{aligned}
1 &= \frac{1}{d} \int\int u^2\, e^{-(au^2+buv+cv^2)}\, \mathrm{d}u\, \mathrm{d}v , \\
1 &= \frac{1}{d} \int\int v^2\, e^{-(au^2+buv+cv^2)}\, \mathrm{d}u\, \mathrm{d}v , \\
\rho &= \frac{1}{d} \int\int uv\, e^{-(au^2+buv+cv^2)}\, \mathrm{d}u\, \mathrm{d}v , \\
1 &= \frac{1}{d} \int\int e^{-(au^2+buv+cv^2)}\, \mathrm{d}u\, \mathrm{d}v .
\end{aligned}$$

These integrals can be solved by the method of Exercise 3.4. The result is a system of four equations in the four unknowns a, b, c, d. Since the detailed procedure, though laborious, is not particularly difficult or instructive, here we report only the final result. If desired, one can easily check afterwards the correctness of the result, which, if $\rho \neq \pm 1$, turns out to be:

$$a = c = \frac{1}{2(1-\rho^2)} , \quad b = -\frac{\rho}{(1-\rho^2)} , \quad d = 2\pi\sqrt{1-\rho^2} . \tag{4.37}$$

Equation (4.36) then becomes:

$$g(u, v; 0, 1) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right] . \tag{4.38}$$

It is easy to see that this density gives rise, as required, to two standard Gaussian marginal densities *for any* $\rho \neq \pm 1$. In fact, by adding and subtracting the term $\rho^2 u^2$

in the exponent and integrating in v, one obtains:

$$g_u(u;0,1) = \frac{e^{-\frac{1}{2}u^2}}{2\pi\sqrt{1-\rho^2}} \int \exp\left[-\frac{(v-\rho u)^2}{2(1-\rho^2)}\right] \mathrm{d}v$$
$$= \frac{e^{-\frac{1}{2}u^2}}{2\pi} \int e^{-\frac{1}{2}z^2}\, \mathrm{d}z = \frac{e^{-\frac{1}{2}u^2}}{\sqrt{2\pi}}\,, \tag{4.39}$$

where $z = (v-\rho u)/\sqrt{1-\rho^2}$, $\mathrm{d}v = \sqrt{1-\rho^2}\,\mathrm{d}z$ and the basic integral (3.29) of Exercise 3.4 has been used. The same result is obtained also for the v marginal density.

Returning from reduced to normal variables, from Eq. (4.38) we finally get the density:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2}\gamma(x,y)}\,, \tag{4.40}$$

$$\gamma(x,y) = \frac{1}{1-\rho^2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right],$$

which represents the general form of the two-dimensional Gaussian density.

At this point we note a remarkable property of Gaussian variables: *when the linear correlation coefficient is null, they are independent*. In fact, setting $\rho = 0$ in Eq. (4.40), a density corresponding to the product of two Gaussians in x and y is obtained. According to Theorem 4.1, this is the density of two independent random variables. The condition $\rho = 0$, which in general is only necessary for the independence, in this case also becomes sufficient. Therefore, we have the following theorem.

Theorem 4.3 (On the Independence of Gaussian Variables) *The necessary and sufficient condition for the independence of two jointly Gaussian random variables is that their linear correlation coefficient is zero.*

We therefore see that the presence of a null covariance (which implies $\rho = 0$) between Gaussian variables ensures the statistical independence between them (see Fig. 4.2).

Reconsidering the marginal distributions (4.39), we note that, passing from standard to normal variables, two one-dimensional Gaussian distributions (3.28) in x and y are obtained, in which the correlation coefficient does not appear. Another important fact comes up here: the knowledge of two marginal distributions, that is, of the projections of the Gaussian on the x and y axes, is not enough for a complete knowledge of the two-dimensional density, because both correlated and uncorrelated variables have Gaussian projections. The knowledge of the covariance

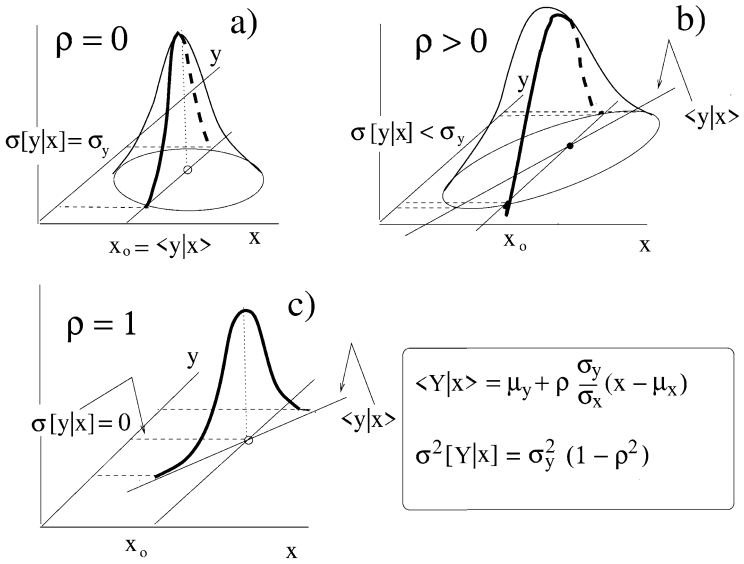


Fig. 4.4 Two-dimensional Gaussian for uncorrelated variables (**a**), partially correlated variables (**b**), totally correlated variables (**c**). Also the marginal distributions for a given x_0 value, the regression line $\langle Y|x\rangle$ and the corresponding dispersion $\sigma[Y|x]$ are shown

or of the correlation coefficients is therefore essential for the complete determination of the statistical distribution of variables.

Let us now make some further considerations on the shape of the two-dimensional Gaussian. Figure 4.4 reports three cases: uncorrelated, partially correlated and totally correlated variables. If we imagine cutting the curve with planes of constant height, the intersection gives rise to a curve of equation:

$$u^2 - 2\rho uv + v^2 = \text{constant} \tag{4.41}$$

for standard variables and to a curve:

$$\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} = \text{constant} \tag{4.42}$$

for normal variables. These curves are named concentration ellipses, centred on the point (μ_x, μ_y). If $\rho = 0$, the ellipse has the principal axes parallel to the reference axes, and it degenerates to a circumference when $\sigma_x = \sigma_y$. Finally, if $\rho = \pm 1$, the variables are completely correlated and the ellipse degenerates into a straight line of equation:

$$\frac{(x-\mu_x)}{\sigma_x} \pm \frac{(y-\mu_y)}{\sigma_y} = \text{constant}\,. \tag{4.43}$$

In this case the normal density is completely flattened on a plane, as shown in Fig. 4.4c). We then have to do with a single random variable (X or Y), which is completely dependent on the other through an analytical equation.

At this point it is necessary to focus on an important property of the two-dimensional Gaussian: when the principal axes of the ellipse are parallel to the reference axes, then $\rho = 0$, the variables are uncorrelated and therefore, by Theorem 4.3, independent. It is then always possible, by operating an axis rotation, *to transform a pair of dependent Gaussian variables into independent Gaussians variables*. This property, which, as we will see, can be generalized to any number of dimensions, is the basis of the χ^2 test applied to hypothesis testing in statistics. The rotation formula can be found, quite simply, by operating a generic rotation by an angle α of the reference axes on the standard variables (4.34):

$$\begin{aligned} A &= U\cos\alpha + V\sin\alpha \\ B &= -U\sin\alpha + V\cos\alpha \end{aligned} \tag{4.44}$$

These simple equations can be found in any basic textbook of physics or geometry. To be uncorrelated and independent, the new variables A and B must have their covariance (and, therefore, their correlation coefficient) equal to zero. We then impose:

$$\begin{aligned} \langle AB\rangle &= \langle (U\cos\alpha + V\sin\alpha)(-U\sin\alpha + V\cos\alpha)\rangle \\ &= -\sin\alpha\,\cos\alpha\left(\left\langle U^2\right\rangle - \left\langle V^2\right\rangle\right) + (\cos^2\alpha - \sin^2\alpha)\,\langle UV\rangle \\ &= -\frac{1}{2}\sin 2\alpha\left(\left\langle U^2\right\rangle - \left\langle V^2\right\rangle\right) + \cos 2\alpha\,\langle UV\rangle = 0\,. \end{aligned}$$

Since, for standard variables, $\left\langle U^2\right\rangle = \left\langle V^2\right\rangle = \sigma^2 = 1$, one obtains the condition:

$$\cos 2\alpha\,\langle UV\rangle = 0 \implies \cos 2\alpha = 0 \implies 2\alpha = \pm\frac{\pi}{2} \implies \alpha = \pm\frac{\pi}{4}\,. \tag{4.45}$$

For this value of α, the system of Eqs. (4.44) becomes:

$$A = \frac{1}{\sqrt{2}}(U+V)\,, \quad B = \frac{1}{\sqrt{2}}(-U+V)\,. \tag{4.46}$$

Also the inverse equations hold:

$$U = \frac{1}{\sqrt{2}}(A-B)\,, \quad V = \frac{1}{\sqrt{2}}(A+B)\,, \tag{4.47}$$

together with the norm conservation in a rotation:

$$A^2 + B^2 = U^2 + V^2\,. \tag{4.48}$$

By substituting Eqs. (4.46–4.48) in Eq. (4.41) we obtain:

$$U^2 - 2\rho UV + V^2 = (1-\rho)A^2 + (1+\rho)B^2 \,. \tag{4.49}$$

With these transformations, the exponent of the standard Gaussian (4.38) assumes, in lowercase notation, the following form:

$$-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2) = -\frac{1}{2}\left[\frac{a^2}{1+\rho} + \frac{b^2}{1-\rho}\right] .$$

Since the Jacobian determinant of the transformation[1] (4.47) has the value:

$$\begin{vmatrix} \frac{\partial u}{\partial a} & \frac{\partial u}{\partial b} \\ \frac{\partial v}{\partial a} & \frac{\partial v}{\partial b} \end{vmatrix} = \begin{vmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{vmatrix} = \frac{1}{2} + \frac{1}{2} = 1 \,,$$

one obtains:

$$\begin{aligned} g(u, v; 0, 1)\, \mathrm{d}u\ \mathrm{d}v &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\right] \\ &= \frac{1}{2\pi} \exp\left[-\frac{1}{2}\left(\frac{a^2}{1+\rho} + \frac{b^2}{1-\rho}\right)\right] \frac{\mathrm{d}a\ \mathrm{d}b}{\sqrt{1-\rho^2}} \\ &= \frac{1}{2\pi} \exp\left[-\frac{1}{2}\left(a_\rho^2 + b_\rho^2\right)\right]\ \mathrm{d}a_\rho\ \mathrm{d}b_\rho \,, \end{aligned} \tag{4.50}$$

where:

$$a_\rho = \frac{a}{\sqrt{1+\rho}} \,, \qquad b_\rho = \frac{b}{\sqrt{1-\rho}} \,. \tag{4.51}$$

Eqs. (4.50, 4.51) contain two important results.

The first one is that, given two Gaussian variables with $\rho \neq \pm 1$, it is always possible to define two new uncorrelated variables (A, B) by passing to the standard variables and rotating of an angle $\alpha = \pm 45°$. This angle corresponds to a rotation that brings the main axes of the concentration ellipse parallel to the reference axes. The angle double sign simply indicates that the rotation can consist of bringing an axis of the ellipse parallel either to the x axis or to the y axis of the reference system.

[1] This quantity, which will be formally introduced in Eq. (5.21), gives the variation of the unit area or volume element during the transformation from one reference system to another one.

The second important fact is that, from the uncorrelated variables (A, B), it is possible to obtain two variables (A_ρ, B_ρ) which are *independent standard Gaussians*. Since the sum of the squares of independent standard Gaussian variables, according to Pearson's Theorem 3.3, is χ^2 distributed, the variable given by the exponent of the two-dimensional Gaussian (4.40):

$$Q = \gamma(X, Y) \,, \tag{4.52}$$

when the X, Y correlation coefficient $\rho \neq \pm 1$, follows the $\chi^2(2)$ distribution.

If $\rho = \pm 1$, there is a deterministic relation between A and B and the degrees of freedom decrease to one. We will also show later that the fundamental Eq. (4.52) holds for any number of dimensions. The fact that Eq. (4.52) represents a χ^2 variable does not require the knowledge of the explicit form of the independent Gaussian variables: it is enough to know that they can always be found, if $\rho \neq 1$. Therefore, the χ^2 calculation is usually performed with the original variables of the problem, even if they are correlated.

Finally, we analyse Gaussian conditional densities. The density $g(y|x)$, for a fixed x, is easily derived by applying Definition (4.13). In our case this requires to divide the density (4.40) by the marginal density $g_X(x)$, which is nothing more than the one-dimensional Gaussian density $g(x; \mu_x, \sigma_x)$ given by Eq. (3.28). After an easy rearrangement, one obtains:

$$\begin{aligned} g(y|x) &= \frac{1}{\sigma_y \sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{1}{2(1-\rho^2)} \left(\frac{y-\mu_y}{\sigma_y} - \rho \frac{x-\mu_x}{\sigma_x} \right)^2 \right] \\ &= \frac{1}{\sigma_y \sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{\left[y - \mu_y - \rho \frac{\sigma_y}{\sigma_x}(x-\mu_x) \right]^2}{2\sigma_y^2(1-\rho^2)} \right] . \end{aligned} \tag{4.53}$$

Since x is constant, this curve corresponds to a one-dimensional Gaussian with mean and variance given by:

$$\langle Y|x \rangle = \mu_y + \rho \frac{\sigma_y}{\sigma_x}(x - \mu_x) \,, \tag{4.54}$$

$$\mathrm{Var}[Y|x] = \sigma_y^2(1-\rho^2) \,. \tag{4.55}$$

These last formulae show that the mean of Y conditioned on x varies with x along a line, called *regression line*. On the contrary, the variance of Y remains constant and depends on x only through the correlation coefficient.

An experimenter who conducts a series of measurements consisting in sampling Y for different values of x kept constant observes a mean moving along the line (4.54) and a constant variance (4.55). The conditional variance $\mathrm{Var}[Y|x]$ measures the dispersion of the data around the regression line and is never larger than the projected variance σ_y^2. Its value depends on the angle that the principal axes of

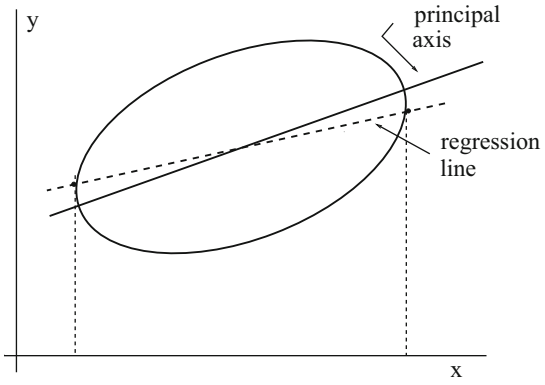


Fig. 4.5 Difference between regression line and principal axis of the concentration ellipse

the concentration ellipse form with the reference axes. If $\rho = 0$ the two sets of axes are parallel, the regression lines coincide with the axes of the ellipse, there is no correlation and the two variances are equal; if $\rho = 1$, $\mathrm{Var}[Y|x] = 0$ and no dispersion is observed along the regression line (see Fig. 4.4). The conditional density $g(x|y)$, for a predetermined value of y, and the corresponding mean, variance and regression line are obtained from the previous formulae by exchanging y with x. Is a non-linear correlation possible between Gaussian variables? The calculations just carried out show that, if the (projected) marginal densities are both Gaussian, *the relation must be linear*. A non-linear relation *distorts* the Gaussian form on at least one of the two axes. We also anticipate that nonlinear dependence will be addressed in detail later, in Chap. 11. Finally, we note that the regression lines are, by construction, the locus of the points of tangency to the ellipse of the lines parallel to the axes and therefore do not coincide with the principal axes of the ellipse (see Fig. 4.5).

In R, there are different ways to study and represent joint distributions. For example, to generate 1000 pairs of normal variables with means $\mu_x = 5$ and $\mu_y = 10$ and covariance matrix with $\sigma_x^2 = \sigma_y^2 = 3$ and $\sigma_{xy} = -2$, one can write:

```
> require(mvtnorm)    # load library
> xy <- rmvnorm(1000,c(5,10),sigma=rbind(c(3,-2),c(-2,3)))
```

To easily obtain two-dimensional density graphs in R, we need to construct two vectors containing bins and frequency matrix. These operations are performed by our routine `HistoBar3D`, which can elaborate both raw data and histograms. The lines of code of `HistoBar3D` used to create vectors and matrices, starting from a matrix of raw data consisting of x_i, y_i pairs ordered by columns, can be of general interests and are detailed here:

```
  # create x and y bins in x.bin and y.bin
x.bin <- seq(0.98*(min(xy[,1])), 1.02*(max(xy[,1])), length=nbinsx)
y.bin <- seq(0.98*(min(xy[,2])), 1.02*(max(xy[,2])), length=nbinsy)
 # fill x.bin and y.bin cells with the number of events
freq <-  as.data.frame(table(findInterval(xy[,1], x.bin),
                             findInterval(xy[,2], y.bin)))
```

```
freq[,1] <- as.numeric(freq[,1])
freq[,2] <- as.numeric(freq[,2])
freq[,3] <- as.numeric(freq[,3])
 # freq2D is the matrix with bin x, bin y, freqxy
freq2D <- matrix(0:0,nrow=nbinsx,ncol=nbinsy)
freq2D[cbind(freq[,1], freq[,2])] <- freq[,3]
 # marginal distributions
xmarg <- apply(freq2D,1,sum) # row sum (x contents)
ymarg <- apply(freq2D,2,sum) # column sum (y contents)
```

The best way to understand these complex instructions is to open the window of R, generate `xy`, for example, with `rmvnorm`, and then interactively study how the above-described `freq2D` matrix is built up. If the input raw data are contained in two separate vectors `x` and `y`, `HistoBar3D` creates the two-column matrix `xy` with the instruction:

```
xy <- matrix(c(x,y),ncol=2,byrow=FALSE)
```

For further details, you can directly examine the `HistoBar3D` routine. Notice also that this code uses the R routine `bkde2D`, which requires raw data and is described in Appendix B.

Exercise 4.3
There are two Gaussian variables, X and Y, where X has mean $\mu = 25$ and standard deviation $\sigma_x = 6$ and

$$Y = 10 + X + Y_R \,, \tag{4.56}$$

while Y_R is normal with parameters $\mu_{y_R} = 0$ and $\sigma_{y_R} = 6$.

Find covariance and correlation coefficient between these two variables. Check the obtained results with simulated data.

Answer The mean and standard deviation of Y are given by:

$$\langle Y \rangle = \mu_y = 10 + \langle X \rangle = 35 \,, \quad \sigma[Y] = \sigma_y = \sqrt{\sigma_x^2 + \sigma_{y_R}^2} = 8.48 \,.$$

By defining $\Delta X = X - \mu_x$, $\Delta Y = Y - \mu_y$, from Eq. (4.24) the covariance between variables can be calculated as:

$$\mathrm{Cov}[X, Y] = \langle \Delta X \, \Delta Y \rangle = \langle \Delta X (\Delta X + \Delta Y_R) \rangle = \left\langle \Delta^2 X \right\rangle = \sigma_x^2 = 36 \,, \tag{4.57}$$

(continued)

Exercise 4.3 (continued)
where the condition $\langle \Delta X \Delta Y_R \rangle = 0$ has been used, since the variables X and Y_R are uncorrelated by construction. The correlation coefficient is given by:

$$\rho = \frac{\mathrm{Cov}[X, Y]}{\sigma[X]\sigma[Y]} = \frac{36}{6 \cdot 8.48} = 0.707 \, .$$

Notice that Eq. (4.55) gives: $\mathrm{Var}[Y|x] = \sigma_y^2(1 - \rho^2) = 8.48^2(1 - 0.707^2) = 6^2 = \sigma_{yR}^2$. Now we can check these results with a simulation.

Opening the R console and proceeding interactively, we can write:

```
> X <- rnorm(20000,mean=25,sd=6)
> Y <- 10+X+rnorm(20000)
> mean(Y)    # the result appears in the console
> mean(X)
> cov(X,Y)
> cov(X,Y)/sqrt(var(X)*var(Y))
> HistoBar3D(X,Y)
```

The results appear in the R console, while after the call to `HistoBar3D` the results of Fig. 4.6 are displayed in the graphics window. Since the correlation between X and Y is linear, the marginal histograms $g(x)$ and $g(y)$ have a Gaussian shape, as can be easily noticed from the graphs of Fig. 4.6. The means and standard deviations of these histograms are an estimate of the corresponding true parameters of the densities $g_X(x)$ and $g_Y(y)$. They can be calculated with Eqs. (2.36, 2.39, 4.23). From 20 000 simulated pairs, we have obtained: $m_x = 24.96$, $m_y = 34.95$, $s_x = 6.02$, $s_y = 8.48$, $s_{xy} = 35.73$, $r = 0.700$. These quantities are indicated in Latin letters because they are sample estimates of the true values (4.17, 4.18).

Since the simulated sample contains a large number of events, all the values we got from the data coincide, within "a few per thousand", with the real ones. The analytical characteristics of this convergence will be considered in Chap. 6.

4.5 The General Multidimensional Case

The generalization of the equations above to the case of more than two variables is quite immediate and not particularly difficult. The joint p.d.f. of n random variables is given by a non-negative function:

$$p(x_1, x_2, \ldots, x_n) \, , \tag{4.58}$$

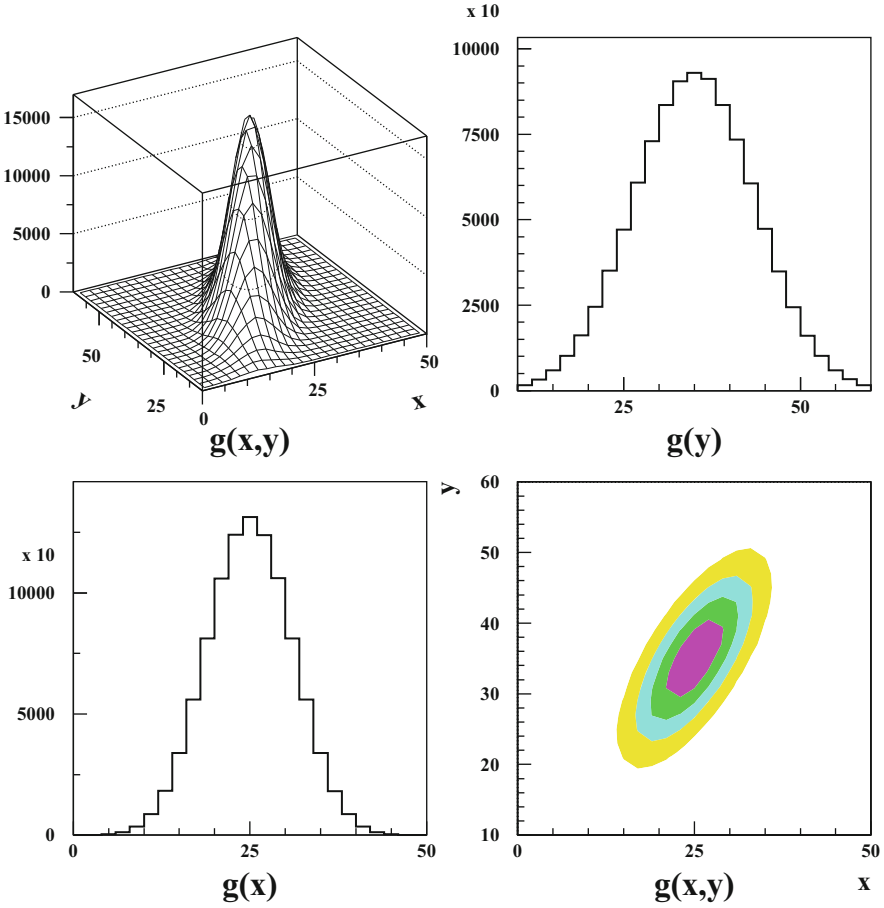


Fig. 4.6 A sample of 20 000 events, computer simulated from the two-dimensional Gaussian population considered in the Exercise 4.3. Two-dimensional histogram of $g(x, y)$, marginal distributions $g(x)$ and $g(y)$ and top view with the density curves (bottom right)

such that, if $A \subseteq \mathbb{R}^n$,

$$P\{(x_1, x_2, \ldots, x_n) \in A\} = \int_A p(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_1\, \mathrm{d}x_2 \ldots\, \mathrm{d}x_n\ .$$

For independent variables, one has:

$$p(x_1, x_2, \ldots, x_n) = p_1(x_1)\, p_2(x_2) \ldots p_n(x_n)\ . \tag{4.59}$$

The mean and variance of X_k (with $1 \leq k \leq n$) are obtained by generalizing Eq. (4.5):

$$\begin{aligned} \langle X_k \rangle &= \int x_k \, p(x_1, x_2, \ldots, x_n) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n = \mu_k \, , \\ \mathrm{Var}[X_k] &= \int (x_k - \mu_k)^2 \, p(x_1, x_2, \ldots, x_n) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \, . \end{aligned} \tag{4.60}$$

In the case of several variables, the covariance can be calculated with Eq. (4.20) for any pair of variables (x_i, x_k):

$$\mathrm{Cov}[X_i, X_k] = \int (x_i - \mu_i)(x_k - \mu_k) \, p(x_1, x_2, \ldots, x_n) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \ldots \mathrm{d}x_n \, . \tag{4.61}$$

Therefore, one has $n(n-1)/2$ different covariances and n variances. They are gathered in a symmetric matrix V, called *covariance matrix*:

$$V = \begin{pmatrix} \mathrm{Var}[X_1] & \mathrm{Cov}[X_1, X_2] & \ldots & \mathrm{Cov}[X_1, X_n] \\ \ldots & \mathrm{Var}[X_2] & \ldots & \mathrm{Cov}[X_2, X_n] \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \mathrm{Var}[X_n] \end{pmatrix} , \tag{4.62}$$

where the diagonal elements are the n variances and the non-diagonal ones are the $n(n-1)/2$ covariances. The matrix is symmetric, since $\mathrm{Cov}[X_i, X_k] = \mathrm{Cov}[X_k, X_i]$; for this reason it is explicitly written only above the diagonal. The number of different elements of the matrix is:

$$\frac{n(n-1)}{2} + n = \frac{n(n+1)}{2} \, .$$

Often the correlation coefficients $\rho[X_i, X_k]$, obtained by generalizing Eq. (4.31), are used. The covariance matrix is then also written as:

$$V = \begin{pmatrix} \mathrm{Var}[X_1] & \rho[X_1, X_2]\sigma[X_1]\sigma[X_2] & \ldots & \rho[X_1, X_n]\sigma[X_1]\sigma[X_n] \\ \ldots & \mathrm{Var}[X_2] & \ldots & \rho[X_2, X_n]\sigma[X_2]\sigma[X_n] \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \mathrm{Var}[X_n] \end{pmatrix} . \tag{4.63}$$

If variables are uncorrelated, all the off-diagonal terms are zero, and the diagonal terms coincide with the variances of the individual variables. Sometimes, instead of the covariance matrix, the correlation matrix is used:

$$C = \begin{pmatrix} 1 & \rho[X_1, X_2] & \ldots & \rho[X_1, X_n] \\ \ldots & 1 & \ldots & \rho[X_2, X_n] \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & 1 \end{pmatrix} , \tag{4.64}$$

which is still symmetric.

Generalizing Eq. (4.24) in several dimensions, the covariance matrix can be expressed as:

$$\left\langle (\boldsymbol{X} - \boldsymbol{\mu})\,(\boldsymbol{X} - \boldsymbol{\mu})^{\dagger} \right\rangle = V \,, \tag{4.65}$$

where $(\boldsymbol{X} - \boldsymbol{\mu})$ is a column vector and the symbol † represents the transposition operation. In this way, the product between a column vector and a row vector gives the symmetric square matrix (4.62).

In R, the calculation of the covariance matrix starting from a set of raw data collected in a matrix `M` and sorted by columns can be done with the routine `cov(M)` or `var(M)`, while the function `cor(M)` must be used for the correlation matrix. Let us open the R console and explore these functions generating 100 triples of three correlated Gaussian variables:

```
> a <- rnorm(100)
> b <- a + rnorm(100)
> c <- a - 2*b + rnorm(100)
> M <- cbind(a,b,c)  # matrix 100 x 3
> cor(M)
           a          b          c
a  1.000000  0.6725680 -0.3832170
b  0.672568  1.0000000 -0.8875424
c -0.383217 -0.8875424  1.0000000
```

The covariance matrix is immediately obtained by typing one of the two equivalent commands `var(M)` and `cov(M)`. The matrices V and C often appear in multivariate probability calculus. It can be shown that they have the fundamental property of being *positive semidefinite*:

$$\boldsymbol{x}^{\dagger}\, V\, \boldsymbol{x} \geq 0 \,, \tag{4.66}$$

where the equality holds when all the n elements of the vector $\boldsymbol{x}$ are null. Notice that in the quadratic form of Eq. (4.66) a row vector $\boldsymbol{x}^{\dagger}$ appears to the left and a column vector $\boldsymbol{x}$ appears to the right, so as to obtain a number (not a matrix) which, if V is diagonal, is a sum of squares of the type $\sum x_i^2 V_{ii}$. Since $V^{\dagger} = V$, $(V^{-1})^{\dagger} = V^{-1}$, if $\boldsymbol{x} = V^{-1}\boldsymbol{y}$ one obtains the equation:

$$\boldsymbol{y}^{\dagger}\, V^{-1}\, \boldsymbol{y} = \boldsymbol{y}^{\dagger}\, V^{-1}\, V\, V^{-1}\, \boldsymbol{y} = \boldsymbol{x}^{\dagger}\, V\, \boldsymbol{x} \geq 0 \,, \tag{4.67}$$

which shows that also the inverse matrix is positive definite. Another useful property is that, for any positive definite matrix V, there exists a matrix H such that:

$$H\, H^{\dagger} = V \implies H\, H^{\dagger} V^{-1}\, H = H \implies H^{\dagger} V^{-1} H = I \,, \tag{4.68}$$

where I is the unit matrix. All these properties are proved in detail in Cramer's classic textbook [Cra51].

It is also possible to generalize Eqs. (4.11, 4.13) to obtain different marginal and conditional distributions of one or more variables by integrating over the remaining ones. This generalization is obvious and is not reported here.

When the $\boldsymbol{X}$ variables are Gaussian, they have a p.d.f. given by the multivariate generalization of the two-dimensional Gaussian (4.40):

$$g(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}} |V|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\dagger V^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]; \ |V| \equiv \det|V| \neq 0 \tag{4.69}$$

where V is the covariance matrix of Eq. (4.62).

The density (4.69) is called *multivariate Gaussian* and can be obtained by extending the procedure which led to the bivariate Gaussian. Obviously, it is easy to verify that the multivariate Gaussian contains the bivariate distribution as a special case. In fact, if $\rho \neq \pm 1$:

$$V^{-1} = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}^{-1} = \frac{1}{\sigma_x^2\sigma_y^2(1-\rho^2)} \begin{pmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{pmatrix},$$

and Eq. (4.40) follows.

If H is a matrix satisfying Eq. (4.68), with the transformation:

$$H\boldsymbol{Z} = \boldsymbol{X} - \boldsymbol{\mu}\,, \tag{4.70}$$

the semidefinite form appearing in the exponent of the Gaussian becomes:

$$\gamma(\boldsymbol{X}) \equiv (\boldsymbol{X}-\boldsymbol{\mu})^\dagger V^{-1}(\boldsymbol{X}-\boldsymbol{\mu}) = \boldsymbol{Z}^\dagger H^\dagger V^{-1} H\boldsymbol{Z} = \boldsymbol{Z}^\dagger\,\boldsymbol{Z} = \sum_{i=1}^{n} Z_i^2\,, \tag{4.71}$$

where the third of Eqs. (4.68) has been used. Equation (4.70) is therefore the equivalent of the two-dimensional rotation (4.44). The new variables $\boldsymbol{Z}$ are still Gaussian because they are linear combinations of Gaussian variables (see Exercise 5.3 in the next chapter); moreover, it is easy to verify that they have null mean and that, according to Eqs. (4.65, 4.70) and the third of Eqs. (4.68), are also uncorrelated standard variables:

$$\begin{aligned} \left\langle \boldsymbol{Z}\boldsymbol{Z}^\dagger \right\rangle &= H^{-1}\left\langle (\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\dagger \right\rangle (H^\dagger)^{-1} \\ &= H^{-1} V\,(H^\dagger)^{-1} = (H^\dagger V^{-1} H)^{-1} = I\,. \end{aligned} \tag{4.72}$$

From Theorem 4.3, it follows that the Z_i variables are mutually independent and that Eq. (4.71) represents a variable $Q \sim \chi^2(n)$. Therefore, Eq. (4.52) also holds for the n-dimensional case.

If $|V| = 0$, there is at least one linear relation between the n variables; it is then necessary to identify a number $r < n$ of linear independent variables and modify Eq. (4.69) accordingly. We think it is useful to remind that two variables are *stochastically* independent (or dependent) if Eq. (4.12) holds (or not); instead, the linear mathematical dependence implies the existence of well-defined constraint equations among the variables. Two linearly dependent or constrained random variables correspond, statistically, to a single random variable. If the equation is linear, the Cauchy-Schwarz theorem 4.2 gives $\rho = \pm 1$. Therefore, when dealing with random variables described by positive semidefinite quadratic forms of the type:

$$(\boldsymbol{X} - \boldsymbol{\mu})^{\dagger} W (\boldsymbol{X} - \boldsymbol{\mu}) ,$$

where $\boldsymbol{X}$ are Gaussian variables and W is a symmetric matrix; it is always necessary to verify if $|W| = 0$. In such condition, it is necessary to reduce the number of variables by determining, if there are p linear equations among the variables, $(n-p)$ new linearly independent variables. Linear systems theory assures that it is always possible to determine these new variables in such a way to obtain a positive definite quadratic form of dimension $(n - p)$, for which Eqs. (4.71, 4.72) still hold. It is therefore possible to express the quadratic form as a sum of $(n - p)$ independent standard Gaussian variables T_i:

$$(\boldsymbol{X} - \boldsymbol{\mu})^{\dagger} W (\boldsymbol{X} - \boldsymbol{\mu}) = \sum_{i=1}^{n-p} T_i^2 . \tag{4.73}$$

At this point, we can generalize the Pearson's Theorem 3.3 as follows:

Theorem 4.4 (Quadratic Forms) *A random variable $Q \sim \chi^2(\nu)$ is given by the sum of the squares of ν stochastically independent standard Gaussian variables or by the positive definite quadratic form (4.73) of Gaussian variables. The degrees of freedom are given in this case by the number of variables minus the number of the linear relations existing between them.*

Multidimensional random variables can be treated as vectors in n-dimensional spaces, thus exploiting many of the results of the theory of n-dimensional $\mathbb{R}^n$ vector spaces. Among these, we recall the scalar product of two variables (vectors) $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} x_i y_i , \tag{4.74}$$

which is simply the generalization of the scalar product between three-dimensional vectors. In a similar way as two vectors are defined to be orthogonal if their scalar product is zero, two sub-spaces A and $A^{\perp}$ are then defined as orthogonal if each vector of A is orthogonal to each vector of $A^{\perp}$. The sub-space $A^{\perp}$ has dimension

equal to $n - k$, if k is the dimension of A. Moreover, we recall that each vector can be written in the form $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$ where $\boldsymbol{x}_1 \in A$ and $\boldsymbol{x}_2 \in A^{\perp}$. Arrays, considered as operators, act on the elements of the vector space. The orthogonal projection operators $P(A) : \boldsymbol{x} \to \boldsymbol{x}_1$, which associates to any $\boldsymbol{x} \in \mathbb{R}^n$ its component on A, are very useful. Projectors have many properties which recall those of the projections of vectors on the three Cartesian axes. For example, $P(A)\boldsymbol{x}$ is the vector of A at the minimum distance from $\boldsymbol{x}$ and the following equations hold:

$$P(A)P(A) = P(A)\,, \quad P(A^{\perp})P(A) = 0\,, \quad I - P(A) = P(A^{\perp})\,. \tag{4.75}$$

The first property, called idempotence, is obvious, because $P(A)\boldsymbol{x} = \boldsymbol{x}$ if $\boldsymbol{x} \in A$, whereas the second one reflects the fact that the projection vector has zero components in the orthogonal subspace. The third property, where I is the identity or unit matrix, holds because $P(A)(I - P(A)) = 0$ for idempotence. The simplest orthogonal projection operator is perhaps the one reducing the non-zero components of the vector:

$$\begin{aligned} P(A)\boldsymbol{x} &= (x_1, x_2, \ldots, x_k, 0, \ldots, 0)\,, \\ P(A^{\perp})\boldsymbol{x} &= (0, 0, \ldots, 0, x_{k+1}, x_{k+2}, \ldots, x_n)\,. \end{aligned} \tag{4.76}$$

Cochran's theorem is based on this decomposition. We will use it several times in the following for Gaussian variables.

Theorem 4.5 (of Cochran) *Let $\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{I})$ be a n-dimensional standard Gaussian random variable and let $A_1, A_2, \ldots, A_k$ be mutually orthogonal vectorial sub-spaces in $\mathbb{R}^n$. Let n_i be the dimension of A_i and $P(A_i)$ be the orthogonal projectors on A_i. Then, the random variables $P(A_i)\boldsymbol{X}, i = 1, 2, \ldots k$ are statistically independent and the variable $|P(A_i)\boldsymbol{X}|^2 = (P(A_i)\boldsymbol{X}, P(A_i)\boldsymbol{X})$ follows the $\chi^2(n_i)$ distribution.*

Proof In essence, the theorem affirms the statistical independence and stability of independent Gaussian variables projected onto orthogonal subspaces. We prove the theorem in the simple case of two subspaces of the type:

$$\begin{aligned} P(A_1)\boldsymbol{X} &= (X_1, \ldots, X_{n_1}, 0, \ldots, 0) \\ P(A_2)\boldsymbol{X} &= (0, \ldots, 0, X_{n_1+1}, \ldots, X_{n_1+n_2})\,. \end{aligned}$$

If $\boldsymbol{Y} = P(A_1)(\boldsymbol{X})$ and $\boldsymbol{Z} = P(A_2)(\boldsymbol{X})$, we have $\mathrm{Cov}[Y_i, Z_j] = 0\ \forall i, j = 1, 2, \ldots, n$ by construction. Since, by assumption, the variables $\boldsymbol{X}$ are independent standard Gaussians, from Theorem 4.3 it also results that $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are independent and that, from Pearson's Theorem 3.3, the variables:

$$\begin{aligned} |P(A_1)\boldsymbol{X}|^2 &= X_1^2 + \ldots + X_{n_1}^2 \sim \chi^2(n_1)\,, \\ |P(A_2)\boldsymbol{X}|^2 &= X_{n_1+1}^2 + \ldots + X_{n_1+n_2}^2 \sim \chi^2(n_2) \end{aligned}$$

follow the χ^2 distribution. It can be shown that one can always be led back to this particular case, through orthogonal transformations that lead subspaces generated respectively by the first n_1 coordinates and by the remaining n_2. In addition, the property $(P(A_i)\boldsymbol{X}, P(A_i)\boldsymbol{X}) = \boldsymbol{X}^\dagger B_i \boldsymbol{X}$ holds, where B_i is a positive semidefinite matrix, so that the theorem can also be formulated in terms of matrices whose rank sum is equal to the dimension of the space. These matrices can be diagonalized according to Eq. (4.71). □

An important application of Cochran's theorem will be described later on, as in Theorem 6.1.

4.6 Multivariate Probability Regions

We now consider multidimensional probability levels. The two-dimensional analogous of the integral (3.94) is given by:

$$P\{a \le X \le b,\ c \le Y \le d\} = \int_a^b \int_c^d p(x, y)\ \mathrm{d}x\, \mathrm{d}y\ , \tag{4.77}$$

which gives the probability that, in an experiment, the numerical realizations of the two random variables lie within fixed limits. In this case, the probability interval turns into a rectangle, as shown in Fig. 4.7a. When the lower bounds are $a = -\infty, c = -\infty$, Eq. (4.77) gives the two-dimensional analogous of the distribution function (2.33). Equation (4.77) can be extended in more dimensions by considering a variable $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ and a region D in the n-dimensional space. The probability that an observation gives a vector in the set D is given by:

$$P\{\boldsymbol{X} \in D\} = \int_D p(x_1, x_2, \ldots, x_n)\ \mathrm{d}x_1\, \mathrm{d}x_2 \ldots \mathrm{d}x_n\ . \tag{4.78}$$

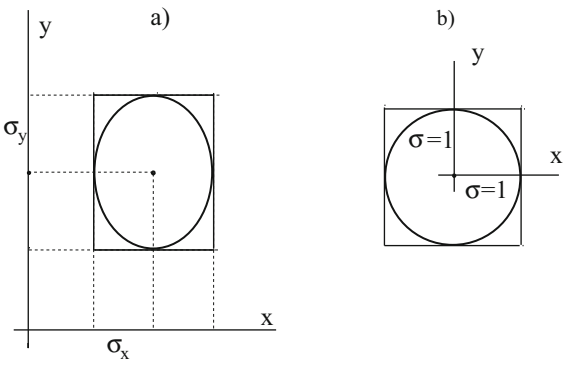


Fig. 4.7 Region of the probability $P\{|X - \mu_x| \le \sigma_x,\ |Y - \mu_y| \le \sigma_y\}$ and corresponding concentration ellipse for uncorrelated normal variables (**a**); for standard variables, the rectangle and the ellipse transform to a square and to a circle of unit half-side or radius centred at the origin, respectively (**b**)

For independent Gaussian variables, Eq. (4.77) results in the product of two integrals of one-dimensional Gaussian densities. When the probability intervals of X and Y are symmetrical and centred on their respective means, the probability levels can be obtained as the product of the functions (3.38):

$$P\{|X-\mu_x| \le a,\ |Y-\mu_y| \le b\} = 2\int_{\mu_x}^{\mu_x+a} \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\frac{(x-\mu_x)^2}{\sigma_x^2}}\,\mathrm{d}x \ \times$$

$$\times\, 2\int_{\mu_y}^{\mu_y+b} \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\frac{(y-\mu_y)^2}{\sigma_y^2}}\,\mathrm{d}y = [2\ \mathrm{E}(a/\sigma_x)]\cdot\left[2\ \mathrm{E}(b/\sigma_y)\right]\,. \tag{4.79}$$

The probability levels corresponding to the one-dimensional 3σ law are therefore obtained as *product of the one-dimensional probability levels (3.35)*. For example, the probability for both variables to be within $\pm\sigma$, i.e. within one standard deviation, is given by:

$$P\{|X-\mu_x| \le \sigma_x,\ |Y-\mu_y| \le \sigma_y\} = 0.683\cdot 0.683 = 0.466\,. \tag{4.80}$$

Let us now consider the standard bivariate Gaussian for the case of uncorrelated variables (4.35) written in polar coordinates:

$$u = r\cos\theta \qquad v = r\sin\theta\,.$$

Since the surface element $\mathrm{d}u\,\mathrm{d}v$ turns into $r\,\mathrm{d}r\,\mathrm{d}\theta$ (the Jacobian of the transformation is r), the function (4.35) becomes:

$$g(r,\theta;0,1) = \frac{1}{2\pi} r e^{-\frac{1}{2}r^2}\,, \quad 0 \le 2\pi\,, \quad 0 \le r < \infty\,. \tag{4.81}$$

This equation shows that, in the case of standard variables, *the concentration ellipse of Fig. 4.7a transforms into a circle centred at the origin* (see Fig. 4.7b). From here, it is then easy to derive the probabilities of falling within the concentration ellipse having the principal axes equal to k times the standard deviations of the two variables:

$$P\{U^2+V^2 \le k^2\} = P\left\{\frac{(X-\mu_x)^2}{\sigma_x^2} + \frac{(Y-\mu_y)^2}{\sigma_y^2} \le k^2\right\} =$$

$$= \frac{1}{2\pi}\int_{-\pi}^{+\pi}\int_0^k r e^{-\frac{1}{2}r^2}\,\mathrm{d}r\,\mathrm{d}\theta = 1 - e^{-\frac{1}{2}k^2}\,. \tag{4.82}$$

From this equation, a "3σ law" in the plane is obtained:

$$P\left\{\frac{(X-\mu_x)^2}{\sigma_x^2}+\frac{(Y-\mu_y)^2}{\sigma_y^2}\leq k^2\right\}=1-e^{-\frac{k^2}{2}}=\begin{cases}0.393 \text{ for } k=1\\ 0.865 \text{ for } k=2\\ 0.989 \text{ for } k=3\,.\end{cases} \tag{4.83}$$

Notice that the rectangular interval $\pm\sigma$ of Eq. (4.80) has a greater probability (0.466 versus 0.393) than that of the corresponding ellipse (4.83), which for $k=1$ has the sides of the rectangle as principal axes. This fact is also shown in Fig. 4.7, where we see that probability rectangle (or square) contains the ellipse (or circle) of concentration.

Let's now consider, instead of two standard variables (U, V), two Gaussian variables (X, Y) of zero mean and having the same variance σ^2. If, we substitute $r \to r/\sigma$ in the integral (4.82) and integrate in θ, we get the function:

$$\rho(r)=\frac{r}{\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right)\,,\quad x^2+y^2=r^2\,, \tag{4.84}$$

which, after integration in $\mathrm{d}r$, gives the bidimensional p.d.f. for two independent Gaussian variables (X, Y) of equal variance. This function is named as Rayleigh density and is the two-dimensional analogue of the Maxwell density (3.75), which is valid in space.

In R, this function is calculated with the routine `rayleigh(scale)` of the library `bayesmeta`, where `scale`$=\sigma$, to which the usual prefixes `d, p, q, r` must be added to have the function, cumulative, quantile or random values (see Table B.2).

You will have noticed that the probability intervals of a certain variable are different depending on the global number of variables and on the type of domain D that is considered in Eq. (4.78). Fixing ideas to the situation of two variables (X, Y), for a certain given probability level (e.g. 68.3%), we can define as many as five probability intervals: an interval $(x \pm \sigma_x)$ for 68.3% probability of finding a value of X for any value of Y and the same by changing X with Y (so that we have two intervals), or the interval $(\langle Y|x\rangle \pm \sqrt{\mathrm{Var}[Y|x]})$ of Eqs. (4.54, 4.55) for the 68.3% probability to find an Y value for a fixed value x_0 and similarly for X for a fixed value y_0 (now we have a total of four intervals). Finally, we can also define a two-dimensional region D which includes 68.3% for the (X, Y) pairs. The densities involved in these five cases are the marginal densities $p_X(x)$, $p_Y(y)$; the conditional densities $p(y|x_0)$, $p(x|y_0)$; and the joint density $p(x, y)$, respectively. The region D of the plane where the multidimensional estimates are calculated is often the concentration ellipse:

$$Q=\text{constant}=(\boldsymbol{X}-\boldsymbol{\mu})^{\dagger}V^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\quad\to\quad\sum_{i=1}^{n}\frac{(X_i-\mu_i)^2}{\sigma_i^2}\,, \tag{4.85}$$

where the expression to the right of the arrow obviously applies in case of uncorrelated variables. This region has some important properties which we will briefly examine. In the case of n Gaussian variables, according to Theorem 4.4, the variable Q is distributed as $\chi^2(n)$. This allows us to use, in the calculation of the probability levels associated with the multidimensional concentration ellipse (4.85), the significance levels of χ^2 density reported in the Tables E.3 and E.4 of Appendix E. Since the marginal density of each variable is Gaussian, the variable:

$$Q_i = \frac{(X_i - \mu_i)^2}{\sigma_i^2} \tag{4.86}$$

follows the $\chi^2(1)$ density. The probability levels associated with $\{Q_i = 1, 4, 9\}$ are the 68.3%, 95.4% and 99.7% Gaussian probability levels corresponding to the $\{\sqrt{Q_i} = 1, 2, 3\}$ values of the T_i variable, since $P\{Q_i \leq \chi^2\} = P\{T_i \leq \sqrt{\chi^2}\}$.

Tables E.3 and E.4 give two different ways of calculating the probabilities for the χ^2 distribution: to obtain the values shown in Table E.4 for a given probability level α, just find the corresponding level $1 - \alpha$ in Table E.3 and multiply the reduced chi square $\chi^2_R(\nu)$ of the table by the ν degrees of freedom required by the problem (note, in fact, that Table E.4 reports the values of the total χ^2, not of the reduced one). The difference between the probability levels of each single variable and the joint ones is shown, for the two-dimensional case, in Fig. 4.8, where the value $\chi^2 = 2.30$ is taken from Table E.4 for $\nu = 2$. Observing the figure, one should pay attention to an important property, which we will apply later: *the projections on the axes of the curve corresponding to $\chi^2 = 1$ define the $1 - \sigma$ intervals of the variables, referred to the respective marginal densities*. This property, obvious in one dimension, also holds for two variables. This can be demonstrated by equating the χ^2 function (4.52) to 1:

$$Q = \gamma(X, Y) = \frac{1}{1 - \rho^2}(u^2 - 2\rho uv + v^2) = 1\ ; \tag{4.87}$$

and evaluating the intersection points of this curve with the regression line of Eq. (4.54) $v = \rho u$. The obtained result, i.e. $u = \pm 1$, corresponds to an interval $(\mu_x \pm \sigma_x)$ and the same, obviously, holds for the projection on the y axis. It can be shown that this property is in general valid also for the n-dimensional case [BR92].

Similar to the one-dimensional case, the Gaussian model is often reasonably valid for a relevant part of multidimensional random phenomena. The use of the concentration ellipse and the χ^2 density then provide a very powerful tool for predicting the probability of an experimental result consisting of a n-tuple of values $(x_1, x_2, \ldots, x_n)$. If, on the other hand, the problem does not allow the use of the Gaussian model, it is necessary to resort to the solution of the integrals (4.78) by defining suitable probability regions, usually hypercubic or elliptical, and taking care of the correlations between variables. Often, to achieve this difficult objective, the Monte Carlo simulation methods described in Chap. 8 are used.

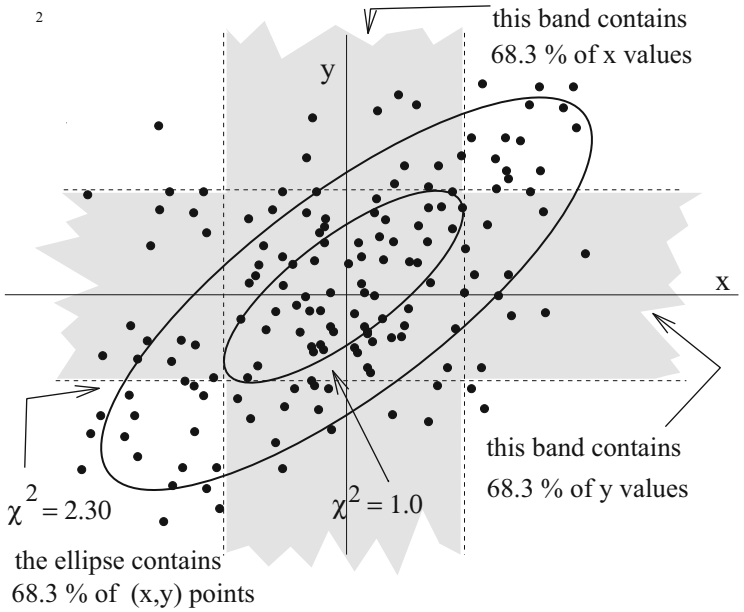


Fig. 4.8 One- and two-dimensional probability regions. The curve corresponding to the value $\chi^2 = 2.30$ contains 68.3% of the variable pairs. The projection on the axes of the curve $\chi^2 = 1$ gives the 1-σ probability intervals for each variable

4.7 Multinomial Distribution

The binomial density (2.29) describes the probability distribution of the variable $I = (I_1, I_2)$, defined as an experimental histogram with N events split into two bins with counts I_1 and I_2, respectively, when the true probabilities of an observation falling into bins 1 and 2 are, respectively, $p \equiv p_1$ and $(1 - p) \equiv p_2$:

$$P\{I = n_1, n_2\} = b(x; N, p) = \frac{N!}{n_1!(N - n_1)!} p^{n_1} (1 - p)^{N - n_1} = \frac{N!}{n_1! n_2!} p_1^{n_1} p_2^{n_2} , \tag{4.88}$$

with $(p_1 + p_2) = 1$ and $(n_1 + n_2) = N$. This equation leads to an immediate generalization to the case of a generic histogram with k bins. The resulting density, called *multinomial*, represents the probability that the random variable $I = (I_1, I_2, \ldots, I_k)$ provides an experimental histogram with n_i events in the i-th bin, when all the a priori probabilities p_i are known:

$$P\{I = n_1, n_2, \ldots, n_k\} = b(\boldsymbol{n}; N, \boldsymbol{p}) = \frac{N!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k} , \tag{4.89}$$

$$(p_1 + p_2 + \cdots + p_k) = 1, \qquad (n_1 + n_2 + \cdots + n_k) = N \,. \tag{4.90}$$

Since each variable, compared to all others, meets the binomial density requirements, the following relations hold:

$$\langle I_i \rangle = N p_i; \qquad \mathrm{Var}[I_i] = N p_i \,(1 - p_i) \,. \tag{4.91}$$

The variables $(I_1, I_2, \ldots, I_k)$ are dependent because of the second of Eqs. (4.90). Their covariance can be directly calculated with Eq. (4.61) or, in a much more simple way, by using the law of transformation of the variance. The calculation will be performed in the next chapter, in Exercise 5.11. The result is the formula (5.86), which we anticipate here:

$$\mathrm{Cov}[I_i, I_j] \equiv \sigma_{ij} = -N p_i p_j \,, \qquad (i \neq j) \,. \tag{4.92}$$

Covariance is negative, because more events in one bin imply less events in the other ones. In the one-dimensional case, the binomial density rapidly tends to the Gaussian density. The same property also applies to the multinomial density, which is a property of great importance for the statistical study of the histograms. Equation (3.24), which is valid for $Np,\ N(1-p) > 10$, can then also be written as:

$$b(n;\, N,\, p) \propto \exp\left[-\frac{1}{2}\frac{(n - Np)^2}{Np(1-p)}\right] = \exp\left[-\frac{1}{2}\left(\frac{(n_1 - Np_1)^2}{Np_1} + \frac{(n_2 - Np_2)^2}{Np_2}\right)\right] \tag{4.93}$$

where the last equality is obtained by setting $p \equiv p_1$, $(1-p) \equiv p_2$, $n \equiv n_1$, $(N - n) \equiv n_2$. This suggests a symmetric form which can be generalized to the case of n variables (a trace of the proof can be found in [Gne76]). We can therefore write the fundamental result:

$$\boldsymbol{b}(\boldsymbol{n};\, N,\, \boldsymbol{p}) \propto \exp\left[-\frac{1}{2}\sum_{i=1}^{k}\frac{(n_i - Np_i)^2}{Np_i}\right] \qquad (Np_i > 10\ \forall i) \,, \tag{4.94}$$

where the sum must be extended *to all* the k variables appearing in the sums (4.90). Therefore, we have approximated the multinomial density *with a multidimensional Gaussian density*, where the sum of the squares of the variables appears in the exponential. However, these variables are correlated through Eq. (4.90) and thus the degrees of freedom are really only $k - 1$. If we now combine this result with Theorem 4.4, we arrive at the important conclusion that we can state as follows.

Theorem 4.6 (Pearson's Sum) *Consider a histogram having k bins and representing a random sample of size N. If p_i is the true probability to observe an event in the i-th bin, the variable:*

$$Q = \sum_{i=1}^{k} \frac{(I_i - Np_i)^2}{Np_i} , \quad (I_1 + I_2 + \cdots + I_k) = N , \tag{4.95}$$

where I_i is the number of observed events in any bin, for $N \to \infty$ tends to the χ^2 density with $k - 1$ degrees of freedom.

This theorem is the key for using χ^2 test in statistics and can already be applied with good approximation if $Np_i > 10$.

We also note that Eq. (4.95) *does not give the sum of squares of standard variables*, since the variance of the I_i variables is given by (4.91) and the variables are correlated. Therefore the theorem non-trivially generalizes the results of Theorem 3.3, which refers to the sums of squares of independent standard Gaussian variables.

4.8 Problems

4.1 If X and Y are two independent uniform random variables, in the intervals $[0, a]$ and $[0, b]$ respectively, find the joint density $p(x, y)$.

4.2 If X_i, $i = 1, 2, 3$ are three independent Gaussian variables with mean μ_i and variance σ_i^2, find (a) the band containing 90% of values of X_1, (b) the ellipse containing 90% of values of the (X_1, X_2) pair and (c) the ellipsoid containing 90% of the n-tuples (X_1, X_2, X_3).

4.3 If X and Y are independent random variables, does the equality $\langle Y|x\rangle = \langle Y\rangle$ hold?

4.4 Calculate graphically, in an approximated way, the correlation coefficient from the concentration ellipse of Fig. 4.6 (top right).

4.5 If $Z = aX + b$ and $U = cY + d$ with $ac \neq 0$, calculate $\rho[Z, U]$.

4.6 The height of a homogeneous population is a Gaussian variable equal to $\langle X\rangle \pm \sigma[X] = 175 \pm 8$ cm for men and $\langle Y\rangle \pm \sigma[Y] = 165 \pm 6$ cm for women. Assuming there is no correlation, find the percentage of couples with men and women higher than 180 and 170 cm, respectively.

4.7 In principle, how would you solve the previous problem in the most realistic case (couples tend to have homogeneous stature) of a correlation coefficient $\rho = 0.5$ between the height of the husband and that of the wife?

4.8 In the independent roll of two dice, the value of the second die is accepted only if the number is odd; otherwise, the value of the first die is assumed as the second value. By indicating with X and Y the pair of values obtained in each test, find the probability density $p(x, y)$, the marginal densities of X and Y, the mean and standard deviations of X and Y and their covariance.

Chapter 5
Functions of Random Variables

Your textbooks fill with triumphs of linear analysis, its failures buried so deep that the graves go unmarked and the existence of the graves goes unremarked.

Ian Stewart, "Does God Play Dice? The New Mathematics of Chaos".

5.1 Introduction

Before moving on to statistics we still need to deal with the following problem: if we consider several random variables $X, Y, \ldots$ defined on the same probability space and combine them into an analytic function $Z = f(X, Y, \ldots)$ (see Eq. (2.8)), *we get a new random variable* Z. If we know the joint probability density $p_{XY\ldots}(x, y, \ldots)$ of the original random variables, what is the density $p_Z(z)$ of the Z variable?

To fix ideas, let us consider the case of two variables X and Y. Two probability densities p_{XY} and p_Z and a function $f(x, y)$ are then involved, according to the following scheme:

$$X,\ Y \sim p_{XY}(x, y) \Longrightarrow Z = f(X, Y) \sim p_Z(z)\ .$$

The density p_{XY} is known, $f(x, y)$ represents an assigned functional relation (sum, product, exponential or others), and our goal is to determine the density p_Z of the new random variable Z. This scheme, which can obviously be extended to any number of variables, represents the core of the problem we want to solve.

Probability densities will always be indicated with the letter $p(\ldots)$, functional relations with the letter $f(\ldots)$. The random variable Z is defined according to the realizations a and b of the random variables X and Y, according to the scheme of Fig. 5.1:

$$Z = Z(a, b) = f(X(a), Y(b))\ ,\quad a \in A\ ,\quad b \in B\ .$$

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2 139, https://doi.org/10.1007/978-3-031-09429-3_5

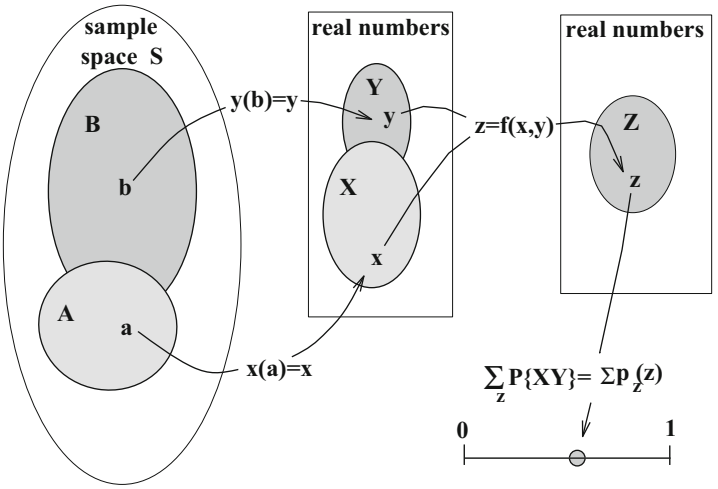


Fig. 5.1 Definition of the random variable $Z = f(X, Y)$

For discrete variables, the probability that Z belongs to a particular numerical set $\mathbb{R}_Z$ is then given, according to Theorem 1.1, by the probability of obtaining $\{X \in \mathbb{R}_X\}$ and $\{Y \in \mathbb{R}_Y\}$ (i.e. $a \in A$ and $b \in B$) added over all cases satisfying the condition $z \in \mathbb{R}_Z$:

$$P\{Z \in \mathbb{R}_Z\} = \sum_{[a \in A, b \in B): f(x,y) \in \mathbb{R}_Z]} p(a, b) = \sum_{[(x \in R_X, y \in R_Y): f(x,y) \in \mathbb{R}_Z]} p_{XY}(x, y) \,. \tag{5.1}$$

If X and Y are independent, one has, from the compound probability Theorem 1.2:

$$P\{z \in \mathbb{R}_Z\} = \sum_{[(x \in R_X, y \in R_y): f(x,y) \in \mathbb{R}_Z]} p_X(x) p_Y(y) \,. \tag{5.2}$$

Figure 5.1 visually shows the meaning of Eqs. (5.1, 5.2).

Let us now consider a simple example: let X and Y be the scores obtained by rolling a dice twice ($1 \leq X, Y \leq 6$), and let $Z = f(X, Y) = X + Y$ the sum of the two scores. Define $\mathbb{R}_Z$ as the set of the results smaller than 5 ($Z = X + Y < 5$), and calculate the probability $P\{Z \in \mathbb{R}_Z\}$. If the two trials are independent, Eq. (5.2) holds and the probability to obtain a generic pair (x, y) is $1/6 \times 1/6 = 1/36$. Eq. (5.1) requires of summing up the probabilities for which $Z < 5$: (1,1), (1,2), (1,3), (2,1), (2,2), (3,1). Since there are six pairs, $P\{Z < 5\} = 6/36 = 1/6$.

The calculation of densities which are functions of random variables, distributed according to the fundamental densities that we have studied so far, is often a very complicated task from the analytic point of view. Sometimes calculations appear to be unsolvable. However, a great help often comes from simple simulation techniques. Suppose, for instance, that we want to determine the distribution of the

variable $Z = \ln(5+X)\cdot\sinh(X)$, where $X \sim N(0,1)$. If we can't find the analytical solution, we can at least know the shape of the distribution with this simple code, which uses the `density` function described in Appendix B:

```
> x <- rnorm(10000)
> z <- log(5+x)*sinh(x)
> plot(density(z,adj=0.01))
```

The curve appearing in the R window shows the behaviour of the solution; we do not have the analytical form, but we can see its trend and calculate its fundamental parameters. For instance, we find that `mean(z)` $\simeq 0.314$ and that `var(z)` $\simeq 7.15$, with an uncertainty that will be calculated in the next chapter, and due to the fact that we have a sample of 10,000 data, but not the parent population. The statistical uncertainty or error can often be made negligible by increasing the sample of simulated data while remaining within reasonable calculation times.

In the following, we will proceed by successive steps, first considering one variable, and then two variables, and finally indicating how to extend the procedure to n variables. We will also show how it is possible, considering only the transformation of the mean and the variance through the function f, to obtain a simple and general solution, although approximate, of the problem. In the most difficult cases, simulation techniques can be used, and they must be part of the basic knowledge of any statistician.

5.2 Functions of a Random Variable

Let X be a continuous random variable with p.d.f. $p_X(x)$, and let Z be a random variable depending on X through the functional relation:

$$Z = f(X)\ . \tag{5.3}$$

To determine the density $p_Z(z)$, the known rules on the change of variable inside a function can be applied. However, attention has to be paid on *how probability intervals are transformed*. It is therefore appropriate to use probability integrals defined by the cumulative functions and use the key Eqs. (2.28, 2.35). In addition, we will also exploit the Leibnitz theorem about the derivation of an integral, the proof of which can be found in many math calculus books. Given a function $z = f(x)$, if $x = f^{-1}(z)$ exists and is differentiable in $[x_1, x_2]$, the equation:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z}\int_{x_1}^{x_2} p(x)\,\mathrm{d}x &= \left(\frac{\mathrm{d}f^{-1}}{\mathrm{d}z}\right)_{z=f(x_2)} p(x_2) - \left(\frac{\mathrm{d}f^{-1}}{\mathrm{d}z}\right)_{z=f(x_1)} p(x_1) \\ &= \frac{p(x_2)}{f'(x_2)} - \frac{p(x_1)}{f'(x_1)}\ ,\end{aligned} \tag{5.4}$$

holds, where the prime symbol indicates the derivative operation.

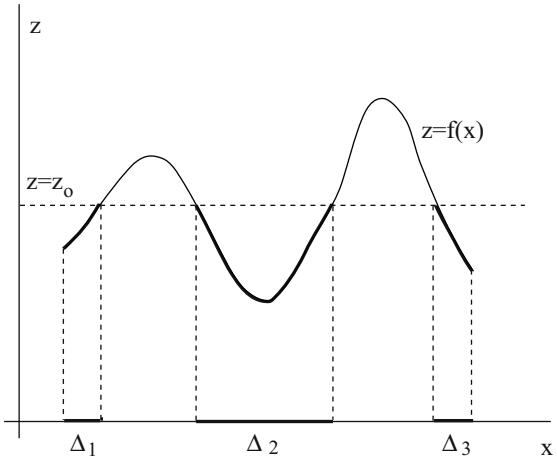


Fig. 5.2 The transformation $z = f(x)$

Let us now consider a generic continuous function $f(X)$, as in Fig. 5.2, and determine the probability that Z is less than a given z_0 value. Basically, we have to find the probability that Z lies below the line $z = z_0$ of Fig. 5.2. From Eqs. (2.28, 5.1) and from the figure above, we obtain:

$$P\{Z \leq z_0\} \equiv F_Z(z_0) = \sum_i \int_{\Delta_i} p_X(x)\,\mathrm{d}x \;, \tag{5.5}$$

where F_Z is the cumulative function of Z and the intervals Δ_i are those of Fig. 5.2. These intervals, except for the first and the last one, have as extremes the real roots $x_1, x_2, \ldots, x_n$ of Eq. (5.3):

$$z_0 = f(x_1) = f(x_2) = \cdots = f(x_n) \;. \tag{5.6}$$

Equation (2.35) shows that, by deriving Eq. (5.5), the required p.d.f. is found. Since from Fig. 5.2 we see that the lower bound of each interval Δ_i does not depend on z (and therefore has a null derivative) or, when it depends on z, always has a negative derivative (the function decreases), applying the Leibnitz formula (5.4) to Eq. (5.5), we can write the density $p_Z(z)$ as the sum of all positive terms, taking the absolute value of the derivatives calculated at the real roots (5.6):

$$\frac{\mathrm{d}F_Z(z)}{\mathrm{d}z} = p_Z(z) = \frac{p_X(x_1)}{|f'(x_1)|} + \frac{p_X(x_2)}{|f'(x_2)|} + \cdots + \frac{p_X(x_n)}{|f'(x_n)|} \;, \tag{5.7}$$

where the right-hand side is a function of $z_0 \equiv z$ through the inverse of Eq. (5.6). The result is always positive, as required for a probability density. When there is only one real root of Eq. (5.3) and $p(x) \geq 0$, this formula coincides with the usual

method of substituting a variable in an integral. This method has already been tacitly applied in the Exercises 3.10 and 3.11 on the Maxwell and Boltzmann distributions.

Let us now apply the fundamental Eq. (5.7) to some other significant case. When Eq. (5.3) is given by:

$$Z = f(X) = aX + b \,, \tag{5.8}$$

Equation (5.6) allows only one solution:

$$x_1 = \frac{z-b}{a} \,.$$

Since $f'(x) = a$, from Eq. (5.7) one obtains:

$$p_Z(z) = \frac{1}{|a|}\, p_X\left(\frac{z-b}{a}\right) \,. \tag{5.9}$$

If, on the other hand, we consider the functional link:

$$Z = f(X) = aX^2 \,, \tag{5.10}$$

Equation (5.6) admits the solutions:

$$x_1 = -\sqrt{\frac{z}{a}} \,, \quad x_2 = +\sqrt{\frac{z}{a}} \,, \tag{5.11}$$

which give rise to the derivatives:

$$|f'(x_1)| = |f'(x_2)| = 2a\sqrt{\frac{z}{a}} = 2\sqrt{az} \,.$$

These results, inserted in Eq. (5.7), allow the determination of the required p.d.f.:

$$p_Z(z) = \frac{1}{2\sqrt{az}}\left[p_X\left(-\sqrt{\frac{z}{a}}\right) + p_X\left(\sqrt{\frac{z}{a}}\right)\right] \,, \quad z \geq 0 \,. \tag{5.12}$$

If the density $p_X(x)$ is the Gaussian of Eq. (3.28) and Eq. (5.8) is applied, Eq. (5.9) becomes:

$$p_Z(z) = \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left[-\frac{[z-(a\mu+b)]^2}{2a^2\sigma^2}\right] \,, \tag{5.13}$$

which is a Gaussian of mean and variance given by:

$$\mu_z = a\mu + b \,, \quad \sigma_z^2 = a^2\sigma^2 \,. \tag{5.14}$$

We note that, since the transformation (5.8) is linear, Eq. (5.14) can also be deduced directly from Eqs. (2.63, 2.64). In the case of the quadratic transformation (5.10) of a Gaussian variable with zero mean, Eq. (5.12) becomes:

$$p_Z(z) = \frac{1}{\sigma\sqrt{2\pi a z}} \exp\left(-\frac{z}{2a\sigma^2}\right) , \quad z > 0 , \tag{5.15}$$

which is the gamma density (3.57) with $k = 1/2$. Mean and variance can be obtained from Eq. (3.58), or by integration by parts of Eqs. (2.54, 2.57), and are:

$$\mu_z = a\sigma^2 , \quad \sigma_z^2 = 2a^2\sigma^4 . \tag{5.16}$$

5.3 Functions of Several Random Variables

We now generalize the results of the previous paragraph to the case of functions of several random variables. Let us first consider the simple case of a single Z function of n random variables:

$$Z = f(X_1, X_2, \ldots, X_n) \equiv f(\boldsymbol{X}) . \tag{5.17}$$

The analogous of the cumulative function (5.5) is now given by:

$$P\{Z \le z_0\} = F_Z(z) = \int \ldots \int_{(\boldsymbol{X} \in D)} p_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_1\, \mathrm{d}x_2 \ldots \mathrm{d}x_n , \tag{5.18}$$

where $p_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n)$ is the p.d.f. of the n variables and D is the set of the n-tuples $\boldsymbol{X} = (x_1, x_2, \ldots, x_n)$ such as $P\{Z \le z_0\}$, according to Eq. (5.1).

It should be noted that the probability density of the variables $\boldsymbol{X}$ appears only as an argument of the integral, while the functional link $Z = f(\boldsymbol{X})$ appears exclusively in the determination of the integration domain D.

In many cases, the derivation of the cumulative (5.18) solves the problem of determining the density of p_z. This method is the generalization of the one used in Sect. 3.8, where we derived Eq. (3.61) to obtain the χ^2 density. As an alternative, to deal with the more general case of n variables $\boldsymbol{Z}$ which are functions of n parent random variables $\boldsymbol{X}$, one can use a well-known formula based on the general theorem of the change of variable in an integrand function. Since this theorem is proved in many mathematical analysis texts, here we report only its statement:

Theorem 5.1 (Change of Variable in Density Functions) *Let* $\boldsymbol{X} \equiv (X_1, X_2, \ldots, X_n)$ *be n random variables with joint density* $p_{\boldsymbol{X}}(\boldsymbol{x})$*, and let* $\boldsymbol{Z} \equiv (Z_1, Z_2, \ldots, Z_n)$

be n variables related to $\boldsymbol{X}$ *by n the functional relationships:*

$$\begin{aligned} Z_1 &= f_1(\boldsymbol{X}) \\ Z_2 &= f_2(\boldsymbol{X}) \\ &\cdots\cdots\cdots \\ Z_n &= f_n(\boldsymbol{X}) \,, \end{aligned} \tag{5.19}$$

which are all invertible and differentiable with continuous derivatives with respect to all arguments (there is a one-to-one correspondence between the two domains of $\boldsymbol{X}$ *and* $\mathbf{Z}$*, for which* $X_1 = f_1^{-1}(\mathbf{Z})$*, etc.).*

The p.d.f. $p_{\mathbf{Z}}$ *is then given by:*

$$\begin{aligned} p_{\mathbf{Z}}(z_1, z_2, \ldots, z_n) &= p_{\boldsymbol{X}}(x_1, x_2, \ldots, x_n)\,|J| \\ &= p_{\boldsymbol{X}}\left(f_1^{-1}(z), f_2^{-1}(z), \ldots, f_n^{-1}(z)\right)|J| \,, \end{aligned} \tag{5.20}$$

where $|J|$ *is the Jacobian, defined as the absolute value of the determinant:*

$$|J| = \begin{vmatrix} \partial f_1^{-1}/\partial z_1 & \partial f_1^{-1}/\partial z_2 & \ldots & \partial f_1^{-1}/\partial z_n \\ \partial f_2^{-1}/\partial z_1 & \partial f_2^{-1}/\partial z_2 & \ldots & \partial f_2^{-1}/\partial z_n \\ \cdots & \cdots & \cdots & \cdots \\ \partial f_n^{-1}/\partial z_1 & \partial f_n^{-1}/\partial z_2 & \ldots & \partial f_n^{-1}/\partial z_n \end{vmatrix} . \tag{5.21}$$

Obviously, the transformation is possible if all the derivatives are continuous and $|J| \neq 0$*. When there is not a unique invertible transformation* f_i $(i = 1, 2, \ldots, n)$*, it is necessary to subdivide the domains of* $\boldsymbol{X}$ *and* $\mathbf{Z}$ *into m disjoint subsets between which there is a one-to-one correspondence and then to sum Eq. (5.20) on these domains:*

$$p_{\mathbf{Z}}(z_1, z_2, \ldots, z_n) = \sum_{L=1}^{m} p_{\boldsymbol{X}}\left(f_{L1}^{-1}(z), f_{L2}^{-1}(z), \ldots, f_{Ln}^{-1}(z)\right)|J_L| \,. \tag{5.22}$$

The theorem can also be applied to the case of the particular transformation (5.17). In fact, let us consider a bivariate Z function:

$$Z = f(X_1, X_2) \,. \tag{5.23}$$

To apply the Jacobian determinant method, we define $Z_1 \equiv Z$ and an auxiliary variable $Z_2 = X_2$. Equation (5.19) then becomes:

$$Z_1 = f(X_1, X_2) \,, \quad Z_2 = X_2 \,. \tag{5.24}$$

The density of $Z \equiv Z_1$ can then be found by applying Eq. (5.20) and then integrating on the auxiliary variable $Z_2 = X_2$. Since the Jacobian is:

$$|J| = \begin{vmatrix} \frac{\partial f_1^{-1}}{\partial z_1} & \frac{\partial f_1^{-1}}{\partial z_2} \\ 0 & 1 \end{vmatrix} = \left| \frac{\partial f_1^{-1}}{\partial z_1} \right| \,,$$

from Eq. (5.20) one obtains:

$$p_{\boldsymbol{Z}}(z_1, z_2) = p_{\boldsymbol{X}}(x_1, x_2) \left| \frac{\partial f_1^{-1}}{\partial z_1} \right| \,. \tag{5.25}$$

The Z p.d.f. is obtained by integration on the auxiliary variable Z_2:

$$p_{Z_1}(z_1) = \int p_{\boldsymbol{Z}}(z_1, z_2) \, \mathrm{d}z_2 \,. \tag{5.26}$$

Recalling that $Z_1 \equiv Z$, $X_2 \equiv Z_2$ and that:

$$p_Z(z) \equiv p_{Z_1}(z_1) \,, \quad \frac{\partial f_1^{-1}}{\partial z_1} \equiv \frac{\partial f_1^{-1}}{\partial z} \,, \quad X_1 = f_1^{-1}(Z_1, X_2) = f_1^{-1}(Z, X_2) \,,$$

we can write Eq. (5.26) as:

$$\begin{aligned} p_Z(z) &= \int p_{\boldsymbol{X}}(x_1, x_2) \left| \frac{\partial f_1^{-1}}{\partial z} \right| \mathrm{d}x_2 \\ &= \int p_{\boldsymbol{X}} \left(f_1^{-1}(z, x_2), x_2 \right) \left| \frac{\partial f_1^{-1}}{\partial z} \right| \mathrm{d}x_2 \,, \end{aligned} \tag{5.27}$$

which represents the requested p.d.f.. This formula provides an alternative to Eq. (5.18) to find the density $p_Z(z)$ when the functional relationship is of the type $Z = f(X_1, X_2)$. If the variables X_1 and X_2 are independent, the density $p_{\boldsymbol{X}}$ factorizes according to Eq. (4.6), and Eq. (5.27) becomes:

$$p_Z(z) = \int p_{X_1} \left(f_1^{-1}(z, x_2) \right) p_{X_2}(x_2) \frac{\partial f_1^{-1}}{\partial z} \, \mathrm{d}x_2 \,. \tag{5.28}$$

When the variable Z is given by the sum:

$$Z = X_1 + X_2 \,, \tag{5.29}$$

the inverse function f^{-1} and its derivative to be inserted in Eq. (5.27) are given by:

$$X_1 = f_1^{-1}(Z, X_2) = Z - X_2 \,, \qquad \frac{\partial f_1^{-1}}{\partial z} = 1 \,, \tag{5.30}$$

and the following result:

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{\boldsymbol{X}}(z - x_2, x_2)\, \mathrm{d}x_2 \tag{5.31}$$

is obtained.

If the two variables X_1 and X_2 are also independent, then the p.d.f. of Eq. (4.3) factorizes in Eq. (4.6), and the previous integral becomes:

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{X_1}(z - x_2)\, p_{X_2}(x_2)\, \mathrm{d}x_2 \,. \tag{5.32}$$

This is called the *convolution integral*. It is often met, both in statistics and in experimental physics, during the analysis of laboratory measurements. In these cases, p_Z is an observed random signal (for instance, an image), p_{X_2} is the true signal (the true image) and p_{X_1} is a blurring or apparatus function. In such conditions, it is necessary to determine p_{X_2} when p_Z is observed and p_{X_1} is known. This is achieved by integral inversion using *deconvolution* algorithms. You can easily imagine the importance and the widespread use of these techniques, from medical diagnostics to astrophysics. We will return to this issue in Sect. 12.15.

After so much mathematics, we also note that this last integral has a simple intuitive explanation: the probability of observing a value $Z = z$ is given by the probability of obtaining a value x_2 times the probability of having a value $x_1 = z - x_2$, so as to satisfy the equality $z = x_1 + x_2$. This probability must be added for all the possible values of X_2. The convolution integral thus appears as a further application of fundamental laws (1.23, 1.24) for continuous variables. We also note that Eq. (5.31) can be derived from the cumulative function (5.18). Indeed, since:

$$\begin{aligned} P\{(X + Y \le z)\} = F_Z(z) &= \int_{(X+Y \le z)} p_{\boldsymbol{X}}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\ &= \int_{-\infty}^{+\infty} \mathrm{d}x_2 \int_{-\infty}^{z-x_2} p_{\boldsymbol{X}}(x_1, x_2)\, \mathrm{d}x_1 \,, \end{aligned}$$

Equation (5.31) can be obtained again by deriving with respect to z and applying Eq. (5.4).

In R there are many possible ways to perform convolution integrals. Our routine `ConvFun` solves Exercise 5.1 and, with few modifications, also the other exercises and in general many simple convolution problems. If we denote by `fun1(x)` and `fun2(x)` two R or user functions which deliver a value according to the input value `x`, the lines of code of `ConvFun` that calculate the convolution integral between `fun1` and `fun2` are given by:

```
> f.X <- function(x) fun1(x)
> f.Y <- function(x) fun2(x)
>   #   $value extracts from integrate the value of the integral
> f.Z <- function(z)
>        integrate(function(x,z) f.X(z-x)*f.X(x),-Inf,+Inf,z)$value
> f.Z <- Vectorize(f.Z)
>   # as an example,
>   # the z vector has limits [-4,+4] in steps of 0.02
> z <- seq(-4,+4,0.02)
> plot(z,f.Z(z),type='l')
```

The first statements formally define the two functions `f.X` and `f.Y`, and then a third function `f.Z` containing the routine `integrate` that performs the convolution. These lines of interactive code are all variables that contain R statements. The `Vectorize` statement is important because it assigns the `f.Z` function to the vector class, so that all R vector functions can be applied to it. After this assignment, if `z` is a vector, the same holds for `f.Z(z)`, which allows it to be used as an argument to `plot` in the next statement. The actual convolution computation occurs within the `plot` call, when you assign the `z` argument to the function `f.Z(z)`. The R online manual contains additional useful information to understand these lines of code.

Exercise 5.1
Find the density of the random variable:

$$Z = X + Y \ ,$$

where $X \sim N(\mu, \sigma^2)$ and $Y \sim U(a, b)$ are independent.

(continued)

Exercise 5.1 (continued)
Answer From Eqs. (3.28, 3.79, 5.32), one has:

$$p_Z(z) = \frac{1}{b-a}\int_a^b \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(z-y-\mu)^2}{2\sigma^2}\right]\mathrm{d}y$$
$$= \frac{1}{b-a}\int_a^b \frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{[y-(z-\mu)]^2}{2\sigma^2}\right]\mathrm{d}y\,. \qquad (5.33)$$

This density is nothing more than a Gaussian with mean $(z-\mu)$ and variance σ^2 integrated within the limits of the uniform density. Using the cumulative Gaussian (3.43), one can rewrite it as:

$$p_Z(z) = \frac{1}{b-a}\left[\Phi\left(\frac{b-(z-\mu)}{\sigma}\right) - \Phi\left(\frac{a-(z-\mu)}{\sigma}\right)\right]. \qquad (5.34)$$

The instrument used to measure physical quantities is often associated with a random uniform dispersion, while the measurement operations are usually associated with a random Gaussian dispersion (see Chapt. 12). In these cases, Eq. (5.34) gives the total smearing of the measure and is therefore important in the study of error propagation, which will be discussed in Sect. 12.9.

The R code lines needed to obtain Fig. 5.3 are:

```
> f.X <- function(x) dnorm(x)
> f.Y <- function(x) dunif(x,min=-2,max=+2)
>   #   $value estracts from integrate the value of the integral
> f.Z <- function(z)
>        integrate(function(x,z) f.X(z-x)*f.X(x),-Inf,+Inf,z)$value
> f.Z <- Vectorize(f.Z)
> z <- seq(-4,+4,0.02)
> plot(z,f.X(z),type='l',lty=2)
> lines(z,f.Y(z),type='l',lty=3)
> lines(z,f.Z(z),type='l',lty=1)
```

This figure shows that the shape of the resulting density is rather similar to a Gaussian. An equivalent result is obtained with a call to our routine `ConvFun(f.X,f.Y,z1=-4,z2=+4)`.

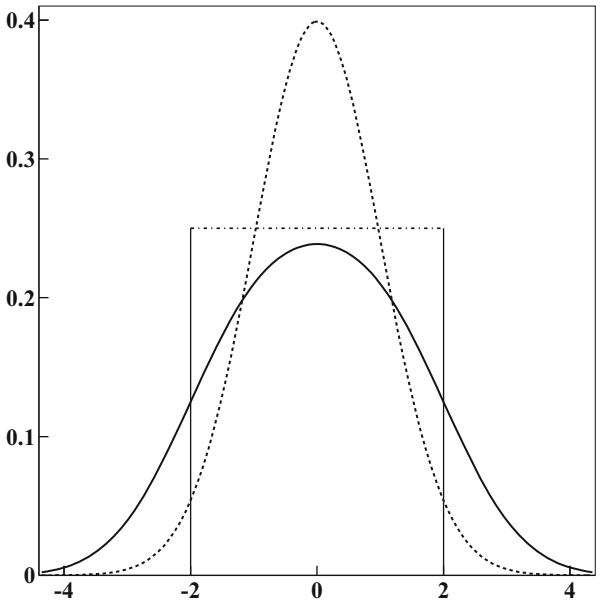


Fig. 5.3 Convolution of a standard Gaussian having $\mu = 0$ and $\sigma = 1$ (dashed line) with an uniform density over $[-2, 2]$ (dash-dotted line). The full curve is the resulting distribution

Exercise 5.2
Find the density of the variable:

$$Z = X + Y \ ,$$

where $X \sim U(0, 1)$ and $Y \sim U(0, 1)$ are independent.

Answer Since $0 \leq X, Y \leq 1$, one has $0 \leq Z \leq 2$. Also in this case, from Eqs. (3.79, 5.32) one immediately obtains:

$$p_Z(z) = \int u_Y(z - x)\, u_X(x)\ \mathrm{d}x \ ,$$

where $u(x)$ is the uniform density:

$$u(x) = \begin{cases} 1 \text{ if } \quad 0 \leq x \leq 1 \\ 0 \text{ otherwise} \ . \end{cases}$$

(continued)

Exercise 5.2 (continued)
The uniform density arguments, which are the variables $(z-x)$ and (x), must therefore lie between 0 and 1. The integral is then composed of two terms:

$$p_Z(z) = \left[\int_0^z \mathrm{d}x\right]_{0\le z\le 1} + \left[\int_{z-1}^1 \mathrm{d}x\right]_{1<z\le 2} ,$$

and gives the result:

$$p_Z(z) = \begin{cases} z & \text{if} \quad 0 \le z \le 1 \\ 2-z & \text{if} \quad 1 < z \le 2 \\ 0 & \text{otherwise} \end{cases} . \tag{5.35}$$

This density is normalized between 0 and 2, triangular and with a maximum in $z = 1$. Also this function will be extensively discussed during the study of the error propagation of two measurements affected by instrumental errors, which will be carried out in Sect. 12.9.

Exercise 5.3
Find the density of the variable:

$$Z = X + Y ,$$

where $X, Y \sim N(\mu, \sigma^2)$ are two independent Gaussian variables.

Answer Also in this case, from Eqs. (3.28, 5.32), one immediately obtains:

$$p_Z(z) = \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{+\infty} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^2} - \frac{(z-x-\mu_y)^2}{2\sigma_y^2}\right] \mathrm{d}x .$$

The integral appearing in this formula can be solved with the method discussed in Exercise 3.4. It is of the type:

$$\int_{-\infty}^{+\infty} e^{-Ax^2+2Bx-C}\, \mathrm{d}x = \sqrt{\frac{\pi}{A}} e^{-\frac{AC-B^2}{A}} ,$$

(continued)

Exercise 5.3 (continued)
where:

$$A = \frac{1}{2}\,\frac{\sigma_x^2+\sigma_y^2}{\sigma_x^2\,\sigma_y^2}\,,\quad B = \frac{\mu_x}{2\sigma_x^2}+\frac{z-\mu_y}{2\sigma_y^2}\,,\quad C = \frac{\mu_x^2}{2\sigma_x^2}+\frac{(z-\mu_y)^2}{2\sigma_y^2}\,. \tag{5.36}$$

One then obtains the density:

$$p_Z(z) = \frac{1}{\sqrt{2\pi}\,\sqrt{\sigma_x^2+\sigma_y^2}}\,\exp\left[-\frac{[z-(\mu_x+\mu_y)]^2}{2(\sigma_x^2+\sigma_y^2)}\right]\,, \tag{5.37}$$

which is a Gaussian with mean and standard deviation given by:

$$\mu_z = \mu_x+\mu_y\,,\quad \sigma_z = \sqrt{\sigma_x^2+\sigma_y^2}\,. \tag{5.38}$$

Since the transformation is linear and the variables are independent, Eq. (5.38) is in agreement with Eqs. (4.8, 4.19). However, this exercise tells us a new and very important fact: *the linear composition of Gaussian variables again generates Gaussian densities*.

Equation (5.37) can also be easily proved using the property (C.4) of the generating functions of Appendix C.

When a density retains its functional form by linear composition of several variables, it is said to be *stable*. Notice that according to the Central Limit Theorem 3.1, the sum of N random variables tends to follow the Gaussian distribution for N large enough, in practice for $N > 10$. Well, if the starting variables are already Gaussian, this condition can be removed and the property holds for any N.

The set of these properties is the basis of the central role that the Gaussian or normal density assumes both in probability theory and in statistics.

Exercise 5.4
Find the density of the variable:

$$Z = X+Y\,,$$

where X and Y are two independent Poissonian variables, with means μ_1 and μ_2, respectively.

(continued)

Exercise 5.4 (continued)
Answer Equation (5.32), for integer variables $X, Y \geq 0$, must be rewritten as:

$$p_Z(z) = \sum_{x=0}^{z} p_X(x)\, p_Y(z-x)\,, \tag{5.39}$$

where both the densities p_X and p_Y are given by the Poisson distribution (3.14). Therefore, one has:

$$p_Z(z) = \sum_{x=0}^{z} \frac{\mu_1^x \mu_2^{z-x}}{x!\,(z-x)!}\, e^{-(\mu_1+\mu_2)}\,.$$

Multiplying and dividing by $z!$ and remembering Newton's binomial formula:

$$(\mu_1+\mu_2)^z = \sum_{k=0}^{z} \frac{z!}{k!(z-k)!}\, \mu_1^k \mu_2^{z-k}\,,$$

one obtains:

$$p_Z(z) = e^{-(\mu_1+\mu_2)}\, \frac{1}{z!} \sum_{x=0}^{z} \frac{z!}{x!(z-x)!}\, \mu_1^x \mu_2^{z-x} = \frac{(\mu_1+\mu_2)^z}{z!}\, e^{-(\mu_1+\mu_2)}\,, \tag{5.40}$$

from which it results that the required density is a Poissonian with mean $(\mu_1+\mu_2)$.

We can get the graph and the values of Poissonian convolutions again using the call to our routine `ConvFun(f.X,f.Y,cont=FALSE)`, which can also deal with discrete distributions by applying Eq. (5.39).

We note that, unlike the Gaussian case, *only the sum, but not the difference* of Poissonian variables, is Poissonian. In fact, if $Z = Y - X$, it is possible to have $Z \leq 0$, and in Eq. (5.40) the term $(z+x)!$ appears instead of $(z-x)!$. It is clear then that the p.d.f. of the difference is not Poisson distributed. This distribution can be studied by changing the sign in Eq. (5.39) or again with the call `ConvFun(f.X, f.Y, cont=FALSE, sign=FALSE)`

In the following two exercises, we will determine the Student and Snedecor's densities, which will be used later in statistics. So, don't skip the exercises (at least read the sentence and the solution), and pay attention.

Exercise 5.5
Find the probability density of a variable Z defined as the ratio between a standard Gaussian variable and the square root of a Q_R variable following the reduced χ^2 density with ν degrees of freedom. These variables are also mutually independent.

Answer Let us denote by X the standard Gaussian variable with density (3.42) and by Y the χ^2 variable with density (3.67). We must then evaluate the density of the variable:

$$Z = \frac{X}{\sqrt{Y/\nu}} = \sqrt{\nu}\,\frac{X}{\sqrt{Y}}\,. \tag{5.41}$$

Since X and Y are independent, we can apply Eq. (5.28), with:

$$x = f^{-1}(z, y) = \frac{1}{\sqrt{\nu}}\, z\sqrt{y}\,, \quad \frac{\partial f^{-1}}{\partial z} = \frac{\sqrt{y}}{\sqrt{\nu}}\,.$$

We then have, by using the product of the densities (3.42, 3.67):

$$p_Z(z) = \frac{1}{\sqrt{2\pi\nu}\,2^{\nu/2}\,\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty y^{\frac{\nu-1}{2}} \exp\left[-\frac{1}{2}\,y\left(\frac{z^2}{\nu}+1\right)\right] \mathrm{d}y\,.$$

Now let us change the variable of integration as:

$$q = \frac{1}{2}\,y\left(\frac{z^2}{\nu}+1\right)\,, \quad y = \frac{2}{\left(\frac{z^2}{\nu}+1\right)}\,q\,, \quad \mathrm{d}y = \frac{2}{\left(\frac{z^2}{\nu}+1\right)}\,\mathrm{d}q\,,$$

which results in:

$$p_Z(z) = \frac{1}{\sqrt{2\pi\nu}\,2^{\nu/2}\,\Gamma\left(\frac{\nu}{2}\right)}\,\frac{2^{\nu/2}}{\sqrt{2}}\,\frac{1}{\left(\frac{z^2}{\nu}+1\right)^{(\nu-1)/2}}\,\frac{2}{\left(\frac{z^2}{\nu}+1\right)}$$
$$\times \int_0^\infty q^{\frac{\nu-1}{2}}\,e^{-q}\,\mathrm{d}q\,.$$

If we recall the definition (3.64) of the gamma function, a direct calculation gives:

$$p_Z(z) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi}}\,\frac{1}{\sqrt{\nu}}\left(\frac{z^2}{\nu}+1\right)^{-\frac{\nu+1}{2}}\,.$$

(continued)

Exercise 5.5 (continued)
This is the well-known Student's density, which takes its name from the pseudonym used by the English statistician W.S. Gosset, who derived it at the beginning of the twentieth century. It is usually written as the density of a variable t, where the identity $\sqrt{\pi} = \Gamma(1/2)$, shown in Eq. (3.65), is also applied.

Therefore, the distribution of a variable t, defined as the ratio between a standard Gaussian variable and the square root of a $\chi^2_R(\nu)$ variable, is:

$$s_\nu(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\nu}} \left(\frac{t^2}{\nu} + 1\right)^{-\frac{\nu+1}{2}} . \tag{5.42}$$

The integral values of the Student's density are shown in Table E.2 in Appendix E. Both from this table and from Fig. 5.4, one can easily verify that this density is very similar to a Gaussian when the number of degrees of freedom is greater than 20–30. The values of the mean and variance can be obtained, as usual, using Eqs. (2.54, 2.57), and are given by:

$$\mu = 0 , \quad \sigma^2 = \frac{\nu}{\nu - 2} . \tag{5.43}$$

The variance is then defined only for $\nu > 2$. For $\nu \leq 2$ the function $s_\nu(t)$ is an example, rather unusual but possible, of a density without variance. In this case, the parametrization of the probability interval (3.94) in terms of standard deviation is no longer possible, and to obtain a given probability level, it is necessary to directly calculate the integral of the density within the assigned limits. These probabilities can also be obtained from Tab. E.2.

In R the function `t( ,df, )` computes the Student's distribution with `df` degrees of freedom. The call sequences use standard R prefixes:

```
dt(x,df) # function value in x
pt(q,df) # cumulative value of the quantile q
qt(p,df) # quantile value of index p
rt(n,df) # vector of n random variates of t
```

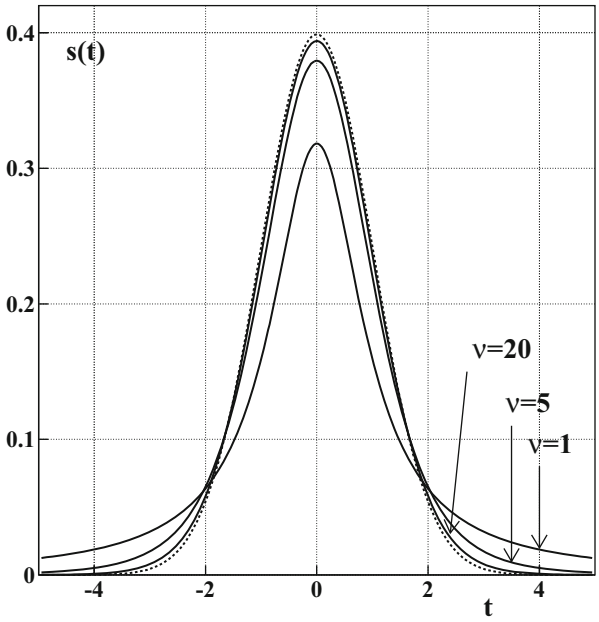


Fig. 5.4 Student's density for 1, 5 and 20 degrees of freedom. The dashed line represents the standard Gaussian

Exercise 5.6
Find the p.d.f of a random variable F given by the ratio of two independent random variables following the reduced χ^2 distribution, with respective degrees of freedom μ and ν:

$$\boldsymbol{F} = \frac{Q_R(\mu)}{Q_R(\nu)} \,. \tag{5.44}$$

According to the statistical practice, the ratio (5.44) should be written in capital letters. We will then denote with F the values assumed by the Snedecor's variable $\boldsymbol{F}$.

Answer We define:

$$\boldsymbol{F} = \frac{Q_R(\mu)}{Q_R(\nu)} \equiv \frac{Y}{X} \equiv f(X, Y) \,,$$

(continued)

Exercise 5.6 (continued)

$$Y = f^{-1}(\boldsymbol{F}, X) = \boldsymbol{F}X \;, \qquad \frac{\partial f^{-1}}{\partial \boldsymbol{F}} = X \;.$$

The reduced χ^2 density with ν degrees of freedom is given by the integrand of Eq. (3.72):

$$p_\nu(x) = a_\nu x^{\frac{1}{2}(\nu-2)}\, e^{-\frac{1}{2}\nu x}\; \mathrm{d}x \;,$$

where:

$$a_\nu = \frac{\nu^{\frac{\nu}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} \;. \tag{5.45}$$

In this case, Eq. (5.28) becomes:

$$\begin{aligned} p_{\mu\nu}(F) &= \int_0^\infty p_\mu\left(f^{-1}(Fx)\right)\, p_\nu(x) \frac{\partial f^{-1}}{\partial F}\; \mathrm{d}x \\ &= a_\mu a_\nu \int F^{\mu/2-1} x^{\mu/2-1} e^{-\mu F x/2}\, x^{\nu/2-1} e^{-\nu x/2} x\; \mathrm{d}x \\ &= a_\mu a_\nu F^{\mu/2-1} \int x^{\frac{1}{2}(\mu+\nu-2)}\, e^{-\frac{1}{2}(\mu F+\nu)x}\; \mathrm{d}x \;. \end{aligned}$$

After the change of variable:

$$\omega = \frac{x}{2}(\mu F + \nu) \;, \qquad \mathrm{d}x = \frac{2}{\mu F + \nu}\; \mathrm{d}\omega \;,$$

one finally obtains:

$$p_{\mu\nu}(F) = \frac{a_\mu a_\nu F^{\mu/2-1}\, 2^{\frac{1}{2}(\mu+\nu)}}{(\mu F + \nu)^{\frac{1}{2}(\mu+\nu)}} \int \omega^{\frac{1}{2}(\mu+\nu)-1}\, e^{-\omega}\; \mathrm{d}\omega \;.$$

Recalling Eq. (5.45) and the integral form of the gamma function $\Gamma[(\mu + \nu)/2]$ of Eq. (3.65), we can write:

$$p_{\mu\nu}(F) = c_{\mu\nu}\, F^{\frac{1}{2}(\mu-2)}\, (\mu F + \nu)^{-\frac{1}{2}(\mu+\nu)} \;, \tag{5.46}$$

$$c_{\mu\nu} = \mu^{\frac{\mu}{2}}\, \nu^{\frac{\nu}{2}}\, \frac{\Gamma\left(\frac{\mu+\nu}{2}\right)}{\Gamma\left(\frac{\mu}{2}\right)\, \Gamma\left(\frac{\nu}{2}\right)} \;.$$

(continued)

Exercise 5.6 (continued)
This result represents the well-known Snedecor's density F of the variable $\boldsymbol{F}$. It is displayed in Fig. 5.5 and is extensively used in the analysis of variance (ANOVA) method, which will be discussed later in Sect. 7.9. Mean and variance are as usual calculated from Eqs. (2.54, 2.57) and are given by:

$$\langle \boldsymbol{F} \rangle = \frac{\nu}{\nu - 2} , \qquad \mathrm{Var}[\boldsymbol{F}] = \frac{2\nu^2(\mu + \nu - 2)}{\mu(\nu - 2)^2 (\nu - 4)} . \tag{5.47}$$

These equation are valid for $\nu > 2$ and $\nu > 4$, respectively. When the degrees of freedom $\mu, \nu \to \infty$, the F density tends to a Gaussian distribution; however, as the figure shows, the convergence toward this function is rather slow.

The values of the ratio $F \equiv F_u(\mu, \nu)$ corresponding to the 95th percentile ($u = 0.95$) and to 99th percentile ($u = 0.99$) and, therefore, at the significance levels of 5% and 1% to the right of the mean, are given by:

$$u = \int_0^{F_u(\mu,\nu)} p_{\mu\nu}(F) \, \mathrm{d}F , \tag{5.48}$$

are the most frequently used in ANOVA. They are reported in Tabs. E.5 and E.6 of Appendix E. The percentiles ($u = 5\%$) and ($u = 1\%$), corresponding to the significance levels of the tails to the left of the mean, are generally not given, because of the following crossing property between quantiles:

$$F_u(\mu, \nu) = \frac{1}{F_{1-u}(\nu, \mu)} . \tag{5.49}$$

This equation can be proved by observing that, by definition:

$$1 - u = \int_0^{F_{1-u}(\mu,\nu)} p_{\mu\nu}(F) \, \mathrm{d}F = \int_{F_u(\mu,\nu)}^{\infty} p_{\mu\nu}(F) \, \mathrm{d}F ,$$

and that, since F is a ratio:

$$\text{if } F \sim p_{\mu,\nu}(F) \quad \text{then} \quad \frac{1}{F} \sim p_{\nu,\mu}(F) .$$

Therefore, one can write:

$$u = P\{\boldsymbol{F} \le F_u(\mu, \nu)\} = P\left\{\frac{1}{\boldsymbol{F}} \le F_u(\nu, \mu)\right\}$$

(continued)

Exercise 5.6 (continued)

$$= P\left\{\frac{1}{\boldsymbol{F}} \geq F_{1-u}(\nu,\mu)\right\} = P\left\{\boldsymbol{F} \leq \frac{1}{F_{1-u}(\nu,\mu)}\right\},$$

and Eq. (5.49) follows.

In R, the function `f( , df1, df2, )` evaluates the F density with `df1` and `df2` degrees of freedom, with the calling sequences:

```
df(x,df1,df2) # function value in x
pf(q,df1,df2) # cumulative value of the quantile q
qf(p,df1,df2) # quantile value of index p
rf(n,df1,df2) # vector of r variates of F
```

To numerically verify Eq. (5.49), it is enough to check the equality of quantile values such as `qf(0.3,df1=3,df2=4)` and `1/qf(0.7, df1=4,df2=3)`; in both cases the obtained value is 0.5038967.

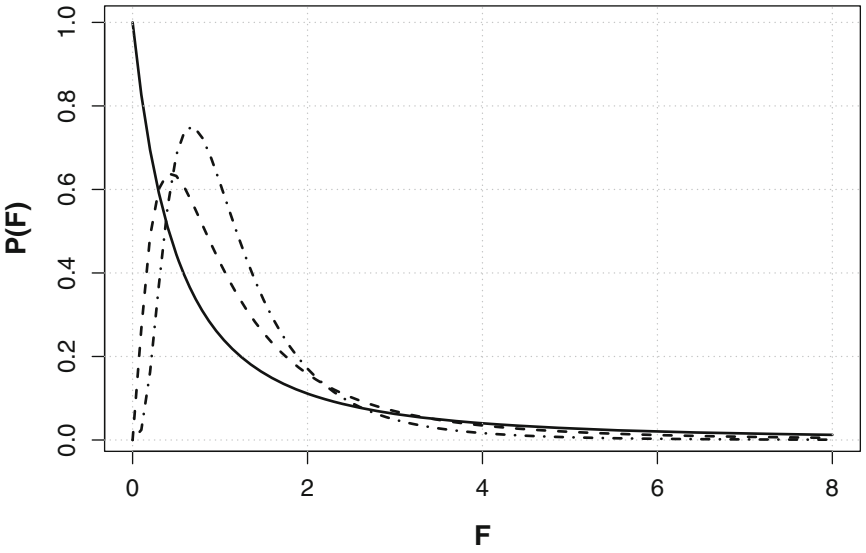


Fig. 5.5 Distribution of the Snedecor's $\boldsymbol{F}$ density with respective degrees of freedom μ and ν. Full curve: $\mu, \nu = 2$; dashed curve: $\mu, \nu = 5$; dash-dotted curve: $\mu, \nu = 10$

One could write some books about functions of random variables (and, as a matter of fact, they have been written ...). If you like mathematical analysis and wish to learn more about this subject, you can refer to the classic text of Papoulis [PUP02].

5.4 Mean and Variance Transformation

As you have seen, the determination of the probability density of a variable expressed as a function of other random variables is a rather complex subject. In the examples developed so far, we have already met some non-trivial mathematical complications, even though we have limited ourselves to the simple case of only two independent variables.

Fortunately, as in the calculation of elementary probabilities, an approximate but satisfactory solution of the problem can almost always be obtained by determining, instead of the complete functional forms, only the central (mean) value and the dispersion (standard deviation) of the density functions under study. This is what we now intend to develop now.

Let us start with the case of a single variable Z which is function of a single random variable X following the known density $p_X(x)$, i.e.:

$$Z = f(X) \; . \tag{5.50}$$

Assuming f to be invertible, $X = f^{-1}(Z)$, and the average of Z is obtained from Eqs. (2.54), (5.7) and differentiating Eq. (5.50):

$$\begin{aligned} \langle Z \rangle &= \int z p_Z(z) \, \mathrm{d}z = \int f(x) p_X(x) \frac{\mathrm{d}z}{|f'(x)|} \\ &= \int f(x) p_X(x) \, \mathrm{d}x \; . \end{aligned} \tag{5.51}$$

It turns then out that the mean of Z is given by the mean of $f(X)$ with respect to the density $p_X(x)$, according to Eq. (2.68), which is the definition of expected value. This result, also valid in a multidimensional space under the conditions of Theorem 5.1, allows to obtain the central value of the density of Z in a correct and quite simple way. In many cases, however, an approximate formula is used, which is obtained by the second-order Taylor expansion of the function $f(x)$ about μ, the mean of the original variable X:

$$f(x) \simeq f(\mu) + f'(\mu)(x - \mu) + \frac{1}{2} f''(\mu)(x - \mu)^2 \; . \tag{5.52}$$

By inserting this expansion into Eq. (5.51), it is easy to verify that the term containing the first derivative vanishes and that, therefore, the approximate relation:

$$\langle Z \rangle \simeq f(\mu) + \frac{1}{2} f''(\mu)\, \sigma^2 \tag{5.53}$$

holds, where σ is the standard deviation of X.

This important equation shows that *the mean of the function f (i.e. of Z) is equal to the function of the mean plus a corrective term that depends on the concavity of the function around the mean of X*. If Eq. (5.50) is linear,

$$Z = aX + b\,,$$

then the second derivative in Eq. (5.53) vanishes and one obtains:

$$\langle Z \rangle = f\left(\langle X \rangle\right)\ . \tag{5.54}$$

In the non-linear case, it is possible to show that $\langle Z \rangle \geq f\left(\langle X \rangle\right)$ if $f''\left(\langle X \rangle\right) > 0$, whereas $\langle Z \rangle \leq f\left(\langle X \rangle\right)$ if $f''\left(\langle X \rangle\right) < 0$.

We now come to the transformation of the variance. As in Eq. (5.51), one can write:

$$\mathrm{Var}[Z] = \int \left[f(x) - \langle f(X) \rangle\right]^2\ p_X(x)\, \mathrm{d}x\ . \tag{5.55}$$

By recalling the approximate result of Eq. (5.53) and using the second-order expansion of Eq. (5.52), one obtains:

$$\begin{aligned}
\mathrm{Var}[Z] &\simeq \int \left[f(x) - f(\mu) - \frac{1}{2} f''(\mu)\sigma^2\right]^2\ p_X(x)\, \mathrm{d}x \\
&= \int \left[f'(\mu)(x-\mu) + \frac{1}{2} f''(\mu)(x-\mu)^2 - \frac{1}{2} f''(\mu)\sigma^2\right]^2\ p_X(x)\, \mathrm{d}x\ .
\end{aligned}$$

Carrying out the square in the integrand and remembering definition (2.59) of the moments of a distribution, after a somewhat long but easy reworking, one obtains:

$$\mathrm{Var}[Z] \simeq \left[f'(\mu)\right]^2 \sigma^2 + \frac{1}{4}\left[f''(\mu)\right]^2 \left(\Delta_4 - \sigma^4\right) + f'(\mu) f''(\mu)\, \Delta_3\ , \tag{5.56}$$

where Δ_i are the moments defined in Eq. (2.59). Since this is still an approximate relation, to have an acceptable estimate of the variance of Z, it is often sufficient to know only the order of magnitude of the moments. If the density of X is symmetrical around the mean, we have $\Delta_3 = 0$. If, in addition to being symmetric, the density is also Gaussian, then, based on Eq. (3.33), $\Delta_4 = 3\,\sigma^4$ and Eq. (5.56) becomes:

$$\mathrm{Var}[Z] \simeq \left[f'(\mu)\right]^2 \sigma^2 + \frac{1}{2}\left[f''(\mu)\right]^2 \sigma^4\ . \tag{5.57}$$

If the density $p_X(x)$ is symmetric but not Gaussian, using Eq. (5.57) usually does not introduce large errors.

When the standard deviation of $p_X(x)$ is small, that is, when $\sigma^2 \gg \sigma^4$, the additional approximation

$$\mathrm{Var}[Z] \simeq [f'(\mu)]^2 \sigma^2 \tag{5.58}$$

holds. This equation becomes exact when between Z and X there exists a linear relation. In this last case, Eqs. (5.54, 5.58) coincide with Eqs. (2.63, 2.64).

It is also useful to verify that, when $Z = aX^2$ and X follow the Gaussian density, Eqs. (5.53, 5.57) give the correct result (5.16).

Let us now deal with the more general situation, consisting of one variable Z which is function of n variables X_i:

$$Z = f(X_1, X_2, \ldots, X_n) \equiv f(\boldsymbol{X}) \,. \tag{5.59}$$

This case can be easily handled if one uses the linear approximation of Eqs. (5.54, 5.58). In other words, we assume that *the mean of the function f coincides with the function of the means and that the variance of Z depends only on the first derivatives of f and on the variances of the distributions of $\boldsymbol{X}$.*

We begin with the simplest situation of two random variables:

$$Z = f(X_1, X_2) \,.$$

If the function f is linearized around the means μ_1, μ_2 of the two original variables X_1, X_2, one obtains:

$$\begin{aligned} z &\simeq f(\mu_1, \mu_2) + \frac{\partial f}{\partial x_1}(x_1 - \mu_1) + \frac{\partial f}{\partial x_2}(x_2 - \mu_2) \\ &\equiv f(\mu_1, \mu_2) + \frac{\partial f}{\partial x_1}\Delta x_1 + \frac{\partial f}{\partial x_2}\Delta x_2 \,, \end{aligned} \tag{5.60}$$

where the derivatives are calculated in $x_1 = \mu_1, x_2 = \mu_2$.

The mean of Z is obtained by extending Eq. (5.51) to two variables:

$$\langle Z \rangle = \int f(x_1, x_2)\, p_{\boldsymbol{X}}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \,, \tag{5.61}$$

and by substituting the expansion (5.60) for $f(x_1, x_2)$. Since the terms of the type $(x_i - \mu_i)\, p_{\boldsymbol{X}}(x_1, x_2)$ are cancelled by the integration, the final result is simply, as always in linear approximation, that the mean of the function coincides with the function of the means:

$$\langle Z \rangle = f(\mu_1, \mu_2) \,. \tag{5.62}$$

The generalization to the case of n variables obviously gives:

$$\langle Z \rangle = f(\mu_1, \mu_2, \ldots, \mu_n) \,. \tag{5.63}$$

The variance of Z is evaluated by considering the generalization of Eq. (5.55):

$$\mathrm{Var}[Z] = \int [f(x_1, x_2) - f(\mu_1, \mu_2)]^2 \; p_{\boldsymbol{X}}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \,, \tag{5.64}$$

after substituting $f(x_1, x_2) - f(\mu_1, \mu_2)$ with the expansion (5.60). We then obtain:

$$\begin{aligned}
\sigma_z^2 \simeq & \left[\frac{\partial f}{\partial x_1}\right]^2 \int (\Delta x_1)^2 p_{\boldsymbol{X}}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\
& + \left[\frac{\partial f}{\partial x_2}\right]^2 \int (\Delta x_2)^2 p_{\boldsymbol{X}}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\
& + 2 \left[\frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2}\right] \int \Delta x_1 \Delta x_2 \; p_{\boldsymbol{X}}(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2 \\
= & \left[\frac{\partial f}{\partial x_1}\right]^2 \sigma_1^2 + \left[\frac{\partial f}{\partial x_2}\right]^2 \sigma_2^2 + 2 \left[\frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2}\right] \sigma_{12} \,.
\end{aligned} \tag{5.65}$$

This result contains some interesting new features: in linear approximation, the variance of z is a function of the variances σ_i^2 of the single variables, of their covariance σ_{12} and of the derivatives in the points $x_1 = \mu_1$, $x_2 = \mu_2$. This law generalizes Eq. (4.19).

If the two variables are independent, they have zero covariance and one obtains:

$$\sigma_{12} = 0 \quad \Longrightarrow \quad \sigma_z^2 = \left[\frac{\partial f}{\partial x_1}\right]^2 \sigma_1^2 + \left[\frac{\partial f}{\partial x_2}\right]^2 \sigma_2^2 \,. \tag{5.66}$$

When $Z = X_1 + X_2$ is given by the sum of two independent variables, the resulting density is given by the convolution integral (5.32), and the explicit calculation of μ_z and σ_z^2 could be rather cumbersome, depending on the complexity of the involved densities. However, in this case Eqs. (5.62) and (5.66) are exact and give the result:

$$\mu_z = \mu_1 + \mu_2 \,, \quad \sigma_z^2 = \sigma_1^2 + \sigma_2^2 \,. \tag{5.67}$$

Therefore, the mean and variance of Z *are known exactly*, even if the explicit form of the final density remains unknown or is too complicated to calculate. Therefore, linear transformations allow to evaluate in an approximate, but simple, way the dispersion of the z values around their mean, by using the criteria of the probability intervals and the 3σ law described in Sects 3.5 and 3.10. In general, this turns out to be an appropriate procedure to solve the problem. Equations (5.62, 5.66) usually

give good (although approximate!) results also when Z is the product or the ratio of independent variables. In this case, Eq. (5.66) gives, both for the product and the ratio, the result:

$$Z = X_1\,X_2\;,\; Z = \frac{X_1}{X_2}\;,\; Z = \frac{X_2}{X_1} \quad \Longrightarrow \quad \frac{\mathrm{Var}[Z]}{\langle Z\rangle^2} = \frac{\mathrm{Var}[X_1]}{\langle X_1\rangle^2} + \frac{\mathrm{Var}[X_2]}{\langle X_2\rangle^2}\;, \tag{5.68}$$

which shows that the relative variance of Z (sometime called the square of the coefficient of variation, $CV = \sigma/\mu$) is the sum of the relative variances of the input variables. However, this is an approximate result, as shown by the following exercise.

Exercise 5.7
Find the exact formula for the variance of the product XY of two independent random variables.

Answer If the two variables are independent, we know, from Eq. (4.9), that the mean of a product is the product of the means. The variance of a product is then given by:

$$\begin{aligned}
\mathrm{Var}[XY] &= \int (xy - \mu_x\mu_y)^2\, p_X(x) p_X(y)\; \mathrm{d}x\,\mathrm{d}y \\
&= \int (x^2y^2 + \mu_x^2\mu_y^2 - 2\mu_x\mu_y xy)\; p_X(x) p_Y(y)\; \mathrm{d}x\,\mathrm{d}y \\
&= \left\langle x^2\right\rangle\left\langle y^2\right\rangle + \mu_x^2\mu_y^2 - 2\mu_x^2\mu_y^2 = \left\langle x^2\right\rangle\left\langle y^2\right\rangle - \mu_x^2\,\mu_y^2\;.
\end{aligned}$$

Recalling Eq. (2.67), we can write:

$$\begin{aligned}
\mathrm{Var}[XY] &= (\sigma_x^2 + \mu_x^2)(\sigma_y^2 + \mu_y^2) - \mu_x^2\mu_y^2 \\
&= \sigma_x^2\sigma_y^2 + \mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2\;,
\end{aligned} \tag{5.69}$$

which is the required solution.

We can compare this equation with Eq. (5.68) if we divide both sides by the product of the squared means:

$$\frac{\sigma_z^2}{\mu_z^2} = \frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} + \frac{\sigma_x^2\sigma_y^2}{\mu_x^2\mu_y^2}\;. \tag{5.70}$$

(continued)

Exercise 5.7 (continued)
This last result shows that Eq. (5.68) holds only if the condition:

$$\frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} \gg \frac{\sigma_x^2 \sigma_y^2}{\mu_x^2 \mu_y^2} \tag{5.71}$$

is verified. This happens when the relative variances are small.

Let us now return to Eq. (5.65), and note that it can be expressed in the matrix form:

$$\sigma_z^2 \simeq \begin{pmatrix} \partial f/\partial x_1 & \partial f/\partial x_2 \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \end{pmatrix} \equiv T\, V\, T^\dagger\,, \tag{5.72}$$

where V is the symmetric covariance matrix (4.62) written for the bidimensional case; T is the derivative matrix, also named gradient or transport matrix ; and † indicates the matrix transposition.

This equation can be interpreted by stating that the variances and covariances of the initial variables are transformed or "transported", by the matrices of the derivatives, through the function f, to obtain the dispersion of the variable Z. Equation (5.72) can be immediately extended to the n-dimensional case of Eq. (5.59):

$$\sigma_z^2 \simeq \begin{pmatrix} \partial f/\partial x_1 & \partial f/\partial x_2 & \dots & \partial f/\partial x_n \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{pmatrix} \begin{pmatrix} \partial f/\partial x_1 \\ \partial f/\partial x_2 \\ \dots \\ \partial f/\partial x_n \end{pmatrix}$$

$$\equiv T\, V\, T^\dagger\,. \tag{5.73}$$

If all the variables are independent, the covariances are zero, and the previous equation directly gives the generalization of Eq. (5.66):

$$\sigma_z^2 \simeq \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2_{x_i = \mu_i} \sigma_i^2\,, \tag{5.74}$$

where the derivatives are calculated, as usual, at the mean values of X_i.

Recall that Eqs.(5.60–5.74) are correct only for linear transformations of the type $Z = b + \sum a_i X_i$ with a and b constant coefficients. However, they provide fairly accurate results even in non-linear cases, when the densities involved are fairly symmetrical, the number of initial variables is large and their relative variances are small.

Fortunately, for complicated cases there is a method, based on simulation techniques, which allows the calculation of the variances of any multivariate function in a very simple and effective way. It will be described in detail in Sect. 8.9.

5.5 Means and Variances for *n* Variables

Let us now deal with the more general case of m variables Z_k that are functions of n variables X_i:

$$\begin{aligned} Z_1 &= f_1(X_1, X_2, \ldots, X_n) \\ Z_2 &= f_2(X_1, X_2, \ldots, X_n) \\ \ldots &= \ldots\ldots\ldots\ldots \\ Z_m &= f_m(X_1, X_2, \ldots, X_n) \, . \end{aligned} \tag{5.75}$$

If we remain within the linear approximation, the m means of the Z variables are obviously given by:

$$\begin{aligned} \langle Z_1 \rangle &= f_1(\mu_1, \mu_2, \ldots, \mu_n) \\ \langle Z_2 \rangle &= f_2(\mu_1, \mu_2, \ldots, \mu_n) \\ \ldots &= \ldots\ldots\ldots\ldots \\ \langle Z_m \rangle &= f_m(\mu_1, \mu_2, \ldots, \mu_n) \, . \end{aligned} \tag{5.76}$$

To determine the variances (and the covariances!) of the variables Z_k, we need to generalize Eq. (5.73). The procedure does not present conceptual difficulties: it is necessary to start from the covariance matrix (4.62) of the variables $\boldsymbol{X}$, to perform the product row by column with the transport matrices T and $T^\dagger$ and to obtain the covariance matrix of the variables $\boldsymbol{Z}$:

$$\begin{aligned} V(\boldsymbol{Z}) \quad &\simeq T \, V(\boldsymbol{X}) \, T^\dagger \, , \\ \mathrm{Cov}[Z_i, Z_k] &\simeq \textstyle\sum_{j=1}^{n} \sum_{l=1}^{n} T_{ij} \, \sigma_{jl} \, T^\dagger_{lk} \, , \quad (i, k = 1, 2, \ldots, m) \, . \end{aligned} \tag{5.77}$$

The covariances $V(\boldsymbol{X})$ and $V(\boldsymbol{Z})$ are square and symmetric matrices with dimension $n \times n$ and $m \times m$, respectively, whereas T is a transport matrix $m \times n$ given by:

$$T = \begin{pmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \dots & \partial f_1/\partial x_n \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 & \dots & \partial f_2/\partial x_n \\ \dots & \dots & \dots & \dots \\ \partial f_m/\partial x_1 & \partial f_m/\partial x_2 & \dots & \partial f_m/\partial x_n \end{pmatrix} , \tag{5.78}$$

and $T^\dagger$ is the $n \times m$ transposed matrix of T. We recall that, from Definition 4.3, $\mathrm{Cov}[X_i, X_i] \equiv \mathrm{Var}[X_i] = \sigma_{ii}$, $\mathrm{Cov}[X_i, X_j] = \sigma_{ij}$. If we introduce the compact notation:

$$\frac{\partial f_i}{\partial x_k} \equiv (\partial_k f_i) , \tag{5.79}$$

we can write Eq. (5.77) as:

$$\mathrm{Cov}[Z_i, Z_k] \simeq \sum_{j,l=1}^{n} (\partial_j f_i)\, \sigma_{jl}\, (\partial_l f_k) . \tag{5.80}$$

This equation allows to calculate, in a fairly simple way, the dispersions of the variables Z and their possible covariances and correlations. For example, when the input variables are independent, all covariances are zero, that is, $\sigma_{ij} = 0$ for $i \neq j$ and Eq. (5.80) becomes:

$$\mathrm{Cov}[Z_i, Z_k] \equiv \sum_{j=1}^{n} (\partial_j f_i)\, \sigma_{jj}\, (\partial_j f_k) . \tag{5.81}$$

Since it is often necessary to calculate covariances of functions of random variables, we want to show you in detail how to do it. We will describe the case of two variables Z_1, Z_2 and X_1, X_2, because the generalization to functions containing a greater number of variables is obvious.

The problem consists, once the variances and covariances of X_1, X_2 have been obtained from the data, in determining the variances and covariances of the variables Z_1, Z_2. As a matter of fact, everything is implicitly contained in Eq. (5.80). The covariance $\mathrm{Cov}[Z_1, Z_2]$ is then given by:

$$\begin{aligned} \mathrm{Cov}[Z_1, Z_2] &\simeq \sum_{j,l=1}^{2} (\partial_j f_1)\, \sigma_{jl}\, (\partial_l f_2) \\ &= (\partial_1 f_1)\, \sigma_1^2\, (\partial_1 f_2) + \end{aligned}$$

$$
\begin{aligned}
&(\partial_1 f_1)\,\sigma_{12}\,(\partial_2 f_2) + \\
&(\partial_2 f_1)\,\sigma_{21}\,(\partial_1 f_2) + \\
&(\partial_2 f_1)\,\sigma_2^2\,(\partial_2 f_2) \\
&= (\partial_1 f_1)\,(\partial_1 f_2)\,\sigma_1^2 + (\partial_2 f_1)\,(\partial_2 f_2)\,\sigma_2^2 + \\
&\quad [(\partial_1 f_1)\,(\partial_2 f_2) + (\partial_2 f_1)\,(\partial_1 f_2)]\,\sigma_{12}\ ,
\end{aligned}
\tag{5.82}
$$

where the equality $\sigma_{12} = \sigma_{21}$ has been used in the last row.

Matrix notation is convenient and compact, but we remind you that Eq. (5.82) can also be directy demonstrated, from the covariance definition (4.24), by expanding the z variables around their mean as Taylor series up to the first order. In this way, one has:

$$
\begin{aligned}
\mathrm{Cov}[Z_1, Z_2] &= \langle (Z_1 - \langle Z_1 \rangle)\,(Z_2 - \langle Z_2 \rangle) \rangle \\
&\simeq \langle\, [\,(\partial_1 f_1)\Delta X_1 + (\partial_2 f_1)\Delta X_2\,]\,[\,(\partial_1 f_2)\Delta X_1 + (\partial_2 f_2)\Delta X_2\,]\,\rangle\ .
\end{aligned}
\tag{5.83}
$$

Going on with the calculation and taking into account that:

$$
\left\langle (\Delta X_1)^2 \right\rangle = \sigma_1^2\ ,
$$

$$
\left\langle (\Delta X_2)^2 \right\rangle = \sigma_2^2\ ,
$$

$$
\langle (\Delta X_1)(\Delta X_2) \rangle = \sigma_{12} = \sigma_{21}\ ,
$$

Equation (5.82) is again obtained.

Exercise 5.8
Two random variables Z_1 and Z_2 depend on two standard independent Gaussian variables X and Y according to the functions:

$$
\begin{aligned}
Z_1 &= X + 3Y\ , \\
Z_2 &= 5X + Y\ .
\end{aligned}
$$

Find the linear correlation coefficient between Z_1 and Z_2.

Answer Even if X and Y are independent, the functional link creates a dependence between Z_1 and Z_2.

(continued)

Exercise 5.8 (continued)

Defining $X_1 \equiv X$ and $X_2 \equiv Y$ and using the notation of Eq. (5.79), one easily finds:

$$(\partial_1 f_1) = 1 \,, \qquad (\partial_2 f_1) = 3 \,,$$

$$(\partial_1 f_2) = 5 \,, \qquad (\partial_2 f_2) = 1 \,.$$

To determine the linear correlation coefficient of Eq. (4.31), it is, at first, necessary to find the variances and the covariance of Z_1 and Z_2.

Since the input variables are independent, to evaluate the variances of Z, we just need to apply Eq. (5.81) and keep in mind that the standard variables have unit variance:

$$\mathrm{Var}[Z_1] = (1)^2\,\sigma_1^2 + (3)^2\,\sigma_2^2 = 10 \,,$$

$$\mathrm{Var}[Z_2] = (5)^2\,\sigma_1^2 + (1)^2\,\sigma_2^2 = 26 \,.$$

Since X and Y are independent standard random variables, the covariance between the Z variables is evaluated through Eq. (5.82), with $\sigma_1^2 = \sigma_2^2 = 1$ and $\sigma_{12} = 0$:

$$\mathrm{Cov}[Z_1, Z_2] = (5 \cdot 1)\,\sigma_1^2 + (3 \cdot 1)\,\sigma_2^2 + (1 \cdot 1 + 3 \cdot 5)\,\sigma_{12} = 5 + 3 = 8 \,.$$

From Eq. (4.31), one finally obtains:

$$\rho[Z_1, Z_2] = \frac{8}{\sqrt{10}\,\sqrt{26}} = 0.496 \,.$$

Exercise 5.9

Two random variables X and Y have known mean, variance and covariance: μ_x, μ_y, σ_x^2, σ_y^2, σ_{xy}. The transformation:

$$Z_1 = 5X + Y \,,$$
$$Z_2 = X \cdot Y \,.$$

is applied to them. Find the covariance $\mathrm{Cov}[Z_1, Z_2]$ between Z_1 and Z_2.

(continued)

Exercise 5.9 (continued)
Answer Let μ_1 and μ_2 be the means of Z_1 and Z_2, respectively. Using Eq. (5.83) and from a series expansion of the two new variables as a function of the old ones, we obtain:

$$\begin{aligned}
\mathrm{Cov}[Z_1, Z_2] &= \langle (Z_1 - \mu_1)\,(Z_2 - \mu_2) \rangle \\
&= \left\langle \left(\frac{\partial Z_1}{\partial X}\,\Delta X + \frac{\partial Z_1}{\partial Y}\,\Delta Y \right) \left(\frac{\partial Z_2}{\partial X}\,\Delta X + \frac{\partial Z_2}{\partial Y}\,\Delta Y \right) \right\rangle \\
&= \frac{\partial Z_1}{\partial X}\,\frac{\partial Z_2}{\partial X} \left\langle (\Delta X)^2 \right\rangle + \frac{\partial Z_1}{\partial X}\,\frac{\partial Z_2}{\partial Y}\,\langle \Delta X\,\Delta Y \rangle \\
&\quad + \frac{\partial Z_1}{\partial Y}\,\frac{\partial Z_2}{\partial X}\,\langle \Delta X\,\Delta Y \rangle + \frac{\partial Z_1}{\partial Y}\,\frac{\partial Z_2}{\partial Y} \left\langle (\Delta Y)^2 \right\rangle \\
&= 5Y\ \mathrm{Var}[X] + X\ \mathrm{Var}[Y] + (5X + Y)\ \mathrm{Cov}[X, Y]\,. \qquad (5.84)
\end{aligned}$$

Since this expansion is made around the mean values, the variables X and Y appearing in the derivatives are the mean values of μ_x and μ_y. Since these values are known, the problem is solved.

Exercise 5.10
The measured coordinates of a point in the x-y plane are considered to be random variables with standard deviations equal to 0.2 cm (for x) and 0.4 cm (for y). These variables are uncorrelated. Determine the covariance matrix in polar coordinates at the point $(x, y) = (1, 1)$.

Answer Since the coordinates are uncorrelated, the covariance matrix of the original variables is:

$$V_{xy} = \begin{pmatrix} 0.04 & 0 \\ 0 & 0.16 \end{pmatrix}\,.$$

The transformation to polar coordinates is given by:

$$r = \sqrt{x^2 + y^2}\,, \qquad \varphi = \arctan \frac{y}{x}\,.$$

(continued)

Exercise 5.10 (continued)
The transport matrix of this transformation is:

$$T = \begin{pmatrix} \dfrac{x}{r} & \dfrac{y}{r} \\ \\ -\dfrac{y}{r^2} & \dfrac{x}{r^2} \end{pmatrix} ,$$

which, at the $(x, y) = (1, 1)$ point, corresponding in polar coordinates to $(r, \varphi) = (\sqrt{2},\ \pi/4)$, becomes:

$$T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} ,$$

The problem is now solved with Eq. (5.77):

$$V_{r,\varphi} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0.04 & 0 \\ \\ 0 & 0.16 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{2} \\ \\ \frac{1}{\sqrt{2}} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 0.100 & 0.042 \\ \\ 0.042 & 0.050 \end{pmatrix} .$$

The square root of the non-diagonal elements of $V_{r,\varphi}$ gives the standard deviations of the variables $r,\ \varphi$. Therefore, given the measurement:

$$x = 1 \pm 0.2\ \text{cm} , \quad y = 1 \pm 0.4\ \text{cm} ,$$

the transformation to polar coordinates provides the values:

$$r = \sqrt{2} \pm \sqrt{0.100} = 1.41 \pm 0.32\ \text{cm} , \quad \varphi = \frac{\pi}{4} \pm \sqrt{0.050} = 0.78 \pm 0.22\ \text{rad} .$$

In addition, the non-zero off-diagonal elements of the matrix signify that the variable transformation has introduced a positive correlation between r and φ.

The problem can also be solved with the following R commands:

```
> x=1;  y=1;
> r=sqrt(x^2+y^2);
> Vxy <- matrix(c(0.04,0.,0.,0.16).byrow=T,ncol=2)
> T <- matrix(c(x/r,y/r,-y/r^2,x/r^2),byrow=T,ncol=2)
> TD <- t(T)   # TD is the transpose
> Vrphi <- T %*% Vxy %*% TD # %*% is the row/column multiplication
```

Exercise 5.11
Prove that the covariance between the variables $(I_i,\ I_j)$ of the multinomial distribution is given by Eq. (4.92).

Answer In a multinomial distribution, there exists a correlation between the bin contents $\{I_i = n_i\}$:

$$f(n_i) \equiv (n_1 + n_2 + \cdots + n_k) = N\ . \tag{5.85}$$

Since N is fixed, $\mathrm{Var}[N] = 0$. If one applies transformation (5.80) to Eq. (5.85), the result:

$$\begin{aligned}\mathrm{Var}[N] &= \sum_{ij} (\partial_i f)\, \sigma_{ij}\, (\partial_j f) \\ &= \sum_{ij} \sigma_{ij} = \sum_i \sigma_i^2 + \sum_{i \neq j} \sigma_{ij} = 0\ ,\end{aligned}$$

is obtained. From the last line of this equation, and from Eqs. (4.91), one then gets:

$$\sum_{i \neq j} \sigma_{ij} = -\sum_i \sigma_i^2 = -\sum_i N p_i (1 - p_i) = -\sum_{i \neq j} N p_i p_j\ ,$$

where the condition:

$$(1 - p_i) = \sum_j p_j\ , \qquad (i \neq j)$$

has been taken into account. The result of Eq. (4.92) is thus derived:

$$\mathrm{Cov}[I_i, I_j] \equiv \sigma_{ij} = -N p_i p_j\ . \tag{5.86}$$

Finally, remember that all the limitations of the linear approximation discussed above apply to the results of this section.

5.6 Problems

5.1 Find the p.d.f. of the variable $Y = -2 \ln X$ where $X \sim U(0, 1)$ is uniform.

5.2 Find the p.d.f. of $Z = X^2$, where the density of X is $p_X(x) = 2(1 - x),\ 0 \leq x \leq 1$.

5.3 Find the density $p_Z(z)$ of the variable $Z = X/Y$ from the known joint density $p_{XY}(x, y)$.

5.4 The densities of the indipendent variables X and Y are $p_X(x) = \exp[-x],\ x \geq 0$ and $p_Y(y) = \exp[-y],\ y \geq 0$, respectively. Determine the density of $Z = X + Y$.

5.5 Find the density of $X = \sum_{i=1}^{n} T_i$, where the variables T_i are independent random times with negative exponential p.d.f..

5.6 The independent variables X and Y have densities $p_X(x) = \exp[-x],\ x \geq 0$ and $p_Y(y) = \exp[-y],\ y \geq 0$, respectively. Determine the density $p_{ZW}(z, w)$ of the variables $Z = X/(X + Y)$, $W = X + Y$. Try to comment on the result.

5.7 Find the density of $Z = XY$, where X and Y are two independent uniform variables $\sim U(0, 1)$, and calculate $\langle Z \rangle$ and $\mathrm{Var}[Z]$.

5.8 Two devices T_1 and T_2, both having a mean life $1/\lambda$, work in parallel. The second device comes into operation only after the failure of the first. If the operating time of the two devices follows the exponential law, find the p.d.f of the operating time T of the system and its mean life.

5.9 Two random variables $Z_1 = 3X + 2Y$ and $Z_2 = XY$ are given, where X and Y are independent random variables. Find the mean, variance, covariance and correlation of Z_1 and Z_2 when (a) X and Y are standardized and when (b) X and Y have unit mean and variance.

5.10 A company produces both shafts, whose diameter is a Gaussian variable with parameters $\langle X \rangle = 5.450$ and $\sigma[X] = 0.020$ mm, and bearings, whose internal diameter is also a Gaussian variable with parameters $\langle Y \rangle = 5.550$ and $\sigma[Y] = 0.020$ mm. The shaft must be seated within the bearing and the coupling is acceptable when the shaft/bearing clearance is between 0.050 and 0.150 mm. Determine the percentage of discarded assemblies.

5.11 On average, μ vehicles transit from A to B in a given time unit, and λ vehicles do the same in the opposite sense, from B to A. Find the p.d.f. of the total number N of vehicles and the probability to observe k vehicles from A to B over a total of n.

5.12 Verify the numerical values obtained in Exercise 5.8 with simulated data.

Chapter 6
Basic Statistics: Parameter Estimation

In which Alinardo seems to give valuable information, and William reveals his method of arriving at a probable truth through a series of unquestionable errors.

Umberto Eco, "THE NAME OF THE ROSE".

6.1 Introduction

In the previous chapters, we have introduced the main results of probability theory.

Let us now enter the fascinating world of statistics by starting with the fundamental question: what is statistics and how does it differ from probability theory? A first answer can be obtained by carefully considering the following two points:

- *A probability problem*: if we attribute to a coin a true probability equal to $1/2$ of getting head in a flip, what is the probability of getting less than 450 heads in 1000 flips? This problem has been solved in Exercise 3.13.
- *The same problem in statistics*: if 450 heads are obtained in 1000 coin flips, what is the estimate that can be given of the true probability of getting heads, that is, the one that would be obtained in an infinite number of flips?

As can be seen, in the probabilistic approach, a *model distribution is assumed* to be the true one describing the studied process. Then the probability of obtaining a certain *experimental result is estimated* on the basis of this premise. In the statistical approach, instead, starting *from the experimental value*, an interval must be evaluated to determine the true value of the probability.

At this point we realize that this estimate lacks of an essential ingredient: the statistical equivalent of the standard deviation. Here we anticipate an approximate result that will be discussed in the next sections: often in statistics the estimation of the standard deviations can be performed by substituting the true parameters with the measured ones: $\sigma \simeq s$. This procedure is sometimes called error *plug-in*. The estimated standard deviation s thus defined is often called, by physicists and engineers (and generally by all who regularly perform laboratory measurements), as *statistical error*. At an international level [fSI93], the recommended term for

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2 139, https://doi.org/10.1007/978-3-031-09429-3_6

Table 6.1 Difference between probability theory and statistics in the simple case of $x = 450$ successes in $n = 1000$ coin tosses

Probability theory	Statistics
Probability of spectrum values	Parameter estimate ($\hat{p}$)
True probability: $p = 0.5$	Frequency: $f = x/n = 0.45$
Expected value: $\langle X \rangle = 500$	Measured value: $x = 450$
Standard deviation: $\sigma[X] = \sqrt{np(1-p)} = 15.8$	Statistical error or uncertainty: $s = \sqrt{nf(1-f)} = 15.7$

measurement results is *statistical uncertainty*. We therefore have three synonyms: estimated standard deviation (mathematical term), statistical error (language of physicists) and statistical uncertainty (term recommended internationally). In the following, we will mainly use *statistical error*.

Using the formulae of Exercise 3.13, we get Table 6.1, which provides the intervals:

$$\mu \pm \sigma = 500.0 \pm 15.8 \simeq 500 \pm 16 = [484, 516] \quad \text{(probability theory)} ,$$

$$x \pm s = 450.0 \pm 15.7 \simeq 450 \pm 16 = [434, 466] \quad \text{(statistics)} .$$

Despite the apparent analogy, these two intervals have a very different meaning: the first, assuming μ and σ to be known, assigns a probability to a set of values of X, while the second provides an estimate for the value of μ.

The above example refers to the *estimation* of a true unknown parameter starting from the data. As we will see, this operation is performed using the *consistent estimators*, defined in Eq. (2.77) as random variables $T_N(X)$. They are functions of a random sample of size N and converge in probability to a given value. We remind you the Definition 2.12 of random sample.

Hypothesis testing, the other field of application of statistics, tries instead to answer questions like the following: if the experiment consisting of a thousand coin tosses is repeated twice and 450 and 600 heads are obtained, how likely is it that the same coin was used in both experiments? In the next chapter, this topic is described in the simplest case, of rejection or acceptance of a single initial hypothesis. Further on, in Chap. 10, we will explain how to optimize the choice among several alternative hypotheses. The concepts so far exemplified with the coin toss can be precisely defined by using, with a slightly different notation, the probability space of Eq. (1.11):

$$\mathcal{E}(\theta) \equiv (S, \mathcal{F}, P_\theta) , \tag{6.1}$$

where the probability P_θ depends on a parameter θ. Then, the random sampling $(X_1, X_2, \ldots, X_N)$ follows the law:

$$P\{X \in A\} = \int_A p(x; \theta)\, \mathrm{d}x . \tag{6.2}$$

In the case of a coin toss, the probability is discrete and $p(x;\theta) = b(x;1000,p)$, with $\theta = p$. Therefore, we will look for an estimate of θ, and we will see how to perform the hypothesis test on the θ parameter, formalizing what has been intuitively explained Exercises 3.13–3.17.

6.2 Confidence Intervals

In the introductory example of Table 6.1, we have intuitively defined an interval of size $2s$ to estimate, on the basis of the result of a thousand tosses, the expected number of heads as:

$$n \pm \sqrt{nf(1-f)} \simeq 450 \pm 16 \,.$$

Now imagine repeating the same experiment to obtain a new interval, which, before performing the coin tosses, is clearly a random interval that we could define as:

$$X \pm S \,.$$

In statistics, the probability for this interval to include the true probability p of getting heads is called *confidence level* and denoted by $CL \equiv (1-\alpha)$, while α is called significance level. In the ideal case, a satisfactory interval should both have a high confidence level and a small width. To meet these requirements, let us try to define a criterion for choosing an interval with the desired CL following the so-called frequentist interpretation, adopted in the experimental sciences on the basis of a famous work published by the statistician J. Neyman in 1937 [Ney37].

Let us consider the density $p(x;\theta)$ of known functional form, and suppose we want to determine the unknown value of the parameter θ. If θ is a position parameter such as the mean, for different values of θ, the density will shift along the x axis, as shown in Fig. 6.1. The unshaded areas of the densities in Fig. 6.1 correspond to the probability levels:

$$P\{x_1 \le X \le x_2;\theta\} = 1-\alpha = \int_{x_1}^{x_2} p(x;\theta)\,\mathrm{d}x \,, \tag{6.3}$$

where α is the sum of the areas of the two shaded tails. The union of all the intervals $[x_1, x_2]$ of Fig. 6.1 creates a region, in the (x,θ) plane, called *confidence band* with confidence level $CL = 1-\alpha$. This band, looking at Fig. 6.1 from above, shows up as in Fig. 6.2. The two curves delimiting it are increasing monotone functions $\theta_1(x)$ and $\theta_2(x)$. In general, this property is not true, but it can be restored by reparametrizing the problem (e.g., by using, as parameter θ, the mean instead of the probability per unit of time when dealing with negative exponential distributions).

Always keeping in mind Figs. 6.1 and 6.2, suppose now to have measured X and obtained a value x. By tracing the line passing from this point and parallel to the

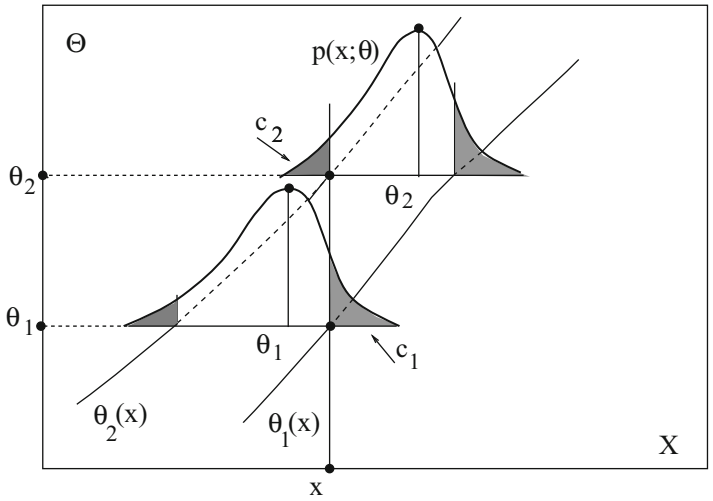


Fig. 6.1 Confidence interval of level CL for the central parameter θ. Unshaded light areas of $p(x;\theta)$ are equal to CL. A variation of the θ value changes the position of $p(x;\theta)$ along the axis of the measured values x. Taking into account the variation of the θ parameter along the vertical axis, one obtains the displayed pattern, where the horizontal width of the Neyman confidence band $[\theta_1(x), \theta_2(x)]$ corresponds to an area equal to CL under $p(x;\theta)$. When an experimental value x is obtained, the points of intersection between the confidence band and the line passing through x and parallel to the θ axis determine the interval $[\theta_1, \theta_2]$ containing the true value of θ with a confidence level $CL = 1 - \alpha$, where $\alpha = c_1 + c_2$ is the sum of the two shaded areas

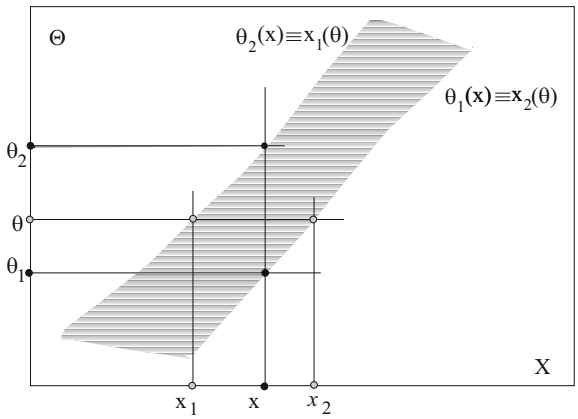


Fig. 6.2 Looking at Fig. 6.1 from above, the Neyman's confidence band for a fixed CL shows up, which allows to determine the confidence interval $\theta \in [\theta_1, \theta_2]$ starting from a measured value x

parameter (vertical) axis, the intersection interval $[\theta_1, \theta_2]$ with the confidence band is obtained. This is the required confidence interval. We note that this is a random interval, since the measured value x varies with each observation. Therefore, we will denote it as $[\Theta_1, \Theta_2]$. If the true value of the unknown parameter is θ, the previous figures show that this procedure leads to the interval $[x_1, x_2]$ on the axis of the measured values. By construction, we then have $P\{x_1 \leq X \leq x_2\} = CL$. Given that when $x = x_1$ on the parameter axis $\theta = \theta_2$ and when $x = x_2$ the condition $\theta = \theta_1$ holds, we have $x \in [x_1, x_2]$ if and only if $\theta \in [\theta_1, \theta_2]$. From these considerations, we finally arrive at the fundamental property of the Neyman *confidence interval* $[\Theta_1, \Theta_2]$ with confidence level $CL = 1 - \alpha$:

$$P\{x_1 \leq X \leq x_2\} = P\{\Theta_1 \leq \theta \leq \Theta_2\} = CL \,. \tag{6.4}$$

We can summarize the previous discussion with the

Definition 6.1 (Confidence Interval) Given two statistics Θ_1 and Θ_2 (in the sense of Definition 2.13) with Θ_1 and Θ_2 continuous variables and $\Theta_1 \leq \Theta_2$ with probability 1, $I = [\Theta_1, \Theta_2]$ is called a confidence interval for a θ parameter, of confidence level $0 < CL < 1$, if, for each θ belonging to the parameter space, the probability that I contains θ is CL:

$$P\{\Theta_1 \leq \theta \leq \Theta_2\} = CL \,. \tag{6.5}$$

If Θ_1 and Θ_2 are discrete variables, the confidence interval is the smallest interval satisfying the condition:

$$P\{\Theta_1 \leq \theta \leq \Theta_2\} \geq CL \,. \tag{6.6}$$

To better highlight the concept of interval "covering" the parameter θ with a certain probability, the confidence level is associated with the terms "coverage" or *coverage probability* [CB90]. The condition (6.6) is said to be *minimum over-coverage*. The *confidence level* therefore coincides with the *coverage* only for continuous variables, and the equal sign in Eq. (6.6) refers to this situation. On the other hand, for discrete variables, the minimum interval ensuring a coverage greater than the requested one must be determined.

Another important feature to be noticed is that the extremes of the confidence interval (6.5) *are random variables, while the θ parameter is fixed*. Consequently, the confidence level refers to the interval $I = [\Theta_1, \Theta_2]$ and indicates *the fraction of experiments that correctly include the true value, in an infinite set of repeated experiments, each of which finds a different confidence interval*. This is equivalent to stating that each particular interval $[\theta_1, \theta_2]$ is obtained with a method that gives the correct result in a fraction CL of the performed experiments.

This frequentist interpretation of the statistical results has a clear operational meaning, very close to what practically happens in laboratory measurements and generally in repeated experiments, and is prevalent in applied sciences. In statistics,

it is legitimate to denote as confidence interval both the random interval $I = [\Theta_1, \Theta_2]$ and its numerical realizations $[\theta_1, \theta_2]$. It is sometimes stated that the CL is the probability that the true value of θ is contained in $[\theta_1, \theta_2]$, but one must always be aware that, in the frequentist interpretation, this phrasing is incorrect, because θ is not a random variable [Cou95].

Let us now assume that θ is a position parameter such as the mean. If a value x of a continuous variable X has been obtained and the density $p(x;\theta)$ of Eq. (6.2) is known, the values $[\theta_1, \theta_2]$ of the random interval $[\Theta_1, \Theta_2]$, for a given CL, can be evaluated with the relations (pay attention to the position of θ_1 and θ_2):

$$\int_x^\infty p(z;\theta_1)\,\mathrm{d}z = c_1\ , \qquad \int_{-\infty}^x p(z;\theta_2)\,\mathrm{d}z = c_2\ , \tag{6.7}$$

where $CL = 1 - c_1 - c_2$. The procedure is also shown in Fig. 6.1. If X is a discrete variable, the integrals must be replaced by the sums over the corresponding spectrum values, as we will shortly see. If a symmetric interval is chosen, then the conditions $c_1 = c_2 = (1-CL)/2$ are valid. The choice of a symmetric interval is obviously not the only possible one, but it is the most common, since it gives the minimum width interval for a symmetric and bell-shaped p.d.f. Sometimes one wants to determine, for a certain CL, only the upper limit of θ, i.e. the interval $(-\infty, \theta]$; in this case, only the second of Eq. (6.7) is used, where $\theta_2 = \theta_U$ and $c_2 = 1 - CL$. For the lower bound, i.e. for the interval $[\theta, +\infty)$, the first of Eq. (6.7) must be used with the conditions $\theta_1 = \theta_L$ and $c_1 = 1 - CL$. The three main types of estimate just described are displayed in Fig. 6.3.

The choice of the interval type is usually determined by the nature of the specific analysed problem. It is however important to keep in mind that this choice must be made *before performing the measurement*. To decide, for example, in the case of rare events, if to determine a symmetrical interval or an upper limit depending on whether the measurement provides results or not (a technique called "flip-flop") leads to an incorrect determination of the levels associated with the confidence intervals [FC98].

Notice that Eq. (6.7) has a general validity since, quite often, the evaluation of a confidence interval requires a consistent estimator T_N satisfying Eq. (2.77) (where $\mu \equiv \theta$). The density of T_N usually depends on the parameter to be estimated and can be denoted as $p(t;\theta)$, with θ considered as a position parameter. Indeed, $\langle T_N \rangle = \int t\, p(t;\theta)\,\mathrm{d}t = \tau(\theta)$, with $\tau(\theta) = \theta$, for unbiased estimators (these concepts will be explored further on, in Chap. 10). If this density is found, the estimate of θ can be performed through Eq. (6.7).

6.3 Confidence Intervals with Pivotal Variables

So far, we have described two ways to determine the confidence interval: the simple graphical method of Fig. 6.1, which however requires the construction of

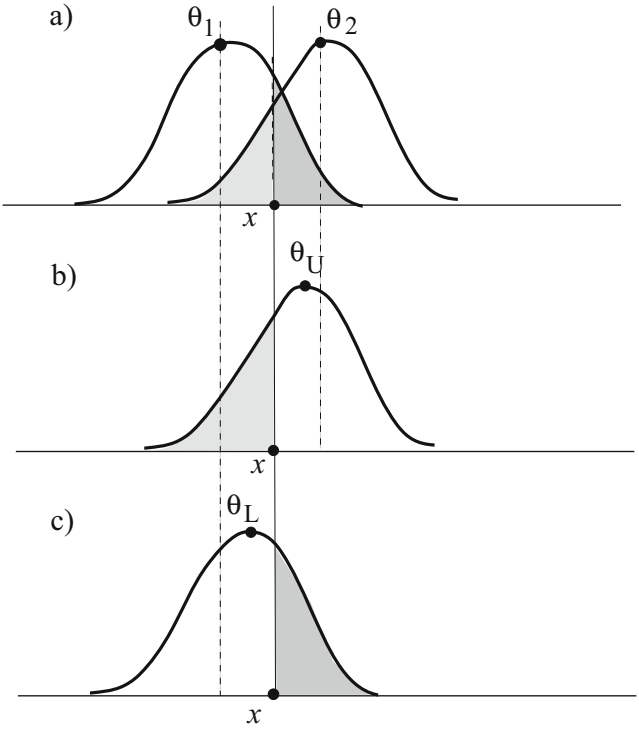


Fig. 6.3 Determination, using the Neyman method, of the bilateral (two-tailed) confidence interval given by Eqs. (6.7) (**a**), of the upper bound θ_U (**b**) and of the lower bound θ_L (**c**) at a measured value x. The sum of the areas of the two shaded tails in (**a**) is $1 - CL$, while each tail in (**b**) and (**c**) has a value $1 - CL$. Therefore, using the same confidence level, one has the situation shown in the figure, with $\theta_2 \geq \theta_{\mathrm{U}}$ and $\theta_1 \leq \theta_{\mathrm{L}}$

the Neyman confidence band, and the calculation of the integrals (6.7). In general, both of them are computationally demanding. Fortunately, if the shape of the density function has some invariance properties with respect to the parameters to be estimated, a particularly simple method can be used. Consider the case where the parameter θ of Fig. 6.1 is the mean of a Gaussian, $\theta \equiv \mu$. Then, the shape of $p(x; \mu)$ is invariant by translation, and the functions $\theta_1(x)$ and $\theta_2(x)$ of Fig. 6.2 are two parallel straight lines. From Fig. 6.4 the following property results:

$$\int_{\mu-t\sigma}^{\mu+t\sigma} p(x;\mu)\ \mathrm{d}x = \int_{x-t\sigma}^{x+t\sigma} p(\mu;x)\ \mathrm{d}\mu\ , \tag{6.8}$$

that can be written as:

$$P\{\mu - t\sigma \leq X \leq \mu + t\sigma\} = P\{-t\sigma \leq X - \mu \leq t\sigma\} = P\{X - t\sigma \leq \mu \leq X + t\sigma\}\ . \tag{6.9}$$

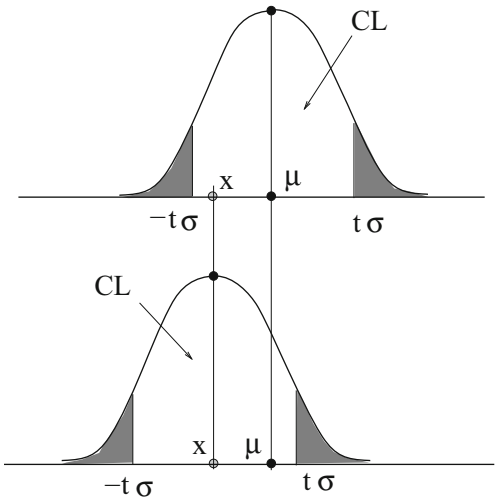


Fig. 6.4 When the shape of the density is invariant for translation of the μ parameter, the confidence interval can be determined with the simple formula (6.9)

In other words, *the probability levels of the interval centred about* μ *coincide with the confidence levels of the random interval centred about* X. The value of X changes at each measurement, but the coverage probability of μ is equal to that of the interval $\mu \pm t\sigma$ (see Fig. 6.4). Using the cumulative function $F(z;\theta)$ (when θ and z are scalar quantities), this property can be written as:

$$F(z;\theta) = 1 - F(\theta;z)\ . \tag{6.10}$$

We therefore have found the following rule of thumb: in the Gaussian case, or in all cases when Eq. (6.10) holds, it is sufficient to centre on the measured value and assume, as confidence levels, the probability levels corresponding to the width of the interval centred on the mean.

For example, we know, considering Table 6.1, that the number x of successes in $n = 1000$ coin flips is a Gaussian variable (because $np,\ n(1-p) \gg 10$). Then, we can estimate the true or expected value of successes as:

$$\begin{aligned} x \pm \sqrt{nf(1-f)} &= 450 \pm 16 \quad (CL = 68.3\%) \\ x \pm 2\sqrt{nf(1-f)} &= 450 \pm 32 \quad (CL = 95.4\%) \\ x \pm 3\sqrt{nf(1-f)} &= 450 \pm 48 \quad (CL = 99.7\%)\ . \end{aligned}$$

As we have just seen, the random variable $Q = (X - \mu)$ includes the parameter μ but has a distribution $N(0,\sigma^2)$, independent of μ. The random variables whose distribution does not depend on the parameter to be estimated are called *pivotal quantities*. Another example occurs when θ is a scale parameter, that is, $p(x;\theta) = h(x/\theta)/\theta$; in this case $Q = X/\theta$ is a pivotal quantity for θ.

Generalizing this argument, we can state that, if $Q(X, \theta)$ is a pivotal quantity, the probability $P\{Q \in A\}$ does not depend on θ for each $A \in \mathbb{R}$. If this distribution has a known density (standard Gaussian, Student, χ^2, ...), the quantiles q_1 and q_2 can be easily determined at a given CL as:

$$P\{q_1 \leq Q(X, \theta) \leq q_2\} = CL \ . \tag{6.11}$$

If the condition $q_1 \leq Q(X, \theta) \leq q_2$ can be solved for θ, one can write, as in Eq. (6.9):

$$P\{q_1 \leq Q(X, \theta) \leq q_2\} = P\{\theta_1(X) \leq \theta \leq \theta_2(X)\} = P\{\Theta_1 \leq \theta \leq \Theta_2\} \ , \tag{6.12}$$

obtaining Eq. (6.5). Therefore, if a pivotal quantity is found which is solvable with respect to θ, according to Eq. (6.12), the confidence interval can be determined *without resorting to integrals (6.7)*. This simple procedure is often the only one reported in elementary texts.

6.4 Mention of the Bayesian Approach

We have just described the basic frequentist method for parameter estimation.

In the Bayesian approach, briefly described in Sect. 1.3, the parameter to be estimated *is considered as a random variable* and the confidence interval represents the knowledge obtained, after the measurement, on the value of this parameter. Let us again assume that 450 heads are obtained in a thousand coin flips. Under the normal approximation and assigning a constant a priori probability to the expected value, it turns out, after the experiment, that the expected value is 450 ± 16 with a degree of credibility (probability or *belief*) of 68%.

In this case the Bayesian approach provides a numerical result equal to the frequentist one but interpreted in a different way, since the Bayesian interval depends on a priori information. In the case of a fair coin ($p = 0.5$), with an uncertainty on the Gaussian balance of, say, $\sigma_p/p = 0.1\%$, we could replace the uniform a priori distribution (constant probability) with the Gaussian distribution $N(p, \sigma_p^2)$, obtaining a result that is numerically different from the frequentist one. We will not elaborate more on these aspects that are treated in detail, from a statistical point of view, in [CB90] and [Gre06].

Finally, we recall that Bayesian analyses have been proposed in physics when it is not easy to find pivotal quantities, as in the case of small counting experiments with background or of samples from Gaussian populations with physical constraints on the measured variables (e.g. if X is a mass, the a priori condition $\{X \geq 0\}$ holds) [Cou95, D'A99]. However, even for these situations, a frequentist approach has been proposed [FC98], which does not require a priori assumptions on the parameter distribution and which has met with the favour of experimental physicists [JLPe00, LW18].

6.5 Some Notations

In the following it will be important to keep in mind the notations used for point and confidence interval estimations. The point estimation of a true value of statistical parameters is obtained using estimator values; for example, the sample mean is a possible estimate of the true mean. The notation we will use is: $m = \hat{\mu}$. We will extensively describe point estimation in Chaps. 10 and 11, while in this chapter we analyze in detail interval estimation. In the Gaussian case, the value θ is contained, with confidence level $1 - \alpha = CL$, in a symmetric interval centred around x when:

$$\theta \in [x - t_{1-\alpha/2}s,\ x + t_{1-\alpha/2}s] \,, \tag{6.13}$$

where $s \simeq \sigma$ and $t_{\alpha/2} = -t_{1-\alpha/2}$ are the standard Gaussian quantiles. It is easy to verify that the quantile indices can be written in terms of CL as:

$$\frac{\alpha}{2} = \frac{1 - CL}{2} \,, \qquad 1 - \frac{\alpha}{2} = \frac{1 + CL}{2} \,. \tag{6.14}$$

In statistics, the 1σ interval is often written as:

$$\theta \in [x - s, x + s] \equiv x \pm s \,. \tag{6.15}$$

The first notation is preferred by mathematicians, the second one by physicists and engineers, who often replace the set membership symbol with that of equality:

$$\theta \in x \pm s \overset{\text{physicists and engineers}}{\rightarrow} \theta = x \pm s \,. \tag{6.16}$$

If the errors to the right and left of the central value are different, the notation of mathematicians obviously does not change, while the other one becomes:

$$\theta \in [x - s_1, x + s_2] \equiv x^{+s_2}_{-s_1} \,. \tag{6.17}$$

It is usually considered improper to assign more than two significant digits to s. For example, if the first significant digit of the error s corresponds to a metre, it makes no sense to give the result with millimetre precision. The following rule of thumb applies, which we report here as:

Statement 6.2 (Significant Digits of the Statistical Error) *Final results of the type $x \pm s$ must be presented with the uncertainty (statistical error) s given with no more than two significant digits and x rounded in the same way. Deviations from this rule must be justified.*

Notice that, in intermediate calculations, more digits can be used to reduce round-off errors; however, in the *final results*, the rule 6.2 should be always followed.

Therefore, the following results are wrong:

$$35.923 \pm 1.407\,, \quad 35.923 \pm 1.4\,, \quad 35.9 \pm 1.407\,,$$

because the first result has too many significant digits, the second and the third ones exhibit a mismatch between result and error. On the contrary, it is correct to write:

$$35.9 \pm 1.4 \quad \text{or} \quad 36 \pm 1\,.$$

6.6 Probability Estimation

Here we consider the following problem: if a Bernoulli test with n trials and x successes is performed and a frequency $f = x/n$ is obtained, what is the estimate of the true probability? From probability theory we know that, after n trials on a system that generates events with constant probability p, we can obtain a number x of successes (spectrum) between 0 and n. However, these results are not equiprobable but are distributed according to the binomial density (2.29).

We should use Eq. (6.7) to solve this problem for discrete binomial random variable. If CL is the required confidence level, the values p_1 and p_2 of the corresponding confidence interval are determined by using the two distributions of Fig. 6.3 a) (where $\theta_1 = np_1$ and $\theta_2 = np_2$). These values can be determined with the so-called Clopper-Pearson equations:

$$\sum_{k=x}^{n} \binom{n}{k} p_1^k (1-p_1)^{n-k} = c_1\,, \tag{6.18}$$

$$\sum_{k=0}^{x} \binom{n}{k} p_2^k (1-p_2)^{n-k} = c_2\,. \tag{6.19}$$

The presence of x in both sum assures the over-coverage condition of Eq. (6.6) for discrete variables. One common choice is the symmetric interval, where $c_1 = c_2 = (1-CL)/2 = \alpha/2$. The solution of these two equations with respect to p_1 and p_2 gives the correct probability estimate from small samples.

The general scheme shown in Fig. 6.3 applies to the determination of the upper and lower limits for a predefined CL. In the case of a discrete binomial distribution, it becomes the one shown in Fig. 6.5.

When $x = 0$ and $x = n$ Eqs. (6.18, 6.19), with $c_1 = c_2 = 1 - CL$, give two important limiting cases:

$$x = n \Longrightarrow \quad p_1^n = 1 - CL\,, \tag{6.20}$$

$$x = 0 \Longrightarrow \quad (1-p_2)^n = 1 - CL\,. \tag{6.21}$$

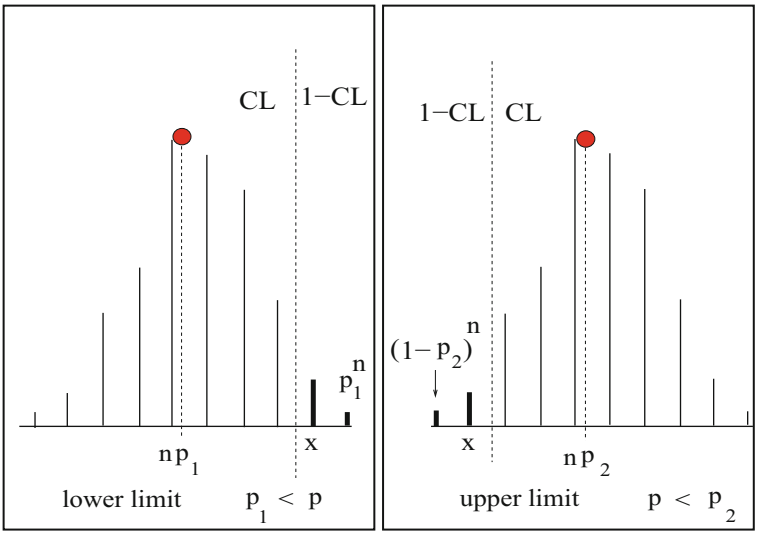


Fig. 6.5 Probability estimation of lower and upper bounds for a predefined CL. In the limiting cases where $x = 0$ and $x = n$, one has $(1 - p_2)^n = 1 - CL$ and $p_1^n = 1 - CL$, respectively

From these equations one obtains, for a fixed CL, the lower bound of a probability when all attempts have been successful:

$$p_1 = \sqrt[n]{1 - CL} = e^{\frac{1}{n}\ln(1-CL)} = 10^{\frac{1}{n}\log(1-CL)} , \tag{6.22}$$

and the upper limit when no success has been recorded:

$$p_2 = 1 - \sqrt[n]{1 - CL} = 1 - e^{\frac{1}{n}\ln(1-CL)} = 1 - 10^{\frac{1}{n}\log(1-CL)} , \tag{6.23}$$

where the use of base-10 or base-e logarithms is useful for large n.

The frequentist interpretation of these limits is that, if the true value were greater (less) than the upper (lower) limit, we would obtain values $\leq$ ($\geq$) than those observed in a fraction of experiments $< CL$.

When n is large and no successes have occurred, p_2 is small. Expanding to the first order the exponential in Eq. (6.23) around the starting point $p_2 = 0$, we obtain the approximation:

$$p_2 \simeq -\frac{1}{n}\ln(1 - CL) , \tag{6.24}$$

corresponding to the equation:

$$e^{-np_2} = (1 - CL) = \alpha , \tag{6.25}$$

which gives the Poissonian probability of getting no events when the mean is $\mu = np_2$.

The formulae just obtained are implemented in the R routine `binom.test(x,n,conf,alt)`, where `x, n, conf` and `alt` are the number of successes, the number of trials, the required confidence level (default value `conf`= 0.95) and the type of interval, respectively. For example, `prop.test(5.20,conf=0.90)` returns the values [0.104, 0.455] in the last four lines of the output message. Further messages (not shown here) refer to the test with a binomial having $p = 0.5$ and should be ignored for the moment. This routine normally provides the Clopper-Pearson bilateral interval (6.18, 6.19), because the variable `alt` is initialized as `alt = "two.sided"`. To obtain the upper limit, one has to set `alt = "less"`, while to get the lower limit the command is `alt = "greater"`. It is instructive, for a given CL, to obtain the values for `alt = "two.sided", "less" and "greater"` and check the situation described in Fig. 6.3.

Exercise 6.1
From an urn containing five black and white marbles in unknown proportions, ten extractions are performed (with replacement), and ten black marbles are extracted. Find the lower limit of the number of black marbles in the urn for $CL = 0.90$. Compare the results with those of Exercise 1.6.

Answer From Eq. (6.22), we get $p = (0.10)^{1/10} = 0.794$. The lower limit for the number of black marbles is $0.794 \cdot 5 = 3.97$. Therefore, we can state that the urn contains at least four marbles with $CL = 0.90$. This result can be obtained also with the R command `binom.test(10,10,conf=0.90,alt="greater")`. An urn with fewer than four black marbles can result in ten consecutive draws of 10 black marbles, but this happens in less than 10% of the experiments.

It is interesting to compare this frequentist solution with the Bayesian result given in Table 1.2: the frequentist estimate is independent of any a priori subjective hypothesis about the initial marble content. Subjective hypotheses usually affect the final results, as shown by the results of Exercise 1.6 and Problem 1.12.

Equations (6.22, 6.23) are important in many reliability problems, as the following examples show.

Exercise 6.2
An emergency pump undergoes a reliability test consisting in 500 "cold starts". If the pump passes the test, what is the probability that it will not start in an emergency situation with a confidence level of 95%?

Answer From Eq. (6.23), one immediately obtains the upper limit:

$$p = 1 - \sqrt[500]{1 - CL} = 1 - \sqrt[500]{0.05} = 0.00597\,,$$

that is about 0.6%. The same result can be obtained with the R command `binom.test(0,500,conf=0.90,alt="less")`. Neyman's interpretation of the result is as follows: a pump with a probability of failure greater than 6 per thousand can start 500 consecutive times, but this happens in less than 5% of tests. Note that this result holds true in the independent tests scheme.

Exercise 6.3
How many consecutive non-failure tests are required to affirm, with $CL = 95\%$, that a device will fail in less than 3% of times?

Answer From Eq. (6.23) using decimal logarithms, one obtains:

$$n = \frac{\log(1 - CL)}{\log(1 - p)} = \frac{\log(1 - 0.95)}{\log(1 - 0.03)} = 98.4\,, \tag{6.26}$$

that is about 100 tests. A device with a probability failure $>$ 3% can successfully pass 100 tests, but this happens in less than 5% of the times.

6.7 Probability Estimation from Large Samples

The Clopper-Pearson formulae (6.18, 6.19), derived in the previous section, are completely general and are valid for both small and large samples. However, they are mathematically laborious to solve in the unknowns p_1 and p_2 and require the use of the R software, so that often approximate formulae are used.

Indeed, we know that the binomial distribution, for np, $n(1 - p) > 10$, rapidly assumes the Gaussian form of mean value np and variance $\sigma^2 = np(1 - p)$. It is therefore extremely important and useful to have simple formulae in Gaussian approximation, which provide practically exact results for large samples.

Consider the frequency $f = x/n$ as the occurrence of the random variable $F = X/n$, which, from Eqs. (2.64, 3.4, 3.6), has mean and variance:

$$\langle F \rangle = \frac{\langle X \rangle}{n} = \frac{np}{n} = p \,, \tag{6.27}$$

$$\mathrm{Var}[F] = \frac{\mathrm{Var}[X]}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \tag{6.28}$$

and define the standard variable (3.37):

$$T = \frac{F-p}{\sigma[F]} \,, \quad \text{which assumes the values} \quad t = \frac{f-p}{\sqrt{\dfrac{p(1-p)}{n}}} \,. \tag{6.29}$$

Under the Gaussian approximation, T can be considered pivotal, and we can thus apply the method described in Sect. 6.3. Since f is known and p is unknown, *using the statistical approach*, we can then determine the values of p for which the value assumed by the standard variable is less than a certain assigned quantile t:

$$\frac{|f-p|}{\sqrt{\dfrac{p(1-p)}{n}}} \leq |t| \,. \tag{6.30}$$

We eliminate the absolute values by squaring both sides and solve, *with respect to the unknown* p, the resulting second degree equation:

$$t^2 \geq \frac{(f-p)^2}{p(1-p)/n} \,, \quad (f-p)^2 \leq t^2 \, \frac{p(1-p)}{n} \,,$$

$$(t^2+n)p^2 - (t^2+2fn)p + nf^2 \leq 0 \,.$$

Since the p^2 term is always positive, the inequality is satisfied for values of p in the range:

$$p \in \frac{(t^2+2fn) \pm \sqrt{t^4 + 4f^2n^2 + 4t^2nf - 4t^2nf^2 - 4n^2f^2}}{2(t^2+n)} \,,$$

from which, in a compact form, one obtains the Wilson formula:

$$p \in \frac{f + \dfrac{t^2}{2n}}{\dfrac{t^2}{n}+1} \pm \frac{t\sqrt{\dfrac{t^2}{4n^2} + \dfrac{f(1-f)}{n}}}{\dfrac{t^2}{n}+1} \,. \tag{6.31}$$

The t parameter indicates any value of the standard variable T; the value $t = 1$ corresponds to one standard deviation. Note that the interval is not centred on the measured frequency f but at the value $(f + t^2/2n)/(t^2/n + 1)$, which is a function

of f and of the number of trials performed. This effect is a consequence of the binomial density asymmetry for small n, as seen in Fig. 2.4. For $n \gg 1$ the interval tends to be centred around the measured frequency and Eq. (6.31), for $CL = 1-\alpha$, becomes:

$$p \in f \pm t_{1-\alpha/2} s = f \pm t_{1-\alpha/2} \sqrt{\frac{f(1-f)}{n}} \,, \tag{6.32}$$

where $|t_{\alpha/2}| = t_{1-\alpha/2}$ are the standard Gaussian quantile values corresponding to the extremes of the interval with an area under the curve of $CL = 1-\alpha$.

Usually the 1σ interval is reported with $t_{1-\alpha/2} = 1$:

$$p \in f \pm s = f \pm \sqrt{\frac{f(1-f)}{n}} \,, \tag{6.33}$$

which is named Wald interval. This interval can also be obtained directly from Eq. (6.30) by replacing in the denominator the true error with the estimated one $\sqrt{f(1-f)/n}$, a technique sometimes called *error plug-in*. By multiplying by the number of trials n, Eq. (6.33) can easily be expressed as a function of the number of successes x. The obtained interval is then related to the expected number of successes μ:

$$\mu \in x \pm \sqrt{x\left(1-\frac{x}{n}\right)} \,. \tag{6.34}$$

This formula, which is used in practice when nf, $n(1-f) > 20, 30$, has been used in the introductory example of Table 6.1. It is easy to remember, because the interval is centred at the measured value, *the variable* $T \sim N(0,1)$ *of Eq. (6.29) is pivotal*, Eq. (6.9) holds, the statistical error is the same as the standard deviation of the binomial distribution (with the probability p replaced by the frequency f) and the confidence levels CL are Gaussian. Finally, we note that the accuracy of Wilson's formula (6.31) can be improved by applying the continuity correction to the frequency $f = x/n$, which generally improves the coverage of the confidence intervals when the variable is discrete:

$$f_{\pm} = \frac{x \pm 0.5}{n} \,. \tag{6.35}$$

In the Gaussian approximation (when $|t_{\alpha/2}| = t_{1-\alpha/2}$), one obtains the following interval estimation:

$$p \in [\max(0, p_-); \min(1, p_+)] \,, \tag{6.36}$$

with

$$p_{\pm} = \frac{f_{\pm} + \dfrac{t^2_{\alpha/2}}{2n}}{\dfrac{t^2_{\alpha/2}}{n} + 1} \pm \frac{|t_{\alpha/2}| \sqrt{\dfrac{t^2_{\alpha/2}}{4n^2} + \dfrac{f_{\pm}(1 - f_{\pm})}{n}}}{\dfrac{t^2_{\alpha/2}}{n} + 1} \,. \tag{6.37}$$

This equation provides a good over-coverage and is, on average, smaller than the interval obtained from Eqs. (6.18, 6.19) [Rot10].

The coverage properties of Eqs. (6.18, 6.19, 6.37), as a function of sample size and confidence levels, will be explored later in the context of simulation techniques, in Sect. 8.11. Its reading is strongly recommended to those interested in probability estimation.

In general, we can state that Eq. (6.37) gives correct results for $np, n(1-p) > 10$, whereas Eq. (6.32) should be used when $np > 100$. Equations (6.31, 6.37) are implemented inside the R routine `prop.test(x,n,alt,conf,corr)`, where `x` and `n` are the successes and the trials, respectively, `alt` is the type of estimate, that is `"two.sided"` (default), `"less"` o `"greater"`, whereas `conf` (default = 0.90) is the confidence level. Finally, `corr` (default = `TRUE`) indicates whether or not the continuity correction is applied. This routine also prints messages related to a hypothesis test with $p = 0.5$ that should be ignored in this context.

Exercise 6.4
During a projection of the election results, 3000 ballots were randomly sampled from the total population of voting cards and examined. The *A* party got 600 votes. Give the final forecast (projection) of the results.

Answer Since $n = 3000$, $f = 600/3000 = 0.20$ and nf, $n(1-f) \gg 10$, Eq. (6.33) can be used, with Gaussian confidence levels. Therefore, one obtains:

$$p \in f \pm \sqrt{\frac{f(1-f)}{n}} = [0.13, 0.27] = (20.0 \pm 0.7)\% \quad CL = 68.3\% \,,$$

$$p \in f \pm 2\sqrt{\frac{f(1-f)}{n}} = [0.186, 0.214] = (20.0 \pm 1.4)\% \quad CL = 95.4\% \,,$$

$$p \in f \pm 3\sqrt{\frac{f(1-f)}{n}} = [0.179, 0.221] = (20.0 \pm 2.1)\% \quad CL = 99.7\% \,.$$

(continued)

Exercise 6.4 (continued)
The R routine `prop.test` can be used with the following lines of code:

```
prop.test(600,3000,conf=0.683),
prop.test(600,3000,conf=0.954),
prop.test(600,3000,conf=0.997),
```

which gives three estimates very close to those previously found with the approximate formula.

Notice the surprising precision obtained even with a limited number of voting cards. Strictly speaking, since the voter population is very large but finite (millions of citizens), a correction should be made to these results, as shown in the next Sect. 6.13, but here it is absolutely negligible. In this type of prediction, the real difficulty lies in obtaining a truly representative sample of the total population. In general, a sample is defined as representative or random when a single individual from any group has a probability of being chosen proportional to the group's size in the total population (see also the Definition 6.4 below). In samples from a natural or physical phenomena, such as those obtained in a physics laboratory, nature itself provides a random sample, if no mistakes or systematic errors are made during the measurements (see also the discussion in Chap. 12). However, the situation is very different in social or biological sciences, where the methods of sampling from a population are so important and difficult to form a special branch of statistics. Those interested in these techniques can consult [Coc77].

Equation (6.32) also allows the determination of the sample size necessary to keep the statistical error below an a priori fixed value. This result is easily reached if we square the statistical error present in the equation and exchange f with the true value p, obtaining the variance:

$$\sigma^2 = \frac{p(1-p)}{n} \,. \tag{6.38}$$

Incidentally, we note that this equation is identical to (3.5), except for the division by the factor n^2, since here we consider the variable $F = X/n$. If we now set to zero the derivative with respect to p, we get:

$$\frac{\mathrm{d}}{\mathrm{d}p}\left[\frac{p(1-p)}{n}\right] = \frac{1}{n}(1-2p) = 0 \quad \Longrightarrow p = \frac{1}{2} \,.$$

Substituting the maximum value $p = 1/2$ into Eq. (6.38), we obtain the upper bound for the true variance, as a function of the number of trials n:

$$\sigma^2_{max} = \frac{1}{4n} \,. \tag{6.39}$$

This formula has the remarkable property to give an upper bound for the variance *regardless of the value of the true probability* p. It is therefore possible to determine a universal formula for the number of trials required to keep the interval estimate $\pm t_{1-\alpha/2}\,\sigma_{max}$ below a certain predetermined value, for a certain confidence level $CL = 1 - \alpha$. Indeed, from Eq. (6.39) one gets:

$$n = \frac{t^2_{1-\alpha/2}}{4\,(t_{1-\alpha/2}\,\sigma_{max})^2} \,. \tag{6.40}$$

Exercise 6.5
Find the number of samples needed to have an absolute interval less than 4 per thousand with a confidence level of 99%.

Answer Since large samples are now considered, we can use Table E.1, that gives a quantile $t_{1-0.005} = 2.57$ for $CL = 99\%$ and a Gaussian tail area of 0.495. The requested interval is $\pm t_{1-\alpha/2}\,\sigma_{max} = \pm 0.002$. By inserting these values in Eq. (6.40) one immediately obtains:

$$n = \frac{(2.57)^2}{4 \cdot (0.002)^2} = 412\,806 \,.$$

Exercise 6.6
In 20 independent Bernoulli trials, 5 events were recorded. What is the probability estimate of the event, with a CL of 90%?

Answer We have now $x = 5, n = 20, f = 5/20 = 0.25$. We are therefore in the case of small samples. If we introduce the data $x = 5, n = 20, CL = 0.90$ in the routine `binom.test(5,20,conf=0.90)` , the following values are obtained:

$$p_1 = 0.104 \,, \quad p_2 = 0.455 \,.$$

(continued)

Exercise 6.6 (continued)
Therefore, according to Eq. (6.17), the interval estimate is:

$$p \in [0.104, 0.455] \equiv 0.25^{+0.20}_{-0.15} \,, \quad CL = 90\% \,.$$

We can also solve the problem in an approximate way, by applying Eq. (6.37) with a value $t = 1.645$, deduced from the usual Table E.1 of the Gaussian probabilities, as an intermediate value between the areas 0.4495 and 0.4505. The approximation consists precisely in this last assumption on the Gaussian levels of t, and not in the use of Eq. (6.37), which is general. From Eq. (6.37) or from the R routine `prop.test(5,20,conf=0.90)`, one obtains the interval:

$$p \in [0.110, 0.458] = 0.25^{+0.21}_{-0.14} \,.$$

In an even more approximate way, we can use Eq. (6.32) with the same t value. The result is:

$$p \in [0.09, 0.410] = 0.25 \pm 0.16 \,.$$

As you can see, the three methods give slightly different results. This fact will be analysed in detail in Sect. 8.11, using simulation techniques.

6.8 Poissonian Interval Estimation

The determination of confidence intervals can also be extended from the binomial to the Poisson case. When a number x of counts is observed, the mean μ can be estimated, in analogy with Eqs. (6.18, 6.19):

$$\sum_{k=x}^{\infty} \frac{\mu_1^k}{k!} \exp(-\mu_1) = c_1 \,, \quad \sum_{k=0}^{x} \frac{\mu_2^k}{k!} \exp(-\mu_2) = c_2 \,, \tag{6.41}$$

where, for $x > 0$, the first equation is equivalent to:

$$1 - \sum_{k=0}^{x-1} \frac{\mu_1^k}{k!} \exp(-\mu_1) = c_1 \,.$$

In the symmetric case, one usually sets $c_1 = c_2 = (1 - CL)/2$. Here too, the over-coverage of the interval, in agreement with Eq. (6.6), is guaranteed by the

presence, in both sums, of the measured value x. For these estimates we can use the R routine `poisson.test(x,conf,alt)`, where `x` is the number of observed events and `conf` and `alt`, as usual, indicate the type of estimate (`"two.sided"` is the default value) and the CL (with 0.95 as default value).

Under the Gaussian approximation, since for the Poissonian $\sigma^2 = \mu$, the confidence interval for the expected value μ can be evaluated from the pivotal quantity:

$$\frac{|x-\mu|}{\sqrt{\mu}} \leq |t_{\alpha/2}| \,, \tag{6.42}$$

which is distributed according to the standard normal p.d.f. Also in this case, as in Eq. (6.37), the values $|t_{\alpha/2}| = t_{1-\alpha/2}$ are the standard Gaussian quantiles at the extremes of the interval with an area under the curve equal to $CL = 1-\alpha$. Introducing, as in Eq. (6.37), the continuity correction:

$$x_{\pm} = x \pm 0.5 \;\text{ if } x \neq 0 \,, \tag{6.43}$$

and solving Eq. (6.42) for μ one obtains:

$$\mu \in x_{\pm} + \frac{t^2_{\alpha/2}}{2} \pm |t_{\alpha/2}| \sqrt{x_{\pm} + \frac{t^2_{\alpha/2}}{4}} \,. \tag{6.44}$$

The knowledge gained from experience with simulated data, which we will discuss later in Sect. 8.11, shows that this interval has excellent over-coverage properties and can be used as an alternative to the correct interval (6.41) for $x > 10$ [Rot10]. When $x > 100$ Eq. (6.44) can be replaced by the asymptotic interval:

$$\mu \in x \pm |t_{\alpha/2}| \sqrt{x} \,, \tag{6.45}$$

that can be obtained directly from Eq. (6.42) with the error plug-in $\sqrt{\mu} \approx \sqrt{x}$. To better understand the practical use of these formulae, we recommend to take a look at Problem (6.12).

The interval of Eq. (6.42) can be obtained with our routine `PoissApp(x, conf, alt)`, where the arguments have the usual meaning. The default value of CL is `conf=0.68`, whereas `alt="two"`.

In line with the scheme of Fig. 6.3, the first of Eq. (6.41), with $c_2 = 1 - CL$, also allows to solve in a general and not approximate way the problem of evaluating, for an assigned CL, the Poisson lower bound when x events have been obtained. The R command, given x observed events, is `poisson.test(x,alt="greater")`. Instead, to find the upper bound with an assigned CL, the second of Eq. (6.41) must be used, with $c_2 = 1 - CL$. The R command is `poisson.test(x,alt="less")`. As in the previous case, if a CL other than 0.95 is required, the command `conf = CL` is necessary. For

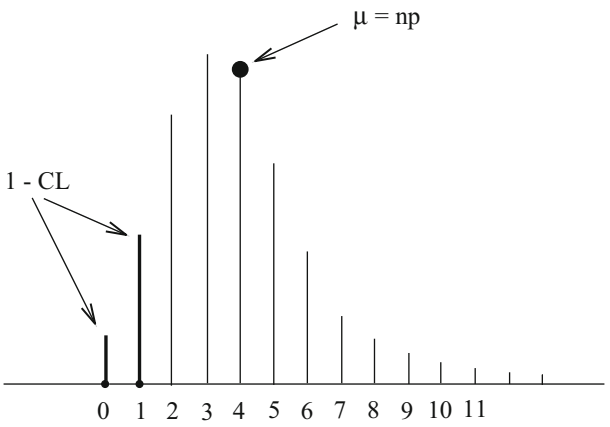


Fig. 6.6 Graphical representation, for $x = 1$, of the second of Eqs. (6.41)

$x = 0, 1, 2$, defining $\mu_2 \equiv \mu$, one has (see also Fig. 6.6):

$$e^{-\mu} = 1 - CL\ ,$$
$$e^{-\mu} + \mu e^{-\mu} = 1 - CL\ ,$$
$$e^{-\mu} + \mu e^{-\mu} + \frac{\mu^2}{2} e^{-\mu} = 1 - CL$$

and so on. Table 6.2 avoids the task of solving the above equations. For example, the table shows that, when $\mu > 2.3$, on average *no events will be observed* in a fraction of experiments <10%. Similarly, if 4.74 is the upper limit for $x = 1$ and $CL = 95\%$, this means that, when $\mu > 4.74$, the values $x = 0, 1$ can be obtained in a fraction of experiments <5%, according to Fig. 6.6.

Instead of Table 6.2, the routine `poisson.test` may be used. For example, when $x = 2$ one has:

`poisson.test(2,conf=0.90,alt="less")` = 5.322,
`poisson.test(2,conf=0.95,alt="great")` = 6.297.

Table 6.2 Poissonian upper limits μ_2 of the mean number of events in correspondence of x observed events, for 90% and 95% confidence levels

x	90%	95%	x	90%	95%
0	2.30	3.00	6	10.53	11.84
1	3.89	4.74	7	11.77	13.15
2	5.32	6.30	8	13.00	14.44
3	6.68	7.75	9	14.21	15.71
4	7.99	9.15	10	15.41	16.96
5	9.27	10.51	11	16.61	18.21

Approximate upper and lower limits can be also determined from Eq. (6.44):

$$\mu_{\mathrm{L}} = x_- + \frac{t_{1-\alpha}^2}{2} - |t_{1-\alpha}|\sqrt{x_- + \frac{t_{1-\alpha}^2}{4}} \tag{6.46}$$

$$\mu_{\mathrm{U}} = x_+ + \frac{t_{\alpha}^2}{2} + |t_{\alpha}|\sqrt{x_+ + \frac{t_{\alpha}^2}{4}}\,, \tag{6.47}$$

where $\alpha = 1 - CL$. The solutions of Eqs. (6.46, 6.47) are calculated by `Poiss.App`, where $x_- = x - 0.25$ and $x_+ = x + 0.5$. The choice x_- has been empirically determined by comparison of the results with those of `poisson.test` when $x < 10$.

Exercise 6.7
In an experiment, 23 counts have been recorded. Find the upper limit at $CL = 0.95$.

Answer From Table E.1, it results that the quantile value corresponding to the tail of area $\alpha = 0.500 - 0.450 = 0.050$ is $|t_{\alpha/2}| = 1.65$. Under the Gaussian approximation, valid for $\mu > 10$, one can write, according to Fig. 6.7:

$$\frac{\mu - 23}{\sqrt{\mu}} = 1.65\,.$$

This corresponds to the second degree equation in $\sqrt{\mu}$:

$$\mu - 1.65\sqrt{\mu} - 23 = 0\,.$$

The positive solution of this equation is $\sqrt{\mu} = 5.69$, therefore the required value is $\mu = 32.37$.

The R routines give the result:

```
poisson.test(conf=0.95,alt="less") = 32.585
```

and, in an approximate way:

```
PoissApp(23,conf=0.95,alt="upp") = 32.410
```

Notice the approximate solution of Eq. (6.45)

$$\mu = 23 + 1.65\sqrt{23} = 30.9\,,$$

which is quite different from the exact result in Gaussian approximation.

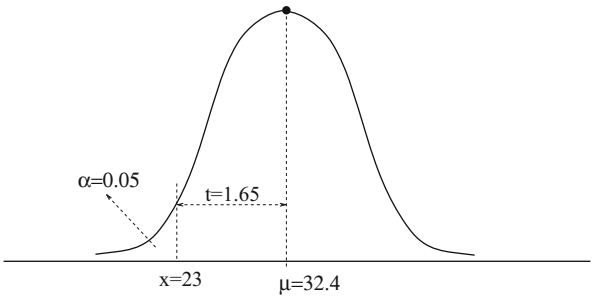


Fig. 6.7 Determination, using the method of Exercise 6.7, of the upper limit μ of a Poissonian process with 23 recorded counts

6.9 Mean Estimation from Large Samples

The estimation of sample mean and variance for any type of random variable is a problem that does not have a general solution. However there are fundamental formulae always valid for large samples, for any variable and when $N > 100$. On the contrary, a general solution exists for Gaussian variables, as we will see shortly in Sect. 6.11. Therefore, the estimation of mean and variance for small non-Gaussian samples ($N < 100$) remains undefined. In this case, if the problem requires great accuracy, Monte Carlo or *bootstrap* simulation techniques are used, as we will see later in the sections dedicated to these topics. Now we deal with the estimation of the mean of large samples in the case of generic variables.

Unlike the true mean μ, which is a fixed quantity, the sample mean is a random variable. In fact, if we produce a random sample of size N from any distribution, calculate the mean:

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i \ , \tag{6.48}$$

and repeat the experiment many times; a different result is obtained for each sample. Therefore, the sample mean is an estimator:

$$M = \frac{1}{N}\sum_{i=1}^{N} X_i \ ,$$

according to Eqs. (2.8, 2.71). Hence, if the operator is applied to the random variable M, from Eqs. (4.19, 5.74) one obtains:

$$\mathrm{Var}[M] = \mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right] = \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{Var}[X_i] \ .$$

Since all the variables X_i belong to the same sample and are independent, $\mathrm{Var}[X_i] = \sigma^2$, where σ^2 is the variance of the parent population. So we get the important and simple result:

$$\mathrm{Var}[M] = \frac{1}{N^2} \sum_{i=1}^{N} \sigma^2 = \frac{1}{N^2} N\sigma^2 = \frac{\sigma^2}{N} \,. \tag{6.49}$$

Therefore, we obtain the 1σ interval:

$$\mu \in m \pm \frac{\sigma}{\sqrt{N}} \simeq m \pm \frac{s}{\sqrt{N}} \,, \tag{6.50}$$

centred on the mean value and of width equal to the statistical error $\sigma/\sqrt{N}$. What is the confidence level of this interval? The Central Limit Theorem 3.1 states that the density of the sample mean is Gaussian for $N \gg 1$, which in practice becomes $N > 10$. In the Gaussian case, the required confidence levels are given, from Eq. (6.9), by the probability levels centred at the mean of the corresponding Gaussian density. The problem is therefore completely solved, at least for large samples.

To explicitly indicate that, for large N, confidence levels are Gaussian, Eq. (6.50) is often cast in the form:

$$P\{X_1 + X_2 + \cdots + X_N \le x\} \simeq \Phi\left(\frac{x - N\mu}{\sigma\sqrt{N}}\right) \,, \tag{6.51}$$

where μ and σ are the mean and the standard deviation of the N variables X_i and Φ is the Gaussian cumulative function (3.43).

Since σ is usually unknown, let us now determine for which values of N it is acceptable to replace it with s, the observed one, in Eq. (6.50). We represent the sample discussed in the previous section for the probability estimation, with a histogram where the values zero and one are assigned to the failure and the success, respectively. Under the approximation $N - 1 \simeq N$, the mean and variance of this sample are obtained from Eqs. (2.53, 2.55):

$$\begin{aligned} m &= \sum_i x_i f_i = 0 \cdot \frac{N-x}{N} + 1 \cdot \frac{x}{N} = \frac{x}{N} \equiv f \,, \\ s^2 &= \sum_i (x_i - m)^2 f_i = (-m)^2 \frac{N-x}{N} + (1-m)^2 \frac{x}{N} \,, \\ &= f^2(1-f) + (1 + f^2 - 2f) f = f(1-f) \,. \end{aligned}$$

This result shows that the histogram mean corresponds to the frequency f and the histogram variance is $f(1-f)$. From Eq. (6.50), one immediately obtains:

$$\mu \equiv p = m \pm \frac{s}{\sqrt{N}} = f \pm \sqrt{\frac{f(1-f)}{N}} \,. \tag{6.52}$$

This result is based on the substitution of the true variance with the measured one and on the approximation for large samples $N \simeq N-1$. Moreover, the frequency given by Eq. (6.33) holds for $N > 100$. The conclusion is that, in Eq. (6.50), $\sigma \simeq s$ is a good approximation for $N > 100$. Therefore, we can write the mean estimate for large sample as:

$$\mu \in m \pm t_{1-\alpha/2} \frac{s}{\sqrt{N}} \,, \quad (N > 100, \text{ Gaussian } CL) \,, \tag{6.53}$$

where $1 - CL = \alpha$ and $t_{1-\alpha/2}$ is the positive Gaussian quantile.

The value m can be calculated with the R routine `mean(x)` for a set of raw data contained in a vector `x` and with our routine `MeanHisto(x,fre)` for a histogram with support `x` and frequencies `fre`. The value s^2 can be obtained, with the same notations, from `var(x)` and `VarHisto(x,fre)`.

The case of small Gaussian samples will be examined in Sect. 6.11.

6.10 Variance Estimation from Large Samples

In statistics we can define two types of variance: one with respect to the true mean, the other with respect to the sample mean:

$$S_\mu^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2 \,, \quad S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - M)^2 \,. \tag{6.54}$$

These two quantities, for $N \to \infty$, tend to the true variance σ^2 in the sense of Eq. (2.77). In general, S^2 denotes the variance with respect to M, which is called sample variance. In the following, we will distinguish the two variances with the notation of Eq. (6.54) only if strictly necessary.

To start, let us account for the $N-1$ factor present in Eq. (6.54). The reason lies in the algebraic relation:

$$\begin{aligned}\sum_i (X_i-\mu)^2 &= \sum_i (X_i - M + M - \mu)^2 = \sum_i [(X_i - M) + (M-\mu)]^2 \\ &= \sum_i (X_i - M)^2 + \sum_i (M-\mu)^2 + 2(M-\mu)\sum_i (X_i - M) \\ &= \sum_i (X_i - M)^2 + N(M-\mu)^2 \,, \end{aligned} \tag{6.55}$$

where in the last row the property $\sum_i (X_i - M) = 0$ has been used. Now take a careful look at Eq. (6.55): it indicates that the dispersion of the data around the true mean is equal to the dispersion of the data around the sample mean plus a term that takes into account the dispersion of the sample mean around the true mean. This sum of fluctuations is just the reason of the term $N-1$. Indeed, by inverting Eq. (6.55), one has:

$$\sum_i (X_i - M)^2 = \sum_i (X_i - \mu)^2 - N(M-\mu)^2 \,. \tag{6.56}$$

We now apply the mean operator to all members of this equation, according to the technique discussed in Sect. 2.11. From Eqs. (2.62, 4.8, 6.49) one obtains:

$$\begin{aligned}\left\langle \sum_i (X_i - M)^2 \right\rangle &= \sum_i \left\langle (X_i - M)^2 \right\rangle = \sum_i \left\langle (X_i - \mu)^2 \right\rangle - N \left\langle (M-\mu)^2 \right\rangle \\ &= \sum_i \sigma^2 - N\,\frac{\sigma^2}{N} = N\sigma^2 - \sigma^2 = (N-1)\,\sigma^2 \,. \end{aligned} \tag{6.57}$$

Basically, this is the justification of the second of Eq. (6.54), because we see that the estimator S^2 is unbiased since it satisfies the property (2.79):

$$\left\langle S^2 \right\rangle = \left\langle \frac{1}{N-1} \sum_i (X_i - M)^2 \right\rangle = \sigma^2 \,. \tag{6.58}$$

The sample variance:

$$\frac{1}{N} \sum_i (X_i - M)^2 \tag{6.59}$$

is an example of a *biased estimator*. Indeed, from Eq. (6.57) it results:

$$\left\langle \frac{1}{N} \sum_{i=1}^{N} (X - M)^2 \right\rangle = \frac{N-1}{N} \sigma^2 \,. \tag{6.60}$$

We recall from Sect. 2.11 that an estimator is biased when the mean of the estimators T_N, calculated from samples of size N, differs from the limit for $N \to \infty$ (2.77) of the estimator. In this case, the mean differs from value of the parameter under estimation, that is, the variance, by a factor $(N-1)/N = 1-1/N$. The term $1/N$ is called *distortion factor* or bias; since it vanishes for $N = \infty$, this type of estimators is defined as *asymptotically correct*, because for large N their mean is very close to the limit in probability. All these aspects will be discussed in detail in Chap. 10.

Equation (6.58) can also be interpreted by recalling the concept of *degrees of freedom*, first introduced in Sect. 3.8 for the χ^2 distribution. If you estimate the mean from the experimental data, the sum of the N squares of the deviations is actually made up of $N-1$ independent terms, because the sample mean establishes a link among the data.

The sample variance is therefore a unique quantity, provided that the sum of the squares of the differences is divided by the degrees of freedom of the statistic. In a rather general way, we then arrive to:

Definition 6.3 (Degrees of Freedom) The number of degrees of freedom ν of a statistic $T = t(X_1, X_2, \ldots, X_N, \hat{\boldsymbol{\theta}})$ which depends on a known set of parameters $\hat{\boldsymbol{\theta}}$ is given by the number N of sample elements minus the number k of parameters $\hat{\boldsymbol{\theta}}$ obtained from the data:

$$\nu = N - k \,. \tag{6.61}$$

Let us now find the variance of the sample variance. This is not a paradox, because the sample variance, like the mean, is a random variable, tending towards the true variance for $N \to \infty$. Similarly to what we did for the mean, we can apply the variance operator to the quantities S_μ^2 and S^2 of Eq. (6.54) and perform the transformation (5.74). The operator formalism, defined in Eqs. (2.60, 2.67) and already applied in Eq. (6.57), greatly simplifies this calculation, at least for S_μ^2. Indeed, we have:

$$\begin{aligned} \mathrm{Var}[S_\mu^2] &= \frac{1}{N^2} \sum_i \mathrm{Var}[(X_i - \mu)^2] = \frac{1}{N^2} \sum_i \left[\left\langle (X_i - \mu)^4 \right\rangle - \left\langle (X_i - \mu)^2 \right\rangle^2 \right] \\ &= \frac{1}{N^2} \left[N \Delta_4 - N \sigma^4 \right] = \frac{1}{N} \left(\Delta_4 - \sigma^4 \right) , \end{aligned}$$

where the fourth-order moment has been indicated with the notation of Eq. (2.59). Substituting the true values Δ_4 and σ^2 by the values estimated from the data, that

is, s_μ^2 and

$$D_4 = \frac{1}{N}\sum_i (X_i - \mu)^4 \,,$$

one obtains the result:

$$\mathrm{Var}[S_\mu^2] \simeq \frac{D_4 - s_\mu^4}{N} \,. \tag{6.62}$$

The 1σ confidence interval for the variance is then given by:

$$\sigma^2 \in s_\mu^2 \pm \sigma[S_\mu^2] = s_\mu^2 \pm \sqrt{\frac{D_4 - s_\mu^4}{N}} \,.$$

A similar calculation can be performed for the variance S^2, but in this case we must also take into account the variance of the sample mean M around the true mean. The final result, quite laborious to obtain, shows that, with respect to Eq. (6.62), corrective terms of order $1/N^2$ appear. Therefore, for large samples, they can be neglected, and Eq. (6.62) can then also be applied now, taking into account that now the degrees of freedom are $(N-1)$ and not N. In the following we will use the notation:

$$\sigma^2 \in \begin{cases} s_\mu^2 \pm \sqrt{\dfrac{D_4 - s_\mu^4}{N}} & \text{known mean,} \\ \\ s^2 \pm \sqrt{\dfrac{D_4 - s^4}{N-1}} & \text{unknown mean .} \end{cases} \tag{6.63}$$

What are the confidence levels of Eq. (6.63)? It can be shown that the sample variance *for any variable* tends to the Gaussian density, but a good approximation is only reached for samples with $N > 100$. If the sample elements are Gaussian, then, as we will see better in the next section, the sampling distribution of the variance is related to the χ^2 density, which converges to the Gaussian density a little faster, roughly for $N > 30$. In general, the random variable "sample variance" tends to a Gaussian distribution much more slowly than the corresponding sample mean. This fact is justifiable if we observe that, in the variance, the combination of the sample variables is quadratic rather than linear.

Once the confidence interval $[s_1^2, s_2^2]$ for the variance is determined, that of the standard deviation can be found by defining $\sigma \in [\sqrt{s_1^2}, \sqrt{s_2^2}]$. The approximate non-linear law (5.56) can be applied, where $x = s^2$ and $z = s = \sqrt{x} = \sqrt{s^2}$. Retaining

only the first term, one has:

$$\mathrm{Var}[S] \simeq \frac{1}{4S^2}\ \mathrm{Var}[S^2] = \frac{D_4 - s^4}{4(N-1)s^2}\ . \tag{6.64}$$

The 1σ confidence interval for the standard deviation is then:

$$\sigma \in s \pm \sqrt{\frac{D_4 - s^4}{4(N-1)s^2}}\ , \quad (N > 100\ ,\ \text{Gaussian}\ CL)\ . \tag{6.65}$$

Our routine `VarEst(x,fre,conf,alt)` estimates the variance and the standard deviation of a data sample with the second of Eqs. (6.63) and (6.65), respectively. As usual, `x` is a row data vector if `fre` is missing, whereas it is the histogram bin value if the frequency vector `fre` is given; `conf` is the CL and `alt` is the estimate type, (`"two"`, `"low"`, `"upp"`). The upper and lower limits for an assigned CL are calculated with the formulae:

$$\sigma_{\mathrm{U}}^2 = s^2 + t_{1-\alpha}\sqrt{\frac{D_4 - s^4}{N-1}}\ ,\ \ \sigma_{\mathrm{U}} = s + t_{1-\alpha}\sqrt{\frac{D_4 - s^4}{4(N-1)s^2}}\ , \tag{6.66}$$

$$\sigma_{\mathrm{L}}^2 = s^2 - t_{1-\alpha}\sqrt{\frac{D_4 - s^4}{N-1}}\ ,\ \ \sigma_{\mathrm{L}} = s - t_{1-\alpha}\sqrt{\frac{D_4 - s^4}{4(N-1)s^2}}\ , \tag{6.67}$$

where $\alpha = 1 - CL$ and t is the Gaussian quantile.

If the sample elements are Gaussian, then relation (3.33), i.e. $\Delta_4 = 3\,\sigma^4$, holds, and Eqs. (6.63, 6.65) become respectively:

$$\sigma^2 \in s^2 \pm \sigma^2\sqrt{\frac{2}{N-1}} \simeq s^2 \pm s^2\sqrt{\frac{2}{N-1}}\ , \tag{6.68}$$

$$\sigma \in s \pm \frac{\sigma}{\sqrt{2(N-1)}} \simeq s \pm \frac{s}{\sqrt{2(N-1)}}\ , \tag{6.69}$$

from which, for $CL = 1 - \alpha$:

$$\frac{s^2}{1 + t_{1-\alpha/2}\sqrt{2/(N-1)}} \le \sigma^2 \le \frac{s^2}{1 - t_{1-\alpha/2}\sqrt{2/(N-1)}}\ , \tag{6.70}$$

$$\frac{s}{1 + t_{1-\alpha/2}\sqrt{1/[2(N-1)]}} \le \sigma \le \frac{s}{1 - t_{1-\alpha/2}\sqrt{1/[2(N-1)]}}\ , \tag{6.71}$$

where $t_{1-\alpha/2}$ are the Gaussian quantiles. The value $N-1$ must be replaced by N when the dispersions are calculated with respect to the true mean.

Equations (6.53, 6.63, 6.65, 6.70, 6.71) solve the problem of estimating mean, variance and standard deviation for large samples. Sometimes, instead of Eqs. (6.70, 6.71), the right sides of Eqs. (6.68, 6.69) are used, where σ is replaced by s in the statistical error formula.

We will shortly give some examples of statistical estimates, after describing the estimates from Gaussian samples.

6.11 Mean and Variance Estimation for Gaussian Samples

The estimation of mean and variance for small samples ($N < 100$) gives intervals and confidence levels that depend on the type of the parent population of the sample. For this reason, at odds with what we have just discussed, it is not possible to obtain formulae of general validity. However, by applying the elements of probability theory so far developed, it is possible to get a simple and complete solution, at least for the most frequent and important case, that of Gaussian samples. The approach is based on two fundamental points. The first point is that, as shown in the Exercise 5.3, the sample mean M is also Gaussian for any N, since it is the sum of N Gaussian variables. Then, due to Eq. (6.49), also the variable:

$$\frac{M-\mu}{\sigma}\sqrt{N}$$

is a standard Gaussian variable for any N, that is, a pivotal quantity. The second point can be summarized in two important theorems:

Theorem 6.1 (Independence of M and S^2) *If $(X_1, X_2, \ldots, X_N)$ is a random sample of size N coming from a Gaussian population $g(x; \mu, \sigma)$, M and S^2 are independent random variables.*

Proof The sample can be considered as an N-dimensional Gaussian vector belonging to the space of Gaussian samples. Therefore, we can consider the subspace formed by the N sample means M of N vectors (samples) belonging to the space: $P(M)\boldsymbol{X} = (M, M, \ldots, M)$. Since this is also a Gaussian vector, we can build the orthogonal subspace with the third of Eq. (4.75):

$$P(M^{\perp})\boldsymbol{X} = [I - P(M)]\boldsymbol{X} = (X_1 - M, X_2 - M, \ldots, X_N - M)\,.$$

Indeed, vectors belonging to these two subspaces have a null scalar product (4.74), as one can easily verify. From Cochran's Theorem 4.5, it results that M and $|P(M^{\perp})\boldsymbol{X}|^2 = \sum_{i=1}^{N}(X_i - M)^2$ are independent. Since $(N-1)S^2 = |P(M^{\perp})\boldsymbol{X}|^2 \sim \chi^2(N-1)$ the theorem is proved. □

Theorem 6.2 (Sample Variance) *If $(X_1, X_2, \ldots, X_N)$ is a random sample coming from a Gaussian population $g(x; \mu, \sigma)$, the variable:*

$$Q_R = \frac{S^2}{\sigma^2} = \frac{1}{N-1} \sum_i \frac{(X_i - M)^2}{\sigma^2} \tag{6.72}$$

follows the reduced χ^2 density (3.72) with $N-1$ degrees of freedom. Therefore, it is a pivotal quantity with respect to σ^2.

Proof From the previous theorem, we deduce that $(N-1)S^2 = |P(M^{\perp})\boldsymbol{X}|^2$; hence, we can apply Cochran's Theorem 4.5 to the vector $(X_i - M)/\sigma$, $(i = 1, 2, \ldots, N)$, so that the theorem is proved.

Notice that if the terms of Eq. (6.56) are rearranged and divided by σ^2, one obtains:

$$\underbrace{(N-1)\frac{S^2}{\sigma^2}}_{N-1} + \underbrace{\frac{(M-\mu)^2}{\sigma^2/N}}_{1} = \underbrace{\sum_i \frac{(X_i - \mu)^2}{\sigma^2}}_{N \text{ degrees of freedom}} ,$$

according to the additivity Theorem 3.4 for the χ^2 variables. □

The confidence interval for the mean can then be determined using the results of Exercise 5.5. Indeed, the variable (5.41), which in this case is:

$$T = \frac{M-\mu}{\sigma}\sqrt{N}\,\frac{1}{\sqrt{Q_R}} = \frac{M-\mu}{\sigma}\sqrt{N}\,\frac{\sigma}{S} = \frac{M-\mu}{S}\sqrt{N}\,, \tag{6.73}$$

turns out to be the ratio between a standard Gaussian variable and the square root of a reduced χ^2 variable. Since these two variables are independent each other, this ratio follows the Student's distribution with $N-1$ degrees of freedom. The Student's quantiles are tabulated in Table E.2 (notice that N appears in Eq. (6.73), but the variable T has $N-1$ degrees of freedom).

Equation (6.73) shows that the Student's variable provides the definition of a pivotal quantity for the mean μ without using the true variance (usually unknown) σ^2 for the determination of the confidence interval. Indeed, if $t_{1-\alpha/2} = -t_{\alpha/2}$ are the T quantile values corresponding to the fixed confidence level $CL = 1-\alpha$ as shown in Fig. 6.8, inverting Eq. (6.73) one obtains:

$$m - t_{1-\alpha/2}\frac{s}{\sqrt{N}} \le \mu \le m + t_{1-\alpha/2}\frac{s}{\sqrt{N}}\,. \tag{6.74}$$

As it is easy to deduce from Table E.2 and Fig. 5.4, for $N > 100$ the Student's density is practically identical to a Gaussian, and Eq. (6.74) coincides with Eq. (6.53). Recall that generally, in the case of a standard Gaussian variable, t is not explicitly indicated or is denoted by z_α.

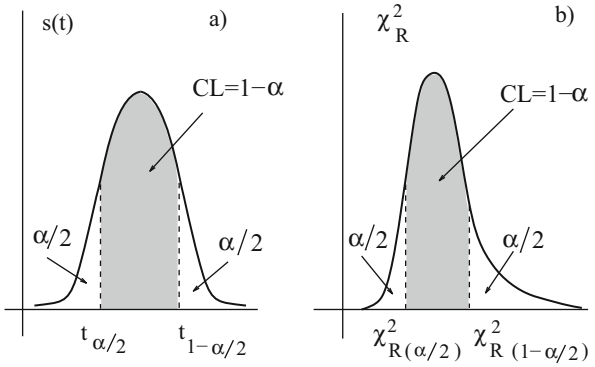


Fig. 6.8 Confidence intervals corresponding to the $t_{\alpha/2} < t_{1-\alpha/2}$ quantile values of the Student's variable (**a**) and of the reduced chi-squared variable (**b**) for the estimates of mean and variance from small Gaussian samples

We postpone the exercises on the use of these formulae to first deal with the estimation of variance. In this case Eq. (6.72) gives the pivotal quantity $S^2/\sigma^2 \sim \chi^2_R(N-1)$. Following the procedure of Eqs. (6.11, 6.12), we start by defining the probability interval:

$$\chi^2_{R\,(\alpha/2)} \leq \frac{S^2}{\sigma^2} \leq \chi^2_{R\,(1-\alpha/2)}\,, \tag{6.75}$$

where $\chi^2_{R(\alpha/2)}$, $\chi^2_{R(1-\alpha/2)}$ are the quantile values of Q_R corresponding to the requested confidence level CL, as shown in Fig. 6.8. Therefore, by inverting the interval (6.75), we obtain the interval for the variance estimation corresponding to the measured value s^2:

$$\frac{s^2}{\chi^2_{R(1-\alpha/2)}} \leq \sigma^2 \leq \frac{s^2}{\chi^2_{R(\alpha/2)}}\,, \quad N-1 \text{ degrees of freedom}\,. \tag{6.76}$$

The probability α is connected to the confidence level CL through Eq. (6.14). The upper and lower limits for $\alpha = 1 - CL$ are given by:

$$\sigma^2_{\mathrm{U}} = \frac{s^2}{\chi^2_{R(\alpha)}}\,, \quad \sigma^2_{\mathrm{L}} = \frac{s^2}{\chi^2_{R(1-\alpha)}}\,, \tag{6.77}$$

and the confidence interval for the standard deviation is simply given by:

$$\frac{s}{\sqrt{\chi^2_{R(1-\alpha/2)}}} \leq \sigma \leq \frac{s}{\sqrt{\chi^2_{R(\alpha/2)}}}\,. \tag{6.78}$$

For values $N > 30$, the χ^2 density is close to the symmetric Gaussian shape, centred around the average $\langle Q_R \rangle = 1$ and with standard deviation:

$$\sigma_v \equiv \sigma\,[Q_R] = \sqrt{\frac{2}{N-1}}\,,$$

in agreement with Eq. (3.71). In this case the quantile values can be written as:

$$\chi^2_{R(\alpha/2)} = 1 - t_{1-\alpha/2}\,\sigma_v\,, \quad \chi^2_{R(1-\alpha/2)} = 1 + t_{1-\alpha/2}\,\sigma_v\,,$$

and the width of the confidence interval with $CL = 1 - \alpha$ becomes:

$$\frac{s^2}{1 + t_{1-\alpha/2}\,\sigma_v} \leq \sigma^2 \leq \frac{s^2}{1 - t_{1-\alpha/2}\,\sigma_v}\,, \tag{6.79}$$

which is again the interval (6.70).

Our routine `MeanEst(x,fre,conf,alt)`, where the argument have been already described before, estimates the mean with Eq. (6.74) using the Student's quantiles. They become practically identical to the Gaussian ones of Eq. (6.53) for ($N > 100$). Our routine `VarEst(x,fre,conf,alt)`, already described above, estimates the variance using Eqs. (6.76, 6.77).

These routines allow a useful comparison between the formulae for large samples and those for Gaussian samples

6.12 How to Use the Estimation Theory

In the previous sections, we have deduced the fundamental formulae for the parameter estimation. They are those commonly used in the analysis of the data that are usually collected in many different scientific fields, from physics to engineering and biology. The overall picture is summarized in Table 6.3, which shows that the formulae derived above solve the estimation problem in a simple and general way for the case of large samples. It is also evident that the estimation of the mean and of the dispersion for small non-Gaussian samples remains not well defined. However, this is a case that occurs quite rarely in practice and for which it is not possible to give a general solution, because both the intervals and the confidence levels depend on the specific distribution involved in the problem. As we have already mentioned, in these cases simulation techniques are often used with success.

From Table 6.3 it also results that, for large samples, all the variances of the estimators for frequency, mean, variance, etc. are of the form σ_0^2/N, where σ_0 is a constant. Therefore, the Kolmogorov condition (2.76), sufficient for the almost certain convergence of all these estimates towards the true values, is fully satisfied. Weak convergence is therefore also verified, as can be directly seen from the

Table 6.3 1σ confidence intervals and corresponding distributions for the determination of the confidence levels (CL) for parameter estimation. Samples with $N > 100$ are usually considered as *large samples*. The symbol ? indicates the lack of a general solution

	Gaussian variables		Any variables	
	Interval	CL	Interval	CL
Probability				
$Nf < 10$	–	–	Equations (6.18, 6.19)	Binomial
$Nf > 10$ $N(1-f) > 10$	–	–	Equation (6.37)	Gauss
Frequency				
$x < 10$	–	–	Equation (6.41)	Poisson
$x > 10$	–	–	Equation (6.42)	Gauss
Mean				
$N < 100$	$m \pm \frac{s}{\sqrt{N}}$	Student	$\simeq m \pm \frac{s}{\sqrt{N}}$	?
$N > 100$	$m \pm \frac{s}{\sqrt{N}}$	Gauss	$m \pm \frac{s}{\sqrt{N}}$	Gauss
Variance				
$N < 100$	$\frac{s^2}{\chi^2_{R1}} \leq \sigma^2 \leq \frac{s^2}{\chi^2_{R2}}$	χ^2_R	$\simeq s^2 \pm \sqrt{\frac{D_4 - s^4}{N-1}}$	?
$N > 100$	Equation (6.70)	Gauss	$s^2 \pm \sqrt{\frac{D_4 - s^4}{N-1}}$	Gauss
Std. Dev.				
$N < 100$	$\sqrt{\frac{s^2}{\chi^2_{R1}}} \leq \sigma \leq \sqrt{\frac{s^2}{\chi^2_{R2}}}$	–	$\simeq s \pm \sqrt{\frac{D_4 - s^4}{4s^2(N-1)}}$	?
$N > 100$	Equation (6.71)	Gauss	$s \pm \sqrt{\frac{D_4 - s^4}{4s^2(N-1)}}$	Gauss

Tchebychev inequality. Indeed, Eq. (3.93) can be written also in the form:

$$P\{|X-\mu| \geq K\,\sigma\} \leq \frac{1}{K^2}\,, \quad P\{|X-\mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}\,,$$

where $K = \epsilon/\sigma$. If $T \equiv T_N$ is an estimator given in Table 6.3, $\mathrm{Var}[T_N] = \sigma_0^2/N$ and we obtain:

$$P\{|T_N - \mu| \geq \epsilon\} \leq \frac{\sigma_0^2}{N\epsilon^2}\,.$$

Therefore, Eq. (2.73) is verified for $N \to \infty$.

Exercise 6.8
A Gaussian sample with N elements has mean $m = 10$ and standard deviation $s = 5$. Estimate, with a confidence level of 95%, mean, variance and standard deviation for $N = 10$ and $N = 100$.

Answer Since the sample is Gaussian, we can use Eqs. (6.74, 6.76, 6.78), which require the determination of the quantiles of the Student's and χ^2 distributions for $\alpha = 0.025$ and 0.975. These distributions are tabulated in Appendix E.

Since the Student's density is symmetric, from Table E.2 it results:

$$-t_{0.025} = t_{0.975} = \begin{cases} 2.26 \text{ for } N = 10 \quad (9 \text{ degrees of freedom}) \\ 1.98 \text{ for } N = 100 \ (99 \text{ degrees of freedom}) \end{cases}$$

From Eq. (6.74) we then obtain the mean estimate:

$$\mu \in 10 \pm 2.26 \frac{5}{\sqrt{10}} = 10.0 \pm 3.6 \quad (N = 10,\ CL = 95\%)\ ,$$
$$\mu \in 10 \pm 1.98 \frac{5}{\sqrt{100}} = 10.0 \pm 1.0 \quad (N = 100,\ CL = 95\%)\ . \quad (6.80)$$

For the dispersion, we must use the quantiles of the χ^2_R density. From Table E.3 one obtains:

$$\chi^2_{R\,0.025},\ \chi^2_{R\,0.975} = \begin{cases} 0.30\ ,\ 2.11 \text{ for } N = 10 \quad (9 \text{ degrees of freedom}) \\ 0.74\ ,\ 1.29 \text{ for } N = 100 \ (99 \text{ degrees of freedom}). \end{cases}$$

From Eq. (6.76) and its square root, the estimates of the variance and of the standard deviation are obtained:

$$\begin{aligned} \sigma^2 &\in \left[\tfrac{25}{2.11}\ ,\ \tfrac{25}{0.3}\right] = [11.9\ ,\ 83.3] \ (N = 10, CL = 95\%) \\ \sigma^2 &\in \left[\tfrac{25}{1.29}\ ,\ \tfrac{25}{0.74}\right] = [19.3\ ,\ 33.8] \ (N = 100, CL = 95\%) \end{aligned} \quad (6.81)$$

$$\begin{aligned} \sigma &\in [\sqrt{11.9}\ ,\ \sqrt{83.3}] = [3.4\ ,\ 9.1] \ (N = 10, CL = 95\%) \\ \sigma &\in [\sqrt{19.3}\ ,\ \sqrt{33.8}] = [4.4\ ,\ 5.9] \ (N = 100, CL = 95\%)\ . \end{aligned} \quad (6.82)$$

(continued)

Exercise 6.8 (continued)
As a useful comparison, we apply the approximate formulae, assuming Gaussian confidence levels. From Exercise 3.8, or directly from Table E.1, we obtain 1.96 when $CL = 0.95$.

For the mean estimate we can use Eq. (6.53):

$$\mu \in 10 \pm 1.96 \frac{5}{\sqrt{10}} = 10.0 \pm 3.1 \quad (N = 10) ,$$

$$\mu \in 10 \pm 1.96 \frac{5}{\sqrt{100}} = 10.0 \pm 1.0 \quad (N = 100) ,$$

which gives (within rounding) a result identical to Eq. (6.80) for $N = 100$, and a slightly underestimated result for $N = 10$. In fact, as can be seen from Fig. 5.4, the Student's density tails subtend areas slightly larger than those subtended by the standard Gaussian.

To roughly estimate the dispersion parameters with $CL = 95\%$, we use Eqs. (6.70, 6.71) again with $t = 1.96$:

$$\sigma^2 \in \left[\frac{25}{1 + 1.96\sqrt{2/9}}, \frac{25}{1 - 1.96\sqrt{2/9}} \right] = [13.0\ ,\ 328.7] \quad (N = 10) ,$$

$$\sigma^2 \in \left[\frac{25}{1 + 1.96\sqrt{2/99}}, \frac{25}{1 - 1.96\sqrt{2/99}} \right] = [19.5\ ,\ 34.8] \quad (N = 100) ,$$

$$\sigma \in [3.6, 18.1] \quad (N = 10) ,$$

$$\sigma \in [4.4, 5.9] \quad (N = 100) .$$

If we compare these results with the correct ones of Eqs. (6.81, 6.82), we notice that the approximate dispersions are only acceptable for $N = 100$. It is also possible to use the right sides of Eqs. (6.68, 6.69) with $s \simeq \sigma$, obtaining the intervals $\sigma^2 \in [1.9, 47.1]$, $\sigma \in [2.7, 7.3]$ for $N = 10$ and $\sigma^2 \in [18.1, 31.9]$, $\sigma \in [4.3, 5.7]$ for $N = 100$. As you can see, for large samples, Eqs. (6.68, 6.69) can also be used. The problem can also be solved with our routines `MeanEst` and `VarEst`.

Exercise 6.9
The analysis of a sample of 1000 electrical resistances (resistors) of 1000 Ω has shown that the values are approximately distributed according to a Gaussian with standard deviation $s = 10\,\Omega$ (actually the production processes of dough resistors well verify the conditions of the Central Limit Theorem 3.1). To keep this production standard constant in time, a quality control was planned by periodically measuring a sample of five resistors with a highly accurate multimeter. Define the statistical limits of the quality control at a 95% confidence level.

Answer We have to assume the nominal value of the resistors as the true average value of production: $\mu = 1000\,\Omega$.

The true dispersion of the electrical resistance values around the mean can be estimated from the data obtained from the sample of 1000 resistors by applying Eq. (6.69) with $s \simeq \sigma$:

$$\sigma \in s \pm \frac{s}{\sqrt{2(N-1)}} = 10.0 \pm 0.2\,,$$

which shows that the sample of 1000 resistors gives an estimate of the dispersion with a relative uncertainty of 2 %. Therefore, we can assume the value $s = 10\,\Omega$ as the true value of the standard deviation.

Since the observed sample is Gaussian, from Table E.1 we can say that the interval $1000 \pm 1.96\,\sigma \simeq 1000 \pm 20\,\Omega$ contains 95% of all values. Basically, only 5 resistors over 100 will fall outside the interval:

$$980\,\Omega \leq R \leq 1020\,\Omega\,. \tag{6.83}$$

The problem is now to establish controls on the produced resistors to verify that these initial conditions remain reasonably constant.

By randomly selecting five resistors, we can set up an adequate quality control using the sample mean. In fact, assuming as true values:

$$\mu = 1000\,\Omega\,, \qquad \sigma \simeq s = 10\,\Omega\,,$$

from Eq. (6.50), we obtain that the sample mean of five elements will be contained in the interval:

$$\mu \pm \frac{\sigma}{\sqrt{5}} = 1000.0 \pm 4.5\,\Omega\,, \tag{6.84}$$

(continued)

Exercise 6.9 (continued)
with nearly Gaussian probability levels. Student's distribution should not be used here, because the true standard deviation value is assumed to be known from the 1000 resistor measurement. On the contrary, we should deduce it from the small five-resistor sample; Eq. (6.84) would still hold, but in this case, instead of σ, we would have to use the the standard deviation s of the five resistors, and the confidence levels would follow the Student's distribution with 4 degrees of freedom, since $N < 10$.

For a first quality control, Eq. (6.84) can be used, with a 95% confidence level, which, from Table E.1, is associated to an interval of about $\pm 1.96\,\sigma$. In this case, the probability of error by judging as poor a good resistor is 5%. A first quality check will then indicate a possible bad production when the sample mean is outside the range:

$$1000 \pm 1.96 \cdot \frac{10}{\sqrt{5}} \simeq (1000 \pm 9)\ \Omega\ ,$$

that is:

$$991\ \Omega \leq m(R) \leq 1009\ \Omega \quad \text{(first quality control, } CL = 95\%)\ . \tag{6.85}$$

The global quality check can be further refined by also verifying that the sample standard deviation does not exceed the value of $10\,\Omega$. Indeed, by inverting Eq. (6.78) and taking its maximum, one has:

$$s \leq \sqrt{\chi^2_{R\alpha}}\,\sigma\ , \tag{6.86}$$

where $\sigma = 10\,\Omega$ and $\chi^2_{R\alpha}$ is the value of $Q_R(4)$, that is, the reduced χ^2 variable with 4 degrees of freedom, corresponding to the required confidence level. For $CL = 95\%$, from Table E.3 one gets:

$$\chi^2_{R\,0.95} = 2.37\ ,$$

and hence:

$$s \leq 10 \cdot \sqrt{2.37} \simeq 15\ \Omega\ .$$

The second quality check will then report one possible bad production when the standard deviation of the sample with five resistors exceeds the limit of $15\,\Omega$:

$$s(R) \leq 15\ \Omega \qquad \text{(second quality control, } CL = 95\%)\ .$$

(continued)

Exercise 6.9 (continued)
If these two quality controls are required to be satisfied at the same time, it is ensured that both the average and the initial dispersion of the electrical resistances are correctly kept within the arbitrarily chosen confidence level.

With a single common $CL = 0.95$, we will have at least one of these limits exceeded with probability $1-0.95^2 = 0.0975$, that is, in about 10% of checks.

A signal outside the confidence band, but within limits of the expected statistic, it is called *false alarm*. The situation is often summarized graphically in the *quality control chart*, in which a zone of normality (or control zone) is chosen; above or below these limits there are two alarm bands and outside of these the forbidden zone. Figure 6.9 shows a possible control chart for the resistor mean value of our problem. The control zone corresponds to the interval (6.85) of width $\pm 1.96\,\sigma$; the alarm zone is from 1.96 to $3.0\,\sigma$, while the forbidden zone is outside $\pm 3.0\,\sigma$. A value in the forbidden zone can occur under normal conditions only 3 times out of 1 000 controls (3σ law), an event that can justify the production suspension and the activation of the machine maintenance processes (warning: this is a subjective decision that can vary from case to case).

As an exercise, with Eq. (6.86) you can also draw a similar chart (S chart) also for the dispersion of the data.

In the alarm zone, on average, we should have 5 values for every hundred checks, corresponding to a priori probability $\gamma = 0.05$. The quality control can then be further refined by detecting if an excessive number of false alarms occur, that is, if there are too many alarms compared to the number of alarms expected when the production quality remains stable. If n is the number of false alarms in N checks, Eq. (6.29) can be applied, since here we assume to know the true probability γ:

$$t_n = \frac{n - \gamma N}{\sqrt{N\gamma(1-\gamma)}} = 4.6\,\frac{n - 0.05\,N}{\sqrt{N}}\,. \tag{6.87}$$

Therefore, the production should be suspended, on the basis of 3σ law, when $t_n > 3.0$. For a Gaussian variable, this value corresponds to a probability of about $\simeq 1.5$ per thousand to wrongly stop a good production. The Gaussian approximation is valid for $n > 10$, which corresponds to $N\gamma > 10$, that is, $N > 200$ in Eq. (6.87). For $N = 200$, it turns out that it is reasonable to proceed with maintenance if $n \geq 19$.

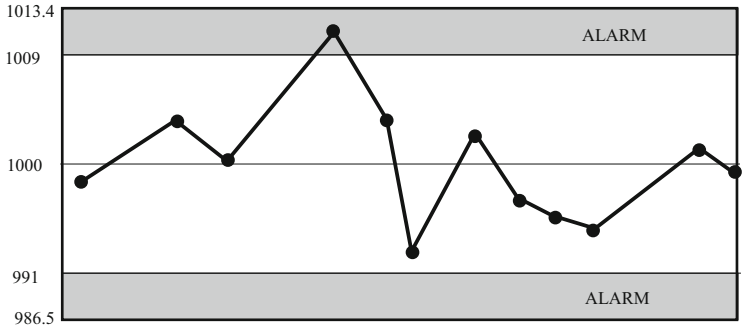


Fig. 6.9 Control chart for the resistor production, as discussed in Exercise 6.9

6.13 Estimates from a Finite Population

In the estimates described so far, we have assumed that the population was made up of an infinite set of elements. The results obtained are also valid for finite populations, provided that, after each draw, the extracted element is replaced into the population (sampling with replacement). However, it is intuitive that there are some changes to be made in the case of sampling without replacement from a finite population, since, if the population were used up, the quantities of interest would become certain and would no longer be statistical estimates. We then begin with the

Definition 6.4 (Random Sample from a Finite Population) A sample S of N elements drawn from a finite population of N_p units is said to be random if it represents one of the $N_p!/[N!(N_p - N)!]$ possible sets, each of which has an equal chance of being chosen.

If X is the random variable contained in the sample S, we can write:

$$M = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}\sum_{i=1}^{N_p} x_i I_i \ , \tag{6.88}$$

where the second sum is over all the N_p units of the population and I_i a dummy variable (see Eq. 2.7), defined as:[1]

$$I_i = \begin{cases} 1 \text{ if } x_i \in S \ , \\ 0 \text{ otherwise.} \end{cases}$$

[1] In order not to overload the notation, we write $x_i \in S$ to indicate that the i-th population unit has been extracted.

I_i is a two-valued binomial variable corresponding to the number of possible successes of a single trial having a probability N/N_p. Therefore, we have:

$$\langle I_i \rangle = \frac{N}{N_p} , \qquad \text{Var}[I_i] = \frac{N}{N_p}\left(1 - \frac{N}{N_p}\right) , \quad \forall i . \tag{6.89}$$

The obtained result:

$$\langle M \rangle = \frac{1}{N}\sum_{i=1}^{N_p} x_i \langle I_i \rangle = \frac{1}{N_p}\sum_{i=1}^{N_p} x_i = \langle X \rangle , \tag{6.90}$$

shows that *also in this case the sample mean is a correct estimator of the true mean*.

To find the variance of the sum of the sampled variables, we can write, using Eq. (5.65):

$$\text{Var}\left[\sum_{i=1}^{N_p} x_i I_i\right] = \sum_{i=1}^{N_p} x_i^2 \,\text{Var}[I_i] + 2\sum_i \sum_{j<i} x_i x_j \,\text{Cov}[I_i, I_j] . \tag{6.91}$$

The covariance estimation (I_i, I_j) requires the knowledge of the mean $\langle I_i I_j \rangle$. Having in mind the general definitions of Sect. 2.8 and only considering the non-zero values, one has:

$$\langle I_i I_j \rangle = P\{x_j \in S | x_i \in S\} P\{x_i \in S\} = \frac{N-1}{N_p - 1}\frac{N}{N_p} .$$

Using Eq. (4.25), we can write the covariance as:

$$\text{Cov}[I_i, I_j] = \langle I_i I_j \rangle - \langle I_i \rangle \langle I_j \rangle = \frac{N}{N_p}\frac{N-1}{N_p - 1} - \frac{N^2}{N_p^2} = -\frac{N(N_p - N)}{N_p^2 (N_p - 1)} .$$

Inserting this result into Eq. (6.91), one has:

$$\text{Var}\left[\sum_{i=1}^{N_p} x_i I_i\right] = \frac{N}{N_p}\frac{N_p - N}{N_p}\sum_{i=1}^{N_p} x_i^2 - 2\frac{N(N_p - N)}{N_p^2 (N_p - 1)}\sum_i \sum_{j<i} x_i x_j . \tag{6.92}$$

Since $(\sum_i x_i)^2 = \sum_i x_i^2 + 2\sum_i \sum_{j<i} x_i x_j$, after some easy algebra one gets:

$$\text{Var}\left[\sum_{i=1}^{N_p} x_i I_i\right] = \frac{N(N_p - N)}{N_p - 1}\left(\langle X^2 \rangle - \langle X \rangle^2\right) , \tag{6.93}$$

where $\langle X^2 \rangle = \sum X_i^2/N_p$ and $\langle X \rangle^2 = (\sum X_i)^2/N_p^2$. Therefore, the variance of the sample mean results:

$$\mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N_p} x_i I_i\right] = \mathrm{Var}[M] = \frac{N_p - N}{N(N_p-1)}\left(\langle X^2\rangle - \langle X\rangle^2\right)$$
$$= \frac{\mathrm{Var}[X]}{N}\,\frac{N_p - N}{N_p - 1} \simeq \frac{\mathrm{Var}[X]}{N}\left(1 - \frac{N}{N_p}\right). \quad (6.94)$$

This equation represents the fundamental result for the estimates from finite populations: the comparison with the analogous formulae for infinite populations (see Table 6.3) shows that the variances of means, frequencies and proportions calculated from samples extracted from finite populations must be corrected with the factor $(N_p - N)/(N_p - 1) \simeq (1 - N/N_p)$. The same factor must be applied if in Eq. (6.94) the true variance σ^2 is replaced by the estimated one s^2. For example, in the case of the Exercise 6.4, considering 30 millions of voters, the correction would be very small and of the order of $\sqrt{1 - 3/30\,000}$. This situation is different from the extraction without replacement from an urn: if the frequency of marbles of a certain type were, for example, $f = 15/30 = 0.50$ and the urn contained 100 marbles, the frequency error would go from $\sqrt{0.5(1-0.5)/30} = 0.09$ (infinite population) to the value $\sqrt{0.5(1-0.5)30}\sqrt{70/99} = 0.08$. The variance would vanish if all 100 marbles were drawn.

The sample variance, unlike the mean, must be corrected for finite populations. In fact, from Eqs. (6.57, 6.94) one has:

$$\left\langle S^2 \right\rangle = \left\langle \frac{1}{N-1}\sum_i (X_i - M)^2 \right\rangle = \frac{1}{N-1}\left[N\sigma^2 - \frac{N_p - N}{N_p - 1}\sigma^2\right] = \sigma^2 \frac{N_p}{N_p - 1}\,. \quad (6.95)$$

It turns out then that the unbiased estimator for the variance of a finite population is:

$$S^2 = \frac{N_p - 1}{N_p}\,\frac{1}{N-1}\sum_i (X_i - M)^2\,. \quad (6.96)$$

The derivation of the correction for the variance $\mathrm{Var}[S^2]$ is more complicated and can be found in [KS73].

6.14 Histogram Analysis

After the analysis of sample mean and variance, we now move to the analysis of the overall sample structure (shape), with the aim of obtaining information on its parent population. Indeed, sample mean and variance are not the only random variables of interest. Usually the sample is presented in the form of a histogram, subdivided into K bins, almost always with fixed width Δx, and each containing a number n_i of events. The quantities n_i give the overall shape of the sample and are of crucial importance if one is interested in studying the density structure of the parent population. *These quantities should be considered as random variables* I_i, because they vary from sampling to sampling. If the histogram is normalized, instead of n_i, the measured frequencies or probabilities $f_i = n_i/N$ are given in each bin, where N is the total number of events in the sample. As an example, in Fig. 6.10 and Eq. (6.97) a non-normalized histogram is shown as obtained from a computer-simulated Gaussian sample of $N = 1000$ events coming from a parent population of true parameters $\mu = 70$, $\sigma = 10$. In Eq. (6.97) x_i and $n_i \equiv n(x_i)$ are the midpoint and the content of a bin, respectively. The bin width is $\Delta x = 5$.

x	$n(x)$	x	$n(x)$
37.5	1	72.5	207
42.5	4	77.5	153
47.5	16	82.5	101
52.5	44	87.5	42
57.5	81	92.5	7
62.5	152	97.5	6
67.5	186		

(6.97)

The R routine `hist(x)` draws the histogram of a raw data set contained in the vector `x`. Without any user input, the bin width and the graphic style of the histogram are automatically set by the routine. If you have a vector `x` containing the abscissas of the bins and a vector `fre` containing the frequencies or the number of events of each bin, you can use our `HistoBar(x,fre)` routine, which draws the histogram as in Fig. 6.10 (top).

If $p(x)$ is the p.d.f. of the population, by defining the random variables I and F related to the events $\{I_i = n_i\}$ and $\{F_i = n_i/N\}$, it follows, from Eq. (2.33), that:

$$\langle I_i \rangle = \mu_i = N p_i = N \int_{\Delta x} p(x)\ \mathrm{d}x \simeq N p(x_0)\ \Delta x\ , \tag{6.98}$$

$$\langle F_i \rangle = p_i = \int_{\Delta x} p(x)\ \mathrm{d}x \simeq p(x_0)\ \Delta x\ , \tag{6.99}$$

where Δx is the bin width and x_0 is a generic point in the bin. The rightmost term in the equations follows from the integral mean value theorem. If the bin width is small

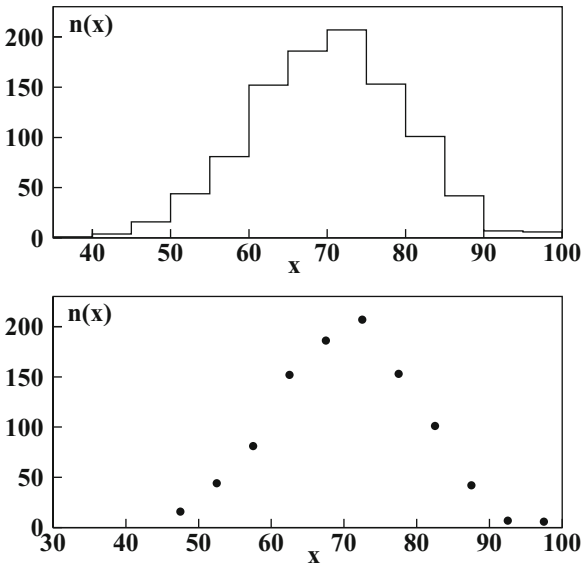


Fig. 6.10 Two possible representations of a histogram obtained with a computer simulation of 1000 events from a Gaussian population with $\mu = 70$ and $\sigma = 10$

enough and the density is a fairly smooth function, one can assume that it varies linearly within the bin width; under these conditions, according to the trapezoidal rule, x_0 is the bin midpoint.

In the case of discrete random variables, the integral over Δx in Eqs. (6.98 6.99) must be replaced by the sum of the true probabilities of the values contained in Δx.

In Sect. 4.7, we have seen that the global probability of having a specific experimental histogram of a random sample of size N, given the true probabilities p_i, $(i = 1, 2, \ldots, k)$ obtained from a p.d.f. $p(x)$ using Eq. (6.99), follows the multinomial distribution (4.89). We already noted, commenting on Eq. (4.89), that the number of I_i events falling in the i-th bin $(x_i, \ , x_{i+1})$ of width Δx follows the binomial distribution. Indeed:

- If the random process is stationary in time, the probability p_i to fall in the i-th bin remains constant.
- The probability to fall in a bin does not depend on the events previously recorded or that will be recorded in other bins.

Therefore, we can state that the random variable I_i (*number of event in a histogram bin*) is given by the occurrence of independent events with a constant probability. If the total number of events N is a predetermined parameter, then the probability for the random variable I_i to take the value n_i will be given by the binomial law (2.29) with elementary probability p_i (see also Eq. (4.89)):

$$P\{I_i = n_i\} = b(n_i; N, p_i) = \frac{N!}{n_i!(N - n_i)!} p_i^{n_i} (1 - p_i)^{N - n_i} , \qquad (6.100)$$

where $(1 - p_i)$ is the probability to fall into any histogram bin different from the i-th one. The standard deviation is:

$$\sigma_i = \sqrt{N p_i \, (1 - p_i)} \, . \tag{6.101}$$

This quantity can be estimated from the data through the uncertainty $s_i \equiv s(n_i)$. If the bin contains more than 20–30 events, Eq. (6.34) can be used, with $x \equiv n_i$ and $n \equiv N$:

$$s_i = \sqrt{n_i \left(1 - \frac{n_i}{N}\right)} \, . \tag{6.102}$$

This approximation is often used even for bin contents above five events.

If the histogram is normalized, Eq. (6.102) must be divided by N, thus obtaining a well-known result, the statistical error (6.33) on the frequency $f_i = n_i / N$:

$$s\left(\frac{n_i}{N}\right) = \sqrt{\frac{n_i}{N^2}\left(1 - \frac{n_i}{N}\right)} = \sqrt{\frac{f_i(1 - f_i)}{N}} \, . \tag{6.103}$$

The two previous formulae are of fundamental importance in the analysis of histograms and are called *random (or statistical) fluctuations of the bin contents*.

If the histogram is not obtained with a fixed total number N of events but is collected considering other parameters, for example, a certain time interval Δt, the number N turns from a constant into a statistical Poissonian variable $\boldsymbol{N}$, and the fluctuations of the bin contents must be calculated in a different way.

We want to explain this rather subtle point with an example. If we look at multiple histograms, each of which refers to the weight of 100 newborns, the fluctuations in the number of babies within a certain weight range (or percentile, as doctors say) will obey Eqs. (6.102, 6.103). If, on the other hand, we collect the histograms of the newborn weights monthly, the fluctuations in the number of babies within a certain weight range will overlap to those of the total number $\boldsymbol{N}$ of babies in a month, which will be Poissonian with a stable average value (if we assume, to simplify, that the births are stable from month to month). To treat this case correctly, it is essential the following:

Theorem 6.3 (On the Binomial and Poissonian Variables) *Let X be the number of successes in N trials. When N is not a fixed parameter but a Poissonian random variable, X follows the Poisson density.*

Proof From the compound probability law, the probability to observe $\{I_i = n_i\}$ events into the i-th bin over a total of N events will be given by the product of the Poissonian probability (3.14) to observe a total of $\{\boldsymbol{N} = N\}$ events, when the mean is λ, times the binomial probability (2.29) to get n_i events in the considered bin,

over a total of N, when the true probability is p_i:

$$P\{I_i = n_i, \boldsymbol{N} = N\} = P\{I_i = n_i | \boldsymbol{N} = N\} P\{\boldsymbol{N} = N\}$$
$$= p(n_i, N) = \frac{N!}{n_i!(N - n_i)!} p_i^{n_i} (1 - p_i)^{N - n_i} \frac{e^{-\lambda} \lambda^N}{N!} .$$

If one now defines $m_i = N - n_i$ and uses the identities:

$$e^{-\lambda} = e^{-\lambda p_i} e^{-\lambda(1-p_i)} , \quad \lambda^N = \lambda^{N-n_i} \lambda^{n_i} = \lambda^{m_i} \lambda^{n_i} ,$$

the probabilities can be written in the form:

$$p(n_i, m_i) = \frac{e^{-\lambda p_i} (\lambda p_i)^{n_i}}{n_i!} \frac{e^{-\lambda (1-p_i)} [\lambda (1 - p_i)]^{m_i}}{m_i!} ,$$

which is the product of two Poissonians, of means λp_i and $\lambda(1 - p_i)$, respectively. From this equation and from Theorem 4.1, one can deduce that the number I_i of events in the i-th channel and the number $(\boldsymbol{N} - I_i)$ of the events contained in the other bins *are both independent Poissonian variables.* In other words, if $\boldsymbol{N}$ is a Poissonian variable and I_i, for fixed $\{\boldsymbol{N} = N\}$, is a binomial variable, then I_i and $(\boldsymbol{N} - I_i)$ are independent Poissonian variables. □

Since we know that the standard deviation of the Poissonian is equal to the square root of the mean, we can immediately change Eq. (6.102) into the form:

$$\sigma[I_i] \equiv \sigma_i = \sqrt{\lambda p_i} \simeq s(n_i) = \sqrt{n_i} ,$$

where the true values have been replaced by the measured ones. The statistical uncertainty of the bin content is then given by:

$$s(n_i) = \sqrt{n_i} , \quad s\left(\frac{n_i}{N}\right) = \frac{1}{N} \sqrt{n_i} = \sqrt{\frac{f_i}{N}} . \tag{6.104}$$

Let us now summarize these results in a coherent scheme. The estimate of the true number of events μ_i (mathematical hope or expected value) in the i-th bin is given by Eq. (6.98), and the corresponding approximate 68.3% level confidence interval is given by:

$$\mu_i \in n_i \pm \sqrt{n_i \left(1 - \frac{n_i}{N}\right)} , \tag{6.105}$$

for histograms with a fixed total number of events N. For histogram where $\boldsymbol{N}$ is a Poissonian variable, one has instead:

$$\mu_i \in n_i \pm \sqrt{n_i} . \tag{6.106}$$

For normalized histograms, Eqs. (6.105 and 6.106) transform respectively as:

$$p_i \in f_i \pm \sqrt{\frac{f_i\,(1-f_i)}{N}}\,, \qquad f_i \equiv \frac{n_i}{N}\,, \tag{6.107}$$

$$p_i \in f_i \pm \sqrt{\frac{f_i}{N}}\,, \tag{6.108}$$

which are the estimates of the true quantities (6.99).

These formulae are valid for $n_i \geq 5,\ 10$, that is, for bins containing at least about 10 events. In this case the Gaussian confidence levels hold.

For bins with less than ten events, Eq. (6.107) should be replaced by Eq. (6.31) (with N instead of n), and Eq. (6.105) should be replaced by Eq. (6.31) multiplied by N. The Poissonian formulae (6.106, 6.108) remain unchanged, but, in these cases, the confidence levels are not Gaussian and must be directly obtained from the binomial and Poisson distributions, depending on whether N is fixed or variable.

All the previous conclusions also give a satisfying intuitive representation of the bin content fluctuations. For example, if we consider a two-channel histogram, the number of events n_1 and n_2 in these two channels is completely correlated when N is constant, since $n_1 + n_2 = N$. In effect, we are dealing with a single random variable, and in this case the statistical errors of the two channels are equal, as is evident from Eq. (6.103), which is symmetric in f and $(1-f)$, or from Eq. (6.102), after a little bit of algebra. In general, a fixed N determines a correlation between channels, given by the covariance (4.92), which statistically, for $n_i,\ n_j > 10$, can be estimated with a good approximation as:

$$s(n_i, n_j) = -N f_i f_j\,. \tag{6.109}$$

On the other hand, when $\boldsymbol{N}$ is a Poissonian variable, any histogram bin behaves as an independent Poissonian event counter with fluctuations equal to the square root of the number of events.

The graphical representation of the histogram, to be complete, must then include statistical errors. By convention, these errors are evaluated using Eqs. (6.106, 6.108), neglecting possible correlation between channels, and are plotted as $\pm s_i$ bars centred on the n_i values. These intervals, called *error bars*, define a band that should contain the true values μ_i or p_i of Eqs. (6.98, 6.99). However, as always, we must remember that confidence levels, if $n_i > 5,\ 10$, follow the 3σ law. Therefore, the total band containing these values with a reasonable certainty is actually three times larger than the error bars shown in the graphs. The histogram of Fig. 6.10, completed with the error bars from Eq. (6.102), is shown in Fig. 6.11. This representation can be obtained with our routine `HistoBar(x,fre,errors="ON")`. In the case of a normalized histogram, the vector `fre` contains the frequencies, and the error request must be completed by the number of events, in order to apply Eq. (6.108). For example, in the case $N = 100$, the call should be `HistoBar(x,fre,errors="ON",nev=100)`.

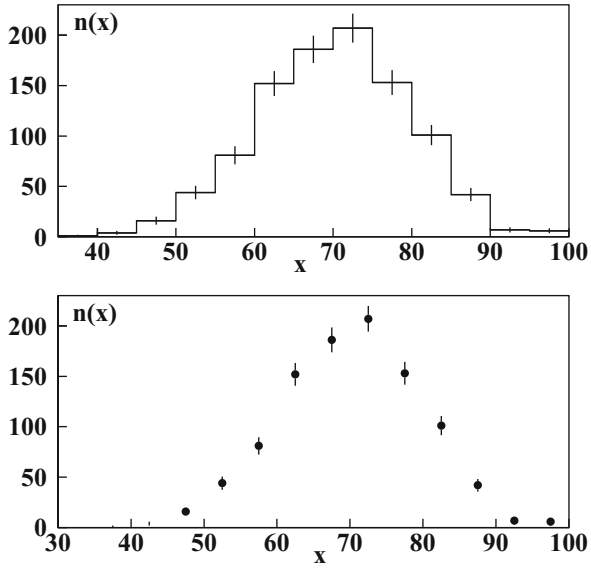


Fig. 6.11 Histogram of Fig. 6.10 with error bars

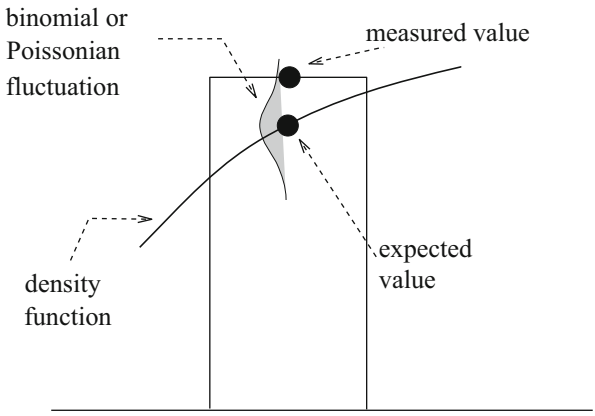


Fig. 6.12 Measured and expected values from a population density model, with the statistical fluctuation of the bin content. The shaded area should be imagined as projected orthogonally to the sheet

We have shown that the *fluctuations* in the number of events contained in a certain histogram bin follows the binomial or Poissonian probability. This is a completely general rule, independent of the density $p(x)$ describing the sample parent population, which can be any. As shown in Fig. 6.12, this density instead determines the overall structure of the sample, which is described by the mean or central values of the bin contents.

On this subject, an important hypothesis testing topic is how to check whether the shape of the sample is or not in agreement with a density model chosen for the population. We will discuss this issue in Sect. 7.5.

6.15 Estimation of the Correlation

The sample correlation coefficient obtained from a finite set of data (x_i, y_i) generally has a non-zero value, even when the variables are uncorrelated. It is therefore necessary to verify if the sample correlation coefficient r evaluated from the data is compatible or not with a null value (hypothesis test) or to estimate the confidence interval within which the true correlation coefficient ρ is located (parameter estimation). The sample estimate of the correlation coefficient (4.31) for a finite set of N elements requires the preliminary definition of the sample covariance $S(X, Y) \equiv S_{XY}$.

Without loss of generality, we can consider a pair of centred variables (with zero true mean), for which the true covariance is:

$$\mathrm{Cov}[X, Y] = \big\langle (X - \mu_x)(Y - \mu_y) \big\rangle = \langle XY \rangle \ . \tag{6.110}$$

To identify possible biases it is necessary, with a procedure similar to that of Eq. (6.57), to find the true mean value:

$$\left\langle \sum_{i=1}^{N} (X_i - M_X)(Y_i - M_Y) \right\rangle \ ,$$

where, in general, $M_X \to m_x \neq 0$, $M_Y \to m_y \neq 0$ even when the true values of the means are zero. By applying the linearity properties of the mean operator and noting that:

$$\sum_i M_X Y_i = M_X \sum_i Y_i = \ \frac{1}{N} \sum_j X_j \sum_i Y_i \ ,$$

and so on, one has:

$$\begin{aligned} \left\langle \sum_{i=1}^{N} (X_i - M_X)(Y_i - M_Y) \right\rangle &= \left\langle \sum_i (X_i Y_i - M_X Y_i - M_Y X_i + M_X M_Y) \right\rangle \\ &= \left\langle \sum_i X_i Y_i \right\rangle - \frac{1}{N} \left\langle \sum_i X_i \sum_j Y_j \right\rangle \\ &= N \, \langle XY \rangle - \frac{1}{N} \left\langle \sum_i X_i \sum_j Y_j \right\rangle \ . \end{aligned} \tag{6.111}$$

The last term of this equation can be rearranged as:

$$\left\langle \sum_i X_i \sum_j Y_j \right\rangle = \left\langle \sum_i X_i Y_i \right\rangle + \sum_{i \neq j} \langle X_i Y_j \rangle \ .$$

Now, X_i and Y_j are independent, because they are coming from different events sampled independently (remember that the correlation exists for the pair (X_i, Y_i), observed in the same event!). From Eq. (4.9) one then can write:

$$\langle X_i Y_j \rangle = \langle X \rangle \ \langle Y \rangle = 0 \ ,$$

since, by assumption, the true means are zero. Since $\langle \sum XY \rangle = N \langle XY \rangle$, Eq. (6.111) becomes:

$$\left\langle \sum_{i=1}^{N} (X_i - M_X)(Y_i - M_Y) \right\rangle = \left\langle \sum_i X_i Y_i \right\rangle - \frac{1}{N} \left\langle \sum_i X_i Y_i \right\rangle = (N-1) \langle XY \rangle \ . \tag{6.112}$$

This result, recalling Eq. (6.110), implies:

$$\mathrm{Cov}[X, Y] = \left\langle \frac{1}{N-1} \sum_{i=1}^{N} (X_i - M_X)(Y_i - M_Y) \right\rangle \ . \tag{6.113}$$

In conclusion, the unbiased sample covariance is:

$$s(x, y) = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - m_x)(y_i - m_y) \ . \tag{6.114}$$

If, instead of raw data, we have a two-dimensional histogram n_{ij} containing the number of pairs (x_i, y_j), the covariance is evaluated as:

$$s(x, y) = \frac{1}{N-1} \sum_{ij} (x_i - m_x)(y_j - m_y) n_{ij} \ . \tag{6.115}$$

The R routine `cov(x,y)` can be used to calculate $s(x, y)$ from a set of raw data while our routine `CovarHisto(x,y,mat)` performs the same operation for data presented in histograms, where `mat` is the matrix n_{ij} of Eq. (6.115).

To estimate the variance of the sample covariance, we can proceed as in the case of the sample variance:

$$\mathrm{Var}[s(x,y)] = \frac{1}{(N-1)^2}\sum_{ij}\mathrm{Var}[(x_i - m_x)(y_j - m_y)] \tag{6.116}$$

$$\simeq \frac{N}{(N-1)^2}\left[\left\langle (x_i - m_x)^2 (y_i - m_y)^2\right\rangle - \left\langle (x_i - m_x)(y_i - m_y)\right\rangle^2\right] ,$$

where the variance properties (2.67) and the additivity formula (5.74) have been used, since the different terms in the sum of $s(x, y)$ have null covariance. The last equality in Eq. (6.116) is approximated, because the exact equation should contain the true means. This inaccuracy can be corrected with some additional terms, discussed in [KS73], which are in the order of $1/N$ with respect to Eq. (6.116) and, therefore, are generally negligible. We have found that these terms affect the error given by (6.116) of 10–15% and only for small samples having about ten events. If `x` and `y` are two experimental data sets, the R coding of the last term of the equation, before the multiplication with $N/(N-1)^2$, is the following one:

```
> meanx = mean(x)
> meany = mean(y)
> mean(((x-meanx)*(y-meany))^2) - (mean((x-meanx)*(y-meany)))^2
```

This coding is used in our routines `CorrelEst` and `CovarTest`.

Let us now consider the sample linear correlation coefficient, that takes the compact form:

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_i (x_i - m_x)(y_i - m_y)}{\sqrt{\sum_i (x_i - m_x)^2 \sum_j (y_j - m_y)^2}} = \frac{\sum_i (x_i y_i)/N - m_x m_y}{s_x s_y} , \tag{6.117}$$

where Eq. (4.25) has been used for the last equality. This value is an occurrence of the random variable "sample correlation coefficient R".

The p.d.f. of R when the true correlation coefficient is $\rho = 0$ has been derived in 1915 by R.A. Fisher. In the case of N pairs of Gaussian variables, this distribution (deduced also in [Cra51]), follows a p.d.f. given by:

$$c(r) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}(1-r^2)^{(\nu-2)/2} , \tag{6.118}$$

where $\nu = N - 2$. In the general case where $\rho \neq 0$, the estimation of the true correlation coefficient starting from the sample data is a difficult problem, the brilliant solution of which is again due to R.A. Fisher, who proved this theorem in 1921:

Theorem 6.4 (Fisher Z Variable) *If R is a sample correlation coefficient obtained from N pairs of Gaussian variables, the variable*

$$Z = \frac{1}{2} \ln\left(\frac{1+R}{1-R}\right) \tag{6.119}$$

follows, for $N \to \infty$, a normal distribution with mean and variance given by:

$$\langle Z \rangle = \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right) , \quad \mathrm{Var}[Z] = \frac{1}{N-3} , \tag{6.120}$$

where ρ is the true correlation coefficient.

Proof The complete proof of the theorem is rather complicated and most texts refer to Fisher's original article or to his famous book [Fis41].

However, we can give a partial proof of the theorem by considering the case $\rho = 0$ and applying the inverse transformation of Eq. (6.119) to the density (6.118):

$$R = \frac{e^{2Z} - 1}{e^{2Z} + 1} = \tanh Z ,$$

where tanh is the hyperbolic tangent. Since:

$$1 - \tanh^2 x = \frac{1}{\cosh^2 x} , \quad \frac{\mathrm{d} \tanh x}{\mathrm{d}x} = \frac{1}{\cosh^2 x} ,$$

indicating with A all the constant coefficients of Eq. (6.118) and taking into account that $\nu = N - 2$, from the equality $c(r)\,\mathrm{d}r \equiv f(z)\,\mathrm{d}z$, one obtains:

$$f(z) = A\,(1 - \tanh^2 z)^{(N-4)/2} \frac{1}{\cosh^2 z} = A\,\frac{1}{\cosh^{(N-2)} z} .$$

We can expand the hyperbolic cosine as:

$$\cosh z = \frac{1}{2}(e^z + e^{-z}) = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \cdots \simeq e^{z^2/2} ,$$

where the last approximation holds for z around unity. Due to the presence of the logarithm, the values of z remain quite limited, except in the extreme case $r \simeq \pm 1$. We can then, with a good approximation, stop this second order expansion and write:

$$f(z) \simeq A e^{-(N-2)z^2/2} .$$

Thus it turns out that the variable Z is approximately normal, with zero mean and variance $\mathrm{Var}[Z] = 1/(N-2)$. However, the correct result is represented by Eq. (6.120). □

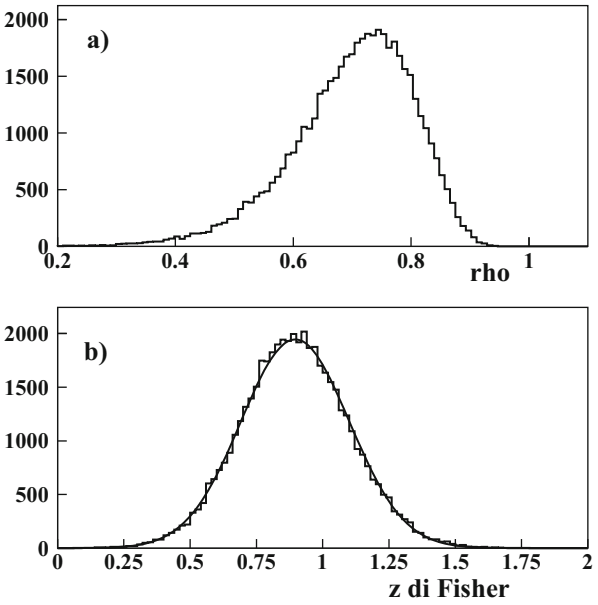


Fig. 6.13 Histogram of 50,000 correlation coefficients between 20 pairs of uniform variables $(X, X + V)$ (**a**); histogram of the corresponding Fisher Z variable and of the Gaussian (full curve) fitting the data (**b**)

In practice, the theorem is very powerful, because it holds for $N > 10$ and gives good results even with non-Gaussian variables. To show you the prodigious properties of the Fisher's transformation, we have shown in Fig. 6.13 the histogram of 50,000 sample correlation coefficients r, each obtained with 20 pairs of uniform variables $(X, X + V)$ of Exercise 4.2, and the corresponding histogram of the variable Z.

For the estimation of the correlation from a dataset $(\{R = r\}, \{Z = z\})$, one performs the transformation (6.119), applies the theory of estimation for Gaussian variables by determining the extremes of the confidence interval and then reverses these extremes by rising to exponential on both sides of Eq. (6.119):

$$e^{2z} = \frac{1+r}{1-r} \quad \Longrightarrow \quad r = \frac{e^{2z}-1}{e^{2z}+1} \,. \tag{6.121}$$

A simpler formula for the estimation interval can be obtained by applying to Eq. (6.121) the error propagation law and Eq. (6.120):

$$s_r = \frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{e^{2z}-1}{e^{2z}+1}\right)\sigma_z = \frac{4\,e^{2z}}{(e^{2z}+1)^2}\,\frac{1}{\sqrt{N-3}} \,.$$

Table 6.4 Height and chest measurement (in cm) of 1665 Italian soldiers of the First World War (the data are reported in: M. Boldrini, Statistica (Teoria e Metodi), Editor A. Giuffrè, Milan 1962 (only in Italian))

		Chest circumferences								Totals
		72	76	80	84	88	92	96	100	
Heights	150		1	7	2	2	1			13
	154		7	27	39	28	6	2		109
	158	3	7	69	118	87	21	5		310
	162		9	110	190	126	46	9	2	492
	166		4	68	145	114	58	12	5	406
	170		1	22	46	69	46	12	2	198
	174		1	4	15	35	22	6		83
	178			5	8	11	16			40
	182				2	10	2			14
Totals		3	30	312	565	482	218	46	9	1665

Substituting the value of z in Eq. (6.119), one then has:

$$s_r = \frac{1 - r^2}{\sqrt{N-3}} \quad \text{giving} \quad \rho \in r \pm \frac{1 - r^2}{\sqrt{N-3}} \quad CL \simeq 68\%. \tag{6.122}$$

Our routine `CorrelEst(x,y,conf,alt)` evaluates both the covariance from Eq. (6.114) and the correlation coefficient from Eq. (6.117) between two raw data vectors `x` and `y`. The variables `conf` and `alt` define CL and the type of estimation `"two"`, `"low"` and `"upp"` according to the scheme of Fig. 6.3. The error of the correlation coefficient is estimated by Fisher's method with Eqs. (6.120–6.121). The standard deviation error is found with Eq. (6.65) and the covariance error with Eq. (6.116) and the bootstrap method, a technique that we will describe later. If the data are represented as a two-dimensional histogram, the routine `CorrelEstH(x,y,mat,conf,alt)` can be used, where `x` and `y` are the bin coordinates and `mat` the two-dimensional matrix containing the number of events in the cell (x, y).

Exercise 6.10
Determine the correlation coefficient between height and chest size from the data of Table 6.4.

Answer The table, graphically reproduced in Fig. 6.14, represents a two-dimensional histogram. The class sizes can be easily deduced from its structure: for example, there are 110 soldiers with chest circumference

(continued)

Exercise 6.10 (continued)
between 78 and 82 cm (central value 80) and height between 160 and 164 cm (central value 162), and so on.

If t_i and s_i are the spectral values of the chest and height, the marginal histograms $n(t_i)$ and $n(s_i)$ have a Gaussian form, as can be easily deduced from the graphs (check this as an exercise). The means and standard deviations of the marginal histograms are an estimate of the corresponding true quantities of the marginal densities of the chest circumference and height. They can be calculated using Eqs. (2.41, 2.42). With obvious notation, one obtains:

$$m_t = \frac{1}{1665}\,(72 \cdot 3 + 30 \cdot 76 + \dots) = 85.71\,,$$

$$s_t = \left[\frac{(72 - 85.71)^2 \cdot 3 + (76 - 85.71)^2 \cdot 30 + \dots}{1664}\right]^{1/2} = 4.46\,,$$

$$m_s = \frac{1}{1665}\,(150 \cdot 13 + 154 \cdot 109 + \dots) = 163.71\,,$$

$$s_s = \left[\frac{(150 - 163.71)^2 \cdot 13 + (154 - 163.71)^2 \cdot 109 + \dots}{1664}\right]^{1/2} = 5.79\,.$$

The two standard deviations s_t ed s_s have been calculated dividing by $(1665 - 1)$, since the sample means have been used.

The 1σ estimate of the chest and height means is given by Eq. (6.50):

$$\mu_t \in 85.71 \pm \frac{4.46}{\sqrt{1665}} \simeq 85.7 \pm 0.1\,,$$

$$\mu_s \in 163.71 \pm \frac{5.79}{\sqrt{1665}} \simeq 163.7 \pm 0.1\,.$$

We now come to the study of correlation. The data should be correlated, because experience shows that short men with huge chests, or vice versa, are very rare. We therefore calculate the sample covariance (6.115):

$$\begin{aligned} s_{st} = \frac{1}{1664}\,[&(72 - 85.71)(158 - 163.71) \cdot 3 + \\ &(76 - 85.71)(150 - 163.71) \cdot 1 + \\ &(76 - 85.71)(154 - 163.71) \cdot 7 + \dots] = 6.876\,. \end{aligned}$$

(continued)

Exercise 6.10 (continued)
The sample correlation coefficient (6.117) is then given by:

$$r_{st} = \frac{s_{st}}{s_s\, s_t} = \frac{6.876}{5.79 \cdot 4.46} = 0.266\,.$$

We now estimate the height-chest correlation coefficient of the soldiers with a $CL = 95\%$. First, we calculate the Z variable corresponding to the sample correlation $r = 0.266$:

$$z = \frac{1}{2} \ln \frac{1 + 0.266}{1 - 0.266} = 0.272\,.$$

This is an approximate Gaussian variable, with standard deviation given by the second of Eq. (6.120). Since the data refer to 1665 soldiers, one has:

$$\sigma = \sqrt{\frac{1}{N-3}} = \sqrt{\frac{1}{1662}} = 0.024\,.$$

The 95% confidence limits for a Gaussian variable are given by the standard variable $t = 1.96$:

$$\mu_z \in 0.272 \pm 1.96 \cdot 0.024 = 0.272 \pm 0.047 = [0.225, 0.319]\,.$$

This interval then contains the value μ_z with $CL = 95\%$. The confidence interval for the true correlation coefficient ρ is finally evaluated by inserting the values (0.225, 0.319) in the second of Eq. (6.121):

$$\rho \in [0.221, 0.309] = 0.266^{+0.043}_{-0.045}\,, \qquad CL = 95\%\,.$$

Therefore, the height-chest data clearly demonstrate the presence of a correlation. This same interval can also be calculated with the approximate formula (6.122).

Moreover, all the previous results can be obtained as well by inserting the data of Table 6.4 into our routine `CorrelEstH()`.

One may ask the question: what chest circumference must a 170 cm tall soldier have to be considered normal? If we take the histogram of Table 6.4 as a reference sample, we can answer the question by estimating, with

(continued)

Exercise 6.10 (continued)
Eqs. (4.54, 4.55), the mean (on the regression line) and the standard deviation of the chest t conditional on the height s:

$$m(t|s) = m_t + r_{st}\,\frac{s_t}{s_s}\,(s - m_s) = 85.71 + 0.204\,(170 - 163.71) \simeq 86.9 \text{ cm}$$

$$s(t|s) = \left[s_t^2\,(1 - r_{st}^2)\right]^{1/2} = [19.89(1 - 0.071)]^{1/2} \simeq 4.3 \text{ cm}.$$

Again, under the assumption of a Gaussian model, we can state that the chest circumference of a normal soldier must be within the limits 86.9 ± 4.3 cm with a probability given by the *3σ law*.

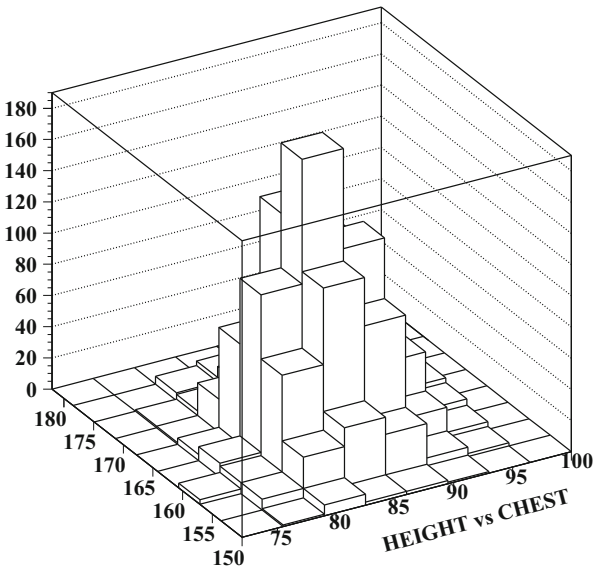


Fig. 6.14 Two-dimensional histogram of the data of Table 6.4

6.16 Problems

6.1 An urn contains 400 red and black marbles in unknown proportions; 30 marbles are drawn at random, including 5 red marbles. Estimate, with $CL = 95\%$, the initial number of red marbles R contained in the urn in the case of (a) extraction with replacement using the approximation for large samples, (b) extraction with replacement using the code `binom.test` and (c) extraction without replacement using the approximation for large samples.

6.2 In an experiment, 20 counts have been obtained. Find the upper bound of the expected value of the counts, with $CL = 90\%$. Use the Gaussian approximation of the Poisson distribution.

6.3 The standard deviation of the weight of a population of adults is $\sigma = 12$ kg. Find the mean $\langle S \rangle$ and the standard deviation $\sigma[S]$ for a sample of $N = 200$ individuals.

6.4 By running an unknown number N of tests, each one with a known a priori success probability $p = 0.2$, $n = 215$ successes have been obtained. Estimate N with $CL = 90\%$.

6.5 If X_i $(i = 1, 2, \ldots, N)$ is a random sample from a normal population, show that the covariance between the mean M and the deviation $(X_i - M)$ is zero.

6.6 25 Gaussian variables with same mean μ (unknown) and $\sigma = 1$ are summed, obtaining 245 as result. Find (a) the standard estimation interval for μ (CL = 68%) and (b) the upper limit of μ with $CL = 95\%$.

6.7 A sample of 40 Gaussian pairs has a sample correlation coefficient $r = 0.15$. Find the interval estimate ρ with $CL \simeq 68\%$.

6.8 A Poissonian count gives $x = 167$ events in 10 s. It is known that the average of the X counts is a function of both counting efficiency ϵ on the signal and of the background b according to the equation $\mu = \epsilon\, \nu + b$, that is:

$$X \sim \text{Poisson}(\epsilon \nu + b) \; .$$

The ϵ efficiency was determined as $\epsilon \pm \sigma_\epsilon = 0.90 \pm 0.10$, while the background is estimated based on a value of $b = 530$ counts in 100 s. Estimate the frequency value ν of the source with error.

6.9 Calculate the number N of bulbs you need to be 95% sure to have 1000 h of light, knowing that each bulb has a negative exponential life expectancy with mean $\tau = 100$ h.

Assume to use one single bulb until it burns out and then to switch immediately on the next one, until the limit of 1000 h is reached. Also assume the Gaussian distribution for the sum of the bulb life time.

6.10 The expected background of a counting experiment (accurately measured during calibration) is 10 counts/s. In an experiment (background plus possible signal) 25 counts are recorded in a second. Using the Gaussian approximation, find the upper limit of the counts for the signal only with $CL = 95\%$.

6.11 A Gaussian sample of $N = 25$ elements has a measured variance $s^2 = 18$. Find the upper limit with $CL = 95\%$.

6.12 An electric cable has 30 defects every 20 km. Find the number of defects/km with its error.

6.13 Having recorded 35 counts, find the lower event limit with $CL = 95\%$ using the routines `poisson.test` and `PoissApp`.

6.14 Create three random Gaussian samples of with 50 elements using the R instructions `x=rnorm(50), y=rnorm(50,5,1)` and `y1=3*x+y`. Find the covariances and the correlation coefficients between the variables `x,y` and `x,y1`, both analytically and using the routine `CorrelEst`.

6.15 Create three random samples with 100 elements from the uniform distribution with the R commands `x=runif(100),y=runif(100)` and `y1=2*x+y`. Find the covariances and the correlation coefficients between the variables `x, y1`, both analytically and with the routine `CorrelEst`.

6.16 In a famous experiment on the efficacy of aspirin [ET93], 104 cases of heart attacks occurred in a sample of 11,037 people who had been taking this drug for several years. In a control sample of 11,034 people who took a placebo, 189 heart attacks occurred. In this kind of studies, the *odds ratio* OR is often considered, i.e. the ratio between the frequencies $f_1 = 104/11037 = 0.00942$ and $f_2 = 189/11034 = 0.01713$. In this case $OR = f_1/f_2 = 0.55$, indicating that aspirin halves the probability of heart attacks. Find the confidence interval at $CL = 90\%$ of this data. (Hint: linearize the problem by applying logarithms.)

Chapter 7
Basic Statistics: Hypothesis Testing

There are two possible outcomes: if the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis, then you've made a discovery.

Enrico Fermi, QUOTED IN: T. JEREMOVICH, "NUCLEAR PRINCIPLES IN ENGINEERING"

No amount of experimentation can ever prove me right; a single experiment can prove me wrong.

Albert Einstein, QUOTED IN: A. CALAPRICE, "THE ULTIMATE QUOTABLE EINSTEIN"

7.1 Testing One Hypothesis

In Exercises 3.13 and 3.17, we have already discussed hypothesis testing in the framework of probability theory. Let us now go deeper into this topic considering the testing (acceptance or rejection) of one hypothesis, called *null hypothesis* H_0 or, more simply, hypothesis. The subject will then be completed later, in Chap. 10, where we will describe the criteria for optimizing the choices among several alternative hypotheses.

We therefore consider the p.d.f. $p(x;\theta)$ of the variable X defined in Eq. (6.2), depending on an unknown parameter θ and suppose that, on the basis of a priori arguments or previous experimental results, the hypothesis H_0 is assumed, corresponding to a value $\theta = \theta_0$. If an experimental value $\{X = x\}$ is obtained, one needs to decide whether or not to accept the model related to this hypothesis, on the basis of this result. The scheme used is typically that of Fig. 7.1: as you can see, the variability interval of X is divided into two regions, a region favourable to the hypothesis and a complementary region, called *critical region*. The shape of these regions strongly depends on the type of problem considered, as detailed in the following. When the critical region is concentrated in only one of the two tails of the distribution, its limit is defined by the quantile values x_α (left tail in Fig. 7.1a) or $x_{1-\alpha}$ (right tail in Fig. 7.1b).

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2 139, https://doi.org/10.1007/978-3-031-09429-3_7

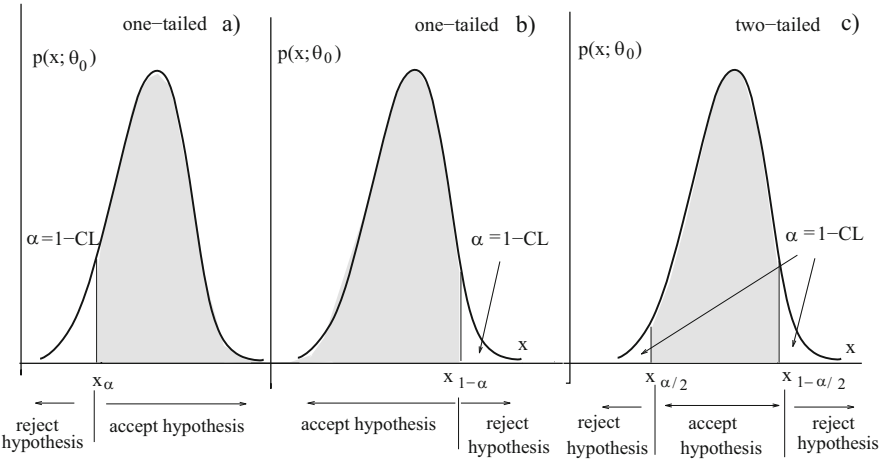


Fig. 7.1 One-tailed (**a**), (**b**) or two-tailed test (**c**). The confidence level CL is the value of the shaded area defined by the quantile values x_α, $x_{1-\alpha}$, $x_{\alpha/2}$ and $x_{1-\alpha/2}$. The test level $\alpha = 1 - CL$ is the value of the tail area outside the confidence interval

If, on the other hand, the critical region is formed by two disjoint subsets, as, for example, when one intends to reject a significant deviation from the mean value, its limits are defined by the quantile values $x_{\alpha/2}$ and $x_{1-\alpha/2}$ (see Fig. 7.1c). As we have already mentioned, the area α of Fig. 7.1 is called *significance level*, and we will denote it by SL. The pre-chosen value of the significance level α defines the α *level of the test*.

When the event $\{X = x\}$ is observed, under the hypothesis H_0 corresponding to $p(x; \theta_0)$, it is then necessary to calculate the conditional probabilities:

$$P\{X \leq x|H_0\} = \alpha_x \ , \ \text{one-tailed test (to the left)} \tag{7.1}$$

$$P\{X \geq x|H_0\} = \alpha_x \ , \ \text{one-tailed test (to the right)} \tag{7.2}$$

$$2\min\left(P\{X \leq x\}, P\{X \geq x\}\right) = \alpha_x \ , \ \text{two–tailed test} \tag{7.3}$$

In the two-tailed test, the area $\alpha_x/2$ is the smallest area, to the left or to the right of the abscissa value, according to Fig. 7.1.

The hypothesis is rejected when $\alpha_x < \alpha$, that is when we obtain by chance values "within the tails". With this rule, the probability that the decision is wrong *when* H_0 *is true* is always less than α. This is called a type I error (to reject a true H_0), or also a *false-positive case*. The other possible error, to accept a false H_0, is known as type II error or *false-negative case*. This second case will be detailed described in the next Chap. 10. The value α is called size of the type I error.

It should be noted immediately that all the previous definitions lose their meaning if we omit to specify *the hypothesis being true*, because, to calculate the significance level, it is necessary to know the p.d.f. corresponding to H_0. Within the frequentist

framework, this requirement is not a trivial aspect but is the very essence of the definition! The probabilities given by Eqs. (7.1–7.3) have not to be confused with the probability that H_0 is true for a given SL, which corresponds to $P(H_0|X \geq x)$. Previously, in Chap. 1, we have already emphasized the dangers of inverting conditional probabilities without going through Bayes' theorem.

The logical approach applied to reject a hypothesis assumed to be true has found an important field of applicability in modern science. In fact, according to the philosopher Karl Popper [Pop59], a scientific theory can be differentiated from the non-scientific ones by the fact that it is *falsifiable*, that is, it can be verified by an experiment. In many cases, this procedure passes through the statistical analysis of experimental data and the frequentist hypothesis test (see also Sect. 12.17). The quotes of Fermi and Einstein, in the epigraph of the chapter, well represent this distinctive scientific thinking.

Finally, it is important to emphasize that H_0 cannot be chosen subjectively but must be determined only on the basis of the currently accepted scientific knowledge (the so-called state of the art). Therefore, before announcing a new discovery, it is necessary to demonstrate that it is not possible to explain the result obtained only starting from past experience. We have already implicitly applied this rule in Exercises 3.13–3.17.

The observed significance level α_x, defined in Eqs. (7.1–7.3), is also called p-value. The meaning assigned to the terms significance level, test level and p-value can be slightly different from text to text. Here, we will adopt the following terminology:

$$\text{significance level} \begin{cases} \nearrow \text{fixed before the experiment} \rightarrow \text{test level} \\ \searrow \text{observed value from an experiment} \rightarrow p\text{-value} \end{cases}$$

Basically, with this terminology, a hypothesis test can be summarized as follows: the hypothesis is discarded if the p-value is less than the level of the test α; it is accepted otherwise.

By performing repeated experiments and calculating the p-value each time, we obtain a sample of the *random variable* p-value. In hypothesis testing it is crucial to know how this variable is distributed: the answer is given by

Theorem 7.1 (On the p-value p.d.f.) *The p-value follows the uniform density, that is, $P \sim U(0, 1)$.*

Proof The proof follows immediately from Theorem 3.5, which states that cumulative variables $C(X)$ are uniform variables. In the one-tailed test, one has $P \sim U(0, 1)$, since the p-value is $P = C(X) \sim U(0, 1)$ or $P = 1 - C(X) \sim U(0, 1)$. In the two-tailed test, the tail smaller area, to the right or to the left of the measured value, is always distributed as $U(0, 0.5)$. Since, in Eq. (7.3), we have defined the p-value to be the double of this value, we get again $P \sim U(0, 1)$. □

A hypothesis test can be performed also with a statistic $T = t(X_1, X_2, \ldots, X_N)$ that estimates the value of the unknown parameter θ, that is, with estimators. In

this case, the density of the estimator $p(t;\theta_0)$ must substitute the sample density $p(x;\theta_0)$ and the test procedure remains unchanged.

There are no fixed rules to establish the test level α. Usually small values, about 1–5%, are chosen, given that, as previously discussed, the probability of making a type I error must, in any case, be kept within an "acceptably low" level. However, when H_0 is based on a theory or physical law that is very well experimentally verified (as, e.g. is the case of Newton's law of universal gravitation), the scientific community requires a very strong contrary evidence to refute it, with a level α which can also reach $\simeq 10^{-6}$. This conservative position essentially serves to avoid a false discovery even in the presence of possible undetected mistakes both in measurement and data analysis procedures (see, e.g. [Lyo13, LW18]).

If the experimental p-value is less than α, it is located in the tails of the reference density. In this situation, it is considered less risky to reject the hypothesis $\theta = \theta_0$ than to accept it. The opposite is interpreted as a *statistical fluctuation*, the deviation is attributed to chance, and the hypothesis is accepted.

Finally, we note that in some cases, the comparison between test level and p-value is ambiguous:

- For a discrete random variable, the p-values of two adjacent values can straddle the outcome of the test. For example, in the one-tailed test on the right, it can happen that:

$$P\{X > x_1\} = p_1 < \alpha < p_2 = P\{X > x_2\}\,,$$

  with $x_1 > x_2$. In this case the hypothesis is accepted if $\{X = x_2\}$ is rejected if $\{X = x_1\}$. Otherwise, if a uniform generator of variables $0 \le U \le 1$ is available (e.g. a computer routine `random`), the test can be randomized with a probability p such that:

$$p_1 + p(p_2 - p_1) = \alpha \quad \Longrightarrow \quad p = \frac{\alpha - p_1}{p_2 - p_1}\,. \tag{7.4}$$

  Provided that the value $\{X = x_1\}$ is obtained, the hypothesis is discarded if the routine `random` gives a value $\{U = u \le p\}$; otherwise it is accepted. In this way, by construction, the experimental p-value exactly coincides with the α level of the test.
- In the case of the two-tailed test, different critical values can correspond to a given level α, that is, pairs of tails with different lengths but with the same area equal to $CL = 1 - \alpha$ are possible. Usually in this case the two left and right extremes are chosen to subtend the same area $\alpha/2$. This convention has been implicitly adopted in Eq. (7.3).

In the following we will describe the three most important test categories: the group of Student's tests (t-test and z-test), the χ^2-test and the test for the significance of variation sources, called ANOVA (*ANalysis Of VAriance*).

7.2 The Gaussian z-Test

This test is based on the standard Gaussian quantile, which is usually denoted as z:

$$z \equiv t = \frac{x - \mu}{\sigma} \,. \tag{7.5}$$

If the standard deviation σ is known, then the problem can be solved with the probability theory, verifying the compatibility of each single value with the values assumed as true. Examples of this technique were given in Exercises 3.13–3.17, and the same method can be applied in the case of *sample means* which, according to the Central Limit Theorem 3.1, can be considered Gaussian for $N >$20–30. However, when some density parameters are unknown, it is often necessary to perform the test using statistics. For instance, σ is usually unknown and is then replaced with the experimental value s. We know, from Sect. 6.11, that for Gaussian samples the quantile follows the Student's density that can be basically considered to be Gaussian when the number N of events is $N > 100$. In these cases the z-test for Gaussian samples can be used. This test can also be extended to two values x_1, x_2 when one needs to evaluate the difference $\mu_d = \mu_1 - \mu_2$ between two means.

Since the variance of the difference of two independent variables is given by Eqs. (5.66, 5.4), one can write:

$$\text{Var}[D] = \text{Var}[X_1] + \text{Var}[X_2] \simeq s_1^2 + s_2^2 \,,$$

where, in the last step, the approximation for large samples was used (see Table 6.3) by replacing the unknown true variances with the measured ones. We can then define the standard value:

$$t_D = \frac{x_1 - x_2 - \mu_d}{\sqrt{s_1^2 + s_2^2}} \tag{7.6}$$

and calculate the significance level according to the Gaussian density. The case of small Gaussian samples where t_D is a quantile of the Student's density will be discussed in the next section. We also remind that (as shown in Exercise 5.3) the difference of two independent Gaussian variables is also Gaussian. The quantiles t_D of Eq. (7.6) have to be identified with x_α, $x_{1-\alpha}$, $x_{\alpha/2}$, $x_{1-\alpha/2}$ of Fig. 7.1, and the evaluation of the corresponding p-value must take into account whether the test is one-tailed or two-tailed.

In R, the cumulative value corresponding to a (negative or positive) Gaussian quantile t is calculated by the `pnorm(t)` routine, which gives the area of the tail on the left of t. Given the symmetry of the distribution, the corresponding p-values are given, in the case of the two-tailed test, by p =`2*(1- pnorm(abs(t)))`, while in the case of the one-tailed test by p =`1-pnorm(abs(t))`. Usually Eq. (7.6) is used, for large samples, to compare two frequencies or two sample means. In the

first case, we can write the two frequencies as $f_1 = x_1/N_1$ and $f_2 = x_2/N_2$, if N_1 and N_2 are the number of trials of the two experiments. From Eq. (6.33), valid for large samples ($N_1, N_2 > 100, x_1, x_2 > 10$), one immediately obtains:

$$t_f = (f_1 - f_2 - p) \left[\frac{f_1(1-f_1)}{N_1} + \frac{f_2(1-f_2)}{N_2} \right]^{-1/2} , \tag{7.7}$$

with Gaussian probability levels, when $p = p_1 - p_2$ is the probability difference under H_0.

When $p = 0$, that is when under H_0 the two samples come from the same binomial population, the frequency obtained by adding the two homogeneous samples is often used for the error calculation:

$$\hat{f} = \frac{x_1 + x_2}{N_1 + N_2} . \tag{7.8}$$

By inserting this estimate into Eq. (7.7) with $p = 0$, we obtain the *pooled* standard variable:

$$t_p = (f_1 - f_2) \left[\hat{f}(1-\hat{f}) \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \right]^{-1/2} . \tag{7.9}$$

If $N_1 = N_2 = N$, the comparison can also be done in terms of number of hits x_1 and x_2 rather than of probabilities, always using Eq. (7.6) with $\mu_d = 0$ and the variance defined in Eq. (6.34):

$$t_x = (x_1 - x_2) \left[x_1 \left(1 - \frac{x_1}{N} \right) + x_2 \left(1 - \frac{x_2}{N} \right) \right]^{-1/2} . \tag{7.10}$$

Our routine `GdiffProp(x,n,p,pool,alt)` performs the test between two frequencies using Eqs. (7.7, 7.9). The variables `x` and `n` are two bidimensional vectors containing the values x_1, x_2 and N_1, N_2, respectively. If `pool=TRUE` (default) Eq. (7.9) is used, which requires $p = 0$ (default value), otherwise Eq. (7.7) is used. If `alt="two"` and `alt="one"`, the two-tailed and one-tailed test p-values are calculated, respectively.

To compare the difference between two true means $\mu = \mu_1 - \mu_2$ with that of two means m_1 and m_2, coming from two different samples, with variances s_1^2, s_2^2 and number of elements N_1, N_2 respectively, from (Eq. 6.53), we get the standard value:

$$t_m = (m_1 - m_2 - \mu) \left[\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right]^{-1/2} . \tag{7.11}$$

Often $\mu_1 = \mu_2$, that is $\mu = 0$, and the true means are unknown, but the compatibility between the two experimental means is tested. As we know, Gaussian probability levels can be applied when $N_1, N_2 > 100$.

The p-value corresponding to Eq. (7.11) is calculated by our routine `GdiffMean(m,s,mu,alt)`, where `m` and `s` are the two-dimensional vectors containing m_1, m_2 and $s_1/\sqrt{N_1}$, $s_2/\sqrt{N_2}$, respectively; `mu=` μ (= 0 by default) and `alt` are defined as in `GdiffProp`. When `m` and `s` contain a single value, `mu` is the true population mean, s is the sample standard deviation, and the test is performed with Eq. (7.5).

Exercise 7.1
In the measurement of the same physical quantity, two groups of experimenters obtained the values:

$$x_1 = 12.3 \pm 0.5\,,$$
$$x_2 = 13.5 \pm 0.8\,,$$

and reported that both measurements are of Gaussian type, since they have been obtained from large samples. Check if the results are compatible to each other.

Answer Since we do not know the size of the samples used by the two groups, we must assume that the approximation for large samples is valid and apply the Gaussian test.

The call to `GdiffMean(c(12.3,13.5),c(0.5,0.8))` gives $t_m = -1.27$ and a two-tailed p-value $p = 0.203$. Notice the call to the R function `c()` used to load the vectors with the experimental results.

Since the probability to be wrong when rejecting the compatibility hypothesis is too high, the two measurements must be considered compatible, i.e. they cannot be questioned only on the basis of statistics.

Exercise 7.2
In an extrasensory perception experiment (ESP), five boxes, numbered from 1 to 5, were prepared, and a target object was placed inside box number 3. Two hundred people were asked to guess which box contained the target and 62 of them indicated just the number 3. In a control test with all the empty boxes, but letting the audience believe that a target was present, 50 persons pointed to box 3. Determine whether the experiment reveals ESP effects or not.

(continued)

Exercise 7.2 (continued)
Answer If we use probability theory, we can proceed as in Exercise 3.17 and assume as a null hypothesis that each of the five boxes has an equal probability of $1/5$ of being chosen through a purely random guess. Then the expected theoretical distribution is binomial, of mean and standard deviation given by:

$$\mu = 40\,, \qquad \sigma = \sqrt{40\,(1 - 40/200)} = 5.65\,.$$

With 62 successes, the standard variable (3.37) is:

$$t = \frac{x - \mu}{\sigma} = \frac{62 - 40}{5.65} = 3.9\,,$$

corresponding, from Table E.1, to a very low significance level, i.e. $< 1 \cdot 10^{-4}$. Therefore, the hypothesis that ESP effects are present seems confirmed by this test.

However, experimental psychology states that the number 3 is psychologically favoured (in general all odd numbers are favoured over the even ones). In other words, in a blank test with boxes numbered from 1 to 5, most of the choices should be on box number 3, *even in the absence of a target*, simply because number 3 is "nicer" than the others.

It is then necessary, in the absence of an a priori model, to abandon the probabilistic approach and solve the problem in within the statistical framework, using only the number of hits with target ($x_1 = 62$) and the one without a target ($x_2 = 50$). The statistical errors to be associated with these observations are obtained from Eq. (6.34):

$$s(50) = \sqrt{50\left(1 - \frac{50}{200}\right)} = 6.12\,, \qquad s(62) = \sqrt{62\left(1 - \frac{62}{200}\right)} = 6.54\,.$$

Now, combining the results of the blank test and of the target test in Eq. (7.10), we have:

$$|t| = \frac{|62 - 50|}{\sqrt{(6.12)^2 + (6.54)^2}} = 1.34\,,$$

corresponding, from Table E.1, to an observed two-tailed significance level (p-value) of

$$P\{|T| > 1.34\} = 2 \cdot (0.5 - 0.4099) = 0.1802 \simeq 18\%\,.$$

(continued)

Exercise 7.2 (continued)
The routine `GdiffProp(c(62,50),c(200,200))` gives the same result; for the one-tailed test, one has `GdiffProp(c(62,50), c(200,200),alt="one")`, which gives the result $p = 0.091$.

Therefore, this analysis shows that in about 18% of times, making guesses and being psychologically biased in favour of number 3, there may be deviations of more than 12 units (in excess or in defect) between the blank test with and the test with the target. The excess occurs in 9% of cases. The result is therefore compatible with pure chance and reveals no possible ESP effects at all.

Exercise 7.3
Evaluate the compatibility between the true and simulated values in Exercise 4.2.

Answer This exercise refers to 10,000 simulated data pairs, from which have been obtained the values:

$$r_1 = 1.308\,10^{-2}\,, \quad r_2 = 0.7050\,, \quad r_3 = -0.7131\,.$$

The true values are:

$$\rho_1 = 0\,, \quad \rho_2 = 0.7071\,, \quad \rho_3 = -0.7071\,.$$

We transform both the experimental and the true values using Eq. (6.119) and the first of Eqs. (6.120), respectively:

$$z_1 = 1.308\,10^{-2}\,, \quad z_2 = 0.8772\,, \quad z_3 = -0.8935\,,$$

$$\mu_1 = 0\,, \quad \mu_2 = 0.8814\,, \quad \mu_3 = -0.8814\,.$$

The z_i values refer to a Gaussian variable with mean μ_i and standard deviation:

$$\sigma = \sqrt{\frac{1}{N-3}} = \sqrt{\frac{1}{9997}} = 0.0100\,.$$

We then can define the three standard variables:

$$t_1 = \frac{|z_1 - \mu_1|}{\sigma} = 1.38\,, \; t_2 = \frac{|z_2 - \mu_2|}{\sigma} = 0.42\,, \; t_3 = \frac{|z_3 - \mu_3|}{\sigma} = 1.21\,,$$

(continued)

Exercise 7.3 (continued)
which all give high significance levels, as you can see from Table E.1. The data differ by 1.38, 0.42 or 1.21 standard deviations from their mean, indicating a good agreement between simulation and expected values.

Exercise 7.4
In a work on the harmfulness of radio frequencies used by cell phones [F+18], a sample of 1631 rats was irradiated for 2 years with radio frequencies of intensity comparable with those of base radio antennas. The pooled results, compared to the control group, were as follows:

	Sample	Heart tumours	Brain tumours
Exposed rats	1631	13	19
Control group	817	2	4

Evaluate whether the data indicate some harmful effect.

Answer In this case, as previously commented, H_0 represents the non-harmfulness of the radio frequencies, that is, the homogeneity of the irradiated and control samples is assumed. For each tumour type, the data are binomial distributed. Due to the small values of x_1, x_2, we cannot apply Eqs. (7.7, 7.9) nor use the Student's density. As often happens, small non-Gaussian samples are generally not analysable with statistical models of general validity. Fortunately, in these cases there are Monte Carlo simulation methods, and in fact we will come back to this problem in Chap. 8.

Nevertheless, we can evaluate preliminarily the results under Gaussian approximation. We use our routine `GdiffProp` to perform the one-tailed test. We obtain the following results:
`GdiffProp(c(13,2),c(1631,817),alt="one")`: $p = 0.049$,
`GdiffProp(c(19,4),c(1631,817),alt="one")`: $p = 0.051$,
indicating a p-value close to the H_0 rejection limit that, in this kind of studies, is usually fixed around 5% or 1%. If we use Eq. (7.10) instead of Eq. (7.9), we obtain even smaller values:
`GdiffProp(c(13,2),c(1631,817),pool=FALSE,alt="one")`: $p = 0.024$,
`GdiffProp(c(19,4),c(1631,811),pool=FALSE,alt="one")`: $p = 0.031$,
We will reconsider the analysis of these data in Sect. 8.13.

7.3 Student's t-Test

From Sect. 6.11, we know that in the case of a Gaussian sample with N events having mean m and standard deviation s, the value:

$$t = \frac{m - \mu}{s/\sqrt{N}} \tag{7.12}$$

is the occurrence of a Student's variable with $N - 1$ degrees of freedom, if μ is the true mean of the parent population. The test must then determine the p-value of the Student's quantile and judge whether or not it is appropriate to discard the hypothesis. We notice that (7.12) is a valid test statistics for $N \geq 2$, since $N = 2$ is the smallest sample size that allows us to calculate s. The only possible test for $N = 1$ is that of Eq. (7.5), which requires the a priori knowledge of μ and σ.

In addition to the case of Eq. (7.12), where a Gaussian sample mean is compared to the population true mean value, the t-test is usually used also in two other situations:

1. To verify the difference between the means of two independent samples with respect to an expected one
2. To check the mean of the differences of two dependent or paired samples with respect to an expected mean difference

Let us now look at the first case.

Given two independent samples with means m_1, m_2, standard deviations s_1, s_2 and number of events N_1, N_2, respectively, the difference test considers the variable t as:

$$t_S = \frac{m_1 - m_2 - \mu}{\sqrt{\dfrac{s_1^2}{N_1} + \dfrac{s_2^2}{N_2}}} \,. \tag{7.13}$$

This equation is formally identical to Eq. (7.11), but in this case, to compute the p-value corresponding to this Student's quantile, it is essential to know the degrees of freedom ν, which is anything but trivial. The following steps explain how to proceed, but they are not essential, and those who are not interested can immediately go to Eq. (7.19).

We know, from Eq. (6.72), that $(N-1)s^2/\sigma^2$ follows a χ^2 distribution with $N-1$ degrees of freedom.

Analogously, if:

$$s_D^2 = \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \,, \quad \sigma_D^2 = \frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} \,,$$

we can affirm that $\nu s_D^2/\sigma_D^2$ is a χ^2 variable with ν degrees of freedom. Since, from Eq. (3.69), we know that $\mathrm{Var}[\chi_N] = 2N$, we can write:

$$\mathrm{Var}\left[\frac{\nu s_D^2}{\sigma_D^2}\right] = 2\nu = \frac{\nu^2}{\sigma_D^4}\,\mathrm{Var}[s_D^2] = \frac{\nu^2}{\sigma_D^4}\left[\frac{\mathrm{Var}[s_1^2]}{N_1^2} + \frac{\mathrm{Var}[s_2^2]}{N_2^2}\right] . \tag{7.14}$$

To find the variances of s_1^2 and s_2^2, we can use again Eq. (3.69):

$$\mathrm{Var}\left[\frac{(N-1)s^2}{\sigma^2}\right] = 2(N-1) \quad \longrightarrow \quad \mathrm{Var}[s^2] = \frac{2\sigma^4}{N-1} , \tag{7.15}$$

so that:

$$\mathrm{Var}[s_D^2] = \frac{2\sigma_1^4}{N_1^2(N_1-1)} + \frac{2\sigma_2^4}{N_2^2(N_2-1)} . \tag{7.16}$$

From the second and third term of Eqs. (7.14), we then have:

$$\frac{2}{\nu} = \frac{1}{\sigma_D^4}\left(\frac{2\sigma_1^4}{N_1^2(N_1-1)} + \frac{2\sigma_2^4}{N_2^2(N_2-1)}\right) , \tag{7.17}$$

and hence:

$$\nu = \frac{\sigma_D^4}{\dfrac{\sigma_1^4}{N_1^2(N_1-1)} + \dfrac{\sigma_2^4}{N_2^2(N_2-1)}} . \tag{7.18}$$

Substituting the unknown quantities σ_1 and σ_2 with the measured ones, we obtain the approximate Welch-Satterthwaite equation [Wel47], according to which the degrees of freedom are given by the integer nearest to:

$$\nu \simeq \frac{\left(\dfrac{s_1^2}{N_1} + \dfrac{s_2^2}{N_2}\right)^2}{\dfrac{s_1^4}{N_1^2(N_1-1)} + \dfrac{s_2^4}{N_2^2(N_2-1)}} . \tag{7.19}$$

Equations (7.13, 7.19), known as Welch's t-test, satisfactorily solve the comparison between the means of two small Gaussian samples having different variances. These formulae have to be used for $N_1, N_2 \geq 2$.

To conclude, we consider now the case where the two means come from samples whose parent populations have the same mean and variance: $\sigma_1 = \sigma_2 \equiv \sigma$. In this

situation, the standard variable of the difference becomes:

$$t_S = \frac{m_1 - m_2}{\sqrt{\dfrac{\sigma_1^2}{N_1} + \dfrac{\sigma_2^2}{N_2}}} = \frac{m_1 - m_2}{\sigma\sqrt{\dfrac{1}{N_1} + \dfrac{1}{N_2}}} \,. \tag{7.20}$$

Using Eq. (6.72), in which the variance is calculated with respect to the sample mean, we define the non-reduced χ^2:

$$\chi^2 = \frac{(N_1 - 1)s_1^2}{\sigma_1^2} + \frac{(N_2 - 1)s_2^2}{\sigma_2^2} = \frac{1}{\sigma^2}\left[(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2\right] \,, \tag{7.21}$$

which, from Theorem 3.4, has $N_1 + N_2 - 2$ degrees of freedom. Based on the results of Exercise 5.5, we can state that the variable (5.41), which now becomes:

$$t_S = \frac{m_1 - m_2}{\sigma\sqrt{\dfrac{1}{N_1} + \dfrac{1}{N_2}}} \frac{\sigma\sqrt{N_1 + N_2 - 2}}{\sqrt{\left[(N_1 - 1)\, s_1^2 + (N_2 - 1)\, s_2^2\right]}} =$$

$$= \frac{m_1 - m_2}{s_{12}\sqrt{\dfrac{1}{N_1} + \dfrac{1}{N_2}}} \,, \quad s_{12} = \sqrt{\frac{(N_1 - 1)\, s_1^2 + (N_2 - 1)\, s_2^2}{N_1 + N_2 - 2}} \,, \tag{7.22}$$

follows the Student's density with $N_1 + N_2 - 2$ degrees of freedom. Notice that, if in Eq. (7.18) one sets $\sigma_1 = \sigma_2$, the result $\nu \simeq N_1 + N - 2$ is *not* obtained, because here the Student's variables have a different structure. Therefore, as a practical rule, Eqs. (7.13, 7.19) must be used when the variances can be different, whereas it is better to use Eq. (7.22), which has the correct degrees of freedom, if it is a priori known that the variances are equal.

The p-value corresponding to the quantile (7.13, 7.19) or to the quantile (7.22) for different and equal variances, respectively, is computed by our routine `TdiffMean(m,s,n,mu,alt,var)`, where `m,s` and `n` are two-dimensional vectors containing the mean, the standard deviation and the number of events of the two samples, respectively. The mean default value `mu`=μ is 0. The two or one-tailed test type is selected through the variable `alt="one"` or `alt="two"`, whereas `var=FALSE` uses Eqs. (7.13, 7.19) and `var=TRUE` Eq. (7.22). As in the case of `GdiffMean`, when `m` and `s` are single-valued parameters, `mu` is the true mean, s is the sample standard deviation, and the test is made performed according to Eq. (7.12).

The R routine `t.test(x,y,mu,alt,var)` compute the p-values when the raw data of two samples are contained in the vectors `x, y`. The variable `alt="two"` (default value) computes the p-value for the two-tailed test. If the value `alt="less"` or `alt="greater"` is set, the left- or right-tailed value is

respectively calculated. The variable `var` is the same as in `TdiffMean`. This routine also includes Eq. (7.12) as a subcase when the `y` array is missing.

Exercise 7.5
Verify Eqs. (7.13, 7.19, 7.22) using the R routine `t.test`.

Answer We use here Monte Carlo methods that are described in the next chapter. In R, the `rnorm(n,m,s)` routine generates a vector of `n` Gaussian deviates with mean `m` and standard deviation `s`. These two vectors are passed to the routine `t.test`, and the variable `p.value` is read with the command `t.test(...)$p.value` (see Appendix B). With the R routine `replicate`, two samples of 10,000 simulated p-values are generated and stored in the vectors `tpst` and `tpsf`, corresponding to the cases of equal and different variances, respectively. From Theorem 7.1, if the method is correct, the p-value must follow the uniform distribution. These two samples are then passed to the R routine `density` which, using numerical methods, finds the shape of the parent population. These densities are finally plotted in the graphical window.
This procedure is included in our routine `TpTest` which is given in the following:

```
TpTest<- function(m1=0,m2=0,s1=1,s2=3,n1=10,n2=5){

  #prepare graph
  pts = seq(0.,1,length=100)  # points of the plot
  plot(pts,2.0*pts,type='n',xlab='p-values',ylab='d(p)')

  # generate p-values according to the two cases of variances
  tpst = replicate(10000,t.test(rnorm(n1,m=m1,s=s1),
                   rnorm(n2,m=m2,s=s2),m1-m2,var=TRUE)$p.value)
  tpsf = replicate(10000,t.test(rnorm(n1,m=m1,s=s1),
                   rnorm(n2,m=m2,s=s2),m1-m2,var=FALSE)$p.value)

  #  add the plots of the p.value densities to the initial one
  lines(density(tpsf),type='l',lwd=2)
  lines(density(tpst),col="red",type='l',lwd=2,lty='dashed')
}
```

Using this routine, and systematically changing the arguments (which we invite you also to do), we have obtained that Welch quantiles were correct for all considered cases, that is, $N_1, N_2 \geq 2$, both for different and equal variances. For small samples ($N_1, N_2 < 5$) and in the case of equal variances, the distribution obtained with the t-test of Eq. (7.22) better follows the uniform density. Surprisingly enough, the quantile (7.22), which assumes equal variances, is reliable even with different variances when $N_1 = N_2$, while it fails when $N_1 \neq N_2$, as shown in Fig. 7.2.

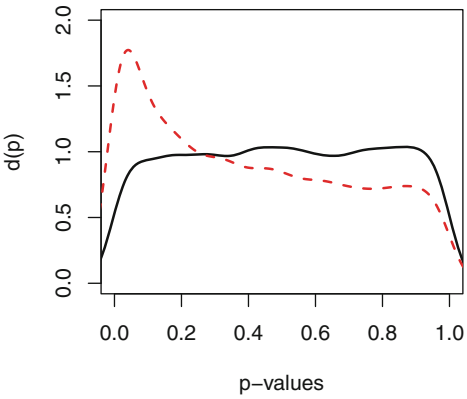


Fig. 7.2 Distribution of 10,000 p-values for the t-test between two samples of different size ($n_1 = 10$, $n_2 = 5$) coming from populations with mean $\mu = 0$ and different variances ($\sigma_1^2 = 1$, $\sigma_2^2 = 3$). The continuous line represents the Welch method using the quantile from Eqs. (7.13, 7.19); the dashed curve represents the p-value corresponding to the quantile (7.22). The plot shows that, in this case, the appropriate distribution is given by the Welch's method

We conclude this section with the use of t-test for paired samples. This procedure can be applied when all the data of the two original samples x_i and x'_i, having the same number of events $i = 1, 2, \ldots, N$, are available. Often this happens when two treatments have been carried out on the same group of objects or individuals. For example, we might have different blood test samples from people who were first given a placebo and then one or more (different) drugs. In this situation, the paired data test analyses the difference in response on a person-by-person basis, thus minimizing the effects due to discrepancies between individuals. With this procedure, a new sample of the differences ($d_i = x_i - x'_i, i = 1, 2, \ldots, N$) is created, with mean and standard deviation m_d and s_d, respectively.

In general, the null hypothesis to be falsified is that of an ineffective treatment. This implies the true mean of the paired difference sample to be null: $\{H_0 : \mu_d = 0\}$. It is therefore necessary check if the variable:

$$t_d = \frac{m_d}{s_d/\sqrt{N}} \tag{7.23}$$

follows the Student's distribution with $N - 1$ degrees of freedom. The smaller the p-value corresponding to t_d, the smaller the probability of making a mistake by discarding H_0 when it is true. When H_0 predicts true means $\mu_1 \neq \mu_2$, defining $\mu_d = \mu_1 - \mu_2$ the t_d variable becomes:

$$t_d = \frac{m_d - \mu_d}{s_d/\sqrt{N}}\,, \tag{7.24}$$

and the test procedure remains unchanged.

In R, the t-test for paired data, when the vectors `x` and `y` have the same dimension, can be performed with the call `t.test(x,y,paired=TRUE)`.

7.4 Chi-Square Test

With both the z and t-tests we can only check the compatibility of a single or a pair of sample values with a parameter of the parent population given under H_0. Using the χ^2-test or chi-square test, we can remove this limitation.

The test is based on the Pearson's Theorem 3.3, which states that the sum of squares of ν independent Gaussian standard variables follows the χ^2 distribution with ν degrees of freedom. In this case, the quantiles χ^2_α, $\chi^2_{1-\alpha}$ used to calculate the significance levels, unlike the Gaussian and Student quantiles, assume different values, due to the asymmetry of the χ^2 distribution. Here, we will also use the notation of Sect. 3.8 and denote with Q a variable following the χ^2 distribution and with the symbol χ^2 (the same of the density) the numerical occurrences of Q. If this variable is divided by the degrees of freedom, the resulting reduced chi-square variable is denoted with Q_R and χ^2_R.

Therefore, the test is based on the value:

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \,, \tag{7.25}$$

where x_i are variates coming from Gaussian densities with means μ_i and variances σ_i^2. The rationale of the test is that the variable (7.25) is distributed according to χ^2 only if μ_i and σ_i *are the correct values*, i.e. those assumed under H_0. The χ^2-test is very flexible and adapts to many different situations. For example, the expected value μ_i may be a theoretical model $\mu_i = f(x_i)$, the expected content of the number of events in a bin of a histogram of central value x_i (as we will see in the next section), the mean $\mu_i \equiv \mu$ of a statistical population from which the values x_i were obtained, and so on.

An important case where the χ^2-test applies is that of N Poissonian counts n_i, which, as we know, can be considered as Gaussians for $n_i > 10$. In this situation, $\sigma_i^2 = \mu_i$, and the test variable becomes:

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \mu_i)^2}{\mu_i} \,. \tag{7.26}$$

Sometimes, when $n_i < 10$, the continuity correction, also known as Yates correction, is used:

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \mu_i - 0.5)^2}{\mu_i} \,. \tag{7.27}$$

Frequently, the value n_i is used in the denominator, thus obtaining the modified χ^2:

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \mu_i)^2}{n_i} \,. \tag{7.28}$$

Equation (7.28) is usually applied when $n_i > 30$. This condition guarantees that this variable is still χ^2 distributed when μ_i is the true expected value. The test is valid also when there is only a single Poissonian sample having $\mu_i = \mu$:

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \mu)^2}{\mu} \,. \tag{7.29}$$

A further simplification of the test occurs when the true value μ, which may be unknown, is replaced by the value of the sample mean m:

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - m)^2}{m} \,. \tag{7.30}$$

In this case, the test only verifies that the sample under examination is Poissonian, without making further assumptions about its true mean. Note that the use of m creates a linear relation between the data and therefore, according to Theorem 4.4, the variable (7.30) has $N - 1$ degrees of freedom. The study of this variable shows that it tends to follow the χ^2 density for samples with number of events $N > 30$, as also shown in the Exercise 7.6 given below.

Exercise 7.6
Verify Eq. (7.30) with simulated data.

Answer We use the same method of Exercise 7.5. With the routine `Chi2Testm` (presented below), a set of Poissonian counts is generated with `rpois`, and then the p-values `pchi` and `pchif`, associated to the χ^2 variables `chi` and `chf` of Eqs. (7.29) and (7.30), are evaluated with `pchisq`. Since the true `lambda` and the experimental values `mexp` are used for the means, the estimators are correct if the Gaussian approximation holds. Then, from Theorem 7.1, the p-values are uniformly distributed.

The simulated distributions are obtained with the R routine `density`, that evaluates the p.d.f. from raw data with smoothing algorithms tuned by the parameter `adj`. With `adj=0.5`, a moderate smoothing is obtained that allows us to clearly recognize, in Fig. 7.3, the uniform shape of the obtained distributions.

(continued)

Exercise 7.6 (continued)

The results show that, for $N = 10$, Eqs. (7.29) and (7.30) follow the χ^2 density. This figure clearly demonstrates the importance to assign the correct number of degrees of freedom in Eq. (7.30), when the sample mean is used and the variables are thus correlated. We suggest you to vary the number $N =$ n of generated variables and check when the p-value distribution differs from the uniform density.

```
Chi2Testm<- function(n=10,Nsim=1000,denom=FALSE,adj=0.5,lambda=10){
   chi   <- seq(0,0,length=Nsim)  #clear vectors
   chif  <- seq(0,0,length=Nsim)
   pchi  <- seq(0,0,length=Nsim)
   pchif <- seq(0,0,length=Nsim)
         #prepare graph
   pts = seq(0.,1,length=100)  # points of the plot
   plot(pts,1.5*pts,type='n',xlab='p-values',ylab='d(p)')
         # simulate Nsim p-values
   for(i in 1:Nsim){
      z <- rpois(n,lambda) # n Poisson data of mean lambda
      mexp=mean(z)
      sigma2 = lambda
      if(denom==FALSE)sigma2=mexp
      for(j in 1:length(z)){
         chi[i]  = chi[i]  + (z[j]-lambda)*(z[j]-lambda)/lambda
         chif[i] = chif[i] + (z[j]-mexp)*(z[j]-mexp)/sigma2
         pchi[i] = pchisq(chi[i],n)      # chi2 p-values
         pchif[i]= pchisq(chif[i],n-1)   # n-1 degrees of freedom
      }
   }
            # final plots
   lines(density(pchi,adjust=adj),type='l',col='red',lwd=2)
   lines(density(pchif,adjust=adj),type='l',lwd=2,l='dashed')
 }
```

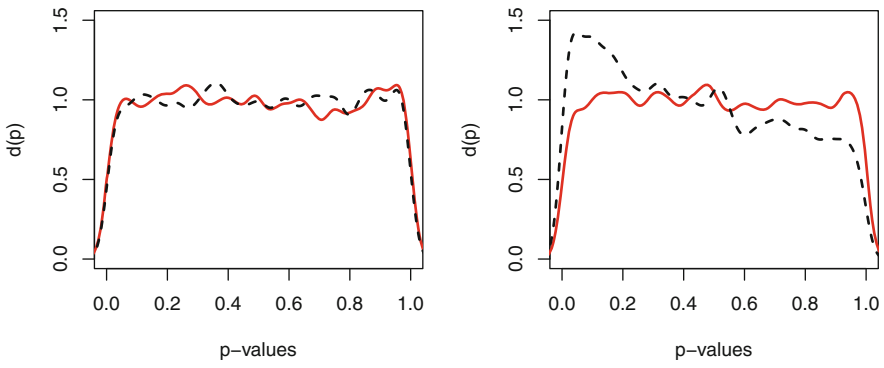


Fig. 7.3 To the left: p-value distribution from a sample with $N = 10$ Poissonian counts with the test variables (7.29) (full line) and (7.30) (dashed line). To the right: the same result is shown when in Eq. (7.30) N degrees of freedoms are wrongly used instead of $N - 1$. In this case, the obtained p-value (dashed line) differs from the uniform density

7.5 Compatibility Check Between Sample and Population

In this section we complete the study of the sample (6.97), already analysed in Sect. 6.14, by using the χ^2 test. The mean and variance calculated from data are:

$$m = \frac{1}{1000}\sum_{i=1}^{10} x_i n_i = 70.09\,, \quad s^2 = \frac{1}{999}\sum_{i=1}^{10}(x_i - m)^2\, n_i = 95.40\,.$$

The 1σ confidence intervals for mean, variance and standard deviation, calculated from Eqs. (6.53, 6.68), are:

$$\mu \in m \pm \frac{s}{\sqrt{N}} = 70.1 \pm 0.3\,,$$

$$\sigma^2 \in s^2 \pm s^2\sqrt{\frac{2}{N-1}} = 95.4 \pm 4.3\,,$$

$$\sigma \in s \pm s\frac{1}{\sqrt{2(N-1)}} = 9.77 \pm 0.22\,.$$

Since, for $N = 1000$, Gaussian confidence levels can be applied, we can say that these results are compatible with the true values $\mu = 70$ and $\sigma = 10$.

Now let us check if the sample comes from a Gaussian density, assuming that the mean and standard deviation values are the true ones, $\mu = 70$, $\sigma = 10$.

The expected or true number of events per bin are then given by Eq. (6.98), where $p(x)$ is the Gaussian, $N = 1000$ and $\Delta x = 5$:

$$\begin{aligned}\mu_i = Np_i &= N\int_{\Delta x}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right]\mathrm{d}x \\ &\simeq 1000\frac{1}{10\sqrt{2\pi}}\exp\left[-\frac{(x_i-70)^2}{200}\right]5\,. \end{aligned} \tag{7.31}$$

In columns (1)–(7) of Table 7.1, we have reported the following quantities: the spectrum value in the bin midpoint (1), the observed number of events (2), the expected number of events from Eq. (7.31) (3), the statistical error s_i from Eq. (6.102) (4), the standard deviation:

$$\sigma_i = \sqrt{\mu_i\left(1 - \frac{\mu_i}{N}\right)}$$

of the number of events (5), the values of the standard variable:

$$t_i = \frac{|n_i - \mu_i|}{s_i}\,,$$

Table 7.1 The data of the histogram of Fig. 6.11 are compared to the values of the Gaussian density (7.31)

1	2	3	4	5	6	7
x_i	n_i	μ_i	s_i	σ_i	t_i	t_i'
37.5	1	1	1.0	1.0	0.00	0.00
42.5	4	4.5	2.0	2.1	0.25	0.24
47.5	16	15.8	4.0	3.9	0.05	0.05
52.5	44	43.1	6.5	6.4	0.14	0.14
57.5	81	91.3	8.6	9.1	1.20	1.15
62.5	152	150.6	11.3	11.9	0.12	0.12
67.5	186	193.3	12.3	12.5	0.59	0.58
72.5	207	193.3	12.8	12.5	1.10	1.10
77.5	153	150.6	11.4	11.3	0.20	0.21
82.5	101	91.3	9.5	9.1	1.02	1.06
87.5	42	43.1	6.3	6.4	0.17	0.17
92.5	7	15.8	2.6	3.9	3.38	2.26
97.5	6	4.5	2.4	2.1	0.90	0.71

related to the observed number of events in each bin and calculated with the statistical error (6), and finally the values:

$$t_i' = \frac{|n_i - \mu_i|}{\sigma_i} ,$$

calculated with the true standard deviation (7). In Fig. 7.4 the histogram of Fig. 6.11 is shown (columns 1, 2 and 5 of Table 7.1) with the addition of the continuous line that represents the expected (true) values coming from the Gaussian (7.31) and computed in column 3 of Table 7.1.

A first evaluation of the agreement between data and model can be made by eye: approximately 68% of the experimental points should "touch", within an error bar, the corresponding true value. In our case, from column 7 of Table 7.1 and from Fig. 7.4, it results that, over a total of 13 points, the agreement with the expected values is found in 9 points (69%) within $\pm\sigma_i$, 12 points (92%) within $\pm 2\,\sigma_i$ and 13 values (100%) within $\pm 3\,\sigma_i$.

As you can see, the fluctuations seem to agree very well with the percentages given by the 3σ law (3.35), indicating that the density assumed as a model gives a correct representation of the data.

We reach the same conclusions also using the statistical errors, that is, the standard variables of column 6. The only anomalous channel is the one corresponding to the value $x_i = 92.5$, which in this case provides an estimated standard value of 3.38, compared to a correct value of 2.26. The disagreement originates from the low content of the channel, which has only seven events. For this reason, if statistical errors are used, discrepancies greater than $3\,\sigma$ are generally accepted in channels having less than ten events.

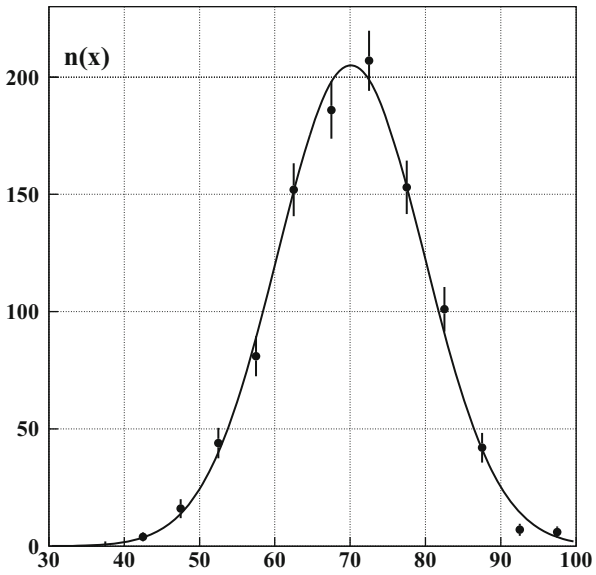


Fig. 7.4 Comparison between the histogram of Fig. 6.11 and the expected values from a Gaussian with $\mu = 70$ and $\sigma = 10$

We now perform the χ^2-test, (7.26) assuming as null hypothesis the true bin probabilities p_i of Eq. (7.31):

$$\chi^2 = \sum_{i=1}^{K} \frac{(n_i - \mu_i)^2}{\mu_i} = \sum_{i=1}^{K} \frac{(n_i - Np_i)^2}{Np_i} \,. \tag{7.32}$$

Equation (7.32) is approximately the sum of squares of K independent standard Gaussian variables when the total number of events $\boldsymbol{N}$ is variable, and the sum is made on bins with more than 5–10 events. Since, for a Poisson distribution, $\sigma_i^2 = \mu_i$, from Pearson's Theorem 3.3, we get that this sum follows the χ^2 density (3.66), with K degrees of freedom. The integral values of the reduced χ^2 density (3.72) are reported in Table E.3.

If, on the other hand, the total number N of events is constant, the variables of the sum (7.32) are correlated, but the Pearson's Theorem 4.6 still assures us that the result is a χ^2 variable, but this time with $(K - 1)$ degrees of freedom.

Note that, when N is constant, it is wrong to write:

$$\chi^2 = \sum_{i=1}^{K} \frac{(n_i - Np_i)^2}{N\, p_i(1 - p_i)} \qquad \text{(wrong!)} \,,$$

because this is the sum of squares of correlated variables.

In conclusion, it is always necessary to add the square of the differences between the observed frequencies and the true ones and divide by the true frequencies, taking care to remember that the degrees of freedom are equal to the number of channels if $\boldsymbol{N}$ is a Poissonian variable, while they must be decreased by one unit if N is constant.

All these rules derive from Pearson's Theorem 4.6 applied to the statistical analysis of histograms.

Often, with this test, one tries only to identify a deviation towards large values, i.e. towards the right tail of the expected distribution, when the null hypothesis is assumed to be true. In this case, the p-value is equal to:

$$SL = P\{Q \geq \chi^2(\nu)\} \,. \tag{7.33}$$

However, as we will discuss below, too small χ^2 values, where the model fits the data very well, are often suspect. In this case it is advisable to perform a two-tailed test (see Eq. 7.1), as shown in Fig. 7.5, doubling the smaller area to the right or left of the quantile value χ^2 or χ^2_R:

$$\begin{aligned} SL &= 2P\left\{Q(\nu) > \chi^2(\nu)\right\} \quad \text{if } P\left\{Q(\nu) > \chi^2(\nu)\right\} < 0.5 \,, \\ SL &= 2P\left\{Q(\nu) < \chi^2(\nu)\right\} \quad \text{if } P\left\{Q(\nu) > \chi^2(\nu)\right\} > 0.5 \,. \end{aligned} \tag{7.34}$$

We now perform the χ^2-test on the data of the histogram (6.97). Since the first bin only contains one event, we group the first two bins and sum over the 11 remaining ones. The value of the reduced χ^2 obtained from the data of Table 7.1 is given by:

$$\chi^2_R(11) = \frac{1}{11}\left[\frac{(n_1+n_2-\mu_1-\mu_2)^2}{\mu_1+\mu_2} + \sum_{i=3}^{13}\frac{(n_i-\mu_i)^2}{\mu_i}\right] = \frac{8.987}{11} = 0.82 \,. \tag{7.35}$$

Since N is fixed, the number of degrees of freedom is the number of the elements in the sum minus one, that is 11.

We can then state that, with the calculated χ^2 value, we have obtained the most comprehensive synthesis, because the results of Table 7.1 and Fig. 7.4, needed to compare data and model, are squeezed into a single value, in our case 0.82. Obviously, if we repeated the experiment, we would get a different result, because the χ^2_R of Eq. (7.35) is the value assumed by a random variable of density given in Table E.3.

We now finally proceed to the χ^2 test. In Table E.3, in the row corresponding to 11 degrees of freedom, we search for the area corresponding to the value of 0.82: we find a value of about 60%. From Eq. 7.34, this area corresponds to a p-value for a two-tailed test of:

$$SL = 2P\{Q_R \leq 0.82\} \simeq 2(1-0.60) = 0.80 \,.$$

In R, the same result is obtained by requiring the cumulative value with the command `2*pchisq(8.987,11)=0.754`. If the model holds, χ^2_R values smaller than this one are possible in at least 40% of experiments. Since a 80% significance level makes the type I error highly probable, the hypothesis must be accepted. In conclusion, we generated an artificial sample of 1000 events from a

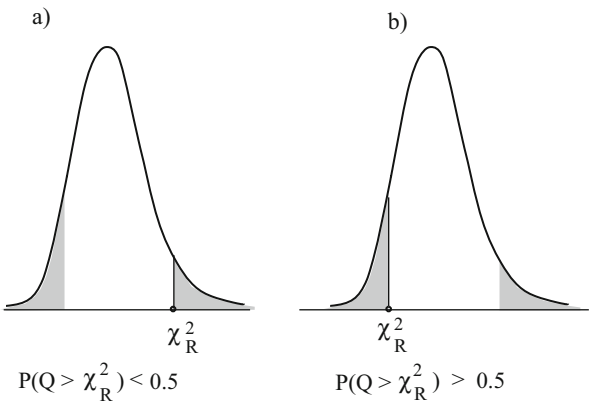


Fig. 7.5 Observed significance level for a two-tailed test (shaded areas) when the experimental χ^2_R value corresponds to a) $P\{Q_R > \chi^2_R\} < 0.5$ probability or b) $P\{Q_R > \chi^2_{R\alpha}\} > 0.5$. The shaded areas are equal

Gaussian distribution with $\mu = 70$ and $\sigma = 10$. Next, we performed statistical tests with respect to the true density. These tests showed good agreement between data and model.

As you can see, the logic of the χ^2-test is exactly the same as that used in the previous tests. The differences just involve only the variable type and the test function. The only point to be careful about is that the χ^2 density, unlike the Gaussian, is not symmetric. It is therefore necessary to take into account both quantile areas $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$. Too high or too low χ^2 values require further considerations: in both cases the test indicates that *the fluctuations of the data around the values assumed to be true are not purely statistical*. When

$$P\{Q_R > \chi^2_R(\nu)\} < 0.01 \,,$$

the reduced χ^2 value is in the right tail of its distribution curve and is too high. In this situation, the inadequacy of the parent population assumed as the model is highly probable. It may also be that the errors assigned to the data have been miscalculated and are underestimated. If errors have been correctly evaluated, the result of the test is the rejection of the hypothesis. Much more rarely, it could happen that the χ^2 value is too small:

$$P\{Q_R < \chi^2_R(\nu)\} < 0.01 \,.$$

This may be the case when the a priori probabilities p_i are evaluated from a density which tends to interpolate the data due to an excessive number of parameters, or when experimental errors have been erroneously overestimated. We will learn more about these concepts in Chap. 11. Often the χ^2 of histogram data is calculated by dividing by the measured frequencies instead of the true ones, thus applying Eq. (7.28):

$$\chi^2 = \sum_{i=1}^{K} \frac{(n_i - Np_i)^2}{n_i} \,. \tag{7.36}$$

Table 7.2 Histogram of the experiment consisting in $N = 100$ trials, each of them made of ten coin tosses (see also Table 2.2). The columns contain the possible number of heads (1); the number of times (successes) in which the number of heads reported in the first column has been obtained (2); the mathematical expectation, or true number of events, given by the total number of trials times the binomial probability of Eq. (2.29) with $n = 10$, $p = 1/2$ (3); the histogram statistical errors calculated with the expected number of events $\sigma_i = \sqrt{\mu_i \left(1 - \frac{\mu_i}{N}\right)}$ (4) and with the measured one $s_i = \sqrt{n_i \left(1 - \frac{n_i}{N}\right)}$ (5); the χ^2 values for each bin, obtained using the true probability (6); and the measured frequency (7)

Spectrum (n. of heads)	Successes	Binomial	Std. dv. "(true)"	Std. dev. (estimated)	χ^2 "(true)"	χ^2 (estimated)
x_i	n_i	μ_i	σ_i	s_i	$\frac{(n_i - \mu_i)^2}{\mu_i}$	$\frac{(n_i - \mu_i)^2}{n_i}$
1	2	3	4	5	6	7
1	0	1.0	1.0	0.0	1.00	—
2	5	4.4	2.0	2.2	0.08	0.07
3	13	11.7	3.2	3.4	0.14	0.13
4	12	20.5	4.0	3.2	3.52	6.02
5	25	24.6	4.3	4.3	0.01	0.01
6	24	20.5	4.0	4.3	0.60	0.51
7	14	11.7	3.2	3.5	0.45	0.38
8	6	4.4	2.0	2.4	0.58	0.43
9	1	1.0	1.0	1.0	0.00	0.00
10	0	0.0	0.0	0.0	0.00	0.00

In this case, the denominator is approximated, even if *model independent*. Therefore, the division by the frequencies expected from the model is more correct and consistent. However, if only channels with more than five to ten events are taken into account in the hypothesis test, the use of the measured frequencies almost always leads to equivalent results. In tests with minimization procedures, which we will describe in Chaps. 10 and 11, the measured frequencies are often used in the denominator. This choice greatly simplifies this type of algorithms since the denominator, being model independent, remains constant during the process of model adjustment to the data.

Exercise 7.7
Analyse the 10 coin experiment of Table 2.2 and Fig. 2.1, assuming independent trials and that all the coins have an a priori probability head/tail of 1/2.

(continued)

Exercise 7.7 (continued)
Answer The data are shown again in the new Table 7.2, where some new values useful for the analysis have been computed. We define the input parameters:

- Number of tosses per trial: $n = 10$
- Number of trials: $N = 100$
- Total number of tosses: $N \cdot n = M = 1000$

The first test is to check the total number of successes that is the total number of heads. From the first two columns of the table, we obtain:

$$x = 2 \cdot 5 + 3 \cdot 13 + \cdots = 521 \quad \text{successes.}$$

Since, under H_0, the expectation value is $Mp = 500$ and the standard deviation, from Eq. (3.5), is $\sigma = \sqrt{500(1-1/2)} = 15.8$, we obtain the standard value:

$$t = \frac{x - Mp}{\sqrt{Mp(1-p)}} = \frac{521 - 500}{15.8} = 1.33 \,,$$

corresponding, in Table E.1, to a p-value:

$$P\{|T| \geq 1.33\} = 2 \cdot (0.5000 - .4082) = 0.1836 \simeq 18.4\% \,.$$

Therefore, we can affirm that, in repeated experiments where 1000 well-balanced coins are flipped independently, in about 18% of times one can observe deviations greater or smaller than the expected average (500) of more than 21 units.

In Exercise 2.6 we computed the mean and variance from the histogram of Table 7.2:

$$m = 5.21 \,, \qquad s^2 = 2.48 \,, \qquad s = 1.57 \,,$$

and the corresponding expected value from a binomial density with parameters $n = 10$ and $p = 1/2$:

$$\frac{Mp}{N} = np = \mu = 5.00 \,, \qquad \sigma^2 = 2.50 \,, \qquad \sigma = 1.58 \,.$$

The test on the mean gives, using Eq. (6.50):

$$t = \frac{m - \mu}{\sigma/\sqrt{N}} = \frac{5.21 - 5.00}{0.158} \simeq \frac{5.21 - 5.00}{0.157} = 1.34 \,.$$

(continued)

Exercise 7.7 (continued)
This result is equal to the one obtained in the previous test on the total number of successes, because the identity:

$$\frac{Mp - x}{\sqrt{Mp(1-p)}} = \frac{Nnp - Nx/N}{\sqrt{\frac{N^2}{N} np(1-p)}} = \frac{np - x/N}{\sqrt{np(1-p)}/\sqrt{N}} = \frac{\mu - m}{\sigma/\sqrt{N}}$$

holds. For the test on the variance, we can use the large sample formula (6.68) and compute the standard value:

$$t = \frac{2.50 - 2.48}{2.50\sqrt{\frac{2}{N-1}}} = \frac{2.50 - 2.48}{0.35} = 0.06 \,,$$

which gives from Table E.1 a p-value:

$$P\{|T| \leq 0.06\} = 0.95 = 95\% \,.$$

In the end, we arrive at the χ^2-test, which is the final test on the overall sample shape. By grouping the first three and last three channels, so as to always have a number of events per bin ≥ 5, we obtain:

$$\chi^2_R(6) = \frac{1}{6}\left[\frac{(5-5.4)^2}{5.4} + \frac{(13-11.7)^2}{11.7} + \cdots + \frac{(7-5.4)^2}{5.4}\right] = \frac{5.18}{6} = 0.86 \,.$$

The number of degrees of freedom is 6, because the total number of events is fixed and there are 7 terms in the sum. Using Table E.3, we determine that the significance level corresponding to this χ^2_R value is:

$$1 - P\{Q_R > 0.86\} \simeq 0.48 \,.$$

As a matter of fact, a value $\{Q_R = \chi^2_R\}$ near the most probable one has been obtained. Also the call `pchisq(5.18,6)` gives a p-value $= 0.48$.

If, in the χ^2 calculation, we had divided by the measured frequencies, as in Eq. (7.36), we would have obtained $\chi^2_R = 7.42/6 = 1.24$ and a significance level of about 28%. The two results, although similar from a statistical point of view, differ significantly due to the rather small sample size ($N = 100$).

Very high significance levels were obtained in all the previous tests. This demonstrates a good statistical agreement between the data and the model assuming independent tosses of fair two-sided coins.

The experimental data with their statistical errors, and the theoretical values given by the binomial distribution, are also shown in Fig. 7.6.

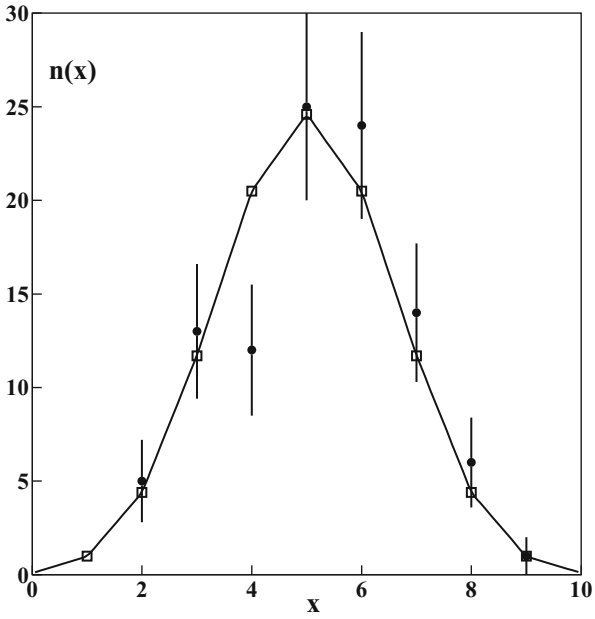


Fig. 7.6 Experimental data with error bars and values of the binomial distribution (empty squares) for 100 trials each consisting in 10 fair coin tossing. To guide the eye, the discrete points of the binomial density have been joined with a line

7.6 Hypothesis Testing with Contingency Tables

So far, we have described how to apply the χ^2-test when comparing a histogram with a parameter-dependent density model. Now let's see how to modify this procedure when comparing two or more samples, without assuming a specific density function for their population. These tests are called *non-parametric*. First of all, we note that, if the experimental data consists of a single frequency corresponding to a number of successes > 10, *the use of the χ^2-test is equivalent to the use of the two-tailed Gaussian test on a standard variable*. Indeed, if we consider the variables

$$T = \frac{X - \mu}{\sigma} , \qquad Q(1) = \frac{(X - \mu)^2}{\sigma^2} ,$$

we see that the first one follows the Gaussian density and the other one the χ^2 density with one degrees of freedom. You can easily check this fact by randomly assigning a value of T and performing both the two-tailed Gaussian test with Table E.1 and the one-tailed test for the variable $Q(1)$ with Table E.3: identical results are obtained.

In the case of a pair of Gaussian variables, the compatibility test can be performed using the Student or Gaussian density, according to the difference method of Sects. 7.2 and 7.3. Alternatively, if the experiment determines how often an event occurs in two independent samples, the analysis of *contingency tables* with the

χ^2-test is often used. This methods requires the creation of a table containing the number of successes n_a and n_b obtained with N_a ed N_b trials, respectively:

	Successes	Failures	Total
Sample A	n_a	$N_a - n_a$	N_a
Sample B	n_b	$N_b - n_b$	N_b
Total	$n_a + n_b$	$N_a + N_b - n_a - n_b$	$N_a + N_b = N$

Assuming that the two samples come from the same stochastic process with true probability p, the expected contingency table is:

	Successes	Failures	Total
Sample A	pN_a	$(1-p)N_a$	N_a
Sample B	pN_b	$(1-p)N_b$	N_b
Total	$p(N_a + N_b)$	$(1-p)(N_a + N_b)$	$N_a + N_b = N$

Each row of the experimental contingency table is a two-bin histogram, and the associated expectation table provides the corresponding expected value of the number of successes and failures. From Pearson's Theorems 4.4 and 4.6 and from the χ^2 additivity Theorem 3.4, the quantity:

$$\chi^2 = \frac{(n_a - pN_a)^2}{pN_a} + \frac{[N_a - n_a - (1-p)N_a]^2}{(1-p)N_a} + \frac{(n_b - pN_b)^2}{pN_b} + \frac{[N_b - n_b - (1-p)N_b]^2}{(1-p)N_b}, \tag{7.37}$$

can be considered as a χ^2 variable with two degrees of freedom.

If p is unknown and the null hypothesis just states that it is the same for the samples A and B, a point estimate can be calculated from the data. If the observed successes and failures are > 10, then one can set:

$$p \simeq \hat{p} = \frac{n_a + n_b}{N_a + N_b} = \frac{n_a + n_b}{N}. \tag{7.38}$$

Since this assumption introduces a further dependency relation between the data, according to the Definition 6.3, Eq. (7.37) represents the values assumed by a χ^2 variable *with a single degree of freedom*. Notice that the method is approximated, because the true probability has been estimated from the observed frequency (7.38). The tendency towards the χ^2 density of the variable (7.37) obviously holds for $N_a, N_b \to \infty$. However, the method is accepted and gives good results for $N > 40$, $n_a, n_b > 10$. This condition assures an approximately Gaussian number

of successes and reliable estimates of the probability. This point has been previously discussed in Sect. 6.7.

Exercise 7.8
To prove the validity of a vaccine, two groups of guinea pigs were studied, one vaccinated and the other unvaccinated. The results are reported in the following contingency table:

	*S*ick	*H*ealthy	*T*otal
Vaccinated	10	41	51
Unvaccinated	18	29	47
Total	28	70	98

Does the vaccine pass the right-tailed test at the 5% level?

Answer If we assume as a null hypothesis H_0 that the vaccine is not effective, then the differences between the two groups of guinea pigs are only due to statistical fluctuations. Under this hypothesis, the best estimate of the probability of contracting the disease will be given by the frequency (7.38):

$$\hat{p} = \frac{28}{98} = 0.286 \simeq 29\% \,.$$

Consequently, the probability to stay healthy is equal to 0.714. We can then construct the expected contingency table (i.e. the table of expected values under H_0: $0.286 \cdot 51 = 14.6$, etc.):

	*S*ick	*H*ealthy	*T*otal
Vaccinated	14.6	36.4	51
Unvaccinated	13.4	33.6	47
Total	28	70	98

From Eq. (7.37) one gets:

$$\chi^2 = \frac{(10-14.6)^2}{14.6} + \frac{(18-13.4)^2}{13.4} + \frac{(41-36.4)^2}{36.4} + \frac{(29-33.6)^2}{33.6} = 4.24 \,.$$

From Table E.3 we deduce that, with one degree of freedom, a value $\chi^2(1) = 3.84$ corresponds to a p-value=5%, whereas the obtained value, $\chi^2(1) = 4.24$, corresponds to $p \simeq 4\%$. We conclude that the vaccine passes the

(continued)

Exercise 7.8 (continued)
efficacy test at the 5% level, because the probability to be wrong by discarding a true H_0 hypothesis, stating that the vaccine is effective, is only 4%.

These calculations can be performed in R with the `chisq.test` routine, which can be used straightforwardly for contingency tables. If the table is loaded by rows with the `rbind` routine, with the simple command `chisq.test(rbind(c(10,41), c(18,29)), corr=F)` the values `X-squared=4.187, df=1, p-value=0.0407` are obtained. The `corr=F` condition excludes the Yates correction (7.27). Since a 2×2 contingency table has been used for this problem, we can alternatively use the method of Sect. 7.2 and the pooled formula (7.9).

With the call `GdiffProp(c(10,18),c(51,47),pool=T)`, we obtain the values: `quantile tz=-2.046, p-value for a two tailed Z test=0.041`. The non-pooled formula(7.7) gives a p-value of 0.038. With this method the expected contingency table is not used. The value $\texttt{tz}^2 = 4.19$ corresponds to a χ^2 with one degree of freedom in agreement with the value 4.24 previously found with the expected contingency table. We would have obtained two identical values if in `GdiffProp` we had used, instead of the experimental frequencies, those of the expected contingency table. It should be noted that both methods are approximate: the method using the expected contingency table implies that the true probabilities coincide with the frequencies calculated from the data under the assumption that the vaccine is ineffective, while the difference method uses the statistical errors calculated from the experimental frequencies instead of the true errors.

We now show how the χ^2 test can be extended to contingency tables of any size. In general, we can consider the following contingency table:

	*C*hannel *1*	*C*hannel *2*	$\cdots$	*C*hannel c	*T*otal
Sample 1	n_{11}	n_{12}	$\cdots$	n_{1c}	$\sum_j n_{1j}$
Sample 2	n_{21}	n_{22}	$\cdots$	n_{2c}	$\sum_j n_{2j}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
Sample r	n_{r1}	n_{r2}	$\cdots$	n_{rc}	$\sum_j n_{rj}$
Total	$\sum_i n_{i1}$	$\sum_i n_{i2}$	$\cdots$	$\sum_i n_{ic}$	$\sum_{ij} n_{ij} = N$

which is composed by r rows (the histogram of the sample) and c columns (the histogram bins). When only one histogram is present, this test can be performed with the method of Sect. 7.5.

As usual, we want to check whether the samples are homogeneous, that is, if they come from the same parent population. If this null hypothesis is true, then we can associate a true (unknown) probability p_j with any column of the table.

After multiplication of any sample row by the total number of elements $\sum_j n_{ij}$, this probability gives, in any cell, the expected number of events $p_j N_i$. Taking into account the Pearson's Theorems 4.4 and 4.6 and the χ^2 additivity Theorem 3.4, we can say that the quantity:

$$\chi^2 = \sum_{ij} \frac{(n_{ij} - p_j \sum_k n_{ik})^2}{p_j \sum_k n_{ik}} = \sum_{ij} \frac{(n_{ij} - N_i p_j)^2}{N_i p_j} \tag{7.39}$$

is χ^2 distributed.

The unknown value of the probability can be estimated from the data using Eq. (7.38) and summing by rows:

$$\hat{p}_j = \frac{\sum_{i=1}^{r} n_{ij}}{N} , \tag{7.40}$$

where N is the total number of events of the table.

Now we come to the calculation of the degrees of freedom. From Pearson's Theorem 4.6, we conclude that every row contributes with $(c - 1)$ degrees of freedom. However, this number, which is $r(c-1)$, must be decreased by the number of Eqs. (7.40) that have been used for the estimation of the true probabilities. These relations are $(c-1)$, because the probability of the last column is fixed by the closure relation:

$$\hat{p}_c = 1 - \sum_{j=1}^{c-1} \hat{p}_j .$$

Therefore, the total number of degrees of freedom is:

$$\nu = r(c-1) - (c-1) = (r-1)(c-1) . \tag{7.41}$$

In conclusion, given a predetermined significance level $SL = \alpha$ (usually $\alpha = 0.01$-0.05), on one or more histograms collected in a contingency table with $(i = 1, 2, \ldots, r)$ rows and $(j = 1, 2, \ldots, c)$ columns, one proceeds as follows. The reduced χ^2:

$$\chi^2_R(\nu) = \frac{1}{\nu} \sum_{ij} \frac{(n_{ij} - N_i \hat{p}_j)^2}{N_i \hat{p}_j} = \frac{1}{\nu} \left(\sum_{ij} \frac{n_{ij}^2}{N_i \hat{p}_j} - N \right) , \tag{7.42}$$

$$\nu = (r-1)(c-1) , \tag{7.43}$$

is calculated, where N_i are the row totals, N is the total number of events of the table and $\hat{p}_j$ is given by Eq. (7.40); H_0 is rejected at a level α if from Tables E.3, E.4 or from the routine `pchisq` a p-value $< \alpha$ is obtained.

Note that, in the non-parametric case, the hypothesis is not rejected when the χ^2 is too small, since the case where the a priori density model contains too many parameters is here inapplicable. The test is therefore always *one-tailed*, on the right tail. However, it should be noted that a too small value of χ^2, corresponding to cumulative values < 0.01, indicates a suspect coincidence between the experimental and expected contingency tables, usually due to *non-random fluctuations* or to correlations among data. In these situations, the null hypothesis should be accepted with caution, especially if there is a large number of degrees of freedom. In this regard, we recommend to solve Problem 7.5.

Could we avoid using the χ^2-test and always resort to the Gaussian difference test as done for the (2×2) contingency tables? The answer is negative, and we want to explain why. In the case of contingency tables with more than two dimensions, we could think of making a sum of frequency differences divided by the relative error, as in Eq. (7.7). If we remove the absolute value of the difference, the sum of these random variables will also be Gaussian, and we could perform the significance test with Table E.1. However, a sum of large fluctuations (positive and negative) could provide a good value of the standard variable even in the case of a definitely wrong hypothesis. If, to avoid this inconvenience, we used the absolute value in the sum of the differences, then we would no longer get at the end a Gaussian variable, and therefore we would not be able to perform the test. On the other hand, the χ^2 variable accumulates all the squared fluctuations being a sum of squares of the data with respect to the true value and is therefore more reliable. In technical language, we can say that the χ^2 test, compared to the Gaussian difference test, minimizes the type II error (to accept a wrong alternative hypothesis) and that therefore, based on the terminology that we will introduce in Chap. 10, is a more powerful test.

Exercise 7.9
In a factory, the production of a rubber timing belt is checked by three operators X, Y and Z. They perform a visual test on the quality of the product and may accept the piece or discard it for a type A or B defect. The work of the operators is summarized in the following table:

	Type A	*Type B*	*Good*	*Total*
Operator X	10	54	26	90
Operator Y	20	50	50	120
Operator Z	30	35	35	100
Total	60	139	111	310

(continued)

Exercise 7.9 (continued)
Determine whether the control criteria of the three operators are homogeneous at a 1% level.

Answer From Eq. (7.40), the following probabilities are obtained:

$$\hat{p}_A = \frac{60}{310} = 0.193\,, \quad \hat{p}_B = \frac{139}{310} = 0.448\,, \quad \hat{p}_{\text{good}} = \frac{111}{310} = 0.358\,,$$

which allow us to evaluate the expected contingency table:

	*T*ype *A*	*T*ype *B*	*G*ood	*T*otal
Operator *X*	17.5	40.4	32.2	90
Operator *Y*	23.2	53.8	43.0	120
Operator *Z*	19.3	44.8	35.8	100
Total	60.0	139.0	111.0	310

When repeating these calculations, you may find some small differences compared to the above values due to rounding effects.

We can use here the R routine `chisq.test`, uploading the data by row with the routine `rbind` through the call:

```
chisq.test(rbind(c(10, 54, 26), c(20, 50, 50), c(30, 35, 35))) ,
```

which gives the results: `X-squared=18.877, df=4, p-value=0.00083`. Indeed, from Eq. (7.43), it results that the degrees of freedom (`df`) are:

$$\nu = (3-1)(3-1) = 4\,.$$

Because of the very small p-value, we can safely conclude that these operators use different test criteria. Therefore, the homogeneity hypothesis must be rejected at the 1% level.

7.7 Multiple Tests

Besides the p-value, the multiplicity is the other crucial parameter of a test. For example, we can generate a vector `x` of 300 Gaussian deviates with the R routine `x <- c(rnorm(300))`. With the χ^2 test, we can easily verify the hypothesis H_0 that this sample originates from the standard normal curve $N(0, 1)$. In fact, given `x`, you can calculate the χ^2 value of Eq. (7.25) with $\mu_i = 0$, $\sigma_i = 1, \forall i$. Then,

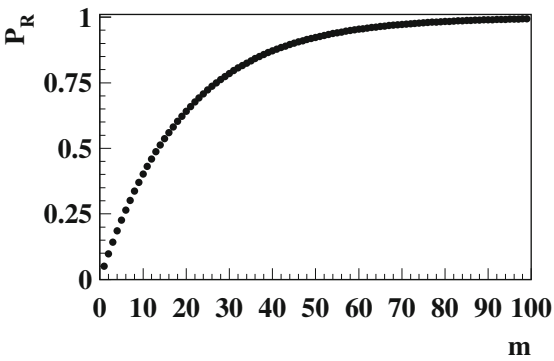


Fig. 7.7 Probability P_R to that at least one test belongs to the rejection region as a function of the number m of tests when all the hypotheses are true

with `xchi <- 1-pchisq(sum(x^2),300)`, you will obtain a large p-value, supporting the validity of H_0. Equivalently, you can also verify that a sample of p-values `xchi` generated in this way will perfectly follow the uniform distribution, according to the Theorem 7.1.

However, we can approach this test as a multiple test of 300 Gaussian variables (7.5). Obviously, this is not convenient now, but it often happens to deal with families of tests, rather than with a single experiment consisting of repeated trials. Generally speaking, suppose we have a test consisting of a *family of hypotheses* $H_1, H_2, \ldots, H_m$ and we want to verify if all the hypotheses of the family must be accepted. Suppose also to have a test level $\alpha = 0.05$. In our previous example, we have 300 p-values, and, although the simulated hypotheses are obviously true, out of 300 hypotheses we will have about 15 p-values $p < \alpha$.

In general, given m hypotheses, the probability for all the test results to be inside the acceptance region is $(1-\alpha)^m$, and therefore that of having at least one element in the rejection region is $P_R = 1 - (1-\alpha)^m$. In Fig. 7.7 the behaviour of P_R is shown as a function of the number m of performed tests, when all the hypotheses of the family are true and $\alpha = 0.05$. As the plot clearly shows, the probability of rejecting at least one true hypothesis and therefore of having false positives increases very rapidly as m increases. To solve this problem, we must observe that now we are mainly interested to evaluate the level of the family test α_F and we must therefore distinguish it from the level α of the single test. We then define the two relations, one inverse of the other:

$$\alpha_F = 1 - (1-\alpha)^m \simeq m\alpha \ , \tag{7.44}$$

$$\alpha = 1 - (1-\alpha_F)^{1/m} \simeq \frac{\alpha_F}{m} \ . \tag{7.45}$$

The exact formula is known as the Sidak correction, whereas the approximation

$$\alpha_F \simeq m\alpha \ , \tag{7.46}$$

which is easily obtained with a Taylor expansion of the term $(1-\alpha)^m \simeq 1 - m\alpha$ around $\alpha \simeq 0$, is named Bonferroni correction.

Therefore, giving a null hypothesis H_0, consisting of one family of m hypotheses H_i associated with m p-values p_i, one of these two equivalent procedures must be chosen to perform a test at the α_F level:

(1) Reject H_0 if, for at least one hypothesis, it results:

$$p_i < \alpha_F/m \ , \ \text{or} \quad p_i < 1-(1-\alpha_F)^{1/m}. \tag{7.47}$$

(2) Transform all p_i according to the rule:

$$p_i \to p_i' = m\, p_i \ , \ \text{or} \quad p_i \to p_i' = 1-(1-p_i)^m \tag{7.48}$$

and reject the hypothesis if there is at least one $p_i' < \alpha_F$.

Normally, the Bonferroni correction is used for simplicity, but, for large p_i values, it is convenient to apply Eq. (7.48), i.e. the non-approximate formula. For example, if $p_i = 0.1$ and $m = 5$, Bonferroni's correction gives $p_i' = 0.50$, the correct formula $p_i' = 0.41$. However, since usually the family test is done for values of $\alpha_F = 0.01$ or $\alpha_F = 0.05$, this discrepancy is often not crucial.

The R routine `p.adjust(pv,method)` uses the procedure (2), where `pv` is the vector of p_i and `method` selects the test type. With the call `pout <- p.adjust(pv, method='bonferroni')`, the routine applies the first of Eqs. (7.48) and gives the vector `pout` containing the p_i' values as output. This allows us to identify the family hypotheses that do not satisfy the test level. Our routine `MultiTest(pv,method,alpha,print), method='sidak')` applies also the Sidak correction using the second of Eqs. (7.48); this routine also checks how many hypotheses do not satisfy the predefined `alpha` level and allows you to check the output with the `print` parameter. Further details on the use of these routines are given in the problems at the end of the chapter. The Sidak-Bonferroni (SB) correction completely solves the problem of multiple tests when all the assumptions of the family are true. However, the goal of tests is usually to identify *which of the family's hypotheses are false*. Consider, for example, a drug or a group of drugs given to different groups compared to a control group that was given a placebo. In these cases, it is assumed as a null hypothesis that all groups of the family are equivalent, and the hypotheses H_i, which do not satisfy this criterion, correspond to the groups to which the effective drug has been given. In this situation, the test is required to be able to identify false hypotheses with the maximum efficiency, since they are connected to the searched effect.

This crucial feature is called *power of the test*. The real situation, summarized in Table 7.3, is then more complex than that considered so far. In the formalism of the table, the power is defined as the mean value of the fraction V/m_1 of the false hypotheses correctly identified.

Table 7.3 Possible results of a multiple test with m hypotheses H_i, of which m_0 true. False positives (FP) are named type I errors; false negatives (FN) are type II errors

	Accept H_i	Reject H_i	Total
H_i true	U true negative	F false positive	m_0
H_i false	W false negative	V true positive	$m_1 = m - m_0$
Total	$m - R$	R	m

When analysed in terms of power, the SB correction is completely unsatisfactory: in fact, the values p_i' of Eq. (7.48) increase linearly with the number of hypotheses present in the test, and many false hypotheses assume quickly p-values above the test level α_F and are therefore not correctly identified. In statistical jargon, the SB correction is not very powerful and conservative (i.e. new effects are easily missed), or it has a low detection potential. For this reason, the technique of multiple tests has been refined in recent years with many other methods, developed with the aim of identifying the false hypotheses H_i present within the H_0 family. For example, in modern genomics, while sequencing complex genomes, multiple tests are used against a null hypothesis requiring hundreds or thousands of p-values. Almost all of these tests are included in the R software. We will here describe the Benjamini-Hochberg [BH95] test, called the BH test, which is one of the most frequently used. To start, we have to introduce two important terms of the multiple test language that describe the probability of getting false positives: the *family-wise error rate* (FWER) and the *false discovery rate* (FDR).

FWER indicates the probability of making at least one type I error, that is, the probability that at least one true hypothesis of the family does not pass the test level, i.e. $P\{F \geq 1\}$, with the notation of Table 7.3. For example, when we have ten hypotheses, $\alpha_F = 0.05$ and Eq. (7.45) is used, FWER gives the probability that at least one test on H_i has $p_i < \alpha_F/m = 0.005$. This fact, if all the hypotheses are true, should occur on average in 5 per thousand of total tested hypotheses, and since the testing procedure checks ten hypotheses, this happens on average in a fraction of the family test exactly equal to α_F. Therefore, if all the hypotheses are true:

$$\text{FWER} \equiv P\{F \geq 1\} = 1 - \left(1 - \frac{\alpha_F}{m}\right)^m \simeq \alpha_F \ , \ m_0 = m \ , \tag{7.49}$$

where the number F of true hypotheses that are rejected is defined in Table 7.3. When a property holds under the condition $m_0 = m$ which, according to Table 7.3 implies that all the hypotheses are true, it is said *valid in the weak sense*. When instead a property holds for $m_0 \leq m$, it is said *valid in the strong sense*.

For example, assuming the existence of false hypotheses within the family, and denoting with i_j, $j = 1, 2, \ldots, m_0$ the indices of the tests corresponding to the true

hypotheses, it can be shown that the following property holds in the strong sense for the Bonferroni correction:

$$\text{FWER} = P\left\{\bigcup_{j=1}^{m_0}(p_{i_j} \leq \alpha_F/m)\right\} \leq \sum_{j=1}^{m_0} P\{p_{i_j} \leq \alpha_F/m\} = \frac{m_0}{m}\alpha_F \,. \tag{7.50}$$

Equation (7.50) is also valid in the general case of mutually dependent hypotheses, because the symbol of set union takes into account the possibility of refusing at the same time two or more hypotheses when their p-values are mutually dependent; the inequality holds on the basis of Eq. (1.17). We can therefore state that the Bonferroni correction of Eqs. (7.44, 7.45) ensures by construction the property of FWER $= \alpha_F$ in the weak sense, and FWER $\leq \alpha_F$ in strong sense, even for correlated hypotheses.[1]

The FDR property instead has been introduced to evaluate the strong properties of the test, when $m_0 < m$. It is defined as the *expected value* of the ratio between the number F of the true hypotheses H_i, falsely rejected, over the total number R of the rejected hypotheses. Hence, FDR $= \langle F/(F+V)\rangle = \langle F/R\rangle$ if $R > 0$, while, to avoid division by zero, FDR $= 0$ otherwise (see, as usual, Table 7.3).

When $m_0 < m$, one obtains, in the strong sense:

$$\text{FDR} = \left\langle \frac{F}{F+V} \right\rangle \leq \langle F \rangle = \sum_{j=1}^{m_0} P\{p_{i_j} \leq \alpha_F/m\} = \frac{m_0}{m}\alpha_F = \text{FWER} \,. \tag{7.51}$$

When $m_0 = m$, FDR can be considered as the mean value of a binary random variable: 0 when $F = 0$ and 1 when $F > 0$, since in this case $V = 0$ and $F/R = 1$. We can then write:

$$\text{FDR} = P\{F = 0\} \cdot 0 + P\{F \geq 1\} \cdot 1 = P\{F \geq 1\} = \text{FWER} \,. \tag{7.52}$$

All previous considerations can be summarized by the following two important properties:

$$\begin{aligned} &\text{FDR} \leq \text{FWER} \;\;,\;\; m_0 \leq m \,, \\ &\text{FDR} = \text{FWER} = \alpha_F \;\;,\;\; m_0 = m \,. \end{aligned} \tag{7.53}$$

To correctly apply these formulae during the following discussion, it is useful to keep in mind that only m and R, among the variables defined in Table 7.3, are known to the experimenter.

[1] Here we define as correlated hypotheses the cases in which test statistics are correlated.

One of the most diffused methods used to increase the power of the test is that of Benjamini-Hochberg, known as the HB method. In 1995 they demonstrated

Theorem 7.2 (of Benjamini-Hochberg) *A family of hypotheses $H_1, H_2, \ldots, H_m$ is given, corresponding to a set of p-values $p_1, p_2, \ldots, p_m$, ordered in increasing order. If k is the largest index i satisfying the inequality:*

$$p_i \leq \frac{i}{m}\alpha \tag{7.54}$$

and all the hypotheses H_i, $i = 1, 2, \ldots, k$ are rejected, a test with FDR $\leq \alpha$ is obtained, for any configuration of false and true hypotheses when they are mutually independent.

Proof For the non-trivial proof of the theorem, based on the principle of induction, we refer to the original article [BH95]. □

The theorem leads to a very simple method, similar to that of Bonferroni based on the first of Eqs. (7.48): instead of multiplying all p_i by m, just sort them in ascending order and multiply them by m/i, where i is the index obtained in the sorting. The value p'_1 is therefore identical to that of Bonferroni, while the last one remains unchanged. The remarkable fact, stated by the theorem, is that, if we exclude all the hypotheses for which $p'_i \leq \alpha$, we obtain an expected value FDR $\leq \alpha$ for the fraction of true hypotheses falsely rejected.

The HB method is implemented in the R routine `p.adjust(pv,method, alpha)` with the command `method="HB"` or equivalently with `method="fdr"`. Finally, it should be kept in mind that in this case the parameter `alpha` (which is 0.05 by *default*) represents the upper bound of FDR rather than the global test level α_F. Therefore, with the HB method, the control of FWER is abandoned. This parameter can then assume large values, even around 0.5. We recall that, from Eq. (7.53), if FWER is controlled, the same happens for FDR; however, the vice versa is not true, because if the power increases, also the number of true hypotheses with small p-values (due to statistical fluctuations) that are rejected inevitably increases.

Exercise 7.10
Generate with R 900 standard normal deviates and 100 variates $S \sim N(3, 1)$ with $\mu = 3$ and test the H_0 hypothesis of origin of the data from the standard Gaussian $N(0, 1)$. Use Bonferroni (SB) and Benjamini-Hochberg (BH/fdr) methods.

(continued)

Exercise 7.10 (continued)
Answer The requested data are generated with the R routine `rnorm`, and the corresponding p-values `pg` are evaluated with `pnorm`. Then the data are analysed with our routine `MultiTest`:

```
g <- c(rnorm(100,mean=3),rnorm(900))
pg <- 1- pnorm(g)
MultiTest(pg,method='bonferroni',alpha=0.05)
```

With a second call to `MultiTest` with the `"fdr"` method, all the test results are obtained.

We note that, without the multiple test techniques, we have now obtained, with $\alpha = 0.05$, 11 data accepted among the 100 in disagreement with the hypothesis (equal to a type II error of 11%) and 45 rejected data among the 900 correct ones (equal to a type I error of 5%, as predicted by the value of α). Since these are simulated data, different simulations will give slightly different values.

With the statement `1- pchisq(sum(g^2),1000)`, where the first argument is simply Eq. (7.25) with $\mu = 0$, $\sigma = 1$ and the second is the number of degrees of freedom, we can verify H_0 with the χ^2-test. A right one-tailed p-value near to zero is obtained, indicating the presence of many wrong hypotheses. With `MultiTest` we can then search for the false hypotheses. We obtain Table 7.4, which clearly shows the power gain which is acquired with the BH method. Notice that in a real experiment only R is known. For a simulated calculation of the parameters FDR and FWER, it would be necessary to repeat the exercise a very large number of times, also reducing the number of elements of the family if necessary. This evaluation can be performed with our `LogpFdr` routine, that you are invited to examine.

Table 7.4 Results of Exercise 7.10 obtained with the routine `MultiTest`. The symbols are those of Table 7.3

Method	F/R	Power V/m_1
Bonferroni	0/22	0.002
BH-fdr	3/66	63/100

7.8 Snedecor's F-Test

Similarly to the case of means, also tests on the compatibility between variances of distributions can be performed.

For samples with less than a hundred events, this test cannot be performed using the Gaussian or Student density. However, for Gaussian samples the Snedecor's density F, introduced in Exercise 5.6, can be used. These tests, called F-tests, have been generalized and used extensively in the ANOVA method, which is illustrated in the next section.

If s_N^2 and s_M^2 are two variances obtained from two independent samples with N and M events, respectively, coming from Gaussian populations having the same variance σ^2, we know, from Eq. (6.72), that the value:

$$F_{N-1,M-1} = \frac{s_N^2/\sigma^2}{s_M^2/\sigma^2} = \frac{s_N^2}{s_M^2} \,, \tag{7.55}$$

is the ratio between two independent reduced χ^2 values. From Exercise 5.6 we then know that this ratio follows the Snedecor's density F of Eq. (5.46) with $(N-1, M-1)$ degrees of freedom. The combined use of Eq. (7.55) and of the F density quantiles of Tables E.5, E.6 is called analysis of variance or F-test.

Exercise 7.11
Two experimenters, who claim to have sampled from the same Gaussian population, have obtained variances equal to

$$s_1^2 = 12.5 \,, \qquad s_2^2 = 6.4 \,,$$

with samples having 20 and 40 events, respectively. Check if the two sample variances are compatible with a single true value σ^2 at the 2% level.

Answer The variable $\boldsymbol{F}$ is given by:

$$F = \frac{12.5}{6.4} = 1.95 \,.$$

Since the (two-tailed) test is at the level of 2%, for the initial claim to be accepted, the experimental ratio F must be smaller than the 99% F value that can be obtained from Table E.6 or with the R statement `qf(0.99,df1=19,df2=39)`:

$$F_{0.99}(19, 39) \simeq 2.41 \,,$$

(continued)

Exercise 7.11 (continued)
and larger than the 1% F value that can be obtained with Eq. (5.49) or with the R statement `qf(0.01,df1=19,df2=39)`:

$$F_{0.01}(19, 39) = \frac{1}{F_{0.99}(39, 19)} \simeq \frac{1}{2.7} = 0.36 \,.$$

Since:

$$0.36 < 1.95 < 2.41 \,,$$

the two values are compatible at the 2% test level.

7.9 Analysis of Variance (ANOVA)

In Sect. 7.3 we applied the t-test to verify the hypothesis that two independent samples have the same mean. This is the simplest example of analysis of variance (ANOVA), that is, the set of procedures used to establish whether groups of elements behave in a similar way or not (i.e. beyond purely statistical fluctuations) under different conditions.

The first step towards a test generalization is to consider more than two groups, i.e. more than two conditions, such as the comparison between a reference drug and alternative drugs for the treatment of a disease, the effect of different teaching methods on learning and the effect of different soldering methods on the quality of printed circuit boards. In the same way, more *levels* of the same treatment can also be compared, such as a chemical reaction at different temperatures or different dosages of the same active ingredient in a drug.

The application examples we have just mentioned indicate that ANOVA applies to programmed experiments, for which there is a specific statistical terminology:

- The *response* is the main parameter of interest (e.g. the maths learning level).
- The *factors* are other quantities that are varied during the experiment, because it is assumed that they can influence the response (e.g. the teaching method).
- The different values that can be taken on by the factors are called *levels*.
- A factor can be *qualitative* or *quantitative*, where in the first (second) case the levels cannot (can) be put in correspondence with values on a scale.

When experiments are planned, it is almost always recognized that there are different factors influencing the response and, to optimize time and material resources, multiple factors are examined at the same time. However, to facilitate our presentation, we will start with an example of a *one-way* ANOVA, with only one single factor.

Table 7.5 Number of breaks in the warp of a fabric according to the tension of the loom

Tension	Breaks								
Low (L)	27	14	29	19	29	31	41	20	44
Medium (M)	42	26	19	16	39	28	21	39	29
High (H)	20	21	24	17	13	15	15	16	28

To produce a fabric, the threads of the weft are intertwined with those of the warp. When weaving, warp threads are stretched parallel on the loom and can break. In one experiment, several fabric samples of equal length were produced by subjecting the warp to high, medium or low tension, and the number of warp breaks in each sample was counted. The results are shown in Table 7.5. The factor of interest is the tension and the response the number of breaks. The basic idea behind ANOVA is to identify the *sources of variations*, in order to disentangle the effects of the factor on the response from pure statistical effects. The way to proceed is suggested by a simple algebraic equivalence: denoting by y_{ij} the number of breaks in the fabric j produced with tension i, by m the number of tension levels and by n the number of fabric samples produced for each tension level, we have in fact:

$$y_{ij} = \bar{y}_{\cdot\cdot} + (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) + (y_{ij} - \bar{y}_{i\cdot}), \quad i = 1, \ldots, m \; ; \; j = 1, \ldots, n \; , \tag{7.56}$$

where $\bar{y}_{i\cdot} = \sum_j y_{ij}/n$ and $\bar{y}_{\cdot\cdot} = \sum_{ij} y_{ij}/(mn)$. For convenience we have associated the tension levels L, M and H to the indices 1, 2 and 3.

Hence, the fabric ij has a number of breaks given by the sum of the general average of all the breaks, adjusted with the difference between the average breaks of the fabrics at tension i and the general average (effect of tension i) and with the difference between the breaks in fabric ij and the average of fabrics at tension i (statistical fluctuation). If we generalize the specific sample of 27 tissues into a population model, we can consider both the general average effects and the tension effect as population parameters and define the following model:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, \ldots, m; \; j = 1, \ldots, n \; . \tag{7.57}$$

Here ε_{ij} are independent random variables with zero mean, and then Y_{ij} is a random variable with mean $\mu+\tau_i$: that is, it is expected that on average the number of breaks is determined by the loom and the specific tension set. Since $\sum_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) = 0$ by definition, it is natural to set $\sum_i \tau_i = 0$. This constraint also ensures that we have m independent parameters to describe the m populations identified by the m tension levels. With the model (7.57), we can investigate whether the tension has an effect on breaks using the hypothesis test:

$$H_0 : \tau_1 = \cdots = \tau_m = 0 \quad \text{against} \quad H_1 : \exists\, i \text{ such that } \tau_i \neq 0 \, , \tag{7.58}$$

which requires an assumption on the distribution of ε_{ij}, that is, $\varepsilon_{ij} \sim N(0, \sigma^2)$. We will discuss later this assumption and that of independence among the ε_{ij}. From the

algebraic equivalence (7.56), the total sum of squares can be decomposed as done before in Eqs. (2.40, 6.55):

$$\sum_{ij}(y_{ij}-\bar{y}_{\cdot\cdot})^2 = n\sum_{i}(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2 + \sum_{ij}(y_{ij}-\bar{y}_{i\cdot})^2 = SS_{Tr} + SS_E \ . \tag{7.59}$$

This quantity can then be partitioned into the sum of squares between groups (breaks at different tension), often denoted as SS_{Tr} (*treatment sum of squares*), representing the variation due to the tension, and the sum of squares within groups (*error sum of squares*), often denoted as SS_E, which gives the statistical measurement errors common to all data. Under H_0, the variables Y_{ij} are iid $N(\mu, \sigma^2)$, and hence also the variables $\bar{Y}_{i\cdot}$ are iid $N(\mu, \sigma^2/n)$. Therefore, due to Theorem 6.2:

$$MS_{Tr} = \frac{SS_{Tr}}{m-1} = \frac{n}{m-1}\sum_{i=1}^{m}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2 \sim \sigma^2\chi^2_R(m-1) \ , \tag{7.60}$$

where MS means *mean squares*. Likewise, $\sum_j(Y_{ij}-\bar{Y}_{i\cdot})^2 \sim \sigma^2\chi^2(n-1)$, and, taking into account that SS_E is the sum of m independent χ^2 distributed variables due to the independent responses, we then have:

$$MS_E = \frac{SS_E}{m(n-1)} = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(Y_{ij}-\bar{Y}_{i\cdot})^2 \sim \sigma^2\chi^2_R(m(n-1)) \ . \tag{7.61}$$

Then, under H_0, both MS_{Tr} and MS_E are two unbiased estimators of σ^2, and we expect their ratio to be not far from 1. Indeed, $\langle MS_{Tr}\rangle = \sigma^2\langle\chi^2_R(m-1)\rangle = \sigma^2$ and similarly $\langle MS_E\rangle = \sigma^2$. Since the scalar product (4.74) $\sum_i[(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})\sum_j(y_{ij}-\bar{y}_{i\cdot})] = 0$ because $\sum_j(y_{ij}-\bar{y}_{i\cdot}) = n\bar{y}_{i\cdot}-n\bar{y}_{i\cdot} = 0$, for the Cochran's Theorem 4.5, MS_{Tr} and MS_E are also independent. Then, from Eq. (7.55), we have:

$$F = \frac{MS_{Tr}}{MS_E} \sim F(m-1, m(n-1)) \ , \tag{7.62}$$

and H_0 will be rejected at the level α if $F > F_{1-\alpha}(m-1, m(n-1))$ or, on the basis of the p-value, which is the probability that $F(m-1, m(n-1))$ is greater than the observed value F: $P\{F(m-1, m(n-1)) > F\}$). We can easily convince ourselves that the test rejection region is the correct one by observing, first of all, that the distribution of MS_E, the mean sum of squares of the tissues at the same tension, is always the same, both below H_0 and below H_1; indeed from the model (7.57), we have:

$$Y_{ij}-\bar{Y}_{i\cdot} = \mu+\tau_i+\varepsilon_{ij}-\mu-\tau_i-\bar{\varepsilon}_{i\cdot} = \varepsilon_{ij}-\bar{\varepsilon}_{i\cdot} \ ,$$

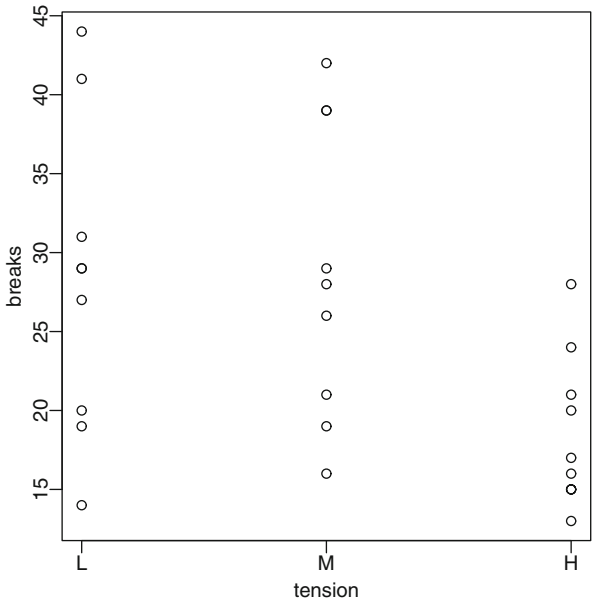


Fig. 7.8 Data of Table 7.5, number of breaks in the warp compared to tension

independently of τ_i. Conversely, the distribution of MS_{Tr} changes under H_1 to a non-central χ^2_R distribution[2] multiplied by σ^2, with expected value $\sigma^2 + n\sum_i \tau_i^2/(m-1)$, and in this case we expect to observe large values of F, at odds with the distribution under H_0.

Let us focus now on the data from the weaving experiment. Before performing any hypothesis tests, it is worth to examine a graphical representation of the data. We want to analyse the relation between tension and breaks; therefore, the representation of the number of breaks with respect to the tension in Fig. 7.8 is adequate. The figure shows a possible difference in mean between the tensions and a dispersion of the measures within the groups around their own mean apparently not completely homogeneous.

To execute the test with R, an object of the type *data frame* should be created and filled with the data of Table 7.5 using the command `data = data.frame(breaks, tension)`, where `breaks` is the vector obtained by concatenating the rows of the table and `tension = rep(c("L", "M", "H"), each = 9)` is obtained replicating nine times the factors `L, M, H`. Then ANOVA is performed with the R routine `aov`, to which the table `data` is loaded, specifying that the first column contains the `breaks` and the third one the factor `tension`. The result can be stored in an object `fit` with the command

[2] The non-central χ^2 distribution designates the variable $Q = \sum_i X_i^2$ where $X_i \sim N(\mu_i, 1)$ and $\sum_i \mu_i^2 = \lambda \neq 0$. The central (standard) χ^2 distribution has $\lambda = 0$.

Table 7.6 ANOVA table: R output with the data of Table 7.5, testing of the tension effect on the number of breaks

	df	SS	MS	F	Pr($>F$)
Tension	2	568.5	284.26	4.059	0.0303
Residuals	24	1680.7	70.03		

Table 7.7 General ANOVA table for the one-way analysis of variance

Source of variation	df	Sum of squares	Mean of squares	F	p-value
Treatment	$m-1$	$SS_{Tr} = n\sum_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$	$MS_{Tr} = \frac{SS_{Tr}}{m-1}$	$\frac{MS_{Tr}}{MS_E}$	$P(>F)$
Residuals	$m(n-1)$	$SS_E = \sum_{ij}(y_{ij} - \bar{y}_{i\cdot})^2$	$MS_E = \frac{SS_E}{m(n-1)}$		
Total	$mn-1$	$SS_T = \sum_{ij}(y_{ij} - \bar{y}_{\cdot\cdot})^2$			

`fit<-aov(breaks tension, data=dat)`. The ANOVA table can then be completed with the function `summary` that analyses the output of `aov`.

The command `summary(fit)` produces Table 7.6. The p-value under five per cent, that can be obtained also with the command `1-pf(4.059,2,24)`, suggests to reject the hypothesis H_0 of the absence of tension effects on warp breaks. In Table 7.7, the equations used by `aov` in Table 7.6 for the calculation of the sum of squares, of the mean squares and of the test function have been collected. The term "residuals" in the table is common in the linear regression models that are discussed in Chap. 11. Indeed, the SS_E value can be also obtained with a least square estimate of the unknown parameters μ and τ_i of Eq. (7.57) by setting $SS_E = \sum_{ij}(y_{ij} - \hat{\mu} - \hat{\tau}_i)^2$, since $\hat{\mu} + \hat{\tau}_i = \bar{y}_{i\cdot}$.

So far so good, but, to outline the main aspects of the procedure, we have postponed the verification of some important assumptions. Taking for granted the validity of the additivity effects of Eq. (7.57), we need to perform the following checks on the collected data:

(1) Are the errors ε_{ij} really random?
(2) Is the variance of the number of breaks really the same for each tension? In other words, are we sure that σ^2 does not depend on the group?
(3) What happens to the F-test distribution if the Gaussian assumption is violated?

Errors are a random sample: this assumption is assured by the *randomization*, that is, by the random allocation of the experimental material to each test and by the random order of the performed tests. This is to prevent factors beyond the investigator's control from systematically influencing the results. In a real-life ANOVA, it is then crucial to know how the experiment was planned and executed.

Constant variance: this assumption is fundamental to ensure that the pooled estimate of σ^2 given by MS_E is valid, independently of Eq. (7.61), which is related to the probability distribution of MS_E. In Fig. 7.8 the high tension seems less dispersed than the other two, but some tests need to be performed for an appropriate verification.

For instance, the Levene's test [Lev60] eliminates the effect of the averages in the deviations by transforming the vector of the responses in each group into a vector of absolute values of deviations with respect to the mean or the median. The use of the absolute value becomes necessary to avoid deleting deviations. Simulation studies have shown that deviations from the medians generally lead to distributions of the ratios between the means of the squares (MS) roughly distributed according to F.

The three medians of the tensions `L`, `M` and `H` can be calculated, and then a new vector `breaks1` can be created to be read by `aov`:

```
breaks1[1:9]<-abs(breaks[1:9]-median(breaks[1:9])
breaks1[10:18]<-abs(breaks[10:18]-median(breaks[10:18]))
breaks1[19:27]<-abs(breaks[19:27]-median(breaks[19:27]))
summary(aov(breaks1~tension))
```

A p-value of di 0.25 is obtained, in good agreement with the hypothesis of variance equality.

In R, the available tests are the Bartlett's test (based on the likelihood ratio, see Chap. 10) and the Levene's test. For what is written above, the latter is to be preferred in case there are doubts about the Gaussianity assumption. The function `bartlett.test(breaks tension, data=dat)` gives a p-value of 0.15; therefore, even with this test, we do not reject the hypothesis of constant variance. The function which performs the Levene's test is present in the library `car`, to be installed and loaded, as it is not included in the basic R distribution: the call `leveneTest(breaks tension, data=dat)` gives a p-value of 0.25, in agreement with that obtained from `aov` by rearranging the data.

Gaussian errors: since the errors ϵ_{ij} are not directly observed, we have to use an estimate of them, such as the residuals $y_{ij} - \bar{y}_{i\bullet}$, to verify their properties. The residuals of different groups are uncorrelated to each other, but this is not true for those in the same group since, as you can easily verify, $\langle (Y_{ij} - \bar{Y}_{i\bullet})(Y_{ik} - \bar{Y}_{i\bullet}) \rangle = -\sigma^2/n$. Therefore, the residuals are not a completely random sample; however, it is a common practice, after sorting, to standardize them with the function `rstandard(fit)`, where `fit` is the output of `aov`, and represent them in a Q-Q plot, as done in Fig. 2.7, versus the expected value of an ordered random sample from a standard Gaussian. The plot thus obtained, shown in Fig. 7.9, is called Gaussian Q-Q plot and can be drawn with the commands `qqnorm(rstandard(fit))` and `abline(0,1)`, where the second command plots the bisector. The points are arranged approximately around the straight line, so we can consider the errors as Gaussian. Also in the case of slight misalignments, we can still use the F test, because it has been shown, by different studies, that the latter tolerates violations of normality rather well.

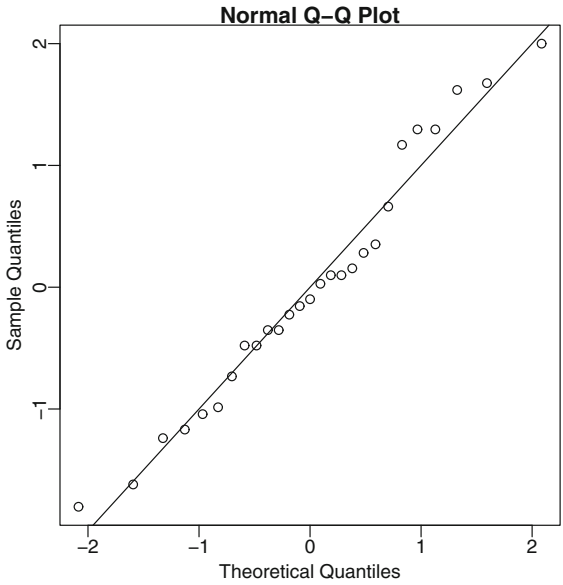


Fig. 7.9 Gaussian Q-Q plot of the ANOVA standardized residuals from the data of the textile experiment of Table 7.5

All the assumptions made are therefore justified, and we can state that we have found that the tension has an effect on the number of defects. What are the tension levels responsible for the test result? The answer to this question is of considerable practical importance, as it indicates the most suitable level of tension to minimize breakage.

From the plot of Fig. 7.8, the high tension group has the smallest mean $\bar{y}_{i\bullet}$, but as usual we have to perform a statistical test to verify whether that this feature is systematic or random. If $\mu_i = \mu + \tau_i$, the comparison between all pairs of expected values, called *Tukey test* [Tuk49] , addresses the question:

$$\begin{array}{l} H_0 : \ \mu_i = \mu_j \\ H_1 : \ \mu_i \neq \mu_j \end{array} \qquad \text{for all } i \neq j \ . \tag{7.63}$$

We have just seen in Sect. 7.7 that, in multiple tests, it would be wrong to perform $m(m-1)/2$ individual tests at the level α assuming that the probability of rejecting at least one hypothesis under H_0 is α. Tukey's criterion for checking the α_F level of the family is to choose a single critical value c for all tests in the family in order to obtain a given α_F level. It is based on the following equivalence of events:

$$\bigcup_{i \neq k} \{|\bar{y}_{i\bullet} - \bar{y}_{k\bullet}| > c\} \iff \{\bar{y}_{\max} - \bar{y}_{\min} > c\} \ , \tag{7.64}$$

where $\bar{y}_{\max}$ and $\bar{y}_{\min}$ are the minimum and maximum values of the within group sample means. Equation (7.64) means that, as is obvious, there is, in absolute values,

at least one difference between means exceeding c if and only if the difference between the maximum mean and the minimum mean exceeds c. Since $\text{Var}[\bar{Y}_{i\bullet}] = \sigma^2/n$, with the correct estimate given by MS_E/n, it is natural to consider the standardized differences divided by $\sqrt{MS_E/n}$ and reformulate the Tukey's criterion as follows:

$$\bigcup_{i \neq k} \left\{ \frac{|\bar{y}_{i\bullet} - \bar{y}_{k\bullet}|}{\sqrt{MS_E/n}} > c \right\} \iff \left\{ \frac{\bar{y}_{\max} - \bar{y}_{\min}}{\sqrt{MS_E/n}} > c \right\} . \tag{7.65}$$

Therefore, by appropriately choosing c, we are able to identify as different all the pairs of means that exceed the threshold with a significance level α_F. The choice of c is based on the p.d.f. under H_0 of the statistic:

$$q = \frac{\bar{Y}_{\max} - \bar{Y}_{\min}}{\sqrt{MS_E/n}} , \tag{7.66}$$

which is the so-called studentized range statistic, that is the distribution of the difference between the maximum and minimum of m independent Gaussian variables having the same mean, i.e. the $\bar{Y}_{i\bullet}$'s, divided by an estimate of its standard deviation obtained from the pooled variance estimate.

The percentiles of the studentized range are tabulated and indicated with $q_\alpha(m, df)$, where m is the number of groups and the degrees of freedom are those of MS_E, so $df = m(n-1)$. In R, they are calculated by the `qtukey` routine and can be easily evaluated using simulation codes as well, as in Problem 8.19. Then we can set $c = q_{\alpha_F}(m, df)$ and claim as different all the pair of means (μ_i, μ_k) for which:

$$|\bar{y}_{i\bullet} - \bar{y}_{k\bullet}| > q_{\alpha_F}(m, df)\sqrt{\frac{MS_E}{n}} , \tag{7.67}$$

with $df = m(n-1)$. From this equation it is immediate to obtain the confidence interval for the expected difference between the two means:

$$(\mu_i - \mu_j) \in (\bar{y}_{i\bullet} - \bar{y}_{j\bullet}) \pm q_{\alpha_F}(m, df)\sqrt{\frac{MS_E}{n}}, \quad \forall i \neq j . \tag{7.68}$$

The function `TukeyHSD(fit)` performs Tukey's test and computes the confidence intervals and the p-values for each difference of means at a significance level $\alpha_F = 0.95$. The result, given in Table 7.8, shows that the mean number of breaks with a high tension level is lower than those with a medium or low tension level and that there is no significant difference between the latter two. When significant, the sign of the difference indicates the most probable ordering of the means.

All the procedures so far described assume that the experimental plan is balanced, i.e. the number of trials for each factor level is constant. This generally ensures greater robustness in the case of assumption violations (such as that of constant

Table 7.8 R output of Tukey's test for the differences between the number of expected breaks for pairs of tension levels. The data are from Table 7.5

	diff	lwr	upr	p adj
L-H	9.444	−0.419	19.296	0.062
M-H	10.000	0.149	19.851	0.046
M-L	0.556	−9.296	10.407	0.989

variance) and greater test power. The formulae to use in unbalanced plans are, for instance, given in [Mon03], from which we have taken much of this section.

Before finishing this example, it is worthwhile to briefly return to the issue of constant variance. It is true that the Bartlett and Levene's tests do not reject this hypothesis, but the p-value of the Bartlett's test is not particularly high and the data graph in Fig. 7.8 indicates a lower variability of the response when the average number of breaks is lower. If the breaks in the tissue sample are randomly distributed and with a smaller and smaller probability as the examined area decreases, we are in the presence of a Poisson process in space, similar to that in time described in Sect. 3.7. This assumption is perfectly reasonable if the loom has been overhauled and tuned before the experiments. The variance of a Poisson variable with mean μ is just μ, and this could explain the lower dispersion at the high tension level. Now we wonder: do we have to redesign an inference method for the Poisson variable from scratch, or can we continue with the one for the Gaussian variable? Using the correct distribution is in principle better, but often it is enough to exploit a standard method and get the needed answers without perfectly describing the process generating the data. In this case, we look for a data transformation that stabilizes the variance, such as y^λ, a solution often adopted. To choose λ, we observe that, using Poisson's hypothesis and by Eq. (5.58):

$$\mathrm{Var}[Y^\lambda] \simeq \left(\frac{\mathrm{d}Y^\lambda}{\mathrm{d}Y}\right)^2 \mathrm{Var}[Y] = (\lambda\mu^{\lambda-1})^2\mu = \lambda^2\mu^{2\lambda-1}\,, \tag{7.69}$$

so that $\lambda = 1/2$ results in a constant group variance. You can redo the data plot, the ANOVA table, the constant variance tests and the residual analysis with the transformed data, and you will find that the F-test has a slightly smaller p-value. The tests on the constant variance have significantly higher p-values, and the comparisons between means with Tukey's test are slightly more significant, confirming the conclusions already reached.

From this digression we have learned that:

- If we can transform the data to bring us back to a standard procedure, we save time and better focus on the fundamental aspects of the problem.
- The chosen transformation depends on the assumptions about the procedure used to generate the data. If we are unable to make reasonable assumptions and we find the correct transformation by trial and error, we will have a perfect analysis,

but we will then have to explain if and how the same conclusions hold for untransformed data.
- When the assumption violation is slight, the fundamental conclusions continue to hold even without applying transformations.

Before dealing with the two-way ANOVA, let us now briefly address the question of choosing the sample size on the basis of the power of the F-test for the one-way ANOVA (we have already introduced the power in Sect. 7.7).

In practice, we would like to know how many data we need to collect to reject H_0, i.e. what is the optimum number of trials n for each level of the factor, assuming the effect has a given value τ_i. The answer comes from the calculation of the power of the test: n must be the smallest value that, under H_1, allows us to reject H_0 with the desired power, for example, 95%, when a test is performed at the assigned level α. The use of this criterion leads to an extremely important consequence, because if we perform too few tests we could accept H_0 even if the effect does exist, missing the discovery.

Therefore, if we require a power $\beta \geq 95\%$, we must calculate:

$$\beta(\tau_1, \ldots, \tau_m) = P\left\{\frac{MS_{Tr}}{MS_E} > F_{1-\alpha}(m-1, mn-m); \tau_1, \ldots, \tau_m\right\}, \tag{7.70}$$

and choose the smallest n value such that $\beta(\tau_1, \ldots, \tau_m) > 0.95$. The second member of Eq. (7.70) shows that the probability to reject H_0 is evaluated from the assigned τ_i values.

The computation of $\beta(\tau_1, \ldots, \tau_m)$ depends not only on τ_i, m and n but also on σ, as can be derived from the considerations on the non-central χ^2_R value following Eq. (7.62). Under H_1, the distribution of MS_{Tr}/MS_E is in fact a non-central F density[3] with expected value:

$$\frac{(mn-m)(m-1+n\sum_i \tau_i^2/\sigma^2)}{(m-1)(mn-m-2)}. \tag{7.71}$$

This value turns into the expected value (5.47) of F when all the τ_i values are zero. From this equation, one deduces that the F distribution depends on the quantities τ_i through $\sum_i \tau_i^2$ or, equivalently, through $\sum_i (\mu_i - \mu)^2/(m-1) = \sum_i \tau_i^2/(m-1)$. This last quantity is a sort of "variance" of the mean effects of the factor levels. This last parametrization is used in the R function `power.anova.test` for the calculation of the power function.

Let us imagine that before carrying out the weaving machines experiment, the effect of the tension would be considered satisfactory if, by varying it, there is a reduction of at least ten breaks. This means that, for example, the power has to be evaluated for $\mu_1 = 35$, $\mu_2 = 25$ and $\mu_3 = 15$. We also assume to have an a priori

[3] The non-central F is the distribution of the ratio of two independent χ^2 random variables, where the χ^2 in the numerator is non-central.

information that σ^2 could be about 60. We consider $m = 3$, the number of levels, as a fixed parameter. With $\alpha = 0.05$ and $n = 9$, the routine `power.anova.test` gives the following output:

```
power.anova.test(groups=3, n=9, between.var=var(c(15,25,35)),
        within.var=60, sig.level=0.05, power=NULL)

     Balanced one-way analysis of variance power calculation

groups = 3
n = 9
between.var = 100
within.var = 60
sig.level = 0.05
power = 0.9976002
```

where we see that the power of the ANOVA test for the chosen values of τ_i, σ^2, n and α is of 99.8%. With this function it is also possible to obtain the value of an argument keeping the others fixed by assigning to it the `NULL` value. For example, one can thus obtain the value of n providing the desired power. Without any precise information on σ, it is possible to conservatively evaluate n for a range of σ values, choosing the value of n corresponding to the maximum value of σ, compatible with the experiment. Using values of σ in $\{5, 6, \ldots, 15\}$, we execute the loop (note the n request through `power.anova.test (...)$n`):

```
for (i in 5:15)
   print(power.anova.test(groups=3, n=NULL, between.var=100,
   within.var=i^2, sig.level=0.05, power=0.95)$n)
```

and interactively obtain the following values of n: `3.2, 4.0, 5.0, 6.1, 7.4, 8.8, 10.4, 12.2, 14.1, 16.2, 18.4`. This example shows that a high n value could be necessary to have a good test power if the experimental variability is high. In particular, with $n = 9$ and $\sigma = 15$, we would get a power of 66%, far from the 95% target.

7.10 Two-Way ANOVA

Now let us move on to the two-way ANOVA. Previously, we stated that experiments are often scheduled to modify more than one factor and also our textile example is not an exception. In fact, the data in Table 7.5 are for a single type of wool, type B, but the same tests with the three tension levels were also repeated with a second type of wool, type A. The `warpbreaks` object in the base distribution of R contains the complete data, as a *data frame*, as also shown by the output of the command `class(warpbreak)`.

The total number of tests is then 54, and the experimental plan is balanced, as displayed in the output of the function `xtabs( wool+tension, data = warpbreaks)` that is given in Table 7.9. This indicates that

Table 7.9 Table of the experimental plan that produced the object of the R library `warpbreaks`: there are nine replicates for each combination of wool and tension

	Tension		
Wool	L	M	H
A	9	9	9
B	9	9	9

a factorial plan has been adopted, that is, all possible combinations of the levels of the two factors were considered, and the whole plan was replicated nine times. The file columns can be extracted with the commands `warpbreaks$tension`, `warpbreaks$wool`, `warpbreaks$breaks`; this last vector contains all the numerical data for the number of breaks. Reasoning in a similar way as we did for Eq. (7.56), we separate the effects of the factors from the statistical ones in order to obtain an algebraic identity, in which we try to identify the general mean, the difference that takes into account the effect of the level of the first factor, that of the second and the statistical fluctuation. This time we have to use three indices: i for the wool, j for the tension and k for each test performed on the level combinations ij. Associating to (A, B) the indices $(1, 2)$ and to (L, M, H) the indices $(1, 2, 3)$, we can write, tentatively:

$$y_{ijk} \stackrel{?}{=} \bar{y}_{\cdot\cdot\cdot} + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}) + (y_{ijk} - \bar{y}_{ij\cdot}) \,, \tag{7.72}$$

but we can easily realize that the identity is not valid and it has to be modified as:

$$y_{ijk} = \bar{y}_{\cdot\cdot\cdot} + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}) + (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot}) + (y_{ijk} - \bar{y}_{ij\cdot}) \,. \tag{7.73}$$

The last formula indicates whether there is a difference between applying the j level of tension when the wool is at the i level and between applying the j tension level by averaging the response of the wool levels. This is an *interaction* effect between the two factors: if the effect of the tension is the same for each type of wool, then its value at the wool i level is equal to the average value and the interaction is absent (apart from statistical fluctuations). With the command `interaction.plot(x.factor=tension, trace.factor=wool, response=breaks, type="b")`, we obtain the trend of the average of the responses to the variation of the tension, for each wool type, that is displayed in Fig. 7.10. We can easily notice that, with type A wool, the effect of the increase in tension is already present at the medium level, while with type B wool, it shows up only at the high level (as confirmed by the Tukey's test performed earlier). In the case of no interaction, the dashed and solid lines would be roughly parallel. As

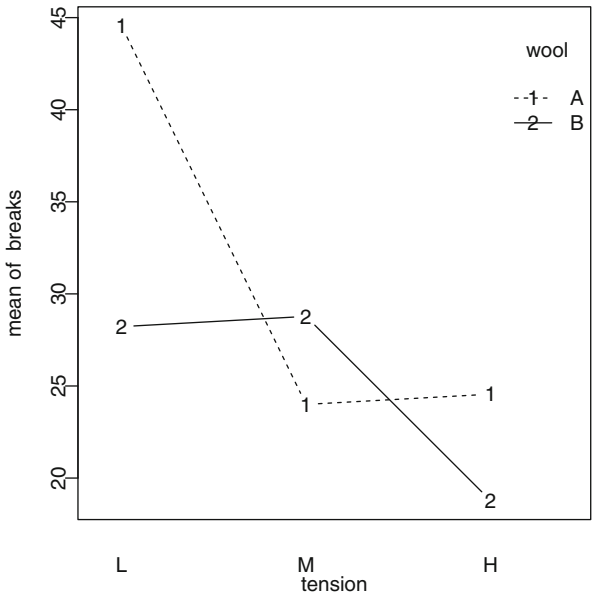


Fig. 7.10 Interaction plot between tension and wool type obtained from the R `warpbreaks` dataset

in the one-way ANOVA, we establish a hypothesis test on the effect of the factors starting from a population model and from Eq. (7.73):

$$Y_{ijk} = \mu + \beta_i + \tau_j + (\beta\tau)_{ij} + \varepsilon_{ijk} \quad \begin{cases} i = 1, \ldots, a \\ j = 1, \ldots, b \\ k = 1, \ldots, n \end{cases} \tag{7.74}$$

where ε_{ijk} are iid $N(0, \sigma^2)$. The notation $(\beta\tau)$ is not the product between β and τ but indicates the interaction parameter between the two factors. In addition to the constraints for the mean effect of the factors, $\sum_i \beta_i = 0$ and $\sum_j \tau_j = 0$, we add those for the interaction parameters, $\sum_i (\beta\tau)_{ij} = \sum_j (\beta\tau)_{ij} = 0$, as suggested by the identities $\sum_i (\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}) = 0$, for $j = 1, \ldots, b$, and $\sum_j (\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}) = 0$, for $i = 1, \ldots, a$. The hypotheses to be verified are therefore whether there is an effect of the first factor (the type of wool), of the second factor (the tension) and of a possible interaction:

$$H_0 : \beta_1 = \cdots = \beta_a = 0 \ \text{ against } \ H_1 : \exists i \text{ such as } \beta_i \neq 0 \tag{7.75}$$

$$H_0 : \tau_1 = \cdots = \tau_b = 0 \ \text{ against } \ H_1 : \exists j \text{ such as } \tau_j \neq 0 \tag{7.76}$$

$$H_0 : (\beta\tau)_{ij} = 0, \ \forall (i, j) \ \text{ against } \ H_1 : \exists (i, j) \text{ such that } (\beta\tau)_{ij} \neq 0 \tag{7.77}$$

As we did in Eq. (7.59), after some tedious algebraic calculations, we obtain the decomposition of the total sum of squares of the SS_T response by separating the

components due to the different sources of variation, with which we can perform the tests of our interest:

$$\begin{aligned}\sum_{ijk}(y_{ijk}-\bar{y}_{\cdot\cdot\cdot})^2 = \; & bn\sum_i(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot\cdot\cdot})^2 + an\sum_j(\bar{y}_{\cdot j\cdot}-\bar{y}_{\cdot\cdot\cdot})^2 + \\ & n\sum_{ij}(\bar{y}_{ij\cdot}-\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot j\cdot}+\bar{y}_{\cdot\cdot\cdot})^2 + \sum_{ijk}(y_{ijk}-\bar{y}_{ij\cdot})^2 \\ = \; & SS_{Tr1}+SS_{Tr2}+SS_{int}+SS_E \; .\end{aligned} \tag{7.78}$$

The useful property of this experimental plan is that we can verify each of the three null hypotheses of Eq. (7.75) independently of the others. Let us take, for example, SS_{Tr1}: substituting y_{ijk} with the right term of Eq. (7.74) and using the zero sum constraints of the parameters, we have:

$$\bar{y}_{i\cdot\cdot} = \frac{1}{bn}\sum_{jk}(\mu+\beta_i+\tau_j+(\beta\tau)_{ij}+\varepsilon_{ijk}) = \frac{1}{bn}\sum_{jk}(\mu+\beta_i+\varepsilon_{ijk}) \; . \tag{7.79}$$

Therefore, $\bar{Y}_{i\cdot\cdot} \sim N(\mu+\beta_i, \sigma^2/bn)$, independently of the presence of the other effects. Then, taking into account that $\bar{y}_{\cdot\cdot\cdot}$ is the mean of $\bar{y}_{i\cdot\cdot}$, one has:

$$MS_{Tr1} = \frac{SS_{Tr1}}{a-1} = \frac{bn}{a-1}\sum_i(\bar{y}_{i\cdot\cdot}-\bar{y}_{\cdot\cdot\cdot})^2 \sim \sigma^2\chi^2_R(a-1) \; , \tag{7.80}$$

under the hypothesis (7.75). With a similar reasoning, we can derive, under H_0, the distributions of the other means of the squares related to the effects of the factors and, as in the one-way ANOVA, verify that the distribution of MS_E is always the same under any of the assumptions (7.75–7.77), that is, $MS_E \sim \sigma^2\chi^2_R(ab(n-1))$. All the obtained results are collected in Table 7.10. Before calculating the ANOVA table with the data of `warpbreaks`, let us verify the hypotheses of variance equality and of Gaussian errors. The groups of observations that should have the

Table 7.10 General table for the two-way ANOVA analysis. Under the H_0 hypotheses (7.75–7.77), MS_{Tr1}, MS_{Tr2} and MS_{int} are distributed as $F(df)$. MS_E is always distributed as $F(df)$, independently of any hypothesis. The test p-value on each factor or on the interaction is the probability that $F(df)$ is greater than the observed F value

Source of variation	df	Sum of squares	Mean of squares	F
Treatment 1	$a-1$	SS_{Tr1}	$MS_{Tr1} = \frac{SS_{Tr1}}{a-1}$	$\frac{MS_{Tr1}}{MS_E}$
Treatment 2	$b-1$	SS_{Tr2}	$MS_{Tr2} = \frac{SS_{Tr2}}{b-1}$	$\frac{MS_{Tr2}}{MS_E}$
Interaction	$(a-1)(b-1)$	SS_{int}	$MS_{int} = \frac{SS_{int}}{(a-1)(b-1)}$	$\frac{MS_{int}}{MS_E}$
Residuals	$ab(n-1)$	SS_E	$MS_E = \frac{SS_E}{ab(n-1)}$	
Total	$abn-1$	SS_T		

Table 7.11 ANOVA table: check of Eqs. (7.75–7.77) with the data of the R library `warpbreaks`

	Df	Sum Sq	Mean Sq	F value	Pr(> F)
Wool	1	2.90	2.902	3.022	0.088542
Tension	2	15.89	7.946	8.275	0.000817
Tension:wool	2	7.20	3.601	3.750	0.030674
Residuals	48	46.09	0.960		

same variance are in this case the six groups identified by the combinations of the two factors as in Table 7.9. To use the routine `bartlett.test` with two classification criteria, we create a factor indicating group membership and then proceed to the test:

```
# character vector with the pairs AL, AM, AH,...
combo = paste(warpbreaks$wool,warpbreaks$tension, sep="")
#  Bartlett's test with the groups identified by combo
bartlett.test(warpbreaks$breaks,combo)
```

obtaining a p-value of 0.023. The tests with type A wool have a different dispersion than those with type B wool. By using again the square root transformation for the break numbers, the Bartlett's test instead has a p-value of 0.289, so let us continue the analysis on the transformed data: $y_{ijk} \equiv \sqrt{y_{ijk}}$. In the call to the `aov` function, we must give the information that we are considering two factors and their interaction, so we will execute `fit = aov(sqrt(breaks) wool*tension,data=warpbreaks)`, where `*` indicates that the potential interaction needs to be taken into account. The Q-Q plot obtained as before with `qqnorm(rstandard(fit))` is in Fig. 7.11 and indicates that the residuals have a distribution with tails slightly different than the Gaussian, but, since the violation is not gross, we continue with the analysis. The ANOVA table obtained with `summary(fit)` is reproduced in Table 7.11. From these data we conclude that the tension effect is highly significant, whereas those of the wool type and of the interaction are marginally significant. Also in this case, we can run the Tukey's test to verify which pairs of tension levels are responsible for the significance. The results obtained with `TukeyHSD (fit,which="tension")` are shown in Table 7.12. If we disregard the type of wool, we conclude that, according to the p-values of Table 7.12, we can be absolutely certain of a significant effect of the tension when passing from L to H, while the different behaviour between two types of wool in passing from L to M and from M to H determines respectively a marginally significant and an insignificant p-value (see again Fig. 7.10 for a cross-check).

We conclude this section on ANOVA with a remark on MS_E. From the comparison of Eq. (7.59) with Eq. (7.78), we can notice that SS_{Tr} and SS_{Tr2} are both constructed using the sum of squares of the deviations of the mean responses for each tension level with respect to the general mean. MS_E is instead evaluated in a different way and, when systematic sources of variation are present and taken into account by the interaction factor, collects a smaller residual variation and therefore produces

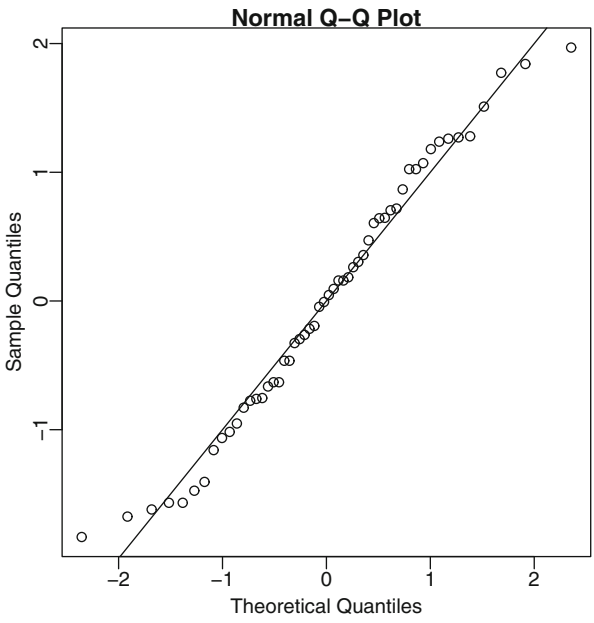


Fig. 7.11 Gaussian Q-Q plot of the standardized residuals of the two-way ANOVA using the R library data `warpbreaks`

Table 7.12 Tukey's test for the difference between expected number of breaks for pairs of tension levels, with the model (7.74) applied to the R library data `warbpreaks`

	diff	lwr	upr	p adj
M-L	−0.830561	−1.620515	−0.04060713	0.0373236
H-L	−1.313572	−2.103526	−0.52361812	0.0005874
H-M	−0.483011	−1.272965	0.30694285	0.3099954

Table 7.13 ANOVA table for the check of Eqs. (7.75–7.76) without interaction effect with the R library data `warpbreaks`

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Wool	1	2.90	2.902	2.273	0.10520
Tension	2	15.89	7.946	7.455	0.00147
Residuals	50	53.29	1.066		

a lower estimate of σ^2. We can check this property performing a fit without interaction, using the sign + instead of `*`, with `fit=aov(sqrt(breaks) wool + tension, data=warpbreaks)` and `summary(anova(fit))`. The result is reported in Table 7.13, where the obtained values coincide with the ones of Table 7.11 except for those depending on the MS_E value.

7.11 Problems

7.1 Two machines produce steel shafts. The average diameter values of two samples with ten shafts each are $\mu_1 \in 5.36 \pm 0.05$ and $\mu_2 \in 5.21 \pm 0.05$ mm. Apply Student's test to verify the homogeneity of the pieces produced by the two machines.

7.2 One group of 70 sick people was given a drug, and another group of 58 sick people assumed simple sugar (placebo). Evaluate whether the drug is effective based on the following contingency table:

	*D*rug	*P*lacebo	Total
Healed	40	28	68
Sick	30	30	60
Total	70	58	128

7.3 The number X of buses arriving at a toll booth in 5-min intervals has been counted for $N = 100$ times. The result of this measurement is given in the following histogram, with the discrete X spectrum divided into 11 channels (from 0 to 10):

No. of busses x_i	0	1	2	3	4	5	6	7	8	9	10
No. of trials n_i	1	5	6	19	20	17	15	8	8	1	0

Perform a two-tailed test to check whether the data are consistent with a Poisson process.

7.4 According to a genetic model, a certain tree should provide a green to yellow pea ratio of 3:1. In a sample of 500, 356 green peas were found. Say whether the model can be accepted at the 5% level.

7.5 A series of 600 rolls with 2 different dice gave the following contingency table for the 6 sides:

	1	2	3	4	5	6	TOTAL
DIE 1	101	105	103	95	90	106	600
DIE 2	99	105	98	103	96	99	600
TOTAL	200	210	201	198	186	205	1200

Check whether (a) the two series of rolls are compatible with each other and (b) dice and rolls are fair.

7.6 After the introduction of a new highway speed limit, car accidents during weekends decreased from 60 to 33. What is the probability to be wrong when affirming that the decrease is due to the new limit?

7.7 A set of cosmic ray detectors collected the following counts in the same time interval:

Detector	1	2	3	4	5	6
Counts	29	19	18	25	17	10

Check whether the flux on the counters is homogeneous at a 5% test level.

7.8 A measurement of the emission rates of two radioactive sources A and B gave the following result:

$$A = 240 \text{ counts in } 10 \text{ s}$$

$$B = 670 \text{ counts in } 10 \text{ s}$$

One experimenter claims to have obtained the result

$$C = 10\,500 \text{ counts in } 100 \text{ s}$$

in a new measurement with the A and B sources combined. Evaluate if this statement is correct.

7.9 The following 47 split times refer to the working intervals observed between consecutive failures of an electronic device:

Interval	$0-120h$	$120-240h$	$240-360h$	$360-480h$
Frequency	22	12	7	6

Verify whether the data support the hypothesis of an exponential law with $\lambda = 0.005\ h^{-1}$ at a test level of 1%.

7.10 During a counting experiment, 1000 arrival times are recorded over fixed time intervals. The final result is:

Interval	$[0-1]s$	$[1-2]s$	$[2-4]s$	$>4s$
Trials	368	266	217	149

Verify whether the data are in agreement with a true frequency of 0.5 s^{-1}.

7.11 Four measurements of a given physical quantity give the following result:

$$1.12 \quad 1.13 \quad 1.10 \quad 1.09\,.$$

Verify the compatibility among the measurements assuming that they come from normal densities with the same variance $\sigma^2 = 4\ 10^{-4}$.

7.12 The law sets a limit on the concentration of a certain air pollutant to 55 parts per million (ppm). A series of ten repeated measurements gave an average value of 58 ppm. Knowing that the uncertainty of a single measurement (standard deviation) is ± 10 ppm, check if the data exceeds the allowable limits at a test level of 5%.

7.13 An experiment gives the following result:
13 values < -1, 25 values in $[-1, 0]$, 44 values in $(0, 1]$, 16 values > 1.
Check whether the data are in agreement with the standard normal $N(0, 1)$.

7.14 In six specimen of a 100 meter long high voltage cable

$$18, 14, 10, 10, 21, 17,$$

defective points of insulation have been detected.
Evaluate whether these data can be considered in agreement with a standard of less than 15 defects per 100 meters.

7.15 After the administration of a drug, 15 parameters related to the health of the tested subjects were followed and compared with a control group. The 15 p-values from the t-tests are, in ascending order, as follows [BH95]:

$$0.0001,\ 0.0004,\ 0.0019,\ 0.0095,\ 0.0201,\ 0.0278,\ 0.0298,\ 0.0344$$
$$0.0459,\ 0.3240,\ 0.4262,\ 0.5719,\ 0.6528,\ 0.7590,\ 1.000\,.$$

Use the routine `MultiTest` to find which parameters are sensitive to the drug, at a test level of 5% and 1%.

7.16 Perform the t-test on the data on the breaks in the warp depending on the tension of the loom given in Table 7.5.

7.17 Calculate the p-values of Table 7.8 with the R routine `ptukey`.

7.18 Two creams A and B and a placebo P were given to 25 patients with blisters. The days needed for the healing are shown in the table:

*T*reatment	*D*ays								
A	5	6	6	7	7	8	9	10	
B	7	7	8	8	9	10	10	11	
P	7	9	9	9	10	10	11	12	13

Make the ANOVA analysis of the data.

Chapter 8
Monte Carlo Methods

It is the powerful development and intensive use of the simulative function that, in my view, characterizes the unique properties of man's brain. And this is at the most basic level of the cognitive functions, those on which language rests and which it probably reveals only incompletely.

Jacques Monod, "CHANCE AND NECESSITY: ESSAY ON THE NATURAL PHILOSOPHY OF MODERN BIOLOGY".

8.1 Introduction

The term "Monte Carlo methods" or "MC methods" generally refers to all those techniques that make use of artificial (i.e. computer generated) random variables to solve mathematical problems using random samples drawn from the corresponding populations.

Undoubtedly, this is not a very efficient way to obtain the solution of a problem, as the (often) time-consuming simulated sampling procedure gives a result that is always affected by the statistical error. In practice, however, we are often faced with situations in which it is too difficult, if not impossible, to use the standard numerical or analytical procedures, and in all these cases, Monte Carlo methods become the only available alternative.

The application of these methods is not limited only to purely statistical problems, as it might seem from the use of probability distributions, but includes all those cases in which a connection can be found between the problem under consideration and the behaviour of a certain random system: for example, the value of a definite integral, which is certainly not a random quantity, can also be calculated using random numbers.

The theoretical foundations of the Monte Carlo methods (or simulation methods) have been known for a long time, and the first example of the use of random numbers for the resolution of definite integrals is even found in a book (*Essai d'aritmetique moral*) written in 1777 by Georges-Louis Leclerc, Comte de Buffon,

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2 139, https://doi.org/10.1007/978-3-031-09429-3_8

French mathematician and naturalist, in which a procedure for the approximate calculation of the π value of is outlined (see Problem 8.13).

For over a century and a half, however, it was used only sporadically and above all for didactic purposes. Its first systematic application took place only in the first half of the 1940s in Los Alamos, by the team of scientists, led by Enrico Fermi, who developed the project of the first atomic bombs. In this same period, the term "Monte Carlo" was also born, which obviously refers to the town famous for its casino and, more precisely, to roulette, one of the simplest mechanical devices that can be used to generate random variables.

The authorship of the name is in particular attributed to the mathematicians J. von Neumann and S. Ulam, who adopted it as the code name of their secret research, conducted using random numbers, on the processes of diffusion and absorption of neutrons in fissile materials (for more historical information, see [HH64]).

After 1950, these methods passed in a few years from the role of mathematical curiosity to that of an indispensable tool for scientific research thanks to the advent of the computers. This has happened not only because computers provide a rapid execution of the long calculations that are often necessary to obtain a meaningful result but also, as we will see later, because they can easily generate random numbers. Currently, there are applications in many different research fields, from nuclear physics to chemistry, from statistical and quantum mechanics to economics.

In this chapter we give a description of the fundamental principles of these methods and the most important technical details necessary to create Monte Carlo codes for the solution of statistical problems. Other significant applications are explained later in Chap. 9.

8.2 What Is Monte Carlo?

Any computer library system includes service routines, generating numbers uniformly distributed in $[0, 1]$, which we can consider random. As we will discuss shortly, this particular distribution is needed to perform any simulation. The realization of this type of routine, which may seem simple on the surface, actually constitutes a complex mathematical problem, as can be verified by consulting references [BS91, Cha75, Com91, FLJW92, Jam90, MNZ90, PFTW92], whose description goes beyond the scope of this book. In R, as we have already mentioned, the uniform generator is the routine:

```
x = runif(n,min=0, max=1)
```

that, by default, produces `n` values in $[0, 1]$. Numbers within any range can be generated by changing the values of `min` and `max`. For example, to generate a vector of 2000 numbers and make their histogram, just write:

```
> hist(runif(2000))  ,
```

or you can use our `HistoBar` routine, which offers some options, such as the calculation of error bars, and other possibilities that you can find commented and described in the code:

```
> HistoBar(runif(2000),grid=TRUE,errors='ON')  .
```

To draw curves, one can also use the `plot` function, and the `density` routine, which is described in Appendix B:

```
> plot(density(runif(2000),adj=0.01)) ,
```

where the degree of smoothing is tuned by the `adj` parameter.

The R library provides routines to generate random numbers extracted from all distributions of current use. As explained in Appendix B, the prefix `r` must be used before the name of the required density. All the routines that generate random numbers must be initialized with one or more integer numbers, called seeds. If seeds are not changed, at each new call of the program, the results (even if random) will always be the same. In other words, the same seed will *always generate the same sample*. In R, the seeds are automatically renewed at each simulation, taking the current value provided by the system software. If you want to fix the seed, which can sometimes be useful in the testing phase of a code, the instruction to be placed at the beginning of your program is the following:

```
> set.seed(123432) ,
```

where the seed `123432` is arbitrary and must be changed when a different sequence is requested. One can easily verify that by typing:

```
> set.seed(13567); runif(10)
  > set.seed(13567); runif(10)  ,
```

exactly the same sequence of ten random numbers is repeated twice. From now on:

- With the symbol $\boldsymbol{\xi}$ (with or without indexes), we will always denote a uniform random variable in the interval [0, 1].
- With ξ (with or without indexes), we will denote the values assumed by $\boldsymbol{\xi} \sim U(0, 1)$ in one or more specific trials. In practice, these are the uniform variates or numerical values supplied by the `runif` routine.

To simulate the simplest experiment, the flip of a coin, then just divide the unit interval in half and, for each generated number ξ, define the event as "head" when $0 \leq \xi \leq 0.5$ and as "cross" when $0.5 < \xi \leq 1$ (of course, the inverse convention would work just as well). The R lines, contained into our routine `MCcoin`, are:

```
# simulation of nsim tosses of 10 coins
heads <- seq(0,0,length.out=nsim)
for( j in 1:nsim}{
   x <- runif(10)
   heads[j] = length(which(x<0.5))
}
HistoBar(heads,nbins=11,minx=0,maxx=11,errors='ON')
```

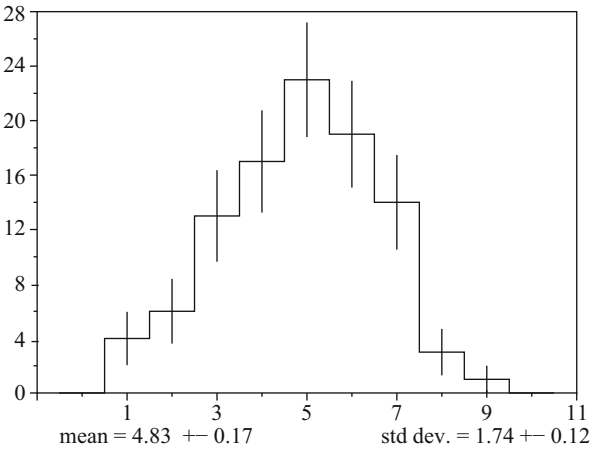


Fig. 8.1 Number of heads obtained in 100 simulated flips of 10 coins. Compare the result with that of the real experiment in Fig. 7.6

Notice the use of the R function `which`, that extracts a vector containing the positions of the values < 0.5 from the vector `x`; the length of this vector is the number of heads obtained, which is stored in `heads`. At the end of the loop, if we denote by n_i the contents of a generic bin of the histogram, the ratio n_i/N gives us, for each bin, the estimate of the probability. The histogram obtained from `HistoBar` routine is shown in Fig. 8.1. It should be noted that by simulating the ten coin flipping, we do nothing but generate a random variable from the binomial distribution (2.29), and, for $N \to \infty$, the frequency histogram we obtain tends to be just $b(x; 10, 0.5)$. You can easily verify this, and obtain again the same histogram of Fig. 8.1, with the instructions:

```
> x <- rbinom(100,size=10,prob=0.5)
> HistoBar(x,min=0,maxx=10,nbins=11,errors='ON')
```

where the R routine `rbinom` has been used to generate binomial variables. Since, for obvious reasons, we can repeat this algorithm only for a finite number of times, the simulation gives frequencies affected by a certain error, as evidenced by the error bars displayed in Fig. 8.1 and calculated with the procedure of Sect. 6.14. Of course the accuracy of our estimate increases with N, but the improvement we get is not very high: the histogram error bars, calculated with Eq. (6.106), show that, by quadrupling the number of events, we only halve the error and, as we will demonstrate shortly, this trend is not related to this particular example but is a general characteristic of all MC calculations.

In the next section, we will give a theoretical justification of MC methods, rewriting the rather qualitative indications we have developed so far in a more general and mathematically correct way.

8.3 Mathematical Aspects

We can view a variable T associated with any stochastic phenomenon as a function of k random variables $(X_1, X_2, \ldots, X_k)$:

$$T = f(X_1, X_2, \ldots, X_k) \; . \tag{8.1}$$

As we explained in the previous chapters, the whole process is, ultimately, characterized by the T mean value and dispersion parameters; the latter are in turn expressible, on the basis of Eq. (2.67), as the difference of mean values. The phenomenon under consideration can therefore always be described on the basis of the expected values (2.68), through the solution of one or more integrals (or sums) of the type:

$$I = \langle T \rangle = \int_D f(x_1, x_2, \ldots, x_k) \, p(x_1, x_2, \ldots, x_k) \, \mathrm{d}x_1 \ldots \mathrm{d}x_k \; , \tag{8.2}$$

where $p(x_1, x_2, \ldots, x_k)$ is the p.d.f. of the variables $(X_1, X_2, \ldots, X_k)$, defined in $D \in \mathbb{R}^k$ and with:

$$\int_D p(x_1, x_2, \ldots, x_k) \, \mathrm{d}x_1 \ldots \mathrm{d}x_k = 1 \; . \tag{8.3}$$

From a strictly formal point of view, all MC calculations, which are simulations of stochastic processes in which events are randomly generated according to specified probability distributions, are equivalent to the approximation of the value of a definite integral or sum.

If we generate N values of the random variable T, using several independent sets of random numbers *extracted from the density* p, we know, from the properties of the mean of a sample (see Sects. 2.11 and 6.9), that the quantity:

$$T_N = \frac{1}{N} \sum_{i=1}^{N} f(x_{1i}, x_{2i}, \ldots, x_{ki}) \; , \tag{8.4}$$

is a correct and unbiased estimate of I.

If we assume that the distribution of the random variable T, as is almost always verified in practice, has a finite variance σ_T^2, by applying the Chebyshev's inequality (3.92) and the formula (6.49) for the variance of the mean, we obtain the relation:

$$P\left\{ |T_N - I| \leq \frac{\sigma_T}{\sqrt{N}} \epsilon \right\} \geq 1 - \frac{1}{\epsilon^2} \qquad \epsilon \geq 1 \; . \tag{8.5}$$

Another very useful property (always assuming a finite σ_T^2) is provided by the Central Limit Theorem 3.1, according to which the p.d.f. of T_N tends "asymptotically" to a Gaussian with mean I and standard deviation $\sigma_T/\sqrt{N}$. As we have already noticed, the asymptotic requirement is actually satisfied for N low enough ($N \geq 10$); so we can almost always write that:

$$P\left\{|T_N - I| \leq \frac{3\sigma_T}{\sqrt{N}}\right\} \simeq 0.997\,. \tag{8.6}$$

Equations (8.5, 8.6) show that the simulated values converge in probability towards the quantity to estimate I since their variance is σ_T^2/N. It therefore turns out that (except in the cases in which, as we will mention later, it is possible to "manipulate" T by reducing its variance) the only way to increase the precision of the T_N estimate is to increase the number N of simulated events. This slow convergence of MC estimators can require considerable computational time in the simulation of very complex systems.

8.4 Generation of Discrete Random Variables

If the density function of the discrete variable to be generated is known, then, bearing in mind that:

- It is possible to construct, with Eq. (2.27), the cumulative function.
- Equations (3.87, 3.89) hold.
- The `runif` routine is available.

it is immediate to realize that the most efficient method for simulating discrete variables consists *in extracting a random variate* $0 \leq \xi \leq 1$, *considering it as a cumulative variable and determining the quantile value corresponding to the extracted cumulative value*.

Indeed, let us consider a segment of unit length, divide it into k intervals and then assign to each of them a length p_i equal to the probability for the corresponding event $\{X = x_i\}$ to occur (see Fig. 8.2). Since the probability for a uniform variate $0 \leq \xi \leq 1$ to fall within a particular interval is exactly equal to the length of that

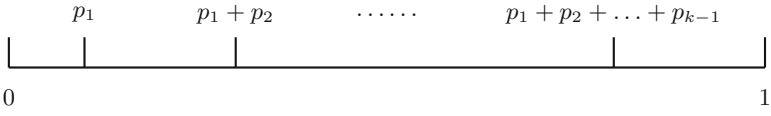


Fig. 8.2 Generation of discrete random variables. The unit interval is divided into k segments of length $p_1, p_2, \ldots, p_k$, and the subinterval where a random number ξ is located identifies the generated event

interval:

$$P\{0 \leq \boldsymbol{\xi} \leq p_1\} = p_1 \,, \tag{8.7}$$

$$P\{p_1 < \boldsymbol{\xi} \leq p_1 + p_2\} = p_2 \,, \tag{8.8}$$

$$\ldots\ldots\ldots \,,$$

$$P\{p_1 + p_2 + \ldots + p_{k-1} < \boldsymbol{\xi} \leq 1\} = p_k \,, \tag{8.9}$$

the value x_j corresponding to the upper extreme $p_1 + p_2 + \ldots + p_j$ of the interval which contains the random number ξ from the `runif` routine must be considered as extracted. The use of the cumulative *is the basic method common to all MC simulations*. This procedure can be summarized as follows:

Algorithm 8.1 (Generation of Discrete Variables) *To generate a discrete random variable X, which can take a finite set of spectral values $x_1, x_2, \ldots, x_k$ with probabilities $p_1, p_2, \ldots, p_k$, is necessary:*

- *To evaluate the cumulative function F_j:*

$$F_j = \sum_{i=1}^{j} p_i \qquad (j = 1, 2, \ldots k) \tag{8.10}$$

- *To generate a random variate* $0 \leq \xi \leq 1$
- *To determine the index* j $(1 \leq j \leq k)$ *satisfying the inequality:*

$$F_{j-1} < \xi \leq F_j \tag{8.11}$$

  (when $j = 1$, one defines $F_{j-1} = 0$)
- *To set* $\{X = x_j\}$

This algorithm, based on inequality (8.11), requires that the vector containing the cumulative data has zero in the first position and one in the last. Therefore, if the discrete values of the spectrum form a vector of dimension k, as in Eq. (8.10), the vector of the cumulative must have $k + 1$ values.

To minimize the computational time needed to solve Eq. (8.11), it is necessary to have an efficient routine to find the index $j - 1$ corresponding to the value extracted from `runif(1)`, that is, a search algorithm on the intervals "closed on the left". Obviously, the least time-efficient method is the sequential search, which is never used by the routines present in statistical software.

The available algorithms mostly belong to the so-called binary (or dichotomic) search family on which there is a very large literature [KR88, PFTW92]. The computation starts from the subinterval corresponding to the central index of the vector; if the target index is greater (smaller), the first (second) half of the vector F is eliminated, and the search continues on the centre of the remaining half, again starting from the middle index.

The number of comparisons needed to find this index is equal to the minimum integer m verifying the inequality $2^m \geq k$ (if $k = 1000$, for example, $m = 10$). Therefore, the search time goes as $t \simeq k/2$ in the case of the sequential search and as $t \simeq \ln(k-2)$ for the dichotomic one.

In R, it is possible to use different search strategies to implement Eq. (8.11). Since this can be an important aspect for all those who perform simulations, let us analyse the performance of the routines that scan a vector `x` *sorted in ascending order* to determine the interval containing a `y` value: `findInterval(y,x)` and `max(0, which(y>x))`. The former is perhaps the most used R routine for dichotomic searches; the second is a possible workaround: `which` finds all positions of the vector `x` that have values $< y$ and then `max` selects the maximum of the list since `x` is a vector sorted in ascending order. The computation time of the two methods can be evaluated with the use of the `system.time` routine with the following in line code:

```
> k = 100
> x <- runif(k)
> x <- sort(x)
> n = 500000
> test<-function(n,x)for(j in 1:n){y=runif(1);z=findInterval(y,x);}
> test1<-function(n,x)for(j in 1:n){y=runif(1);z=max(0,which(y>x));}
> system.time(test(n,x))
> system.time(test1(n,x))
```

With these values and on the PC we have used, `findInterval` takes 2.89 s, whereas `max-which` takes 2.34 s. Obviously, these times linearly increase with the number of comparisons `n` while, changing the number of positions `k` from 100 to 500, `findInterval` takes 3.32 s and while `max-which` takes 3.67 s. The time increment roughly follows the dichotomic search rule $t \simeq \ln(k-2)$ for both algorithms, but we see that the pair of routines `max-which` seems to be a little faster for low values of `k` and slightly slower for large values. Both solutions are efficient, after all.

Exercise 8.1
Write a code to simulate the rolling of a pair of fair dice.

Answer We must first evaluate the probability distribution and the cumulative function F associated with each score. If we denote with by S the sum of the points obtained in a single roll, these probabilities are represented by Table 8.1. To solve the problem, we use our code `MCdices`:

```
MCdices<- function(nsim=1000,grid=TRUE){
 cumul <- c(0.,1/36,3/36,6/36,10/36,15/36,21/36,
              26/36,30/36,33/36,35/36,1.)
 cores <- seq(2,12,length.out=11)
```

(continued)

Exercise 8.1 (continued)

```
    tosses <- seq(0,0,length.out=11)

  for( j in 1:nsim){
     x = runif(1)
     ind = findInterval(x,cumul)
     tosses[ind] = tosses[ind] + 1 }
  meansc=MeanHisto(scores,tosses)
  stdsc=sqrt(VarHisto(scores,tosses))
  ermean = stdsc/sqrt(sum(tosses))
  erstd = stdsc/sqrt(2*sum(tosses))
  output <- paste("mean = ",round(meansc,digits=3)," +-",
                           round(ermean,digits=3),
                  " sigma = ",round(stdsc,digits=3),"+-",
                            round(erstd,digits=3))
  HistoBar(scores,tosses,errors='ON',grid=TRUE, xex=output)
  if(grid==TRUE) grid()
 }
```

which produces the results of Fig. 8.3, relative to 20 000 rolls. From Eq. (8.6), we can estimate from the histogram the score mean value $\langle S \rangle$ of a pair of dice as:

$$\langle S \rangle = 7.00 \pm 0.03 \,, \quad CL \simeq 99.7\% \,.$$

Table 8.1 Probability distribution $p_i \, (i = 1, \ldots, 11)$ and cumulative function F_i $(F_i = \sum_{j=1}^{i} p_j)$ of the score S_i resulting from the roll of a pair of dice

i	S_i	p_i	$F_i = \sum_{j=1}^{i} p_j$
1	2	1/36	1/36
2	3	2/36	3/36
3	4	3/36	6/36
4	5	4/36	10/36
5	6	5/36	15/36
6	7	6/36	21/36
7	8	5/36	26/36
8	9	4/36	30/36
9	10	3/36	33/36
10	11	2/36	35/36
11	12	1/36	1

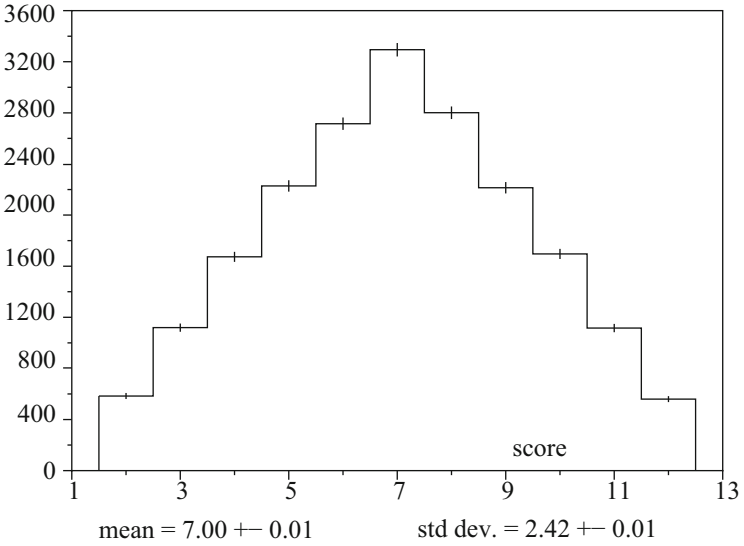


Fig. 8.3 Histogram of 20,000 simulated rolls of a pair of dice

8.5 Generation of Continuous Random Variables

In the case of continuous variables, it is sufficient to invoke again Theorem 3.5 to immediately arrive at the same type of procedure that we have just used for discrete variables.

Let us consider a generic p.d.f. $p(x)$, having cumulative $F(x)$ and defined on any interval $[a, b]$. If we randomly generate $\{\boldsymbol{\xi} = \xi\}$ and calculate the corresponding value:

$$x = F^{-1}(\xi) \,, \tag{8.12}$$

we obtain, from Eq. (3.87), a random generation of X according to $p(x)$.

To generate random variables from any density, it is therefore sufficient to solve an integral and invert the obtained function, using an algorithm that we can be stated as follows:

Algorithm 8.2 (Inverse Transformation) *To generate a continuous random variable X, distributed as $p(x)$ and defined in the interval $[a, b]$, it is necessary:*

- *To generate a variate $0 \leq \xi \leq 1$*
- *To solve, with respect to x, the equation:*

$$\int_a^x p(t)\, dt = \xi \tag{8.13}$$

The following exercises help you to get familiar with this procedure, which is of fundamental importance for the MC methods. We suggest also to look again at Exercise 3.12.

Exercise 8.2
Generate a uniform random variable X within the interval [a,b].

Answer From Eq. (3.79), one has:

$$p(x) = \frac{1}{(b-a)} , \tag{8.14}$$

and Eq. (8.13) becomes:

$$\int_a^x \frac{\mathrm{d}x}{(b-a)} = \xi . \tag{8.15}$$

Hence, the final result is:

$$x = a + \xi(b-a) . \tag{8.16}$$

In R, Eq. (8.16) can be implemented with the calling string: `x = runif(1, min=a, max=b)`.

Exercise 8.3
Randomly generate points uniformly distributed in a circle of radius R with constant density ρ (points/cm^2).

Answer To define the position of a generic point P within a circle, it is convenient to use the pair of polar coordinates r and φ (with $0 \leq r \leq R$ and $0 \leq \varphi \leq 2\pi$).

To determine their probability densities $p(\varphi)$ and $q(r)$, we first observe that $p(\varphi)\,\mathrm{d}\varphi$ is given by the ratio between the number of points contained in the infinitesimal dashed sector of Fig. 8.4a and the total number of points contained in the circle surface:

$$p(\varphi)\,\mathrm{d}\varphi = \frac{\rho R^2\,\mathrm{d}\varphi/2}{\rho\pi R^2} = \frac{\mathrm{d}\varphi}{2\pi} . \tag{8.17}$$

(continued)

Exercise 8.3 (continued)
Analogously, for $q(r)$, we obtain (see Fig. 8.4b):

$$q(r)\,\mathrm{d}r = \frac{\rho 2\pi r\,\mathrm{d}r}{\rho\pi R^2} = \frac{2r}{R^2}\,\mathrm{d}r\ . \tag{8.18}$$

The corresponding cumulative functions are:

$$\xi_1 = P(\varphi) = \int_0^{\varphi} p(\varphi)\,\mathrm{d}\varphi = \frac{\varphi}{2\pi}\ , \tag{8.19}$$

$$\xi_2 = Q(r) = \int_0^{r} q(r)\,\mathrm{d}r = \frac{r^2}{R^2}\ . \tag{8.20}$$

Applying again Algorithm 8.2, we obtain:

$$\begin{cases} \varphi = 2\pi\xi_1 \\ r = R\sqrt{\xi_2}\ , \end{cases} \tag{8.21}$$

where ξ_1 and ξ_2 are two independent uniform random numbers.

These equations can be quickly checked with the following code, where $R = 2$:

```
> phi <- 2*pi*runif(1000)  # pi is pigrec in the R software
> rho <- 2*sqrt(runif(1000))
> x <- rho*cos(phi)
> y <- rho*sin(phi)
> plot(x,y,pch='.',cex=2.5)
```

It is interesting to notice that an isotropic point distribution in the circle implies uniformity in φ but not in r. This can be intuitively justified by looking at Fig. 8.4b. Uniformity in r would mean to have the same number of points for two circular sectors with different radii and then a higher density for the one closer to the centre. Since $\xi < 1$, the square root operation "pulls" points closer to the circumference to fulfil the isotropy condition.

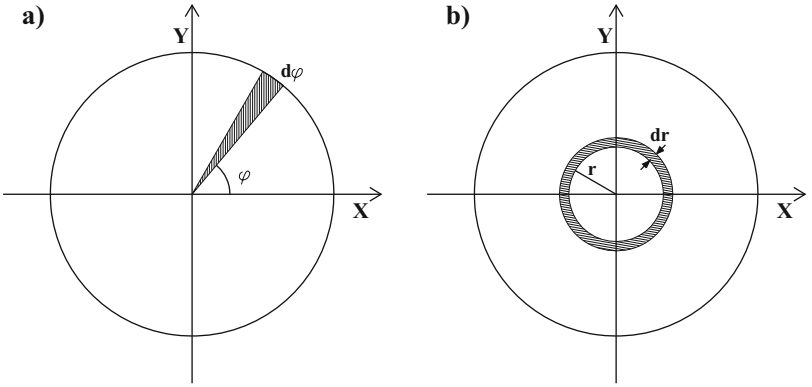


Fig. 8.4 (**a**) When points are uniformly distributed in the circle, $p(\varphi)\,\mathrm{d}\varphi$ is equal to the ratio between the infinitesimal dashed area and the total number of points on the circular surface. (**b**) As in the previous figure, but for $q(r)\,\mathrm{d}r$

Exercise 8.4
Generate points isotropically distributed on a spherical surface of radius R and uniformly within the spherical volume.

Answer The position of a generic point P located on a spherical surface (that, without losing generality, we assume to be at the centre of a system of Cartesian axes) is usually defined (as in Fig. 8.5a) by the three coordinates (r, φ, ϑ) with:

$$\begin{cases} 0 \le \varphi \le 2\pi & (\varphi = \text{azimuthal angle}) \\ 0 \le \vartheta \le \pi & (\vartheta = \text{polar angle}) \\ 0 \le r \le R & (r = \text{radial distance})\ . \end{cases} \tag{8.22}$$

The formulae giving the transformation of the point coordinates to the orthogonal system XYZ are:

$$\begin{cases} x = r \sin\vartheta \cos\varphi \\ y = r \sin\vartheta \sin\varphi \\ z = r \cos\vartheta\ . \end{cases} \tag{8.23}$$

Now let us consider the infinitesimal spherical volume:

$$\mathrm{d}V = r^2\,\mathrm{d}\Omega\,\mathrm{d}r = r^2 \sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi\,\mathrm{d}r\ . \tag{8.24}$$

If we now denote by n_{tot} and $\mathrm{d}n$ the total number of points on the sphere and in $\mathrm{d}V$, respectively, for the isotropy condition, we must assume that the

(continued)

Exercise 8.4 (continued)
density points in the sphere are constant and that the probability $p(V)\,\mathrm{d}V$ for any point to be within $\mathrm{d}V$ is:

$$p(V)\,\mathrm{d}V = \frac{\mathrm{d}n}{n_{tot}} = \frac{\mathrm{d}V}{V} = \frac{r^2 \sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi\,\mathrm{d}r}{\frac{4}{3}\pi R^3}\,. \tag{8.25}$$

The probabilities $p(\varphi)\,\mathrm{d}\varphi$, $q(\vartheta)\,\mathrm{d}\vartheta$ and $\rho(r)\,\mathrm{d}r$ for a point to be inside the interval $[\varphi, \varphi + \mathrm{d}\varphi]$ for any ϑ and r, inside the interval$[\vartheta, \vartheta + \mathrm{d}\vartheta]$ for any φ and r and inside the interval $[r, r + \mathrm{d}r]$ for any ϑ and φ, are evaluated with the corresponding marginal densities, defined in Eq. (4.11):

$$p(\varphi)\,\mathrm{d}\varphi = \frac{1}{V}\,\mathrm{d}\varphi \int_0^R \int_0^\pi r^2 \sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}r = \frac{1}{2\pi}\,\mathrm{d}\varphi\,, \tag{8.26}$$

$$q(\vartheta)\,\mathrm{d}\vartheta = \frac{1}{V}\sin\vartheta\,\mathrm{d}\vartheta \int_0^R \int_0^{2\pi} \mathrm{d}\varphi\,\mathrm{d}r = \frac{\sin\vartheta}{2}\,\mathrm{d}\vartheta\,, \tag{8.27}$$

$$\rho(r)\,\mathrm{d}r = \frac{1}{V}r^2\,\mathrm{d}r \int_0^{2\pi} \int_0^\pi \sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi = \frac{3}{R^3}r^2\,\mathrm{d}r\,, \tag{8.28}$$

where $V = (4/3)\pi R^3$. The corresponding cumulative functions are:

$$\xi_1 = P(\varphi) = \frac{\varphi}{2\pi}\,, \tag{8.29}$$

$$\xi_2 = Q(\vartheta) = \frac{1 - \cos\vartheta}{2}\,, \tag{8.30}$$

$$\xi_3 = R(r) = \frac{r^3}{R^3}\,. \tag{8.31}$$

Finally, the formulae for the random generation of ϕ, ϑ and r are respectively given by:

$$\begin{cases} \varphi = 2\pi\xi_1 \\ \vartheta = \mathrm{acos}(1 - 2\xi_2) \\ r = R\,\xi_3^{1/3}\,. \end{cases} \tag{8.32}$$

If all three equations are considered, a uniform distribution within the sphere is obtained. If we set $r = R$ constant, the first two formulae give the angles that, inserted in the first two of Eqs. (8.23), generate the isotropic distribution of points on the spherical surface.

(continued)

Exercise 8.4 (continued)

So, remember the following not very intuitive consideration: isotropy on the spherical surface means uniformity in φ but not in ϑ. In fact, from the previous equations, it is easy to realize that the isotropy condition is satisfied when $\xi = (1 - \cos\vartheta)$ (or $\xi = \cos\vartheta$), and not ϑ, is a uniformly distributed variable. As before, this property can be qualitatively understood from a geometric point of view if we consider the two hatched surfaces S_1 and S_2 of Fig. 8.5b, which are respectively included between $[\vartheta_1, \vartheta_1 + \mathrm{d}\vartheta]$ and $[\vartheta_2, \vartheta_2 + \mathrm{d}\vartheta]$. A uniform generation in ϑ would give roughly the same number of points on the two surfaces, but, since the area of S_2 is much larger than that of S_1, the resulting point density would be much greater at the poles than at the equator. When the generation is uniform inside the sphere volume, the cubic root transformation for the generation of r acts in the same way as in the circle case.

These formulae can be applied and verified with our `MCsphere` routine. Figure 8.6 shows a result obtained with the generation inside a spherical volume.

Just considering the previous exercises, Algorithm 8.2 would seem to solve all random generation problems. In reality, the situation is not so simple, because the density $p(x)$ to be integrated could be known only numerically, or the integral appearing in the left-hand term of Eq. (8.13) might result in a function that is not analytically invertible. In all these cases, there is a wide variety of alternative procedures to be used. Most of these methods are already implemented in R, as well as in the other statistical software, to generate random numbers from many different distributions.

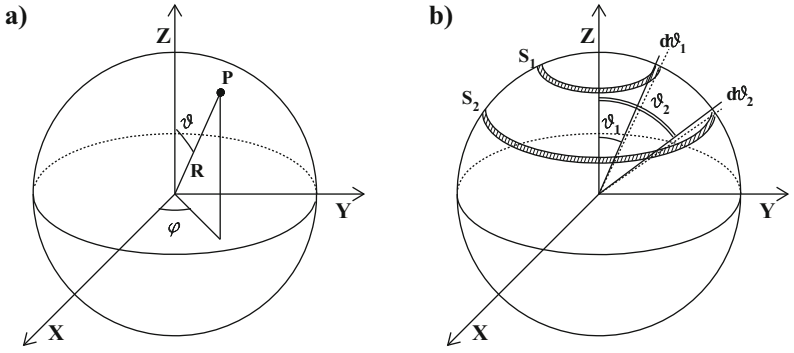


Fig. 8.5 (**a**) The generic point P on the spherical surface is identified by the three coordinates (R, ϑ, φ). (**b**) To obtain the same point density on the surfaces S_1 and S_2, a uniform variable $(1 - \cos\vartheta)$ or $\cos\vartheta$ must be sampled. The uniform sampling of ϑ would give a different point density

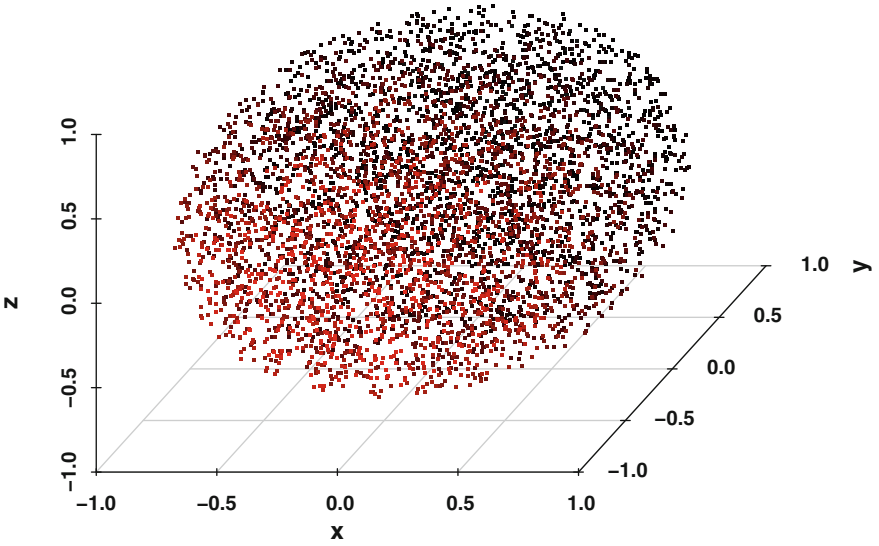


Fig. 8.6 Uniform simulated point distribution inside a spherical volume from the `MCsphere` routine

In the next sections, we will have a detailed look at some of the more commonly used random generation techniques. This will enable you to deal with problems where particular densities, not included in the most used statistical packages, may be involved.

8.6 Linear Search Method

When the cumulative $F(x)$ cannot be represented analytically, we can always numerically compute the integral:

$$F(x) = \int_a^x p(x)\,\mathrm{d}x \tag{8.33}$$

and obtain this function at N different points $x_1, x_2, \ldots, x_N$, with $x_1 = a$ and $x_N = b$ (see Fig. 8.7):

$$\begin{aligned} F(x_1) &= 0 \\ F(x_2) &= F(x_1) + \textstyle\int_{x_1}^{x_2} p(x)\,\mathrm{d}x\ , \\ F(x_3) &= F(x_2) + \textstyle\int_{x_2}^{x_3} p(x)\,\mathrm{d}x\ , \\ &\cdots\cdots\cdots\cdots \\ F(x_N) &= F(x_{N-1}) + \textstyle\int_{x_{N-1}}^{x_N} p(x)\,\mathrm{d}x = 1\ . \end{aligned} \tag{8.34}$$

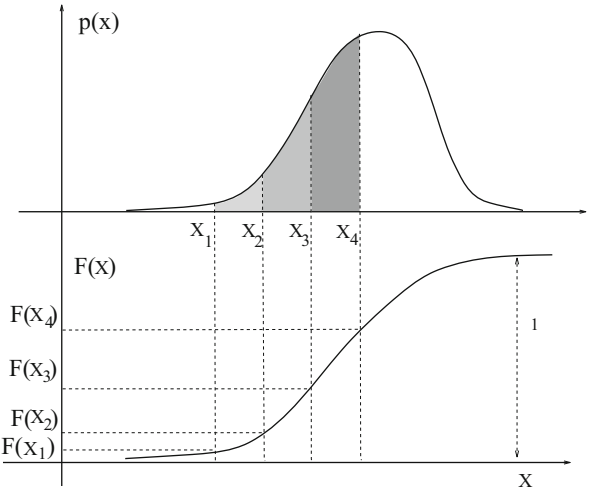


Fig. 8.7 Generation of random variables using the linear search method. The cumulative function $F(x)$ is evaluated on a set of points to go back to the case of the generation of a discrete variable

In this way, we have returned to the random generation of a discrete variable, already described in Sect. 8.4. Indeed, through the relation:

$$F_{j-1} = \sum_{i=1}^{j-1} p(x_i) < \xi_i \leq \sum_{i=1}^{j} p(x_i) = F_j \ , \qquad (1 \leq j \leq N) \tag{8.35}$$

we are immediately able to determine the particular subinterval $[x_{j-1}, x_j]$ containing the random variable we have to generate.

Assuming $F(x)$ to be linear within each subinterval, we can then obtain the desired random value with the linear interpolation between the limits x_{j-1} and x_j of the selected subinterval:

$$x = x_{j-1} + \frac{\xi - F_{j-1}}{(F_j - F_{j-1})}(x_j - x_{j-1}) \ . \tag{8.36}$$

Since the cumulative $F(x)$ is a monotonically increasing regular function, the approximation (8.36) generally gives correct results. All these considerations are summarized in the following procedure:

Algorithm 8.3 (Linear Search) *To generate a random variable X, distributed as $p(x)$ and defined in the limited or unlimited interval $[a, b]$, it is necessary:*

- *To calculate numerically by points N values of the cumulative function $x_1, x_2, \ldots, x_N$ ($x_1 = a$; $x_N = b$):*

$$F_j = \int_a^{x_j} p(x)\, dx \quad 1 \leq j \leq N \tag{8.37}$$

- *To generate a variate* $0 \leq \xi \leq 1$

- *To determine the index j such as $F_{j-1} < \xi \leq F_j$*
- *To calculate x through Eq. (8.36)*

With this method, the cumulative does not need to be inverted, and numerical methods can be used when the integral is difficult or impossible to compute analytically.

However, a relevant number of points could be needed to precisely reproduce the $F(x)$ behaviour. Moreover, a non-negligible time may be required by the integral calculation or by the search required to determine the correct index j.

8.7 Rejection Method

This method is based on the property of the definite integrals to be the area between the integrand function and the x axis. The procedure starts by randomly choosing uniformly distributed points within a rectangle (delimited by the vertices A, B, C, D in Fig. 8.8), with base $(b - a)$ and height h, which encloses the considered probability density $p(x)$ (obviously h must be greater than or equal to the maximum p_{max} assumed by $p(x)$ in $[a, b]$). From Eq. (8.16), the generic coordinates (x_i, y_i) of these points are:

$$\begin{cases} x_i = a + \xi_1 (b - a) \\ y_i = \xi_2 h \ , \end{cases} \tag{8.38}$$

with $0 \leq \xi_1, \xi_2 \leq 1$.

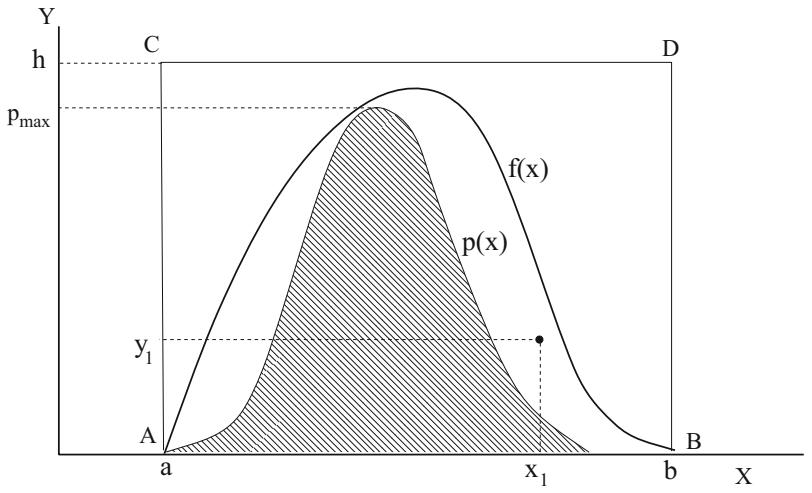


Fig. 8.8 Generation of random variables with the rejection method. The probability density $p(x)$ is bounded within a rectangle (or within a function $f(x)$) and a uniformly generated point within it is "accepted" if it is in the region defined by $p(x)$ and the abscissa axis (hatched area)

Let us now consider the probability for the uniform random variable $a \leq X \leq b$ to fall within the infinitesimal interval $[x_i, x_i + \mathrm{d}x]$:

$$P\{x_i < X < x_i + \mathrm{d}x\} = P\{x_i < a + \boldsymbol{\xi_1}(b-a) < x_i + \mathrm{d}x\} = \frac{\mathrm{d}x}{(b-a)}\,, \tag{8.39}$$

and the conditional probability that, for a given x_i, a uniform variable $0 \leq Y \leq h$ is less than $p(x_i)$:

$$\begin{aligned} P\{Y \leq p(x) | x_i < X < x_i + \mathrm{d}x\} &\simeq P\{Y \leq p(x_i)\} \\ &= P\{\boldsymbol{\xi_2} \leq p(x_i)/h\} = \frac{p(x_i)}{h}\,. \end{aligned} \tag{8.40}$$

According to Theorem 1.21 of compound probabilities, the probability for both the previous events to occur is:

$$\begin{aligned} &P\{x_i < X < x_i + \mathrm{d}x, Y \leq p(x_i)\} \\ &= P\{Y \leq p(x) | x_i < X < x_i + \mathrm{d}x\} \cdot P\{x_i < X < x_i + \mathrm{d}x\} \\ &\simeq \frac{p(x_i)\,\mathrm{d}x}{h(b-a)} = \varepsilon p(x_i)\,\mathrm{d}x\,. \end{aligned} \tag{8.41}$$

Apart from the constant factor $\varepsilon = 1/[(h(b-a)]$, this formula coincides with the probability for X to be in $(x_i, x_i + \mathrm{d}x)$.

These simple considerations are the basis of the rejection technique, which applies Eq. (8.41) by trial and error and which we can thus state as:

Algorithm 8.4 (Simple Rejection Algorithm) *To generate a continuous random variable X, having p.d.f. $p(x)$ defined in the finite interval $[a, b]$ with maximum value p_{max}, it is necessary:*

- *To uniformly generate a random point $x \in [a, b]$: $x = a + \xi_1(b-a)$.*
- *To calculate $p(x)$.*
- *To uniformly generate a random point y within 0 and h ($h \geq p_{max}$): $y = \xi_2 h$.*
- *If $y \leq p(x)$, then x is accepted; otherwise, it is rejected, and the procedure is restarted from the beginning.*

Clearly, to apply this procedure, it is just necessary to know the analytic expression of $p(x)$, while no information on its cumulative is needed; however, the price to pay is that at least two uniform random numbers are required to generate a variable from the density $p(x)$.

The constant $\varepsilon = 1/[h(b-a)]$ is the ratio between the number of accepted points and the totality of those generated and represents the generation efficiency or, equivalently, the inverse of the average number of attempts required to accept a point. It is also equal to the ratio between the area under $p(x)$ and that of the rectangle ABCD of Fig. 8.8. To optimize the method, it would then be necessary to generate all points (to be accepted or discarded) no longer within a rectangle but

within a curve $f(x)$, which contains $p(x)$ and mimic its behaviour, in such a way as to maximize the ratio between the respective areas (see Fig. 8.8). This generalization makes the rejection method also valid for functions defined in an unlimited range as it suffices to find a function $f(x)$ also defined in the same range. To have a simple (and above all fast) sampling procedure, it is necessary for $f(x)$ to have an analytically invertible cumulative $F(x)$. If so, using the equations:

$$\begin{cases} x_i = F^{-1}(\xi_1) \\ y_i = \xi_2 f(x_i) \, , \end{cases} \tag{8.42}$$

a point is sampled within $f(x)$ which, as we have just seen, is accepted if $y_i \leq p(x_i)$.

To formally demonstrate this new procedure, Eqs. (8.39) and (8.40) must be rewritten, which now become, respectively:

$$P\{x_i < X < x_i + \mathrm{d}x\} \simeq \frac{f(x_i)\,\mathrm{d}x}{\int_a^b f(x)\,\mathrm{d}x} \, , \tag{8.43}$$

and:

$$P\{Y \leq p(x) | x_i < X < x_i + \mathrm{d}x\} \simeq P\{\boldsymbol{\xi_2} f(x_i) \leq p(x_i)\} = \frac{p(x_i)}{f(x_i)} \, . \tag{8.44}$$

After inserting these last two relations in Eq. (8.41), one obtains the result:

$$P\{x_i < X < x_i + \mathrm{d}x, Y \leq p(x_i)\} \simeq \frac{p(x_i)\,\mathrm{d}x}{\int_a^b f(x)\,\mathrm{d}x} \propto p(x_i)\,\mathrm{d}x \, . \tag{8.45}$$

The efficiency ε of this generalized method is obtained by integrating the previous relation over all x values:

$$\varepsilon = \frac{\int_a^b p(x)\,\mathrm{d}x}{\int_a^b f(x)\,\mathrm{d}x} = \frac{1}{\int_a^b f(x)\,\mathrm{d}x} \, . \tag{8.46}$$

In this case, the constant ε, which in Eq. (8.41) was the inverse of the area $h(b-a)$ of the rectangle ABCD, now is the inverse of the integral of $f(x)$, i.e. the area under this function.

The previous algorithm can then be generalized as:

Algorithm 8.5 (Optimized Rejection) *To generate a random variable X, distributed as $p(x)$, which is defined in the (limited or unlimited) interval $[a, b]$ and bounded by a function $f(x)$ having an analytically invertible cumulative function $F(x)$, one needs:*

- *To randomly generate a point $x \in [a, b]$ having density proportional to $f(x)$: $x = F^{-1}(\xi_1)$.*

- *To calculate* $p(x)$.
- *To uniformly generate a random point* y *between* 0 *and* $f(x)$*:* $y = \xi_2 f(x)$.
- *If* $y \leq p(x)$*,* x *is accepted; otherwise, it is rejected, and the procedure restarts from the beginning.*

It is also possible to formulate a third version of the rejection method (devised by the American mathematician J. von Neumann in the 1950s) when the function $p(x)$, from which the variable X must be generated, can be factored as the product of two functions:

$$p(x) = g(x)h(x) \,, \tag{8.47}$$

where $h(x)$ has an analytically invertible cumulative $H(x)$ and $g(x)$ is limited within the interval $[a, b]$. In this case, we can write:

$$H(x) = \frac{\int_a^x h(x)\,\mathrm{d}x}{\int_a^b h(x)\,\mathrm{d}x} \,, \tag{8.48}$$

and:

$$0 \leq g(x) \leq G \,. \tag{8.49}$$

As usual, we sample two random uniform numbers ξ_1 and ξ_2 and define the conditions:

$$\begin{cases} x_i = H^{-1}(\xi_1) \\ \xi_2 \leq \dfrac{g(x_i)}{G} \,. \end{cases} \tag{8.50}$$

Equations (8.43) and (8.44) now become:

$$P\{x_i \leq X < x_i + \mathrm{d}x\} \simeq \frac{h(x_i)\,\mathrm{d}x}{\int_a^b h(x)\,\mathrm{d}x} \,, \tag{8.51}$$

$$P\{\boldsymbol{\xi_2} \leq g(x)/G | x_i \leq X < x_i + \mathrm{d}x\} = P\{\boldsymbol{\xi_2} \leq g(x_i)\big/G\} \,, = g(x_i)/G \tag{8.52}$$

and hence:

$$P\{x_i \leq X < x_i + \mathrm{d}x, \boldsymbol{\xi_2} \leq g(x)/G\} \simeq \frac{1}{G\int_a^b h(x)\,\mathrm{d}x} h(x_i)g(x_i)\,\mathrm{d}x \,. \tag{8.53}$$

Also in this case, the efficiency ε is obtained, by integration over all x values:

$$\varepsilon = \frac{\int_a^b h(x)g(x)\,\mathrm{d}x}{G\int_a^b h(x)\,\mathrm{d}x} \leq \frac{G\int_a^b h(x)\,\mathrm{d}x}{G\int_a^b h(x)\,\mathrm{d}x} = 1\ . \tag{8.54}$$

If $p(x)$ is normalized, then $\int_a^b h(x)g(x)\,\mathrm{d}x = 1$, and the efficiency simply results in:

$$\varepsilon = \frac{1}{G\int_a^b h(x)\,\mathrm{d}x} < 1\ . \tag{8.55}$$

Equation (8.47) can be rewritten as:

$$p(x) = G\cdot\int_a^b h(x)\,\mathrm{d}x\cdot\frac{h(x)}{\int_a^b h(x)\,\mathrm{d}x}\frac{g(x)}{G} \equiv Ch^*(x)g^*(x)\ , \tag{8.56}$$

where $h^*(x)$ is a normalized density and $0 \leq g^*(x) \leq 1$. The constant C, if $p(x)$ is normalized, is the inverse of the efficiency, and therefore it must satisfy the condition $C \geq 1$.

The algorithm then becomes:

Algorithm 8.6 (Weighted Rejection) *To generate the values of a random variable X, distributed as $p(x)$, which is defined in the (limited or unlimited) interval $[a, b]$ and factored as $p(x) = Cg(x)h(x)$, where $C \geq 1$, $0 \leq g(x) \leq 1$ and where $h(x)$ is a p.d.f. with analytically invertible cumulative $H(x)$, it is necessary:*

- *To generate $0 \leq \xi_1 \leq 1$.*
- *To randomly generate a point $x \in [a, b]$ from the density $h(x)$: $x = H^{-1}(\xi_1)$.*
- *To calculate $g(x)$.*
- *To uniformly generate $0 \leq \xi_2 \leq 1$.*
- *If $\xi_2 \leq g(x)$, then x is accepted; otherwise, the procedure restarts from the beginning.*

This algorithm can be easily kept in mind if we consider $h(x)$ as the base density of events and $g(x)$ as a "weight" function: a generated point $x_i = H^{-1}(\xi_1)$ will be more important ("heavy") the closer $g(x_i)$ is to one. This condition is taken into account in the second generation, when the event is accepted only if $\xi_2 \leq g(x_i)$. We finally note that if, in Eq. (8.56), we define the weight function as $g(x) = p(x)/h$ and $h(x) = 1/(b-a)$, one gets Algorithm 8.4, whereas if the weight function is $g(x) = p(x)/h(x)$, one gets Algorithm 8.5.

Exercise 8.5
Generate a random variable within the interval $[0, \pi/2]$ with p.d.f.:

$$p(x) = x\sin x\,\mathrm{d}x\ . \tag{8.57}$$

(continued)

Exercise 8.5 (continued)
Answer In this case we cannot apply the inverse cumulative method (algorithm 8.2) because the equation:

$$\int_0^x x \sin x = \sin x - x \cos x = \xi \tag{8.58}$$

is not analytically invertible for x. We therefore apply the rejection technique using the three procedures we have just derived.

- Algorithm 8.4
  We delimit $p(x)$ within the $[0, \pi 2] \times \pi/2$ square, as in Fig. 8.9.
  Referring to Eq. (8.38), now we have $a = 0$; $b = \pi/2$; $h = \pi/2$. Therefore, the variable is simulated with an extraction efficiency of about 40% ($\varepsilon = 4/\pi^2$), with the R code:

```
pig05 = 0.5*pi   # pi is 3.1415...  in the R software
#  basic rejection method
xv <- seq(0,0,length.out=nsim)  # nsim is the number of points
for(k in 1:nsim) {
  csi = 1
  px = 0
  while(csi > px){
      x = pig05 *runif(1)
      px = x*sin(x)
      csi = pig05*runif(1)
      xv[k] = x
  }
}
```

- Algorithm 8.5
  If we generate points uniformly distributed under the curve $f(x) = x$ (see Fig. 8.9), we double the generation efficiency of the previous algorithm as the ratio between the areas under $p(x)$ and $f(x)$ results:

$$\varepsilon = \int_0^{\pi/2} p(x)\,dx \Big/ \int_0^{\pi/2} f(x)\,dx = \left(\frac{8}{\pi^2}\right) \simeq 81\% \,. \tag{8.59}$$

  To implement this method, $f(x)$ has to be normalized, calculating its area in the interval $[0, \pi/2]$:

$$\int_0^{\pi/2} x\,\mathrm{d}x = \frac{\pi^2}{8} \,, \tag{8.60}$$

  and then the equation:

$$\int_0^x \left(\frac{8}{\pi^2}\right) x\,\mathrm{d}x = \xi_1 \,, \tag{8.61}$$

(continued)

Exercise 8.5 (continued)
must be solved. The result is:

$$x = \left(\frac{\pi}{2}\right)\sqrt{\xi_1}\,. \tag{8.62}$$

In this way the random abscissa x_i has been sampled, whereas the corresponding ordinate is obtained with a random uniform sampling between 0 and $f(x) = x$, through the equation $y = x\xi_2$. Therefore, the variable generation loop is:

```
#  optimized rejection method
 xv1 <- seq(0,0,length.out=nsim)
 for(k in 1:nsim) {
    csi = 1
    px = 0
    while(csi > px){
        x = pig05 *sqrt(runif(1))
        px = x*sin(x)
        csi = x*runif(1)
        xv1[k] = x
    }
 }
```

- Algorithm 8.6
  Based on Eq. (8.60), the factorization of $p(x)$ is obtained as:

$$p(x) = \left(\frac{\pi^2}{8}\right)\left(\frac{8}{\pi^2}\right)x\,\sin x\,, \tag{8.63}$$

and the factors are identified as $C = \pi^2/8$; $g(x) = \sin x$ and $h(x) = (8/\pi^2)x$. In this case the loop becomes:

```
#  weightedrejection method
 xv2 <- seq(0,0,length.out=nsim)
 for(k in 1:nsim) {
    csi = 1
    px = 0
    while(csi > px){
        x = pig05 *sqrt(runif(1))
        px = sin(x)
        csi = runif(1)
        xv2[k] = x
    }
 }
```

You can find the complete solution of the exercise and the generation of the distributions of the values `xv, xv1, xv2` in our `MCxsinx` routine.

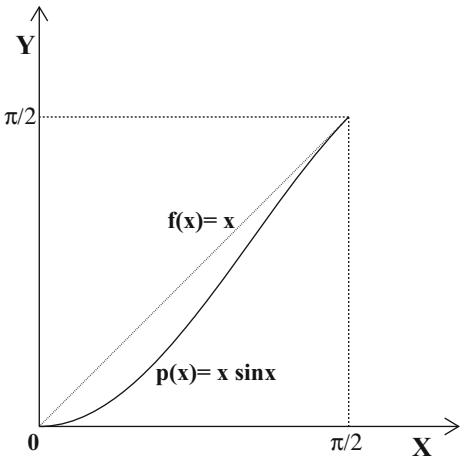


Fig. 8.9 Behaviour of the functions $f(x) = x$ and $p(x) = x \sin(x)$ in the interval $[0, \pi/2]$

8.8 Particular Random Generation Methods

In some cases none of the algorithms discussed so far can generate values of random variables in a simple or sufficiently rapid way.

Fortunately, several well-established algorithms have been existing since a long time to efficiently solve "ad hoc" many special random generation problems.

As an example, here we will show some of the methods used to generate Gaussian and Poisson density variables while, to have a complete review of all (or almost all) the different random generation algorithms, we suggest to consult the references [Fis96, Knu81, Mor84] and [Rub81]. Although these algorithms are already implemented inside the R routines that generate random numbers from all common distributions, we think it is equally instructive to take a look at these methods, to give you additional hints useful both to solve non-standard random generation problems and to review some important probabilistic and statistical concepts.

(a) Gaussian random variates generation.

As we have seen in Sect. 3.5, it is possible to obtain, from a Gaussian density $g(x)$, with any μ and σ, a standard Gaussian value or deviate: (with $\mu = 0$ and $\sigma = 1$):

$$t = \frac{x - \mu}{\sigma} \tag{8.64}$$

coming from the standard Gaussian:

$$g(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \qquad (-\infty \leq t \leq +\infty)\,. \tag{8.65}$$

If we were able to sample a deviate from this density, with the inverse transformation of Eq. (8.64):

$$x = t\sigma + \mu \,, \tag{8.66}$$

we would obtain any other Gaussian variate. However, it is not possible to analytically derive the cumulative of $g(t)$ and, consequently, use Algorithm 8.2.

Given the infinite range of variation of t, Algorithm 8.4 cannot be used either, while Algorithms 8.5 (see Exercise 8.5), 8.6 and 8.3 are all applicable (see, e.g. [TC93]). However, other methods, simpler or more efficient, are usually preferred. One of these, based on the Central Limit Theorem, exploits the properties of the sum of uniform random variables and is described in Problem 8.7. Also our `MCgauss1` routine can be examined.

The procedure that is most frequently used for the Gaussian generation is the one devised at the end of the 1950s by the American mathematicians G.E.P. Box and M.E. Muller, who proved that, contrary to what one might intuitively assume, it is easier to generate not one but two independent Gaussian variables simultaneously. Let us consider the standard bivariate Gaussian in polar coordinates (r, φ) of Eq. (4.81):

$$g(r, \varphi)\, \mathrm{d}r\, \mathrm{d}\varphi = \frac{r}{2\pi} e^{-r^2/2}\, \mathrm{d}r\, \mathrm{d}\varphi = p(r) q(\varphi)\, \mathrm{d}r\, \mathrm{d}\varphi \,, \tag{8.67}$$

and calculate the cumulative functions of the modulus and angle of the polar vector:

$$\xi_1 = P(r) = \int_0^r p(r)\, \mathrm{d}r = 1 - e^{(-r^2/2)} \,, \tag{8.68}$$

$$\xi_2 = Q(\varphi) = \int_0^\varphi q(\varphi)\, \mathrm{d}\varphi = \frac{\varphi}{2\pi} \,. \tag{8.69}$$

Since both $p(r)$ and $q(\varphi)$ have analytically invertible cumulative functions, by applying Algorithm 8.2, one easily gets:

$$\begin{cases} r = \sqrt{-2 \log \xi_1} \\ \varphi = 2\pi \xi_2 \,. \end{cases} \tag{8.70}$$

Finally, going to the Cartesian coordinates (z_1, z_2):

$$\begin{cases} z_1 = r \cos\varphi = \sqrt{-2 \log \xi_1} \cos(2\pi \xi_2) \\ z_2 = r \sin\varphi = \sqrt{-2 \log \xi_1} \sin(2\pi \xi_2) \,, \end{cases} \tag{8.71}$$

the variates of a pair of independent standard Gaussian variables are obtained. Notice that just two random numbers ξ_1 and ξ_2 have been used now.

To speed up the algorithm, an ingenious expedient, described in [Knu81], can be used: if we randomly generate a point of Cartesian coordinates (v_1, v_2) inside the unit circle centred on the origin, the sum $s = v_1^2 + v_2^2$ is a random uniform variate between 0 and 1 (we leave the simple proof of this statement as an exercise). We can use this number instead of ξ_1, while the angle defined by this point and by the abscissa axis represents the random angle $2\pi\xi_2$. In this way the direct calculation of the trigonometric functions is avoided since the cosine and the sine appearing in the Eqs. (8.71) are calculated through the ratios $v_1/\sqrt{s}$ and $v_2/\sqrt{s}$. This advantage is partially counterbalanced by the disadvantage to use the rejection technique to obtain the coordinates (v_1, v_2), since we must first generate two numbers v_1, v_2 uniformly distributed in the interval $[-1, 1]$ and then accept only the pairs satisfying the condition: $v_1^2 + v_2^2 \leq 1$. However, the efficiency of this operation ($\simeq 78\%$, equal to the ratio between the unit circle and the square circumscribed about it) is quite high, and the latter procedure resulted on our computer about 20% faster than the "classical" method described by Eqs. (8.71). The Gaussian generation algorithm can be summarized as follows:

Algorithm 8.7 (Gaussian Generation) *To generate two independent normalized Gaussian variates z_1, z_2 it is necessary:*

- *To generate two independent uniform variates $0 \leq \xi_1, \xi_2 \leq 1$.*
- *To define $v_1 = 2\xi_1 - 1$; $v_2 = 2\xi_2 - 1$ and to calculate the sum $s = v_1^2 + v_2^2$.*
- *If $s > 1$, the procedure is repeated from the beginning.*
- *If $s \leq 1$, the events $\{Z_1 = z_1\}$, $\{Z_2 = z_2\}$ are generated as:*

$$\begin{cases} z_1 = v_1 \sqrt{\dfrac{-2\log s}{s}} \\ z_2 = v_2 \sqrt{\dfrac{-2\log s}{s}} \end{cases} . \tag{8.72}$$

This algorithm is implemented in the `MGgauss` routine, here reported, which, as a result, gives the histogram of Fig. 8.10 (upon request) and the two independent Gaussian variates `g1` and `g2`.

```
MCgauss<- function(nsim=1000,mu=0,sigma=1,plot=TRUE,grid=TRUE){
 index <- seq(1,nsim,by=2)
 for(j in index) {
     s=2
     while(s>1){
       v1 = 2.*runif(1) - 1
       v2 = 2.*runif(1) - 1
       s = v1^2 + v2^2
     }
     ls = sqrt(-2*log(s)/s)
     z1 = v1*ls
     z2 = v2*ls
     g[j]   = mu + sigma*z1
```

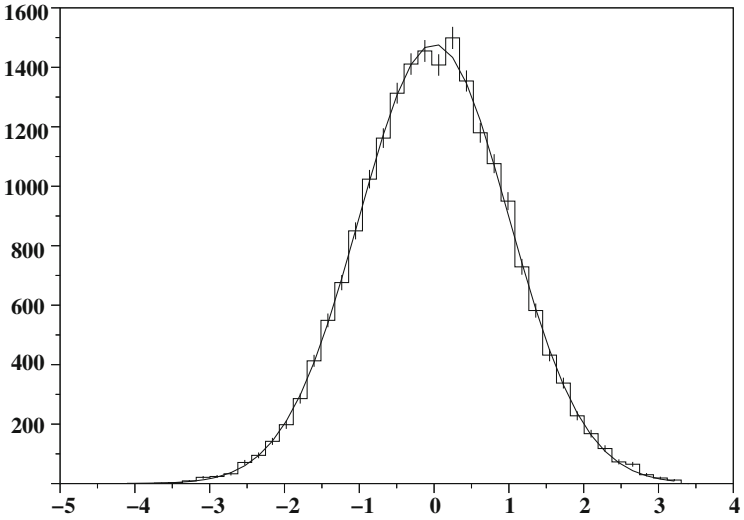


Fig. 8.10 Histogram of a sample of 20,000 random numbers from the `Gauss2` routine. The continuous curve represents the standard Gaussian

```
    g[j+1] = mu + sigma*z2
}
    # plot results
if(plot==TRUE && nsim>2){
     # $x.val is the binning of the histogram
  xplot <- HistoBar(g,nbins=30,errors='ON',
                    xex=' ',yex=' ',out=TRUE)$x.val
  dx = (xplot[2]-xplot[1])
  yplot <- nsim*dnorm(xplot,mean=mu,sd=sigma)*dx
  lines(xplot,yplot,type='l')
  if(grid==TRUE) grid()
}
 rval = list(g1=g[1],g2=g[2])
 return (rval)
}
```

(b) Generation of Poissonian variates.

Each stochastic process in which discrete and independent variables are generated with probability λ constant over time is characterized by two fundamental properties:

- The time interval between two consecutive events is a random variable with exponential density.
- The number of events that occur within a given time interval of a predetermined length Δt is a Poissonian variable of mean $\mu = \lambda \Delta t$.

In Exercise 3.12, we have seen that the time τ between two consecutive stochastic events can be simulated with the equation:

$$\tau = -\frac{1}{\mu}\log\xi \ . \tag{8.73}$$

To obtain the number x of events occurring within $\Delta t = 1$, it is then sufficient to generate a sequence $\tau_0, \tau_1, \ldots, \tau_x, \tau_{x+1}$ of time intervals (with $\tau_0 = 0$) until the inequality:

$$\sum_{i=0}^{x} \tau_i \leq 1 < \sum_{i=0}^{x+1} \tau_i \qquad x = 0, 1, \ldots \tag{8.74}$$

is verified. Taking into account Eq. (8.73), this inequality can be rewritten as:

$$-\sum_{i=0}^{x} \log\xi_i \leq \mu < -\sum_{i=0}^{x+1} \log\xi_i \qquad x = 0, 1, \ldots \tag{8.75}$$

or, without logarithms:

$$\prod_{i=0}^{x} \xi_i \geq e^{-\mu} > \prod_{i=0}^{x+1} \xi_i \qquad x = 0, 1, \ldots \tag{8.76}$$

In conclusion, we arrived to the following:

Algorithm 8.8 (Poissonian Generation) *To generate a random Poissonian variate, it is necessary:*

- *To define* $k = 0$*;* $s = 1$.
- *To generate a uniform variate* $0 \leq \xi_k \leq 1$.
- *To set* $s = s\,\xi_k$.
- *If* $s < e^{-\mu}$*, then set* $x = k$*; otherwise, set* $k = k + 1$ *and go back to the second step.*

This algorithm is implemented in our `MCpoiss` routine. It is easy to deduce that the computation time of this routine increases proportionally to μ; in fact, we have found that the execution times of `MCpoiss` are much longer than those of the R routine `rpois`, which uses a limiting formula of the binomial distribution.

A useful exercise with the `rpois` routine is to check Table 6.2 of Sect. 6.6. By setting $\mu = 2.3$ and generating larger and larger samples, you will observe that the number of times in which zero events are drawn will tend to 10%, that setting $\mu = 4.74$ the number of successes $x \leq 1$ will tend to 5% and so on.

To simulate random variables having distributions not considered here, it is often sufficient to resort to the definitions and theorems of probability theory. As an example, to simulate a $Q(\nu) \sim \chi^2(\nu)$ variable, just remember Theorem 3.3 and add

the squares of ν standard Gaussian variables. By following the standard definitions, correct algorithms are certainly obtained, although not always particularly fast and efficient.

8.9 Monte Carlo Analysis of Distributions

One of the most interesting and astonishing applications of the MC methods is probably the determination of whatever complicated statistical distributions. As we have seen in Chap. 5, the study of the distribution of random variables that are functions of other random variables requires rather laborious mathematical techniques, so much so that the analytic solution, in most cases, may only be within the reach of skilled mathematicians or not even be obtainable in practice. Monte Carlo methods have revolutionized this branch of applied statistics, as they provide, using elementary procedures, the required solution, even if in the approximate form of histograms and not in an appropriate analytical form. From the simulated histogram, it is then possible to evaluate all the requested parameters (such as mean, variance, area under the tails) with negligible statistical errors in the context of the problem under study.

The procedure is sketched in Fig. 8.11: the random variates of the variables $(X_1, X_2, X_3, \dots)$ are generated according to their densities, and, at each generation, the variate of the variable $Z = f(X_1, X_2, X_3, \dots)$ is calculated. After this algorithm has been repeated a sufficient number of times, at the end of the generation loop, the result is usually presented as a histogram of Z, whose binning can be varied at will to be consistent with the problem under study. Finally, the fundamental statistical quantities characterizing the distribution are calculated from this artificial histogram. It is also possible to perform the histogram *best fit* (with the approaches described in the next Sect. 10.7) using empirical functions such as exponentials,

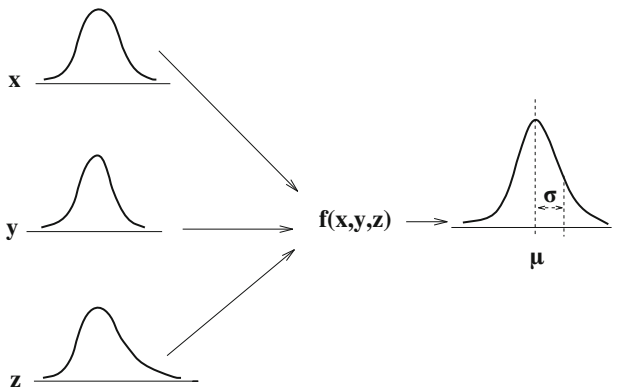


Fig. 8.11 Simulation of a distribution with MC methods

polynomials and sums of Gaussians, thus obtaining an analytic form, independent of the histogram bin, *that interpolates the true solution* with the desired degree of accuracy. The use of the computer has made this method so simple that it can be fully illustrated with an elementary example.

Suppose one wants to determine the distribution of the random variable:

$$Z = \frac{\sin X}{\sin Y} \,, \tag{8.77}$$

where X and Y are Gaussian angles having mean and standard deviation equal to:

$$\begin{aligned} \theta_x &= 20^\circ \,, \quad \sigma_x = 3^\circ \,, \\ \theta_y &= 13^\circ \,, \quad \sigma_y = 3^\circ \,. \end{aligned} \tag{8.78}$$

By using the approximated Eqs. (5.63, 5.66) of Sect. 5.4, the mean and variance of the unknown distribution can be estimated as:

$$\langle Z \rangle = \frac{\sin \theta_x}{\sin \theta_y} = 1.52 \,, \tag{8.79}$$

$$\sigma[Z] = \left[\left(\frac{\cos \theta_x}{\sin \theta_y} \right)^2 \left(\frac{\pi}{180} \right)^2 \sigma_x^2 + \left(\frac{\sin \theta_x \cos \theta_y}{\sin^2 \theta_y} \right)^2 \left(\frac{\pi}{180} \right)^2 \sigma_y^2 \right]^{1/2} = 0.41 \,, \tag{8.80}$$

where the variance has been converted in radians, since z must be expressed in the decimal system. As discussed in Sect. 5.4, both these two formulae are approximate: Eq. (8.79) holds only if the relationship between $Z(X, Y)$ is linear, while Eq. (8.80) requires both the validity of the linear dependency and of small percentage errors.

Since all these conditions are drastically violated by Eqs. (8.77, 8.78), in this case we are not at all certain of either the validity of Eqs. (8.79, 8.80) or of the distribution of the Z variable. Indeed, our knowledge of probability theory allows us to assume a non-Gaussian distribution for Z. Let us see how the simulation is able to easily solve all these issues.

Using the `MCrefrac` routine given below, we generate 20 000 Gaussian pairs (8.78) which are after combined using Eq. (8.77): the result is the histogram of Fig. 8.12a.

```
MCrefrac<- function(nsim=20000,theta1=20,theta2=13,errtheta=3){
  rad = 180/pi   # degrees of a 1 radian angle
  #angles in radiants
  ang1 = theta1/rad; ang2 = theta2/rad; errang = errtheta/rad;
  for(k in 1:nsim){
    t=100
    while(t>4.5){  # truncation at t=4.5 to avoid a long tail
                   # of negligible values
```

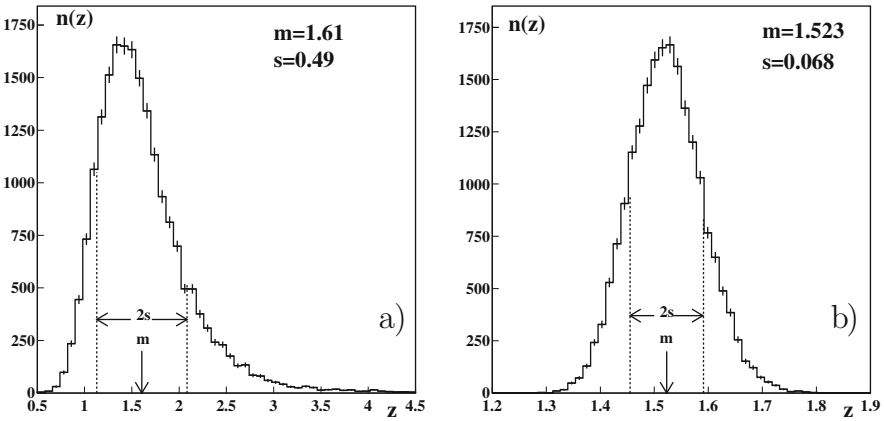


Fig. 8.12 Histogram of 20,000 variates of Z from Eq. (8.77), where X and Y are Gaussian variables given by (**a**) from Eqs. (8.78); (**b**) from Eq. (8.81)

```
      X = rnorm(1,mean=ang1,sd=errang)
      Y = rnorm(1,mean=ang2,sd=errang)
      t = sin(X)/sin(Y)
      }
      g[k] = t
    }
    HistoBar(g,errors='ON',nbins=20,grid=TRUE)
}
```

As can be easily noticed, the parameters μ and σ, deduced from the histogram:

$$\langle Z \rangle \simeq 1.61 \,, \qquad \sigma[Z] \simeq 0.49 \,,$$

have values rather different from the predictions of Eqs. (8.79, 8.80). Their statistical error, according to the fundamental formulae of Table 6.3, is of the order of 0.002–0.003 and can therefore be neglected in this context. In any case, it can be reduced at will simply by increasing the number of simulated variables. This simulation also shows that *the shape of the density deviates significantly from the normal curve*. For instance, it can be easily checked that the number of events in the interval:

$$\mu \pm \sigma \simeq m \pm s = 1.61 \pm 0.49 = [1.12, 2.10]$$

is equal to 14 895, corresponding to a percentage of 75%.

It is then clear that, apart from the lack of knowledge of the exact analytical form of the histogram density of Fig. 8.12a, all other information is easily accessible with the simulation. It is also interesting to decrease the dispersion of X and Y and see what happens. Intuition tells us that we should eliminate the long tail of density, which is due to the non-linearity of the sine function. We then operate as before, but replacing the conditions of Eq. (8.78) with:

$$\begin{aligned} \theta_x &= 20^\circ\ , \quad \sigma_x = 0.5^\circ\ , \\ \theta_y &= 13^\circ\ . \quad \sigma_y = 0.5^\circ\ . \end{aligned} \tag{8.81}$$

Equation (8.79) remains unchanged, whereas Eq. (8.80) in this case gives the value:

$$\sigma[Z] \simeq 0.068\ .$$

The simulation result is shown in Fig. 8.12b. Now the density is "almost" normal, with mean and standard deviation very close to the approximations of Eqs. (8.79, 8.80). Within the interval $m \pm s$, 13 723 events are found, corresponding to a percentage of 68.6%, a value in perfect agreement with 3σ law.

8.10 Evaluation of Confidence Intervals

MC techniques also play a fundamental role in the determination of confidence intervals, because they allow us to solve the integrals (6.7) in an approximate way even when it is difficult or impossible to find the functional form of the density $p_Z(z;\theta)$, where $Z = f(X_1, X_2, \ldots)$ is a random variable that is a function of other primary variables of known distributions. The parameters θ are usually related to the distributions of the variables $X_1, X_2, \ldots$.

Assuming we need to estimate a parameter $\theta \in \Theta$, the method follows Definition 6.1: if z_{meas} is the measured value (often an estimator of θ obtained from a sample of size n), *for each* value of θ, with the methods described in the previous paragraph, a histogram representative of the density $p_Z(z;\theta)$ must be obtained. The conditions defined by the integrals (6.7) are approximated by counting, for each generated histogram, the fraction f (frequency) of events for which $z \le z_{\text{meas}}$. The extremes of the confidence interval $[\theta_1, \theta_2]$ are found when:

$$1 - f \simeq \int_{z_{\text{meas}}}^{\infty} p_Z(z;\theta_1)\,\mathrm{d}z = c_1\ , \quad f \simeq \int_{-\infty}^{z_{\text{meas}}} p_Z(z;\theta_2)\,\mathrm{d}z = c_2\ . \tag{8.82}$$

If the parameter is continuous, a discrete grid is considered, choosing a step $\Delta\theta$ small enough to meet the required precision.

In Eq. (8.82), the equality is replaced by the symbol $\simeq$, which means *values within the statistical error*. If the required precision is defined in term of standard deviation, from Eq. (6.27), it follows that a number n of observations must be simulated until the errors:

$$\sqrt{\frac{c_1(1-c_1)}{n}}\ , \quad \sqrt{\frac{c_2(1-c_2)}{n}} \tag{8.83}$$

assume the requested value.

Let's suppose, for instance, that one needs to determine a standard symmetric confidence interval with $CL = 0.683$; in this case $(1 - CL)/2 = c_1 = c_2 = 0.158$.

If we generate a histogram with 100 000 events, about 15 800 events will fall under the tails, and, from Eq. (8.83), the statistical error of f would be about 0.001.

When the density $p_Z(z, \boldsymbol{\theta})$ depends on a set of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, one builds a multidimensional grid for all parameters and records the sets of values for which, as in Eqs. (8.82), the chosen confidence levels are satisfied.

The method based on Eqs. (8.82), known as the *grid method*, is fully general but often rather cumbersome. However, in some cases it is possible, when the Z density is invariant with respect to the parameters $\boldsymbol{\theta}$, to proceed much easier and more directly.

Let us first recall the results obtained in Sect. 6.2. If Eq. (6.10) holds, the second of Eqs. (8.82) becomes:

$$c_2 = \int_{-\infty}^{z_{\text{meas}}} p(z; \theta_2)\, \mathrm{d}z = F(z_{\text{meas}}; \theta_2) = 1 - F(\theta_2; z_{\text{meas}}) \ .$$

This equation, when F is invertible, can be solved with respect to θ_2:

$$\theta_2 = F^{-1}(1 - c_2; z_{\text{meas}}) \ . \tag{8.84}$$

Analogously, for θ_1 one obtains:

$$\theta_1 = F^{-1}(c_1; z_{\text{meas}}) \ . \tag{8.85}$$

In other words, we need to evaluate the quantiles of order $1 - c_2$ and c_1 of the F distribution with parameter z_{meas}. When F is not analytically invertible, θ_1 and θ_2 can be easily obtained by simulating a histogram of this distribution.

Equations (6.9, 6.10) hold in the Gaussian case. Therefore, if Z has mean θ (as in the case of an estimator of the mean), Eq. (6.10) becomes:

$$F(z_{\text{meas}} - \theta) = 1 - F(\theta - z_{\text{meas}}) \ ,$$

where F is the cumulative function of a zero mean Gaussian, which is symmetric around zero. If $CL = 1 - \alpha$ and $c_1 = c_2 = \alpha/2$, Eqs. (8.84) and (8.85) give:

$$\theta_1 - z_{\text{meas}} = F^{-1}(\alpha/2), \quad \theta_2 - z_{\text{meas}} = F^{-1}(1 - \alpha/2) \ ,$$

that is:

$$\theta_1 = z_{\text{mas}} + t_{\alpha/2}, \quad \theta_2 = z_{\text{meas}} + t_{1-\alpha/2}$$

(see Fig. 8.13, with $\mu = \theta_1$ and $x = z_{\text{meas}}$). Basically, to solve the problem with the MC technique, we focus on the measured value and find the location of θ_2 from the density tail, simulating a sample with a Gaussian density $g(z; z_{\text{meas}})$ and evaluating θ_2 through the histogram. We proceed in a similar way for θ_1.

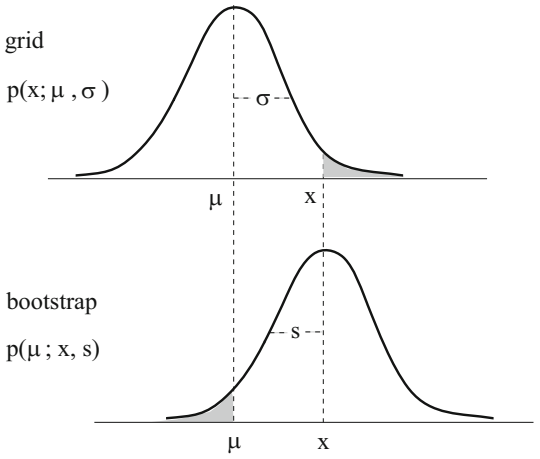


Fig. 8.13 Determination of the confidence interval, when the property (6.10) holds, that is in the case of symmetry and shape invariance of the density with respect to parameters under study (here μ). With the grid method (upper curve), the value of μ is found when the shaded area matches with the assigned CL (usually $1 - CL$ o $(1 - CL)/2$). With the bootstrap method (lower curve), a histogram from the density which has as parameter the measured value x is obtained. Then, the confidence interval of μ is found as if it were a probability interval. The shaded area under the tail is the same

This procedure (see [Buc84, DH99, Rip86]), which replaces the true parameters with the estimated ones, is part of a class of MC algorithms known as "bootstrap" which we will in more detail in Sect. 8.12.

Let us now apply these general principles to the concrete case of Eq. (8.77). If Eqs. (8.78) refer to a measurement, the true quantities θ_x and θ_y must be replaced with the experimental values t_x and t_y. We will suppose that the quantity $z \equiv n_{\text{meas}} = (\sin t_x / \sin t_y) = 1.52$ is the value of the refractive index n obtained measuring the directions of the light rays with an optical goniometer that has a resolution of $\pm 3°$ or of $\pm 0.5°$. The result, as is usually the case for laboratory measurements, must be given at a confidence level of 68.3% centred on the measured value $n_{\text{meas}} = 1.52$. Let us consider, to begin with, the case of large errors ($\sigma = 3°$). To apply the grid method, since the measured angles of the Gaussian variables are independent, it is necessary to calculate the distribution of $\sin x / \sin y$ where X and Y are Gaussian variables with means θ_x and θ_y and standard deviation equal to the measurement error of 3° and determine the two refractive indices for which the measured value $n = 1.52$ is respectively the quantile of value α and $1-\alpha$, with $\alpha = 0.1585$. The calculation must be performed for all possible values of θ_x and θ_y. The result, obtained with the `MCgrid` routine and also shown in Fig. 8.14, gives the interval:

$$n \in [1.10, 1.95] = 1.52^{+0.42}_{-0.43}\,, \qquad CL \simeq 68.3\%\,. \tag{8.86}$$

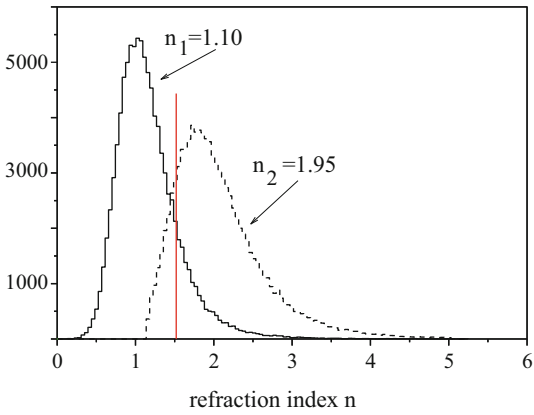


Fig. 8.14 Simulated samples from the density $p(n;\theta_x,\theta_y)$ for the values of θ_x and θ_y giving the upper and lower limits of the interval (8.86). The vertical line represents the measured value

Now let us apply the bootstrap method: in Eqs. (8.84, 8.85), the values θ_x and θ_y of Eq. (8.78) are replaced by the measured angles t_x and t_y, and the histogram of $Z = \sin X/\sin Y$ is generated with X and Y sampled from the Gaussian densities $g(x;t_x,\sigma_x)$ and $g(y;t_y,\sigma_y)$. This histogram is nothing but the one already displayed in Fig. 8.12a that should be considered as sampled from a population of density $p(n;t_x,\ t_y)$. Using 100 000 simulated events, with the `Bootgrid` routine , 15 850 events have been obtained inside the intervals $(-\infty, 1.17]$ and $[2.03, +\infty)$:

$$n \in [1.17, 2.03] = 1.52^{+0.51}_{-0.35} \quad , \qquad CL \simeq 68.3\% \, . \tag{8.87}$$

This result is different from the correct one given by Eq. (8.86). The reason for the discrepancy is the strong asymmetry of the distribution, as seen from Fig. 8.14.

In the situation shown in Fig.8.12b, corresponding to an error of $\pm 0.5°$, the measurement density assumes an invariant form within the considered angular range and the confidence interval becomes symmetrical around the measured value. In this case, both the methods previously described, grid and bootstrap, give the same result:

$$n \in [1.45, 1.59] = 1.52 \pm 0.07 \, , \quad CL \simeq 68.3\% \, ,$$

where the uncertainty of the interval extremes, which is of some part per thousand, is neglected. The simulation shows that, in order to obtain a reliable measurement, it is advisable to have measurement errors of the order of half a degree. This is the typical accuracy of ordinary optical goniometers. In Sect. 12.9 we will apply simulation techniques to the extremely important case of the propagation of measurement errors.

8.11 Simulation of Counting Experiments

In Sects. 6.6 and 6.8, we have introduced the confidence intervals for the estimation of probabilities and frequencies. Here we want to explore, using simulation, the statistical coverage properties of these intervals. Over the past 10 years, these results have changed the approximate formulae to be used in counting experiments that are presented in many statistics books.

Given a measured frequency $f = x/n$, obtained from x successes in n trials, the interval containing the true value of the probability with a confidence level CL can be evaluated with Eqs. (6.32, 6.18, 6.19, 6.37) that we report again here for convenience:

- The Clopper-Pearson general formula for the binomial case, which is the frequentist estimate $p \in [p_1, p_2]$, where $CL = 1 - c_1 - c_2$ and p_1, p_2 are the solutions of the equations:

$$\sum_{k=x}^{n} \binom{n}{k} p_1^k (1-p_1)^{n-k} = c_1 \,, \qquad (6.18)$$

$$\sum_{k=0}^{x} \binom{n}{k} p_2^k (1-p_2)^{n-k} = c_2 \qquad (6.19)$$

- The Gaussian approximation with continuity correction, where $f_\pm = (x \pm 0.5)/n$ gives the Wilson interval:

$$p \in \frac{f_\pm + \dfrac{t^2_{\alpha/2}}{2n}}{\dfrac{t^2_{\alpha/2}}{n} + 1} \pm \frac{|t_{\alpha/2}| \sqrt{\dfrac{t^2_{\alpha/2}}{4n^2} + \dfrac{f_\pm(1-f_\pm)}{n}}}{\dfrac{t^2_{\alpha/2}}{n} + 1} \qquad (6.37)$$

- The large sample approximation, which gives the well-known Wald interval, presented in any elementary statistics handbook:

$$p \in f \pm |t_{\alpha/2}| \sqrt{\frac{f(1-f)}{n}} \qquad (6.32)$$

Given a number of attempts n and a probability p, the simulation code `MCbinocov` (given below) randomly extracts a number of successes x with the `rbinom` routine, calculates the confidence intervals with the formulae just considered and counts how many times there is a success, that is, when the true value p is included in the interval.

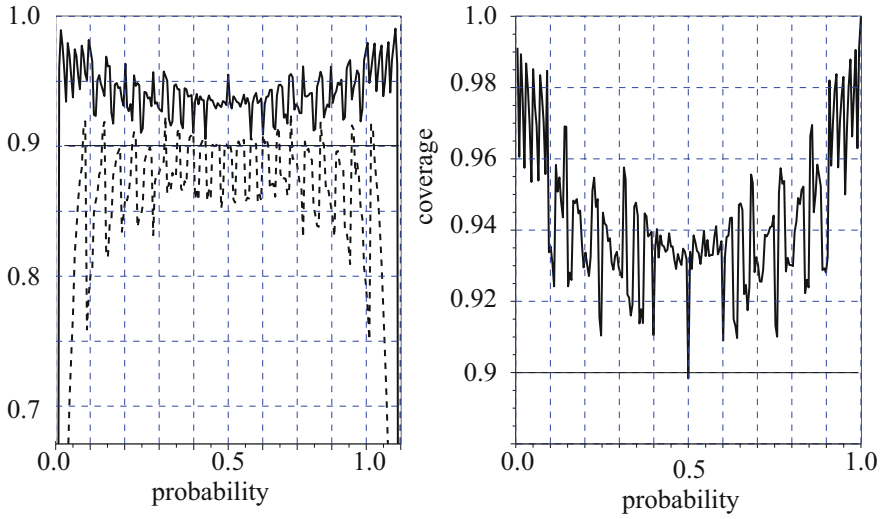


Fig. 8.15 Coverage curves for the binomial distribution, with $n = 30$ and $CL = 0.90$, for the frequentist (Clopper-Pearson) confidence intervals (6.18, 6.19) (right box), for the Wilson interval (6.37) (full curve in the left box) and for the approximated (Wald) interval (6.32) (dashed curve in the left box). The Clopper-Pearson and the Wilson formulae provide equivalent coverages

The coverage curve is obtained by repeating the procedure 10,000 times for each value of p and plotting, with the `plot` routine, the fraction of inclusions. The values p_1 and p_2 of the frequentist formulae (6.18, 6.19) are obtained by inversion of the cumulative of the binomial distribution from the R routine `binom.test`. The numerical methods performing this inversion are also described in [ZeaPDG20, PFTW92]. For the limiting cases $x = 0$ and $x = n$, Eqs. (6.22, 6.23) are used.

The coverage curves are shown in Fig. 8.15 for a binomial distribution with $n = 30$ for intervals with $CL = 0.90$, which correspond to a Gaussian quantile $|t_{\alpha/2}| = t_{1-\alpha/2} = 1.645$, where $\alpha = 1-CL$. The structure of the curves appears irregular due to the discrete value of the variable examined, namely, the number x of successes.

The result, quite surprising, shows how the approximate formula (6.32) (dashed curve in the left box of Fig. 8.15) provides an absolutely unsatisfactory coverage, almost always well below the assigned confidence level. Contrary to usual practice, this formula should therefore only be used for quite large samples, at least with $n > 300$ [Rot10]. The frequentist interval of Eqs. (6.18, 6.19), shown in the right box, despite the irregular behaviour of the curve, provides a correct over-coverage, with values always above the chosen confidence level. As noticed before, this over-coverage is simply due to the presence of the x value in both sums (6.18, 6.19).

The simulated results also clearly show the difference between confidence level and coverage in the case of discrete variables. Another interesting result is that the Wilson interval with the continuity correction given by Eq. (6.37), which

can be easily calculated, provides a coverage equivalent to that of the correct frequentist formulae, which, on the contrary, require the use of statistical software or ad hoc programs to be calculated. In conclusion, the simulation shows that the use of Eq. (6.37) should be much more widespread than it is now, because *this formula, unlike Eq. (6.32), provides reliable confidence intervals already for* $n > 10$ [BCD01, Rot10]. Here you find the routine used for the previous tests. We suggest you to try it with different input parameters for a comprehensive check of all these approaches.

```
#   MCbinocov(nsim,N,conf,grid,scale,wald,wilson,clopper):
#   check of the coverage of the
#   Clopper-Pearson, Wilson and Wald formulae
#   INPUT:
#   nsim = number of simulated events
#   N = number of trial of the binomial
#   conf = [0,1] confidence level
#   grid = when TRUE a grid is made on the plots
#   scale = scale*conf is the lower limit of the final plots
#   wald, wilson, clopper = when TRUE the plots are drawn
#    OUTPUT:  plots of the coverages
#
MCbinocov<- function(nsim=1000,npts=100,N=20,conf=0.68,grid=TRUE,
                     scale=0.8,wald=TRUE,wilson=TRUE,clopper=TRUE){
 t = qnorm(0.5*(1+conf))
 p   <- seq(0.0,1.,length.out=npts)
 cov <- seq(0.,0.,length.out=npts)
 waldcov <- seq(0.,0.,length.out=npts)
 wilscov <- seq(0.,0.,length.out=npts)
 for(j in 1:npts){    # points of the plots
   for(k in 1:nsim){  # event simulated at each point
     x = rbinom(1,size=N,prob=p[j])
     f= x/N
     fp=min(1,(x+0.5)/N)
     fm=max(0,(x-0.5))/N
     b1 = binom.test(x,N,conf.level=conf)$conf.int[1] # Clopper
     b2 = binom.test(x,N,conf.level=conf)$conf.int[2]
     w1 = f - t*sqrt(f*(1-f)/N) # Wald
     w2 = f + t*sqrt(f*(1-f)/N)
     ww1 = (fm+(t^2/(2*N))-
                    t*sqrt(fm*(1-fm)/N + t^2/(4*N^2)))/(1+t^2/N)
     ww2 = (fp+(t^2/(2*N))+
                    t*sqrt(fp*(1-fp)/N + t^2/(4*N^2)))/(1+t^2/N)
     ww1 = max(0.,ww1)  # wilson
     ww2 = min(1.,ww2)
     if(b1 <= p[j] && p[j] <= b2) cov[j]=cov[j]+1
     if(w1 <= p[j] && p[j] <= w2) waldcov[j]=waldcov[j]+1
     if(ww1 <= p[j] && p[j] <= ww2) wilscov[j]=wilscov[j]+1
  }
 }
 cov <- cov/nsim          # coverages
 waldcov <- waldcov/nsim
 wilscov <- wilscov/nsim
 if(clopper==TRUE){             #plots
   title = paste("_____ Clopper  ---- (red) Wald --- (blue) Wilson")
```

```
    plot(p,cov,type='l',lwd=2,ylim=c(scale*conf,1),main=title)
    if(wald==TRUE)lines(p,waldcov,lty=2,lwd=2,col='red')
    if(wilson==TRUE)lines(p,wilscov,lty=2,lwd=2,col='blue')
  }
  else if(clopper==FALSE && wilson==TRUE){
    title = paste("  ----(red) Wald       _______Wilson")
    plot(p,wilscov,type='l',lwd=2,ylim=c(scale*conf,1),main=title)
    if(wald==TRUE)lines(p,waldcov,lty=2,lwd=2,col='red')
  }
  else{
    title = paste("  _______ Wald ")
    plot(p,waldcov,type='l',lwd=2,ylim=c(scale*conf,1),main=title)
  }
  if(grid==TRUE)grid()
  abline(h=conf,col='red')
}
```

If we need to estimate a counting frequency μ starting from a measured count x, the equations to use are (6.41, 6.44, 6.45), that, for convenience, we write again here:

- The general formula for the Poisson case is the frequentist estimate of the interval $\mu \in [\mu_1, \mu_2]$, where the values μ_1, μ_2 are the solutions of the equations:

$$\sum_{k=x}^{\infty} \frac{\mu_1^k}{k!} \exp(-\mu_1) = c_1 \ , \quad \sum_{k=0}^{x} \frac{\mu_2^k}{k!} \exp(-\mu_2) = c_2 \ . \quad (6.41)$$

- Using the Gaussian approximation with continuity correction ($x_\pm = x \pm 0.5$), one obtains:

$$\mu \in x_\pm + \frac{t_{\alpha/2}^2}{2} \pm |t_{\alpha/2}| \sqrt{x_\pm + \frac{t_{\alpha/2}^2}{4}} \ . \quad (6.44)$$

- When x is large, Eq. (6.44) can be replaced by:

$$\mu \in x \pm |t_{\alpha/2}| \sqrt{x} \ . \quad (6.45)$$

The coverage curves for these three intervals, for $\mu \leq 30$, are shown in Fig. 8.16. These plots have been obtained with our `MCpoisscov` routine, not reported here, since its structure is similar to `MCbinocov` that has been detailed just above.

As in the binomial case, we also notice here that the coverage given by the approximate formula (6.45) is absolutely unsatisfactory. The frequentist formulae and the one under Gaussian approximation with continuity correction are practically equivalent and give a good over-coverage (full curves in Fig. 8.16). Equation (6.44) gives a result more appropriate than the approximate formula (6.45), which should only be used when the average exceeds a few dozen events [Rot10].

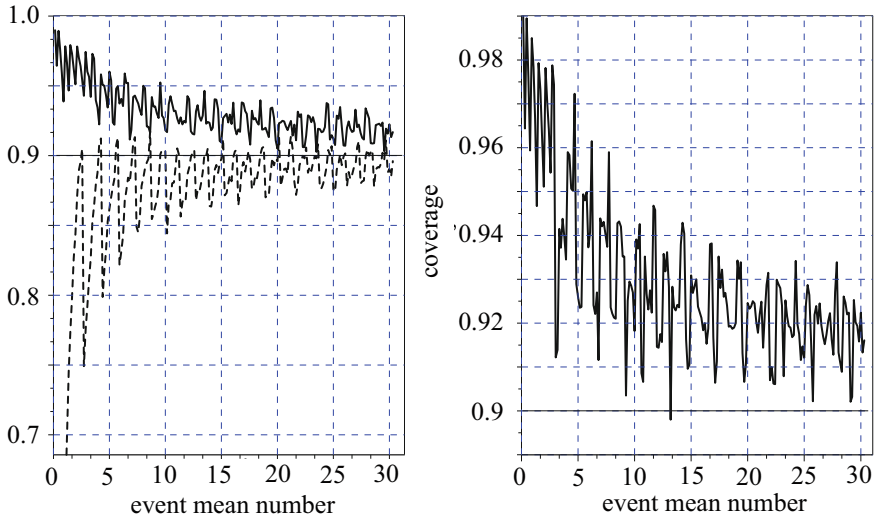


Fig. 8.16 Coverage curves, for the Poisson density with $CL = 0.90$, for the frequentist confidence intervals (6.41, 6.19) (right box), for the interval of Eq. (6.44) (full curve in the left box) and for the approximated interval (6.45) (dashed curve in the left box). The frequentist formulae and Eq. (6.44) give equivalent coverages

8.12 Non-parametric Bootstrap

In the previous sections, we have seen how to numerically derive the properties of a stochastic variable through random samples obtained with parametric probability density models. Often, if no information is available, the estimated values instead of the true ones are assigned to the model parameters (means, variances or other). This method, known as *parametric bootstrap*, can be further generalized to those problems where *there is no information on the probability density of the random variable to be examined* [DH99, DE83, Efr79, Efr82, ET93, PFTW92]. To fully understand this *non-parametric* bootstrap technique, let us generate a sample of 100 standard Gaussian variables and calculate their mean:

```
> x <- rnorm(100)
> mean(x)
> 0.01365425
```

We now proceed to the bootstrapping of the sample, generating N new samples of the same dimension of the original one ($n = 100$). We sample, with replacement, the elements of the original set of values. In R this operation can be performed by many routines, `sample` among others, through the call `sample(x,size=n,replace=TRUE)`. This routine is also often used to permute elements of a vector, with the call `sample(x, size=length(x))`, when the `replace` parameter is set by default at the value `FALSE`. It should

be immediately noticed that the bootstrap samples thus generated differ from the original one due to the presence of repeated elements. Actually, this apparently trivial fact is the core of the method.

We now proceed by generating $N = 1000$ bootstrap samples, loading their means in a vector `boot`, with the in line statements:

```
> boot <- seq(0,0,length.out=1000)
> for(j in 1:1000) boot[j] = mean(sample(x,size=100,replace=TRUE))
```

Now let us calculate the mean and standard deviation of the sample of the means:

```
> mean(boot)
> 0.01502075
> sqrt(var(boot))
> 0.1005019
```

The first result coincides, within the statistical error, with the mean `mean(x)` of the initial set of values, while the second one represents the surprise: *it corresponds to the statistical error of the sample mean* given by Eq. (6.50) as it can be easily check with the R command `var(x)/sqrt(99)`. In other words, without applying any statistical theory, the standard deviation of the means of the bootstrap samples gives the standard deviation of the mean of `x`. This example indicates two essential aspects of the non-parametric bootstrap:

1. The mean values of the bootstrap samples usually do not give new information, because they are distributed around the initial experimental values. However, significant differences, larger than the statistical error, can sometimes occur between the original and the bootstrap mean. In this case, the difference is called bias and can be used to correct the confidence interval, as will be discussed later.
2. The dispersion of the bootstrap samples is an estimate of the dispersion of the parameter under consideration. Thus, bootstrap is very useful, for example, when studying the variances of some complicated quantities.

Let us now check the reliability of the method in the more difficult case of the variance. We generate a sample of $n = 100$ Gaussian variates with $\mu = 0$ and $\sigma = 2$:

```
> y <- rnorm(100,sd=2)
> var(y)
> 4.761325
```

As previously mentioned, both in this and in the other similar cases, if you repeat the exercise, you would get slightly different values due to the statistical fluctuations present in the simulated data. Let us now generate a bootstrap sample of variances:

```
> bootv <- seq(0,0,length.out=1000)
> for(j in 1:1000) bootv[j] = var(sample(y,size=100,replace=TRUE))
```

From this sample we find the two quantile values $q_{0.158}, q_{0.841}$, corresponding to a confidence interval with $CL = 0.683$ (equal to 1σ in the Gaussian case) using the R routine `quantile`:

```
> mean(bootv)
> 4.719306
> quantile(bootv,c(0.158,0.841),names=FALSE)
> 4.1585526  5.259678
```

Again one has `mean(bootv)` $\simeq$ `var(y)` within the statistical error, while the quantile values are very close to those of the exact formula (6.76):

```
> 99*var(y)/qchisq(0.841,df=99)
> 4.171343
> 99*var(y)/qchisq(0.158,df=99)
> 5.54933
```

Finally, let us go back to the Exercise 6.10 in which we determined a confidence interval for the true correlation coefficient of the chest/height data pairs given in Table 6.4, under the assumption that the bivariate probability density $p(x, y)$ of the data pairs was given by the two-dimensional Gaussian function. Let us now try to determine a confidence interval for the chest/height correlation coefficient without assuming a Gaussian density for these two variables and, therefore, not using Fisher's transformation.

With the bootstrap method, it is assumed that the obtained sample represents the *true* probability density of the data from which random samples are generated to estimate the parameters to be determined. Then, correlated pairs (s_i, t_i) of chest/height values are extracted (with replacement) from the experimental sample until a new "virtual" sample of 1665 elements is obtained. Since only the histogram data of Table 6.4 are available, to obtain bootstrap samples, it is first necessary to generate an approximate sample of raw data, which maintains the original data set structure. This is done in our `BootCor` routine by duplicating the chest and height data a number of times equal to the content of each bin of the two-dimensional frequency histogram. For example, pairs of 88 cm chest and 166 cm height will be repeated 114 times and so on. From this sample of 1665 "original" data, bootstrap samples are then created to estimate of the correlation coefficient r^*_{st}, obtained through Eq. (6.117) with the same operations of Exercise 6.10.

By repeating this operation for a sufficiently large number of times, we obtain a fairly large sample of coefficients r^*_{st}, which allows the determination of the confidence interval for the correlation coefficient.

The histogram obtained with 10 000 different simulated values of r^*_{st} with our `BootCor` routine is shown in Fig. 8.17. From these data the following confidence interval is obtained:

$$r^*_{st} \in [0.221, 0.309] = 0.266^{+0.043}_{-0.045} \quad (CL \simeq 95.4\%) \,,$$

which is exactly the same as that found in Exercise 6.10!

After these examples, it is time to ask ourselves when and why bootstrap works. The generation of fictitious or artificial samples starting from the real sample (and having the same size n) is equivalent to replace the unknown probability density $p(\boldsymbol{x})$ with a discrete probability distribution $p^*_n(\boldsymbol{x})$ with n components and to

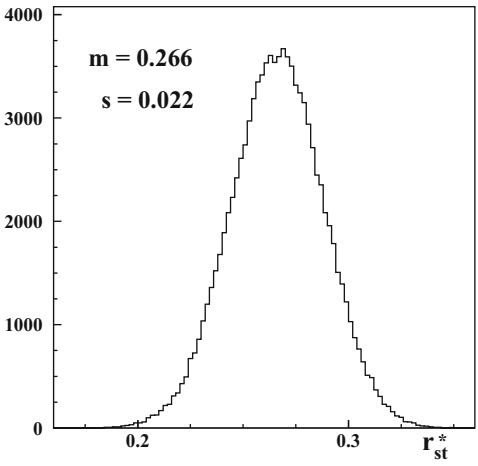


Fig. 8.17 Histogram of r^*_{st} obtained with 10,000 bootstrap samples

assign the same probability $1/n$ to each observed value. From the artificial sample $(\boldsymbol{x}^*_1, \ldots, \boldsymbol{x}^*_n)$ obtained from this distribution, a summary value t^* is obtained. By repeating this operation r times, a sample of values $t^*_1, t^*_2, \ldots, t^*_r$ is collected, which allows to estimate the property of the statistic $T = t(\boldsymbol{X_1}, \ldots, \boldsymbol{X_n})$, which encloses the information about the θ parameter.

Consider now the cumulative function $F^*_n(\boldsymbol{x})$ of $p^*_n(\boldsymbol{x})$, which assumes the values $(0, 1/n, 2/n, \ldots, n/n)$. In a formal way, it can be written as:

$$F^*_n(\boldsymbol{x}) = \frac{\#(\boldsymbol{x}_i \leq \boldsymbol{x})}{n} \,, \tag{8.88}$$

where the symbol # denotes the number of times the condition in bracket is verified. Using Eq. (8.88), it is easy to recognize that F^* follows a binomial distribution with $p = F(\boldsymbol{x})$, where F is the cumulative of the true, but unknown, density $p(\boldsymbol{x})$. From these considerations, it follows that:

$$\begin{aligned} E[F^*(\boldsymbol{x})] &= nF(\boldsymbol{x}) = np \,, \\ \mathrm{Var}[F^*_n(\boldsymbol{x})] &= \frac{1}{n}\{F(\boldsymbol{x})\,[1 - F(\boldsymbol{x})]\} = \frac{p(1-p)}{n} \,. \end{aligned} \tag{8.89}$$

Remembering the Central Limit Theorem, it is easy to derive that, when n increases, $F^*_n(\boldsymbol{x})$ tends both to be Gaussian distributed and to better and better approximate the true cumulative $F(\boldsymbol{x})$. The correctness of the bootstrap method relies on this simple property, which at first glance may seem almost miraculous![1]

[1] Indeed, due to its outrageous simplicity, this method, after the well-known works of Efron [Efr79, Efr82], took some time to be adopted by statisticians.

As we have just seen, in non-parametric bootstrap $F(\boldsymbol{x})$ is replaced by $F_n^*(\boldsymbol{x})$. The error introduced by this type of approximation is both due to the difference $(\hat{t} - \theta)$ between the true value θ of the parameter and its correct statistical estimate $\hat{t}$, and to the difference $(\hat{t} - \hat{t}^*)$, where $\hat{t}^*$ is evaluated from the simulated bootstrap sample. For the variance estimation, the method is based on the validity of the condition:

$$\hat{t} - \theta \simeq \left\langle \hat{t}^* \right\rangle - \hat{t}^* \,, \tag{8.90}$$

that is on the approximated similarity between the bootstrap data dispersion around their mean and the estimator dispersion around the true value. The method also allows us to verify a possible discrepancy between the average of the bootstrap data and the estimated value $\hat{t}$, determined by the difference, called *bias*:

$$\hat{t} - \left\langle \hat{t}^* \right\rangle \,, \tag{8.91}$$

which may be due both to the finite size of the experimental and bootstrap samples (n and r, respectively) and to the particular estimator T used. For example, if T is not a linear function, it is usually not true that, asymptotically, $\hat{t}^* = \hat{t}$, and this leads to a systematic bias in the bootstrap estimate.

The size n of the bootstrap samples is usually set equal to that of the initial experimental sample, while the number r of bootstrap samples (called replication), from which the parameter variance is estimated, is generally gradually increased until the statistical error becomes negligible for the problem under study, and the solution appears stable. Usually a number of replications $300 < r < 1000$ is adequate for this purpose. The systematic error or bias can be corrected (see, for instance, [DH99]) but is independent of r.

To ensure the validity of Eq. (8.90), it would be necessary to find pivotal estimators (that have the same distribution both with respect to $F_n^*(\boldsymbol{x})$ and to $F(\boldsymbol{x})$), but this is not always feasible. A list of the various methods that are successfully applied in these cases can be found in [DH99]. In conclusion, whenever the statistical properties of a certain parameter to be estimated from a sample are not known, *the confidence interval can be estimated with the bootstrap method*. However, keep in mind that you cannot know in advance what are the "bad" bootstrap samples, and this ultimately remains the main limitation of the method. These applicability issues are discussed in detail in [DH99]. We suggest, in order to check if the bootstrap method is applicable to your specific case, to test the procedure on simulated data, verifying how much the results obtained are compatible with the true values, which are known in this case. An example of this procedure can be found in our `CovarTest` routine, where we compare the variances of the covariance of simulated data calculated using *bootstrap* and Eq. (6.116).

8.13 Hypothesis Test with Simulated Data

As we have seen, for large samples and for Gaussian samples of any size, there are general methods for both estimations and hypothesis testing. On the contrary, the field of small non-Gaussian samples, for which it is not possible to formulate a general theory, remains still open. It is then not surprising that simulation techniques are generally used to solve this type of problems, even in the case of hypothesis testing.

As an example, let us consider the t-test on means or pairs of values with the *permutation test*. Suppose we want to check the compatibility of the means m_x and m_y of two samples $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_m)$ generated from an unknown parent population. We construct a vector $\boldsymbol{z} = (x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m)$ of dimension $(n + m)$, simply obtained by joining the two initial vectors. Under the hypothesis H_0 of equality of the true means, the sample $\boldsymbol{z}$ represents a homogeneous sample coming from common mean populations. At this point, the unknown distribution of the difference between the means is estimated with the following steps:

(a) Permute the vector $\boldsymbol{z}$.
(b) Calculate the means of both the first n elements $(z_i, i = 1, ..n)$ and of the last m elements $(z_i, i = n + 1, ..n + m)$ of the permuted vector z. The difference between these means is calculated, and its (absolute) value is recorded.
(c) Repeat operations (a) and (b) R times to obtain a sample of the differences $d^* = |m_x^* - m_y^*|$. This sample is assumed as the difference sample drawn from the population under H_0.
(d) The p-value of the difference $d = |m_x - m_y|$ is estimated from the difference sample generated in (c) with the formula:

$$p = \frac{\#(d^* > d)}{R} , \tag{8.92}$$

where, as before, # is the number of times the condition in bracket is verified.

This method is generally classified among the non-parametric bootstrap techniques, because the reference population is estimated using real data. Since the test is usually two-tailed, the absolute value of the differences in the algorithm is considered.

We also note that in the permutation test the pure difference has been considered, without dividing it by the total standard deviation, because the method assumes that two original samples have comparable variances, so that they can be exchanged without affecting the results. Small differences can be tolerated by the method when the data are mixed during the permutations.

Many R routines allow to perform the permutation test, one of them is `twot.permutation`, from the library `DAAG`. We wrote the routine `BootPermTest`, whose core is the following:

```
n1 = length(vec1)
n2 = length(vec2)
texp= abs(mean(vec1)-mean(vec2)) # experimental value
if (tmedian==TRUE)  texp= abs(median(vec1)-median(vec2))
pool <- c(vec2,vec1)  # global pooled vector under H_0
for(j in 1:Nperm){
 permut <- sample(pool)   #  permutation of the pooled vector
 if(median == TRUE){
  pdiff[j] = abs(median(permut[1:n1])-median(permut[(n1+1):(n1+n2)]))
 }
 else {
  pdiff[j] = abs(mean(permut[1:n1])-mean(permut[(n1+1):(n1+n2)]))
 }
}
   # p-value: sum  the TRUE cases over the total
pval = sum(pdiff > texp)/Nperm
```

The routine receives in input the two raw data vectors `vec1` and `vec2` and also offers the possibility to evaluate the difference between the medians, through the `tmedian` parameter. This possibility, due to the flexibility of the simulation methods, is often useful in the case of long-tailed distributions, since in this case the median is a more stable (robust) parameter than the mean. We also note that, to obtain correct results, it is important to calculate the two means *from the same permutation*. If you try to compare this routine with the *t*-test of `t.test`, you will see that the results are practically similar for large samples and for Gaussian or quasi-Gaussian samples. However, if we sample from a negative exponential distribution with `vec1 <- rexp(5,rate=1)` and `vec2 <- rexp(5,rate=7)`, we see that the Student's test of `t.test` fails because it gives *p*-values around 15%, while `BootPermTest` gives *p*-values around 4% for both the mean and the median, a clear hint of a possible systematic difference between samples.

Sometimes, in hypothesis testing, simulated artificial data are needed instead of real ones. We explain this case by reconsidering Exercise 7.4, where we have to deal with small and binomial-distributed samples that are quite far from the Gaussian case. We can solve the problem by inserting the data into two-valued vectors `x` (sick rats) and `n` (all rats) and generating `N` results from two binomials having the same pooled probability `fhat` of Eq. (7.9). These steps are coded in the R instructions:

```
fhat = sum(x)/sum(n)    # pooled frequency
b1 <- rbinom(N,n[1],fhat)
b2 <- rbinom(N,n[2],fhat)
if(x[1]>x[2]) pval= sum(b1-b2>=x[1]-x[2])/N
if(x[1]<=x[2] pval= sum(b1-b2<=x[1]-x[2])/N
```

which are part of our `MCDiffProp` routine. As can be seen, this routine counts the events where the data generated from the binomials have a difference greater than the observed difference `(x[1]-x[2])`. The *p*-values obtained with this method, for example, with the call `MCDiffProp(x=c(4,19), n=c(817,1631))` (see the table in Exercise 7.4), are around 7% for both the

considered tumours, a value significantly higher than those obtained before with the Gaussian approximation in Exercise 7.4.

We think that the previous examples clearly explain this technique and will enable you to apply it in other similar cases.

8.14 Problems

8.1 Simulate the behaviour of a player who bets on lottery numbers that have been delayed for more than 90 weeks. Does this strategy increase the chance of winning?

8.2 Solve Monty Problem 1.1 with a simulation code.

8.3 Solve the encounter problem 1.7 with a simulation code.

8.4 Write a simulation code that calculates the probability that the distance between two randomly drawn points uniformly within a circle is less than the radius.

8.5 Generate, with Algorithm 8.4, random variates x following the so-called half-normal probability:

$$p(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2} \qquad \forall x \geq 0$$

by using the function $f(x) = ke^{-x}$. Find the k value that gives the highest generation efficiency. With the value of x thus obtained, generate a variable $Z \sim N(0, 1)$.

8.6 Determine the confidence levels of Eq. (6.31) for $n = 5$, $p = 0.25$ and $t = 1, 2, 3$ with a simulation code.

8.7 Write an algorithm for the generation of Gaussian deviates using the Central Limit Theorem.

8.8 Generate a pair of standard normal variables X and Y having linear correlation coefficient ρ.

8.9 How can the rejection algorithm be used to randomly generate points uniformly distributed inside a circle of radius R?

8.10 Generate the histogram of the variable $Z = Y/X$ where X and Y are standard Gaussian variables.

8.11 The height of a homogeneous population is a Gaussian variable with $\langle X \rangle \pm \sigma[X] = 175 \pm 8$ cm for men and $\langle Y \rangle \pm \sigma[Y] = 165 \pm 6$ cm for women. Knowing that there is a positive correlation with $\rho = 0.5$ between the height of husband and wife (couples tend to have a similar height), find, by using a simulation code, the percentage of couples with male and female taller than 180 and 170 cm, respectively (see also Problems 4.6, 4.7).

8.12 Solve Problem 5.9 with a simulation code.

8.13 Parallel lines are drawn on an infinite plane at a unitary distance. A unit length needle is randomly thrown on the plane. It can be shown that the probability that the needle falls across a line is $p = 2/\pi$ (Buffon's needle problem [Gne76]). Estimate the value of π with a Monte Carlo code.

8.14 In the right-left (or top-bottom) asymmetry problem, events can occur "to the left" with probability P or "to the right" with probability $1 - P$. Simulate 5000 left-right experiments and count, over $N = 50$ events, the number of times with n_s events on the left and $n_d = N - n_s$ events on the right. Make the histogram of the asymmetry $A = (n_s - n_d)/N$ and compare the standard deviation of the data with the true one evaluated in the Problem 12.8.

8.15 Using the `MCbinocov` routine with $CL = 95\%$, find the smallest value of n such that the approximate Eq. (6.32) gives a difference between coverage and $CL < 5\%$.

8.16 Using the `MCpoisscov` routine with $CL = 95\%$, find the smallest value of μ such that the approximate Eq. (6.45) gives a difference between coverage and $CL < 2\%$.

8.17 The average prices (in \$) of the shares (s) and the bonds (b) of the New York Stock Exchange in the years 1950–1959 are shown in the following table from [Spi61]:

	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
s	35.2	39.9	41.9	43.2	40.1	53.3	54.1	49.1	40.7	55.2
b	102.4	100.9	97.4	97.8	98.3	100.1	97.1	91.6	94.85	94.65

An economic theory predicts a negative correlation between stock and bond prices. Use the non-parametric bootstrap method to test the model at the 5% level.

8.18 Using the bootstrap method, find the confidence interval with $CL = 90\%$, for the odds ratio of Problem 6.16.

8.19 Write an MC code for the calculation of Tukey quantiles (7.66).

Chapter 9
Applications of Monte Carlo Methods

THUMB'S FIRST POSTULATE

It is better to solve a problem with a crude approximation and know the truth, plus or minus 10 percent, than to demand an exact solution and not know the truth at all.

Arthur Bloch, "MURPHY'S LAW BOOK TWO".

9.1 Introduction

The fields of application of the Monte Carlo (MC) methods are practically unlimited, and it is really difficult to deal organically with such a wide variety within the limited space the of a chapter.

However, we think it is useful to work out the detailed solution of a few general problems, starting from the overall framework up to the extreme detail of the simulation code. In this way, you will acquire a complete mastery of the procedures, so that you will be able to successfully implement and adapt these methods to your specific problems.

The examples presented in this chapter show the variety of contexts where MC methods can be applied: the process of particle diffusion in matter (a typical problem of experimental physics, chemistry or engineering), the calculation of the optimal number of workers in a plant (an instructive application to industrial management), some applications of the Metropolis algorithm (study of systems with a large number of identical components, which are of interest in economics, engineering, physics and chemistry) and, finally, the calculation of the value of definite integrals (here we are in the fields of theoretical physics and mathematics).

9.2 Study of Diffusion Phenomena

The propagation of particles in a given material, such as neutrons in a nuclear reactor, electrons in a metal or in a semiconductor and photons in the atmosphere,

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3_9

is usually referred to as *diffusion*. In this type of processes, each particle, starting from an initial state, follows a predetermined trajectory during a certain time period up to a random instant in which it undergoes an *impact*, i.e. an interaction, with the traversed material, with a possible production, in specific situations, of secondary particles. As a result, the particle changes, always randomly, direction and velocity modulus and continues its path until the next collision. The general scheme is that of Fig. 9.1.

From a theoretical point of view, this system can be described by the Boltzmann (or transport) equation, whose derivation is given in many advanced physics, chemistry or engineering texts, such as [Lam66].

Unfortunately, as J. Lamarsh correctly commented in the above reference, "it is much easier to derive the Boltzmann equation than to solve it", since it has a complex integro-differential form, where, due to the complicated configuration of real systems, there are rapidly varying parameters depending both on space and particle energy.

MC methods are thus almost always the only available approach to these problems, and, despite the great variety of diffusion processes, it is also possible to outline a simulation scheme of almost universal validity which, for example, is used by the simulation codes widely used in particle and nuclear physics, such as GEANT [A$^+$03] and MCNP [W$^+$18].

Let us then consider, as a useful practical example, the case of neutron diffusion within a material. Here, as for the γ rays and for the other electrically neutral particles, the calculation of the trajectories is very simple, since neutrons follow straight line paths between two successive collisions with the atomic nuclei of the traversed medium.

Let us consider, as in Fig. 9.1, a point source isotropically emitting neutrons with kinetic energy of 0.0038 eV,[1] located at the centre of a homogeneous sphere of infinite radius and composed of the diffusing material. The MC procedure that we present can be divided into various steps, which are common to all diffusive calculations (see also [RvN63]):

(a) *Geometry, materials and interactions* Since the medium in which the neutrons propagate is infinite and homogeneous, our system is very simplified, and it is therefore not necessary to introduce any geometric information. Instead, we need to know which parameters define the neutron interactions with atomic nuclei. As for the other particles, these processes are described by a quantity, called "total macroscopic cross section" (or, more briefly, Σ_T), which gives the probability, per unit path, of having any type of interaction. Σ_T is an intrinsic property of any material, is constant in homogeneous media and depends on the particle kinetic energy but not on the previous particle path. In the case of

[1] Electron-volt, or eV, is the energy unit widely used in atomic and nuclear physics. By definition, it is the kinetic energy acquired by an electron in passing through two points where a potential difference of 1 Volt exists. It is easy to show that $1\text{eV} = 1.60210^{-12}$erg. Neutron of 0.0038 eV are in thermal equilibrium with matter at the room temperature of 27 °C.

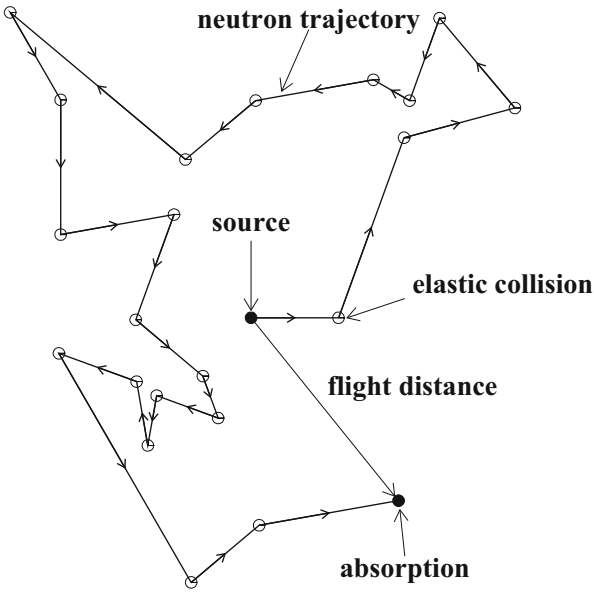


Fig. 9.1 Schematic representation of the trajectory followed by a thermal neutron during the diffusion process

a non monoenergetic source, it is therefore necessary to know, with sufficient precision, the behaviour of Σ_T in the considered energy interval. Suppose, as it is reasonable to do for thermal neutrons, that the possible interactions are of two types: absorption or elastic scattering (change of direction without loss of energy); with this hypothesis the total macroscopic cross section is obviously decomposed as:

$$\Sigma_T = \Sigma_a + \Sigma_{el} \ , \tag{9.1}$$

where Σ_a and Σ_{el} are the microscopic cross sections of the considered processes. Since the neutrons we are considering are monoenergetic, we need only two numbers, read as initial parameters, to quantitatively define their interactions. Other data that will be needed (and which will always be read at the beginning of the programme) are the mass number of the target nucleus (necessary for the calculation of the trajectories) and the speed of the neutrons, as we are also interested in determining the time needed for their absorption.

(b) *Kinematics* The neutron emission point is fixed and, for convenience, is set at the centre of the coordinate system. The flight direction of the neutrons exiting the source is determined, on the basis of Eq. (8.32), by randomly generating the angles φ and ϑ through the relations:

$$\begin{cases} \varphi = 2\pi\xi_1 \\ \vartheta = \arccos(1 - 2\xi_2) \ . \end{cases} \tag{9.2}$$

Then, the direction cosines are calculated:

$$\begin{cases} \alpha = \sin\vartheta\cos\varphi \\ \beta = \sin\vartheta\sin\varphi \\ \gamma = \cos\vartheta \end{cases} \tag{9.3}$$

which, with the knowledge of the starting point, allow us to unambiguously define the initial trajectory parameters of the particle.

(c) *Tracking* Based on the scheme of Fig. 9.1, a tracking step coincides with the flight distance, since the neutron does not have any other interactions between two successive collisions. This parameter can be calculated by noting that, according to the very definition of macroscopic cross section, the neutron interaction in any of the possible ways is a stochastic process having the same characteristics as those we have described in Sect. 3.7. In fact, the neutron, during its path in the traversed material, undergoes collisions (i.e. "generates"events) that are uncorrelated, discrete and with a constant probability Σ_T; then, similarly to Eq. (3.48), the probability of having a distance x between two successive collisions is:

$$p(x)\ \mathrm{d}x = \Sigma_T\, e^{-x\Sigma_T}\ \mathrm{d}x\,. \tag{9.4}$$

From this distribution, (whose mean $1/\Sigma_T$ is the mean free path between two successive collisions), and using Eq. (3.91) rewritten as:

$$d = -\frac{1}{\Sigma_T}\ln\xi_3\,, \tag{9.5}$$

it is possible to associate a distance d to each neutron, equal to the length of the simulation "step" between the starting point and the one where the next collision occurs.

Since we know both the emission angles and the covered distance, we are able to calculate the coordinates of the interaction point between the neutron and a nucleus of the diffusing material. Next, we need to evaluate the effects of this interaction on the trajectory of the particle, knowing that the probability to be absorbed is given by Σ_a/Σ_T, while that of being diffused is Σ_{el}/Σ_T. The choice between these two alternatives is made by drawing a new random number ξ_4: if $0 \le \xi_4 \le (\Sigma_a/\Sigma_T)$, the neutron is absorbed, and the current event ends; if, instead, $(\Sigma_a/\Sigma_T) < \xi_4 \le 1$, an elastic scattering occurs and a new flight direction is generated, as shown in Fig. 9.1.

For thermal neutrons, the collision process is isotropic with respect to the direction of incidence in the neutron-nucleus centre of mass system.[2] The azimuthal (φ_{cm}) and polar (ϑ_{cm}) emission angles must then be generated in this system:

$$\begin{cases} \cos\vartheta_{cm} = 1 - 2\xi_5 \\ \varphi_{cm} \quad\quad = 2\pi\xi_6 \end{cases} \tag{9.7}$$

and transformed into the laboratory system [Lam66]:

$$\begin{cases} \varphi \quad\quad = \varphi_{cm} \\ \cos\vartheta = \dfrac{1 + A\cos\vartheta_{cm}}{\sqrt{A^2 + 2A\cos\vartheta_{cm} + 1}}\,, \end{cases} \tag{9.8}$$

where A is the mass number of the target nucleus. Afterwards, the direction cosines of the new flight direction (α', β', γ') of the neutron are calculated as:

$$\begin{cases} \alpha' = \mu\alpha + a(\alpha\gamma\sin\varphi + \beta\cos\varphi) \\ \beta' = \mu\beta + a(\beta\gamma\sin\varphi - \alpha\cos\varphi) \\ \gamma' = \mu\gamma - a(1-\gamma^2)\sin\varphi \end{cases} \tag{9.9}$$

where:

$$a = \sqrt{\frac{1-\mu^2}{1-\gamma^2}} \quad ; \quad \mu = \cos\vartheta \quad ; \quad |\gamma| \neq 1\,. \tag{9.10}$$

If $|\gamma| = 1$, one instead has:

$$\begin{cases} \alpha' = \gamma b\cos\varphi \\ \beta' = b\sin\varphi \\ \gamma' = \gamma\mu \end{cases} \tag{9.11}$$

[2] Consider an set of N particles of masses $m_1, m_2, \ldots, m_N$ having velocities $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N$ respectively, in a certain reference coordinate system. The point defined with the equation:

$$\mathbf{r}_{cm} = \frac{\sum_{i=1}^{N} m_i \cdot \mathbf{r}_i}{\sum_{i=1}^{N} m_i}$$

(where $\mathbf{r}_i$ is the distance of the i-th particle from the origin of the coordinate system) is called centre of mass. Similarly, its velocity can be defined as:

$$\mathbf{v}_{cm} = \frac{\sum_{i=1}^{N} m_i \cdot \mathbf{v}_i}{\sum_{i=1}^{N} m_i}\,. \tag{9.6}$$

A reference system where $\mathbf{v}_{cm} = 0$ is called centre of mass system of the particle set. It allows, as in our case, to describe many nuclear reactions in a simple way.

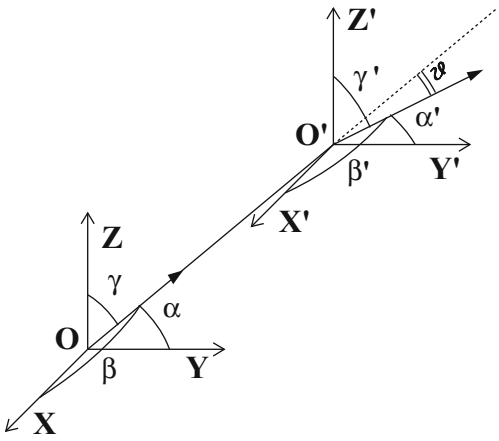


Fig. 9.2 The direction cosines (α, β, γ) and $(\alpha', \beta', \gamma')$ represent the neutron flight direction before and after the collision, respectively. The collision occurs at the point O', whereas ϑ is the polar angle of the scattered neutron with respect to the initial direction

with $b = \sqrt{1-\mu^2}$. The geometry of the collision process in the laboratory system is shown in Fig. 9.2. The demonstration of these transformations is in our web pages [RPP].

(d) *Event storage:* In the case of elastic scattering, the most relevant parameters of the neutron trajectory are updated before returning to the previous point and continuing to follow the particle path:

- The total distance d_i travelled between the source and the point where the last collision occurs (which we suppose to be the i-th one):
- The projections of d_i on the Cartesian axes:

$$x_i = x_{i-1} + d_i \alpha \tag{9.12}$$

$$y_i = y_{i-1} + d_i \beta \tag{9.13}$$

$$z_i = z_{i-1} + d_i \gamma \tag{9.14}$$

(obviously $x_0 = y_0 = z_0 = 0$)

- The flight time d_i/v (v is the velocity modulus of the considered neutrons; in our case $v = 2.2 \cdot 10^5$ cm/s);
- The total number of collisions.

On the contrary, when the neutron is absorbed, after k collisions, the quantity:

$$r = \sqrt{x_k^2 + y_k^2 + z_k^2} \tag{9.15}$$

is calculated, which is the distance between the source and the final absorption point (see Fig. 9.2).

The procedure ends when the predetermined number of particles has been generated. You can perform this simulation using our code `MCneutrons`. The

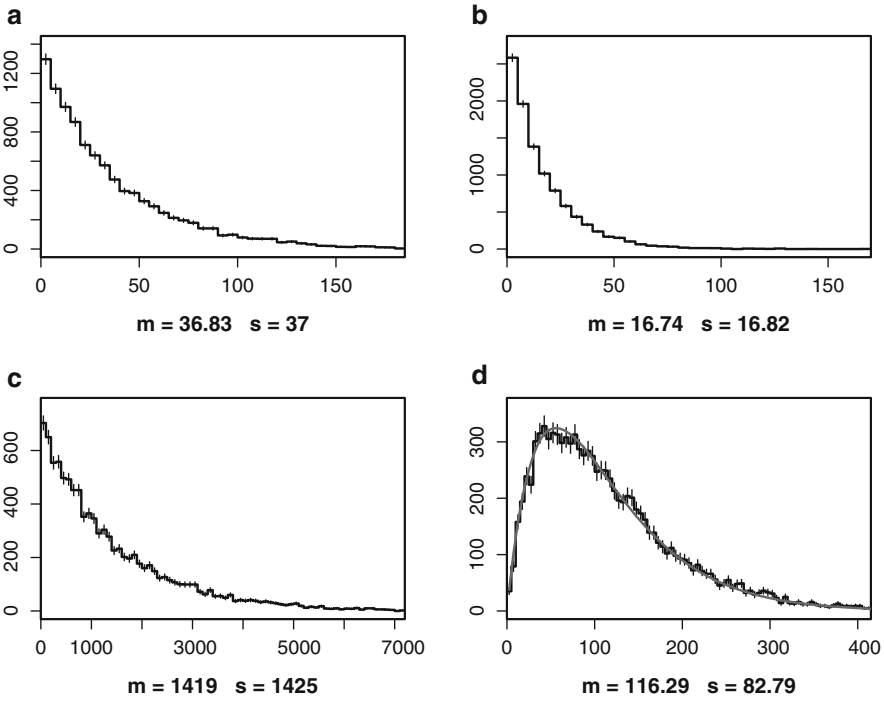


Fig. 9.3 Final histograms from the simulation of 10,000 neutrons in Carbon. (**a**) Zigzag distance (m). (**b**) Time (ms). (**c**) Number of collisions. (**d**) Flight distance (cm)

results provided by the programme, from 10,000 simulated events, are shown in the histograms of Fig. 9.3 when the target nucleus is carbon ($A = 12$, $\Sigma_{el} = 0.3851\,\mathrm{cm}^{-1}$, $\Sigma_a = 2.728 \cdot 10^{-4}\,\mathrm{cm}^{-1}$).

From the analysis of the histogram distributions, we can both obtain the main parameters characterizing the neutron diffusion in carbon and also infer some important additional information on this process.

Let us examine histogram (a) of Fig. 9.3 (zigzag total neutron travelled distance): the error on the mean m from the histogram is $s/\sqrt{10\,000} \simeq 36$ (s is the histogram standard deviation). Within the statistical error, there is then coincidence between mean and standard deviation. This is an indication that the parent population is a negative exponential density (see Sect. 3.3). It is easy to understand this feature if we observe that the only quantity influencing the travelled path is Σ_a. Recalling the same considerations that led us to Eq. (9.4), the initial probability density must therefore be:

$$p(x) = \Sigma_a e^{-x\,\Sigma_a}\,, \tag{9.16}$$

and in fact, always within the statistical errors, the following identity is verified:

$$\mu_H = \sigma_H = \frac{1}{\Sigma_a} . \tag{9.17}$$

Similar considerations also apply to histogram (b), which represents the elapsed time between emission and absorption. This quantity is nothing else than the ratio between the total travelled distance x and the velocity v of the neutrons, so this distribution has the same characteristics as the previous one. To explicitly obtain the parent density $p(t)$ (with $t = x/v$) of this histogram, we just need to apply the transformation law (5.7):

$$p(t) = p(x(t))\frac{\mathrm{d}x}{\mathrm{d}t} = v\Sigma_a e^{-tv\Sigma_a} . \tag{9.18}$$

Let us now consider the distribution of the number of collisions that each particle had before absorption (histogram (c)). At first sight, taking into account that each interaction can be considered a rare, uniform and stationary event, one might expect this histogram to be Poisson distributed. However, also in this case, there is a coincidence between the mean and standard deviation, a characteristic indication of a negative exponential distribution.

Let us examine the situation in more detail. Contrary to what happens, for example, in the flip of a coin, where the probability of having heads or tails is not affected by previous tosses, the alternative elastic scattering-absorption has an effect on following interactions since, if absorption occurs, the neutron path ends, and the possibility of having other collisions is thus excluded. On the basis of these considerations, it is easy to realize that only the number of collisions per unit path length is Poissonian, since this quantity only depends on the elastic scattering, a process where each collision is certainly independent from the previous one. From these considerations, it is easy to determine that the distribution of the total number of collisions follows the geometric law (3.7) with probability of success, i.e. of absorption, $p = \Sigma_a/\Sigma_T$. As we noted in Sect. 3.7, this distribution, when $p \ll 1$ (as in our case), is practically indistinguishable from the exponential distribution with parameter p of Eq. (3.50); for this reason, always within the statistical errors, the mean and variance of the histogram are equal to Σ_T/Σ_a.

Finally, to interpret histogram (d), we note that the flight distance (9.15) is the modulus of a vector whose components x_k, y_k, z_k are realizations of a sum of independent random variables from populations with finite variance:

$$x_k = \sum_{i=1}^{N_{\mathrm{coll}}} d_i\alpha_i \; ; \quad y_k = \sum_{i=1}^{N_{\mathrm{coll}}} d_i\beta_i \; ; \quad z_k = \sum_{i=1}^{N_{\mathrm{coll}}} d_i\gamma_i \; . \tag{9.19}$$

Due to the Central Limit Theorem, x_k, y_k and z_k are Gaussian variables; then on the basis of Pearson's Theorem 3.3, the p.d.f. $p(r)$ of the flight distance r could be Maxwellian distributed. However, in this specific case, a condition of this theorem is

not verified: although $d_i\alpha_i$, $d_i\beta_i$ and $d_i\gamma_i$ are independent and have finite variance, in sums (9.19) the number of collisions per event N_{coll} *is a random variable*. Then $p(r)$ can be considered as a superposition of different Maxwellian functions depending on the N_{coll} values. These types of variables are known as *stochastic sums*; their p.d.f. are derived in some books as [PUP02]. We omit the derivation of the analytic solution, known as Fick's law and report the final result [Lam66]:

$$p(r) = \frac{r}{L^2} e^{-r/L} , \qquad L = \sqrt{\frac{\Sigma_{el}}{3\,\Sigma_a\,\Sigma_t^2}} , \tag{9.20}$$

where the parameter L (in our case $L = 56.3\,\text{cm}$) is called *diffusion length*, whose square is proportional to the mean square flight distance travelled by a neutron from the source to its absorption point. In this graph we have also drawn with a line the function calculated by assuming the formula (9.20) as the model of the parent population of the histogram and applying Eq. (6.98). The χ^2 test to check the consistency between sample and population, applied as explained in Sect. 7.5, provides values $\chi^2_R \simeq 1$ and p-values much greater than 5%, indicating, as is also evident from the figure, a good agreement between sample and population.

9.3 Simulation of Stochastic Processes

Also the study of the time evolution of any stochastic process becomes very convenient, if MC methods are used. As an example, we will consider a typical operational research[3] topic: the study of waiting phenomena.

Queues of people in front of a service desk are the best known example of the systems we are about to examine: from them the theory of waiting phenomena has taken its name and most of the terminology, but the classes of processes that can be analysed in this framework are very numerous, and many of them are also closely related to several moments of our daily life. The regulation of urban car traffic or of a train station, the scheduling of airline flights or the number of checkout counters in a supermarket and the operation of a warehouse or the maintenance system of a factory are just some of the problems that can be interpreted and solved through queuing theory. Its knowledge is therefore essential when dealing with methodologies related to industrial, economic or social management.

In a schematic way, the structure of a waiting phenomenon includes a certain number of "service stations"or "channels"(which can be clerks or workers assigned to a certain task, communication lines, parking lots, etc.) carrying out the work requested by "customers"(people waiting for a service, machines to be repaired,

[3] Operational research is a discipline that applies mathematical methods to the study and analysis of problems involving complex systems in order to find suitable solutions for their optimal organization.

goods to be shipped, etc.). The serving system can be in two different states when a new customer arrives:

(a) There is at least one free station: the customer's request is immediately taken over, and this operation occupies one or more stations for a certain period of time.
(b) All stations are busy: the customer waiting to be "served"is put in a "queue", whose characteristics are very different depending on the considered system.

The purpose of this study is to evaluate the overall performance of the production process taking into account the costs related to both customer waiting times and the number or periods of inactivity of the serving stations (for a more complete discussion, see [CS61]).

The more complicated part in simulating this type of process is to correctly define a "clock"that reproduces the continuous event timeline. This problem can be solved in two different ways, which in the literature are called synchronous or continuous simulation and asynchronous or discrete simulation [BFS87].

In the synchronous simulation, the clock problem is solved in an extremely simple way: a unit of time (second, minute, hour, etc.) is defined in an arbitrary way, as long as it is small compared to the characteristic event rates of the system, and at each programme cycle, the clock is advanced by one unit. Simulation is thus discrete, but it is also possible to simulate a continuous process as the unit of time increment in a cycle is small compared to the transition times of the system.

Then, knowing the time probabilities $0 < \lambda_i \ll 1$ (in the chosen unit) of each possible event or change of state of the system, a random extraction of a set of uniform random variates $0 \leq \boldsymbol{\xi_i} \leq 1$ is performed, and the event occurs if $\xi_i < \lambda_i$; otherwise the system is left unchanged. For each predetermined time interval Δt, the temporal averages and the standard deviations of the characteristic quantities describing the system are then calculated, and, if necessary, graphs and histograms are created at the end of the simulation. As an example, Fig. 9.4 shows how to simulate the failure of a machine with a mean of five failures per hour, choosing the second as a time unit.

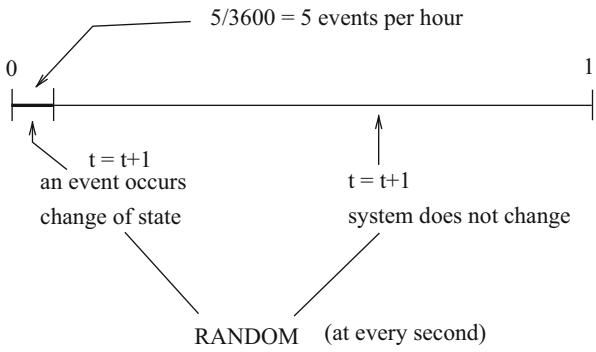


Fig. 9.4 Synchronous simulation of an event with a five event/hour probability

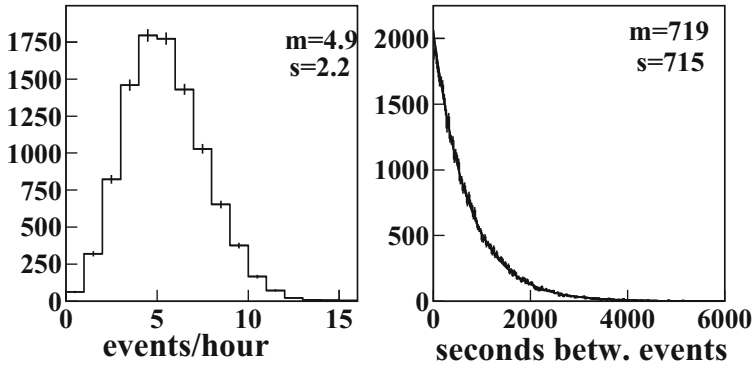


Fig. 9.5 Histograms of the average number of events per hour and of the time gaps between two successive events generated with the synchronous simulation of Fig. 9.4. The solid line shown in the time histogram is the fit of the data with the negative exponential law

Let us now ask ourselves the fundamental question: is the synchronous simulation in agreement with the general law of independent stochastic phenomena, which predicts a Poissonian event distribution and adjacent events separated by exponential time intervals? The answer is yes, as long as the condition $\lambda_i \ll 1$ holds. Indeed, in this case the probability of having n elementary time instants between adjacent events is given by the geometric density (3.50), which, as repeatedly noted, becomes an excellent approximation of the exponential density (3.50) with parameter p when $p \ll 1$. In support of this assertion, we show in Fig. 9.5 the histograms of the simulation, for 10,000 h, of the number of service requests per hour and the number of seconds between two adjacent events for the process displayed in Fig. 9.4, where $\lambda = 5/3600\ \mathrm{s}^{-1}$: as you can see, the number of events per hour is Poissonian, and the time gaps between two events follow the negative exponential density. Note that, even with an average time between two events of about $3600/5 = 720$ s (12 min), you can have waiting times as long as two hours! This is one of the reasons for the large fluctuations in temporal averages that often occur in stochastic phenomena, a feature that simulation is able to accurately reproduce.

We now come to the other type of simulation, the asynchronous or discrete one [Bun86, Ros96], often referred to simply as Monte Carlo simulation. In this case, one exploits the fact that the system changes its state only when very specific events take place (e.g. the arrival of a new customer or the end of a station's service), otherwise remaining substantially unchanged.

Thus the time of the simulation clock does not advance regularly but, at variable intervals, obtained through Eq. (3.90), which marks the arrival of a new event. At this time all the indicators describing the state of the system to be studied are updated (let us assume again for simplicity that each system change occurs immediately). The time instant in which a new event occurs is determined by sorting a "list", where all the possible types of events that can happen in the system, as well as their instant of occurrence, must be recorded.

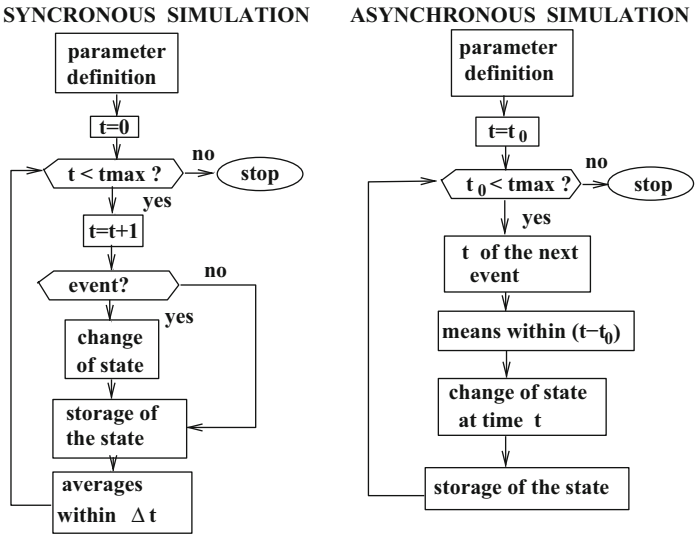


Fig. 9.6 Flux diagrams for the synchronous and asynchronous simulations of waiting phenomena

Generally, synchronous simulation results in computational codes much simpler than those using the asynchronous one, since the latter must ensure the correct ordering of arrival times. This task is often difficult, especially if the system is complicated. However, asynchronous simulation is sometimes absolutely preferable, as the computational time required by synchronous simulation, which runs a fixed cycle even when the system remains unchanged, can be unacceptable. Typical flux diagrams for synchronous and asynchronous simulations are displayed in Fig. 9.6.

The goal of this type of simulations is usually the determination of the averages of some characteristic quantities within predetermined time intervals. To fix ideas, we might be interested in the number of customers waiting at a gas station or in a supermarket checkout line, averaged within an hour (taken as a unit of time). To obtain these averaged quantities, it is necessary to record the time instants t_i of the system modifications (e.g. when a new customer is added to the queue) and to calculate, for each variation, the quantity:

$$x_i(t_i - t_{i-1}) \equiv x_i \Delta t_i \ , \tag{9.21}$$

where x_i is the value of the variable (discrete or continuous) *before* the variation at t_i.

If we observe the process for a long time of t hours (as an example, for a day), we can define an average quantity as:

$$m_t = \frac{\sum_i x_i \Delta t_i}{t} \ . \tag{9.22}$$

If we divide the measurement period in rather small and equal intervals (as in the synchronous simulation) $\Delta t_i \equiv \delta_t$, so as to contain at maximum one change of state, and we assign to each δ_t the value x_i of the last variation, this formula is nothing more than the normal mean of a sample of $n = t/\delta_t$ events.

The mean and variance of X can be obtained using a simulated sampling on N cycles of duration t (e.g. several days, a week or a month) and applying the standard statistical formulae:

$$\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} (m_t)_i \ , \quad \left\langle x^2 \right\rangle = \frac{1}{N} \sum_{i=1}^{N} (m_t)_i^2 \ , \tag{9.23}$$

$$s^2(x) = \frac{N}{N-1} \left[\left\langle x^2 \right\rangle - \langle x \rangle^2 \right] \ . \tag{9.24}$$

In these processes, the estimate of the X variance must be evaluated with Eq. (9.24), by progressively computing the partial means, since the final average value will be known only at the end of simulation.

Notice that fluctuations on the final result now depend not only on the statistical error but also on the variability over time of the parameters that describe the state of the process under study. However, almost always, this type of systems reach, after a certain operational time, a steady state in which the quantities (9.23, 9.24) no longer have appreciable variations in time. Determining when the steady state is reached is not generally an analytically solvable problem, since it depends on the particular characteristics of each individual process, such as the intensity of customer traffic, the shape of the distribution of arrivals and service times. For this reason, the results are usually printed at regular intervals of simulated time in such a way as to empirically observe the variations and the convergence to the equilibrium value of the state variables of the system.

Even the evaluation of the statistical error is quite complicated since all the events are correlated (the waiting time of a generic customer depends,e.g. on those of the immediately preceding customers), so that the use of formulae valid for independent observations would lead to wrong results.

A general result of statistics, which remains valid, is that of the error of the mean $\langle X \rangle$, in the case of steady state, will decrease as the square root of the number of sampling cycles, which are proportional to the total simulated time T_{sim}:

$$\sigma_{MC} \propto \frac{1}{\sqrt{T_{sim}}} \ . \tag{9.25}$$

One of the most used algorithms for the statistical error calculation is the *batch means* method which we will be shortly discussed in Sect. 9.7. Alternatively, two simple procedures can be applied:

- Once the simulated time interval of the process under study has been determined, the entire programme is repeated, with different sets of random numbers, a

number of times (at least 15 or 20) large enough to apply the Central Limit Theorem for the calculation of the error on the mean of the obtained results.
- On the other hand, if one wishes to find the minimum simulated time interval necessary to obtain a predetermined precision on the final results, it is necessary to carry out a short preliminary test to obtain the error on the result for a small value of T_{sim}. The requested value is then easily obtained by exploiting the proportionality law (9.25) on the statistical error.

9.4 Number of Workers in a Plant: Synchronous Simulation

To better exemplify and clarify all the previous considerations, we solve, using synchronous simulation, the problem of determining the optimal number of workers to be assigned to the control and maintenance of a certain number of industrial machines.

Suppose that the main characteristics of the system to be studied, deduced from empirical observations, are the following:

- In the plant there are ten machines, each of them requires on average 3 interventions per hour both for the failure repair and during the standard operational work phase.
- The duration of each intervention (which requires the activity of only one person) follows a negative exponential law with an average of two interventions completed in 1 h by each worker, regardless of the person who carries it out and of the machine requiring it.
- The hourly cost linked to the inactivity of a worker (c_1) is estimated at 70 euros, while that linked to the inactivity of a machine (c_2) is 20 euros.

First of all, it is necessary to begin with a concise (but complete!) description of the system. We remind you that, on average, there are three intervention requests per hour per machine and that, on average, each worker repairs or maintains two machines per hour. The objective of the optimization is to minimize the cost of the system, knowing that a machine waiting for intervention costs 20 euros/hour and an inactive operator costs 70 euros/hour. The simulation will therefore have to determine the average number of inactive operators $\langle N_{oi} \rangle$ and of waiting machines $\langle N_{mi} \rangle$ and minimize the average hourly cost:

$$C = 70 \cdot \langle N_{oi} \rangle + 20 \cdot \langle N_{mi} \rangle \ . \tag{9.26}$$

The short description of the system is reported in Table 9.1.

The second step is the execution of the routine `MCsinc`, which is partially reported below. The chosen time unit of measurement is the second, and, therefore, the probabilities of intervention request and end of repair are, respectively, 3/3600 and 2/3600. At each second, the programme examines the state of the machines (initially all running, variable `Ms[k]=0`). If a particular machine works, it requires

Table 9.1 Logical description used for the simulation of the system "number of workers in a plant". The status variable of a machine can take on three values: running (0), under repair (1), awaiting repair (2)

Machine status	Change of state	Inactive workers	Variation inact. workers	Machine status
Working	Works	Unchanged	–	0
	Intervention needed	If > 0	-1	1
		If $= 0$	–	2
In repair	Continues	Unchanged	–	1
	Intervention finished	Increases	$+1$	0
Awaiting repair	–	If > 0	-1	1
	–	If $= 0$	–	2

action if the uniform random variable $\xi \leq 3/3600$; if instead the machine is being serviced by a technician, an end of service occurs if $\xi \leq 2/3600$. Finally, if the machine is waiting for intervention and there are free technicians, it is served (`Ms[k]=1`); otherwise it remains waiting (`Ms[k]=2`). This is the phase of state change, indicated in the second column of Table 9.1 and also in the block diagram of Fig. 9.6. The part of the routine that performs the status sampling is as follows:

```
Hour = 60         # unity of measure: one hour= 60 minuts
Minuts = Hour*H   # minuts of the simulation
Mp = 3/Hour       # 3 request of interventions/hours
Op = 2/Hour       # 2 end of interventions/hours
Ol = On   # Ol = number of workes in stand-by
Oi = 0    # Oi = cumulative number of workers in stand-by
Mi = 0    # Mi = number of machines waiting for an intervention

     # simulation for H hours in steps of 3600/hour seconds

  for(n in 1:Minuts){          # beginning of general loop
     for(k in 1:Mn){           # scan the machines
        if(Ms[k] == 0){
           if(runif(1) < Mp){ # request of intervention
             if(Ol>0){        # there are workers in stand-by
               Ms[k]=1        # machine number k under repair
               Ol = Ol-1      # update the number of workers
             }                #                   in stand by
             else{ Ms[k] = 2 }  # machine number k waiting
           }                    #                 for repair
        }
        else if(Ms[k] == 1){  # machine k under repair
           if(runif(1) < Op){ # end of intervention?
              Ms[k] = 0       # machine repaired
              Ol = Ol + 1     # update the number of free workers
           }
        }
        else if(Ms[k] == 2){  # machine k is waiting for repair
           if(Ol != 0){       # are there free workers?
              Ms[k] = 1       # machine k under repair
```

```
            Ol = Ol - 1      # update the number of workers
          }                  #                    in stand by
      }
  }
  for(k in 1:Mn){            # number of machines waiting repair
    if(Ms[k]==2){Mi = Mi + 1}
  }
  Oi = Oi + Ol               # cumulative of workers in stand by
```

The number of inactive operators `Oi` and the number of machines awaiting intervention `Mi`, which are the variables to be considered, are updated every second. These quantities are averaged every 24 h according to Eq. (9.22), and the corresponding means and variances are progressively calculated with Eqs. (9.23, 9.24). Figure 9.7 shows the trend of `Oi` and `Mi` for a simulated observation period of 500 days.

The means and the final variances of these quantities are reported in Table 9.2, with the evaluation of the overall cost through Eq. (9.26). An examination of the

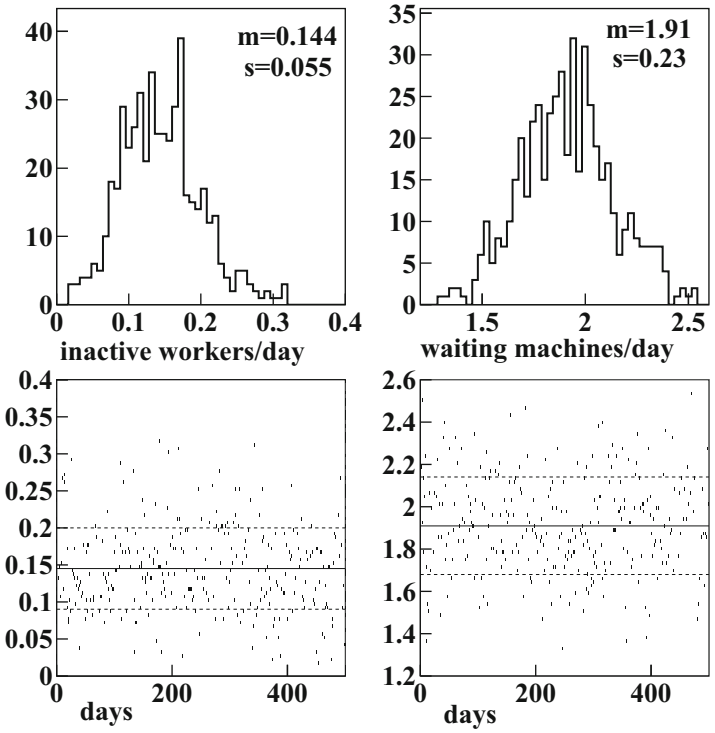


Fig. 9.7 The lower plots show the daily trend of the number of machines awaiting service and the number of inactive workers per hour for the "optimal number of workers in a plant" problem, when five persons are employed; the projection on the ordinates of the daily values is shown in the upper histograms. The mean m and the $m \pm s$ values of the top histograms are the solid and dashed lines of the bottom plots

Table 9.2 Changes in the global cost of the plant with the number of the employed workers. The errors are the sample standard deviations, not errors on the mean

Number of workers	Inactive workers/hours ($\langle N_{oi} \rangle \pm s_{oi}$)	Waiting machines/hour ($\langle N_{mi} \rangle \pm s_{mi}$)	Global cost per hour
3	$1.5 \cdot 10^{-3} \pm 3 \cdot 10^{-3}$	5.00 ± 0.21	100.1
4	$2.21 \cdot 10^{-2} \pm 1.84 \cdot 10^{-2}$	3.38 ± 0.26	69.1
5	0.148 ± 0.055	1.91 ± 0.23	48.6
6	0.53 ± 0.12	0.87 ± 0.15	52.4

figure and of the table shows relevant fluctuations around the average values. In fact, the plant is not very efficient, because, given ten machines with an average of three requests for intervention per hour and an average service time of half an hour, there will always be a large number of inactive machines (about five or six) even with ten workers, one per machine. However, given the exponential law assumed for both intervention and service request times, which are also comparable with each other, it is natural to expect large fluctuations in the daily quantities observed. In any case, since these features are given a priori, the simulation still solves the cost minimization problem, showing that the optimal number of workers is five, as shown in Table 9.2. In Fig. 9.8, the average number of inactive workers and the number of machines awaiting repair for each hour of operation of the plant are reported. The fluctuations present at the beginning of the simulation are due both to the specific initial conditions assumed for the system, far from the stationary regime, and to the insufficient quantity of data. From this figure, it can also be deduced that stationarity is reached quite quickly.

It should also be noted that the means and standard deviations of Table 9.2 are derived from a 500-day observation. Considering as an example the optimal case of five workers, from the formulae of Table 6.3, we can estimate the errors on these quantities, using Eq. (9.25), as:

$$\sigma\left[\langle N_{mi} \rangle\right] \simeq \frac{1.91}{\sqrt{500}} = 0.08\,, \quad \sigma\left[s(N_{mi})\right] \simeq \frac{0.23}{\sqrt{1000}} = 0.007$$

$$\sigma\left[\langle N_{oi} \rangle\right] \simeq \frac{0.148}{\sqrt{500}} = 0.007\,, \quad \sigma\left[s(N_{oi})\right] \simeq \frac{0.12}{\sqrt{1000}} = 0.004\,.$$

9.5 Number of Workers in a Plant: Asynchronous Simulation

The synchronous simulation has the disadvantage of carrying out many unnecessary cycles, in which the system does not change state. To speed up the synchronous code, we changed the unit of time from second to minute, obtaining results different

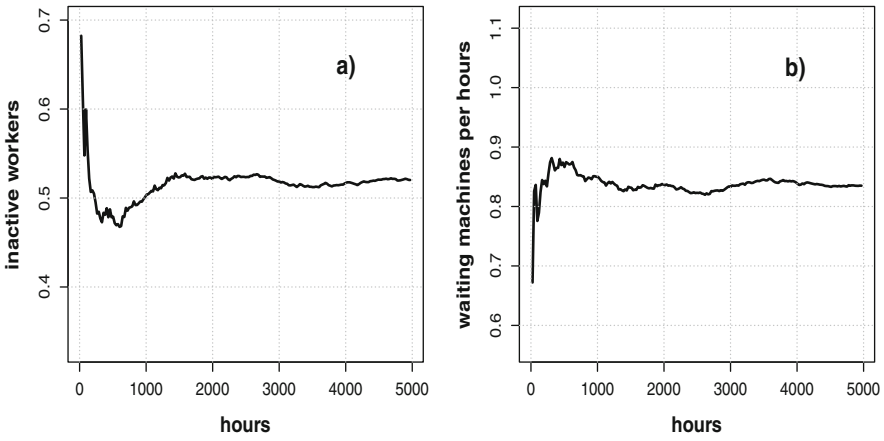


Fig. 9.8 Variation, as a function of the simulation time, of the average number of inactive workers (**a**) and of the fraction of machines awaiting repair (**b**) for each hour of plant operation and in the case of six workers

only by a few percent from those of Table 9.2. Furthermore, in this scheme, only exponential time distributions for intervention and repair are allowed. A more general and quicker way to describe the system is to use asynchronous simulation.

In this case, time does not flow at regular intervals but in uneven jumps and only at event occurrence times. In the considered system, there are two types of events: request for intervention and end of intervention. In the simulation code `MCasinc`, it is then necessary to define the vectors `Tmac[k]` and `Top[k]` which contain, for each machine, the intervention request and end-of-service times, respectively, evaluated with the R `rexp` routine by using the exponential law. During the initialization phase, a series of request times are extracted while the end-of-service times are "frozen"with an "infinite"`BIG` time. The core of `MCasinc` which generates the simulated events is the following:

```
for( k in 1:Mn){
  Tmac[k]=rexp(1,rate=Mp)  # exponential intervention request
  Top[k] =BIG           # all workers are available
}
#  Days  of simulations
while (t<=TIME){
     #   awaiting machines Mi and free workers Ol in (t-tprec)
  tprec=t;
  Mi=0;
  for(j in 1:Mn){ if(Ms[j]==2){ Mi=Mi+1} }
  Ol = On - occ;
#search for indices  of the minimum time and of the state change
  kmac = which.min(Tmac)
  kop  = which.min(Top)
  if(Tmac[kmac]<Top[kop]){   #  one machine is out
     t= Tmac[kmac]
     if(occ<On) {            # there are free workers
```

```
      occ = occ + 1
      Top[kmac] = t + rexp(1,rate=Op)
      Ms[kmac] = 1
      Tmac[kmac] = BIG
    }
    else {                          # awaiting machine
      Ms[kmac] = 2
      Tmac[kmac] = BIG
    }
  }
  else {                            # one worker is free
    t = Top[kop]
    occ = occ -1;
    Ms[kop] = 0;
    Tmac[kop] = t + rexp(1,rate=Mp)
    Top[kop] = BIG
    n=1
    flag =0
    while (flag==0 && n<=Mn){          #  use a free worker
        if(Ms[n]==2) {         #  if there are awaiting machines
           Ms[n] = 1
           occ = occ + 1
           Top[n] = t + rexp(1,rate=Op)
           flag = 1
        }
        n = n + 1
     }
  }
  oinh = oinh + Ol*(t-tprec)     # daily data of Ol and Mi
  minh = minh + Mi*(t-tprec)     # moving means
```

The simulation clock moves forward by searching for the minimum of the time instants contained in these two vectors, using the `which.min` routine. If a machine is waiting for repair or a worker becomes inactive, the request or end-of-service times must be stopped by introducing into the `Tmac` and `Top` vectors the time `BIG`.

Now suppose that we have already considered a certain number of events; the next one will be either a new request for intervention or the end of a worker's service. If the event is a request for intervention on the machine `kmac`, the simulated time is updated, an operator is appointed (if available), the machine "is frozen"(`Tmac[kmac]=BIG`) and the end of intervention time is simulated (`Top[kmac ]=t + texp(Op)`); if there are no free operators, the machine is set in state 2 (`Ms[kmac]=2`). If, on the other hand, the event is the end of the operator's service on the machine `kop`, a new request of intervention time is generated `Tmac[kop]=t + texp(Mp)`, the machine repair time is set as (`Top[kop]=BIG`) and the vectors and status indices are updated, as commented in the programme. If there are other machines in the queue, the operator who has just finished his job immediately takes care of a waiting machine (the number `n`), with an intervention that will end at the time `Top[n ]=t + texp(Op)`. The number of waiting machines `Mi` and the number of free operators `Ol` are multiplied by the time between two events `(t-tprec)`, obtaining the new variables `oin` and `min`. This is an important point: this time interval multiplies `Ol` and `Mi` *set at*

the previous time `tprec`, since between `tprec` and `t`, the system state has not varied after the change at `tprec`. This part of the code is equivalent to that of the synchronous simulation where the same quantities were summed up every second, even if unchanged. The daily averages of (9.22) are then updated by dividing `oin` and `min` by the 24 h interval. The moving averages and the relative variances are then calculated as in the case of the synchronous simulation. The simulation ends when the time `t` reaches a predetermined value.

As can be seen, the logic and structure of the code are more complicated than in the case of synchronous simulation. However, the execution time was about 15 times shorter, because now the system state transition is only calculated when an event actually occurs. The code gives the same results of the synchronous simulation that we have already reported in Table 9.2 and Fig. 9.7.

For simple programmes like these, the time gain is not important, but for complex models, in which the execution time of the programmes can be even in the order of hours, a gain of this magnitude is decisive.

You can now modify the code and complicate its structure, trying to describe a more realistic model by directly exploring how it is possible to study complex systems in a very simple way. For example, different types of service requests (breakdown, maintenance, etc.) can be introduced, with different intervention times, the availability of spare parts could be considered, different repair times can be assigned to each worker and so on. The inclusion of these details in an analytical model makes the problem quickly intractable, while, with a simulation code, it is possible to add new details in a modular way without complicating both the structure and the management of the model.

9.6 Kolmogorov-Smirnov Test

Here we complete the hypothesis testing topic, begun in Chapt. 7, with the Kolmogorov-Smirnov (KS) test, which gains simplicity and clarity if is explained with the support of simulation techniques. The KS test is non-parametric and holds for continuous variables.

Given a sample x_i (with $i = 1, 2, \ldots, n$) of n values of the variable X sorted in ascending order, an empirical partition function is defined as:

$$F_n(x) = \begin{cases} 0 & \text{if } x < x_1 \\ \dfrac{k}{n} & \text{if } x_k \leq x < x_{k+1} \\ 1 & \text{if } x \geq x_n \end{cases} . \tag{9.27}$$

This function is constant between two consecutive points and increasing, discontinuously, by $1/n$ at every point. The function $F_n(x)$ is an unbiased estimator of the cumulative function $F(x)$, as we have already shown in Eq. (8.89). Furthermore,

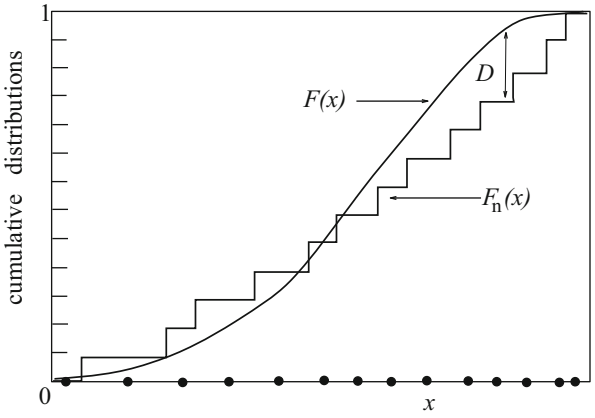


Fig. 9.9 In the Kolmogorov-Smirnov test a sampled cumulative distribution (step function) is compared with the expected distribution $F(x)$ (continuous curve). D is the greatest distance between the two distributions

each point of the cumulative is strongly correlated to the other points, being their moving sum. The D estimator of KS is based on these principles and is simply:

$$D_n = \sup_{-\infty<x<+\infty} |F_n(x) - F(x)| \ . \tag{9.28}$$

To find the maximum of this difference, we need to explore *all* the x values of the density support, as shown in Fig. 9.9

Let us briefly recall the properties of this estimator. First of all, it can be shown that F_n almost certainly converges to $F(x)$ (see Eq. 2.74) [MGB73]. This property is known as the Glivenko-Cantelli theorem.

The second very important property is the *independence of the D distribution from* $F(X)$. This follows directly from the cumulative variable theorem 3.5, according to which we know that if $X \sim p(x)$ and $F(x)$ is the cumulative of $p(x)$, then $F(X) \sim U(0, 1)$. We can then write the relation:

$$F_n(x) = \frac{\#(j : X_j \le x)}{n} = \frac{\#(j : F(X_j) \le F(x))}{n} = \frac{\#(j : U_j \le F(x))}{n} \ ,$$

that, in words, means: the fraction of sample values less than a certain value x is equal to the fraction of values $F(X_j) \le F(x)$, since F is an ascending monotone function. We also know that $F(X_j) \equiv U_j \sim U(0, 1)$, hence the last relation. Based on this property we can write the KS estimator as:

$$D_n = \sup_{-\infty<x<+\infty} \left| \frac{\#(j : U_j \le F(x))}{n} - F(x) \right| = \sup_{0\le u\le 1} \left| \frac{\#(j : U_j \le u)}{n} - u \right| \ , \tag{9.29}$$

from which it follows that the distribution of D can be obtained by randomly extracting n uniform variates and finding the value of the maximum difference between the fraction $\#(U_j \leq u)/n$ and u for $u \in [0, 1]$. We therefore deduce that the distribution of D_n is non-parametric, universal and depends only on n.

The third property is related to the p.d.f. of the D statistic, whose properties have to be known for the p-value calculation to be used in the test. In fact, the concept at the base of the test is that $F_n(X)$ follows the KS statistic (and will not give small p-values) when $F(x)$ is the true parent cumulative distribution of the data. The determination of the true form of this p.d.f. is still an open problem, but there are many empirical solutions. The first of them was proposed in 1933 by Kolmogorov himself [Kol33] during a stay in Rome. He noticed, in a brilliant way, that the fluctuations of $F_n(X)$ around $F(x)$ are the same as in certain types of Brownian motion around zero. Fortunately, as we have just seen, the distribution is universal and depends only on n, so it can be found empirically. In [PFTW92] the formula for the p-value calculation corresponding to an observed value d is reported as:

$$P\{D_n > d\} = Q_K(\lambda) = 2\sum_{j=1}^{\infty}(-)^{j-1}\exp(-2j^2\lambda^2)\,, \tag{9.30}$$

$$\lambda = \left(\sqrt{n} + 0.12 + \frac{0.11}{\sqrt{n}}\right) d\,.$$

This approximation gives good results already from $n \geq 4$.

In R, this test can be performed using the `ks.test` routine, which requests the data vector `x` and the name of the distribution function as input. Our routine `MCKolmoDist(nsim, ndata, type)` simulates `nsim` variates of D_n from a sample of `ndata` data, Gaussian (`type='pnorm'`) or uniform (`type='punif'`). The results are shown in Fig. 9.10, from which we can see the goodness of the approximation given by Eq. (9.30).

In addition to the test between a sample and a distribution, it is also possible to compare two samples, because also in this case the KS test maintains the fundamental property (9.29). The changes to be made to the reference distribution under H_0 (the two samples come from the same population) are minimal: if n_1 and n_2 are the elements of the two samples, just set:

$$n = \frac{n_1 n_2}{n_1 + n_2} \tag{9.31}$$

in Eq. (9.30). The R routine `ks.test` performs also this test if the two experimental samples are given in the calling string.

The pros and cons of the KS test are as follows:

(a) Pros: the KS test is more powerful than the χ^2 test because the latter does not detect anomalies when the points are above or below the model curve in a non-random way but still within the experimental error. This generally does

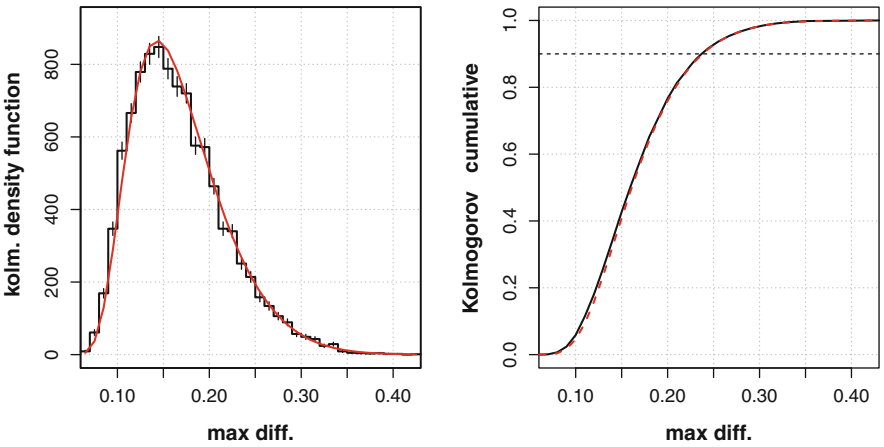


Fig. 9.10 Distribution of D_n for $n = 25$. To the left: p.d.f. of D_n obtained by calculating differences using Eq. (9.30) (full curve) and from a Gaussian sample (histogram). To the right: approximation from Eq. (9.30) (dashed curve) and cumulative function of the simulated data. The result does not change if, for example, the uniform distribution is used in the simulation. The abscissa of the intersection point between the horizontal line and the cumulative gives the 90% quantile value, corresponding to $p = 0.10$

not happen with cumulative data. Furthermore, the distribution is independent of the type of density considered.

(b) Cons: the test can only be applied to continuous data. Hence, it cannot be used for histograms or discrete distributions.

This last point can be partially overcome with simulation techniques, which allow us to find the reference distribution even for histogrammed data. In this case, one has to be careful to retain the same bin number and the sample dimensions equal of the experimental data. As before, simulation allows us to use the good properties of the cumulative distributions.

For instance, our `MCKolmoHist` routine finds the reference distribution from two simulated Gaussian or uniform histograms with n_1 and n_2 events, respectively, and the same number of bins.

The bin content of the cumulative histograms is $n(x)/n$, where $n(x)$ is the number of simulated events inside the bin with mean value x and n is the total number of events. Then, the difference is calculated as:

$$T_{Kn_1n_2} = \sup_{1 \le M < K} \left| \sum_{i=1}^{M} \frac{n_1(x_i)}{n_1} - \sum_{i=1}^{M} \frac{n_2(x_i)}{n_2} \right| . \tag{9.32}$$

After repeating the cycle a large number of times, one gets the graphs of the density and the cumulative function of $T_{Kn_1n_2}$. In Fig. 9.11, the simulated distribution of 10,000 differences is shown and compared with the same function as in Fig. 9.10.

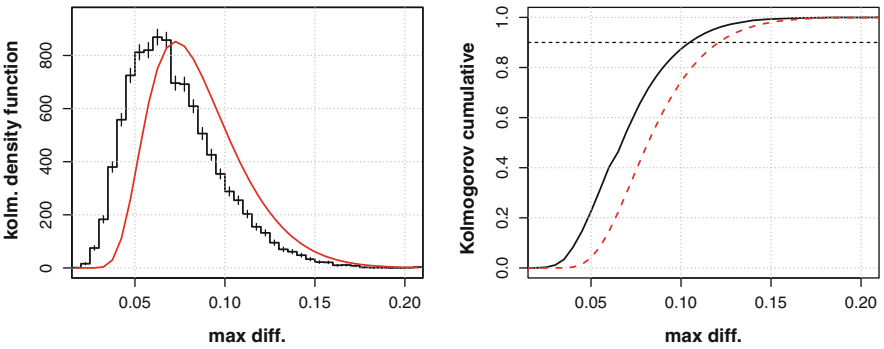


Fig. 9.11 Distribution of 10,000 differences $T_{Kn_1n_2}$ from Eq. (9.32) for $n_1 = n_2 = 200$. To the left: p.d.f. obtained by difference from the approximation (9.30) (full curve) and H_n distribution obtained from two homogeneous Gaussian samples (histogram). To the right: approximation (9.30) (dashed curve) and cumulative function of the simulated data. The simulation clearly shows the deviation of the histogram data from the Kolmogorov-Smirnov model

From this figure, we can clearly deduce that the p-values of a real experiment must not be calculated with the Eq. (9.30), but directly from the histograms simulated under H_0. Our routine also allows the generations of histograms from the uniform distribution. You can check that, with the same total number of events and channels, the simulation results are very similar. However, this is a property that must be verified on a case-by-case basis.

Exercise 9.1
Generate 20 variates from the Gaussian $N(\mu = 0.5, \sigma^2 = 4)$, and perform the KS test in R with a standard Gaussian and with the correct one.

Answer The R code is:

```
> x<- rnorm(20,mean=0.5,sd=2)
> ks.test(x,'pnorm')   # one obtains p=0.011
> ks.test(x,'pnorm',mean=0.5,sd=2)   # one obtains p=0.82
```

from which we see that the first test gives, as expected, a small p-value, while the second, with the correct density, gives a two-tailed p-value corresponding to 0.41 for each tail. The test therefore tends to reject the first hypothesis and accepts the second one.

9.7 Metropolis Algorithm

The Metropolis algorithm is a sophisticated method to generate a sample from distributions that cannot be easily simulated with the techniques described in the previous chapter. It is best applied to functions that can be written as:

$$p(\boldsymbol{x}) = \frac{h(\boldsymbol{x})}{Z} ,$$

where $\boldsymbol{x}$ is a d-dimensional random vector. Due to the normalization conditions, we have $Z = \sum_{\boldsymbol{x}} h(\boldsymbol{x})$ in the discrete case, and $Z = \int h(\boldsymbol{x})\, d\boldsymbol{x}$ in the continuous one. The normalization constant Z is must be known to obtain any quantity related to $p(\boldsymbol{x})$ (such as mean, variance and percentiles). However in some cases its calculation may be impossible in practice. In physics or chemistry, this happens, for example, when systems consisting of a large number d of identical elementary components, such as molecules in a gas or in a crystalline solid, are studied. Typically, $d \simeq 10^{23}$, a value that roughly represents the number of atoms contained in a cm^3 of matter.

In this case, $\boldsymbol{x}$ is a set of parameters (position, velocity, etc.) describing the behaviour of all the elementary system components, and we suppose that each of them can assume k different states. If $g(\boldsymbol{x})$ describes a macroscopic system parameter (such as temperature, pressure, magnetic moment, etc.), the calculation of its mean, $\sum_{\boldsymbol{x}} g(\boldsymbol{x}) p(\boldsymbol{x})$, would require to evaluate $h(\boldsymbol{x})$ for each of the possible k^d system configurations, an effort that is out of reach with the currently available computing resources.

In these cases, the Metropolis algorithm is very powerful since, to generate a sample from $p(\boldsymbol{x})$, it is not necessary to know Z nor to evaluate $h(\boldsymbol{x})$ for all values of $\boldsymbol{x}$. It is only needed to generate a sequence of simulated states whose asymptotic frequency distribution tends to $p(\boldsymbol{x})$.

So, let us imagine a system that is initially in a state x, sampled from the density $p(\boldsymbol{x})$, and that can afterwards "migrate"to another state $\boldsymbol{y}$ according to an *arbitrary* transition probability $t(\boldsymbol{x} \rightarrow \boldsymbol{y})$. Systems in which these probabilities depend only on the current and the previous states are called *Markov chains* and are of fundamental importance in the study of many stochastic processes [RC99]. A sufficient condition for the chain to converge to a state distributed as $p(\boldsymbol{x})$ is that it stabilizes in an equilibrium situation, where each transition occurs with probability equal to the inverse one. One of the conditions for this to happen is given by the so-called detailed balance equation:

$$p(\boldsymbol{x})\, t(\boldsymbol{x} \rightarrow \boldsymbol{y}) = p(\boldsymbol{y})\, t(\boldsymbol{y} \rightarrow \boldsymbol{x}) , \tag{9.33}$$

where the term to the left (right) indicates the probability that the system evolves from $\boldsymbol{x}$ to $\boldsymbol{y}$ (from $\boldsymbol{y}$ to $\boldsymbol{x}$)

The arbitrary function $t(\boldsymbol{x} \to \boldsymbol{y})$ is usually written as:

$$t(\boldsymbol{x} \to \boldsymbol{y}) = q(\boldsymbol{x}, \boldsymbol{y})\, \alpha(\boldsymbol{x}, \boldsymbol{y}) \,. \tag{9.34}$$

For each value $\boldsymbol{x}$ belonging to the spectrum of $\boldsymbol{X}$, the auxiliary distribution $q(\boldsymbol{x}, \boldsymbol{y})$ is required to be a probability distribution on the spectrum of $\boldsymbol{X}$. In other words, for any $\boldsymbol{x}$, $q(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ for each value of $\boldsymbol{y}$ and $\sum_{\boldsymbol{y}} q(\boldsymbol{x}, \boldsymbol{y}) = 1$ in the discrete case or $\int q(\boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y} = 1$ in the continuous one. Another very useful requirement to speed up the simulation is the possibility of quickly generating values from the distribution $q(\boldsymbol{x}, \cdot)$. The probability $\alpha(\boldsymbol{x}, \boldsymbol{y})$ of accepting the value proposed by the auxiliary distribution is instead defined by the Metropolis algorithm to guide the evolution of the system towards increasingly probable states (where, e.g. the total energy, or temperature, or pressure, is minimal). In this way, it can be demonstrated that a stationary regime can always be reached where the macroscopic parameters of the system do not vary with time, and, then, also Eq. (9.33) is satisfied.

The algorithm, proposed by Metropolis and co-workers in 1953 [MRR+53] and generalized by Hastings in 1970 [Has70], consists of N steps; if $\boldsymbol{x}^{(i)}$ is the state value generated at step i, to obtain the next value, one proceeds as follows:

Algorithm 9.1 (Metropolis-Hastings Algorithm)

(1) Generate a value $\boldsymbol{y}$ *from the auxiliary distribution* $q(\boldsymbol{x}^{(i)}, \cdot)$.
(2) Generate a value ξ *from the uniform distribution* $U(0, 1)$.
(3) Compute the probability of acceptance:

$$\alpha(\boldsymbol{x}^{(i)}, \boldsymbol{y}) = \min \left\{ 1, \frac{h(\boldsymbol{y})\, q(\boldsymbol{y}, \boldsymbol{x}^{(i)})}{h(\boldsymbol{x}^{(i)})\, q(\boldsymbol{x}^{(i)}, \boldsymbol{y})} \right\} \,, \tag{9.35}$$

where the ratio inside brackets is known as acceptance ratio*;*
(4) If $\xi \leq \alpha(\boldsymbol{x}^{(i)}, \boldsymbol{y})$, *then* $\boldsymbol{x}^{(i+1)} = \boldsymbol{y}$; *otherwise set* $\boldsymbol{x}^{(i+1)} = \boldsymbol{x}^{(i)}$. *If* $i < N$ *return to step 1.*

It is easy to show that the values obtained with Eq. (9.35) follow a density that satisfying the detailed balance condition. Indeed, since from Eqs. (9.33, 9.34) one has:

$$p(\boldsymbol{x})\, q(\boldsymbol{x}, \boldsymbol{y})\, \alpha(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y})\, q(\boldsymbol{y}, \boldsymbol{x})\, \alpha(\boldsymbol{y}, \boldsymbol{x}) \,,$$

taking into account Eq. (9.35), if $p(\boldsymbol{y})\, q(\boldsymbol{y}, \boldsymbol{x}) / [p(\boldsymbol{x})\, q(\boldsymbol{x}, \boldsymbol{y})] < 1$, asymptotically one has $\alpha(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{y}) q(\boldsymbol{y}, \boldsymbol{x}) / [p(\boldsymbol{x})\, q(\boldsymbol{x}, \boldsymbol{y})]$ and $\alpha(\boldsymbol{y}, \boldsymbol{x}) = 1$; thus the identity:

$$p(\boldsymbol{x})\, q(\boldsymbol{x}, \boldsymbol{y})\, \frac{p(\boldsymbol{y})\, q(\boldsymbol{y}, \boldsymbol{x})}{p(\boldsymbol{x})\, q(\boldsymbol{x}, \boldsymbol{y})} = p(\boldsymbol{y})\, q(\boldsymbol{y}, \boldsymbol{x})$$

is obtained. Hence, given an initial value $\boldsymbol{x}^{(0)}$, we obtain a sample $(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)})$ by repeating N times the steps 1–4, without the need to know Z, as step 3 only depends on the ratio $p(\boldsymbol{y})/p(\boldsymbol{x}^{(i)}) = h(\boldsymbol{y})/h(\boldsymbol{x}^{(i)})$.

But what kind of sample did we get? Obviously, it is not a random sample, since the generation of $\boldsymbol{x}^{(i+1)}$ depends on $\boldsymbol{x}^{(i)}$. Furthermore, the initial value $\boldsymbol{x}^{(0)}$ has little to do with $p(\cdot)$, which it is usually arbitrarily picked.

If we choose to collect the values of $\boldsymbol{x}_i$ for i greater than some value such that sample parameters (usually mean and variance) stabilize, we obtain a sample well approximating the requested stationary distribution. The internal correlations between the different sample elements does not generally prevent the determination of the important parameters of the distribution. In fact, under simple assumptions, one can show [RC99] the validity of the following:

Theorem 9.1 *If $q(\boldsymbol{x}, \boldsymbol{y}) > 0$ for any $\boldsymbol{x}$ and for any $\boldsymbol{y}$ belonging to the $\boldsymbol{X}$ spectrum, the property*

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{N} g(\boldsymbol{x}^{(i)})}{N} = \langle g(\boldsymbol{X})\rangle = \begin{cases} \sum_{\boldsymbol{x}} g(\boldsymbol{x})p(\boldsymbol{x}), & \text{discrete case,} \\ \int g(\boldsymbol{x})p(\boldsymbol{x})\,d\boldsymbol{x}, & \text{continuous case,} \end{cases} \tag{9.36}$$

holds for any initial point $\boldsymbol{x}^{(0)}$. Equation (9.36) remains valid even if the condition $q(\boldsymbol{x}, \boldsymbol{y}) > 0$ is not satisfied for all the $(\boldsymbol{x}, \boldsymbol{y})$ pairs provided that, for any set A of the spectrum with $p(A) > 0$, $q(\boldsymbol{x}, \boldsymbol{y})$ is such that A is reachable with positive probability starting from any $\boldsymbol{x}$.

In its simplest formulation, proposed in 1953 by [MRR+53], the algorithm is used with $q(\boldsymbol{x}, \boldsymbol{y}) = q(\boldsymbol{y}, \boldsymbol{x})$. In this case α depends only on the ratio $p(\boldsymbol{y})/p(\boldsymbol{x})$.

Often $\boldsymbol{X}$ is generated from the uniform distribution, within the support of $p(\boldsymbol{x})$. This is the method used by our test routine `MCmetrop`, applied to the Gaussian case, which encodes the algorithm as follows:

```
for(k in 2:N){
   i[k] = k
              # sampling in +- ks sigma around the mean
   # y= mu-ks*sigma + 2*ks*sigma*runif(1)
   y = runif(1,min=mu-ks*sigma,max=mu+ks*sigma)

        # Metropolis ratio between Gaussians
   alpha= exp( -0.5*( (y-mu)^2 - (x[k-1]-mu)^2 )/sigma^2 )
   u= runif(1)
   x[k] = x[k-1];
   if(u<alpha) x[k]=y
   sumk = sumk+x[k]
   sumk2= sumk2+x[k]^2;
   plotk[k]   = sumk/k;                        # Metropolis for...
   plotk2[k] = sqrt(sumk2/k - plotk[k]^2);    # mean and sigma

      # continuous display of the mean
   plot(i,plotk,type='p',main='mean',lwd=2)
```

```
    grid()

}  # end of Metropolis cycle
```

The parameters `mu`, `sigma` and `ks` are given as input.

It is very instructive to perform some tests with this routine; as an example, we suggest to solve Problem 9.7.

The Metropolis algorithm then provides us with the estimator:

$$\frac{\sum_{i=1}^{N} g(\boldsymbol{X}^{(i)})}{N} \equiv M \tag{9.37}$$

for $\langle g(\boldsymbol{X})\rangle$. The variance of M *is not* $\sum_{i=1}^{N} \mathrm{Var}[g(\boldsymbol{X}^{(i)})]/N^2$, because the simulated random variables are correlated. A method frequently used to circumvent the dependency is to split the simulated sequence $(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(N)})$ into consecutive k blocks of b elements each (with b and k such that $kb = N$):

$$(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(b)}, \boldsymbol{x}^{(b+1)}, \ldots, \boldsymbol{x}^{(2b)}, \ldots, \boldsymbol{x}^{((k-1)b+1)}, \ldots, \boldsymbol{x}^{(kb)}) \,,$$

and to compute the sample mean g of each block:

$$(\langle g(\boldsymbol{x})\rangle^{(1)}, \ldots, \langle g(\boldsymbol{x})\rangle^{(k)}) \,.$$

As the block size increases, non-consecutive blocks are increasingly distant (in terms of iterations) and therefore less and less correlated. It can be shown that also the correlation between consecutive blocks tends to cancel out as b increases, approaching the uncorrelated situation. Hence, it is natural to use the estimate of the sample variance of the sequence of the block averages to estimate the error of $\langle g(\boldsymbol{x})\rangle^{(i)}$. This error is associated with a sample mean of b terms; therefore, to get the error associated with $m \equiv \langle g(\boldsymbol{x})\rangle$, which is an average of $N = kb$ terms, we must divide by k again, finally finding the *batch means estimates* (with a CL of 95.4%):

$$\langle g(\boldsymbol{X})\rangle \in \langle g(\boldsymbol{x})\rangle \pm 2\sqrt{\frac{1}{k(k-1)}\sum_{i=1}^{k}[\langle g(\boldsymbol{x})\rangle^{(i)} - \langle g(\boldsymbol{x})\rangle]^2} \,. \tag{9.38}$$

The previous equation implies the validity of the Central Limit Theorem for the distribution of M given Eq. (9.37). This can be proved, under the hypotheses of Theorem 9.1, for aperiodic Metropolis algorithms, i.e. when there is no partition of the state space that is visited in the same sequence during the simulation. Once the number of iterations N has been set, b must be chosen to calculate the batch means estimate. This is a very complex problem; a commonly used practical rule is to set $b = \sqrt{N}$ (see, for instance, [FJ10]).

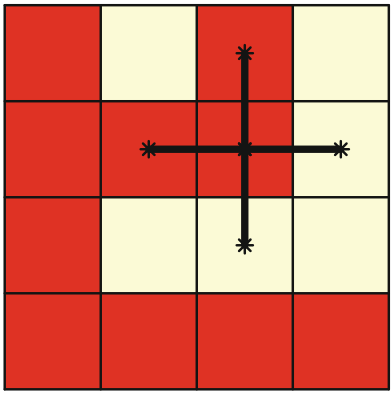


Fig. 9.12 A 4×4 lattice of atoms of a crystalline solid with their spins (-1 dark colour and 1 light colour) and a nearest-neighbour interaction, denoted by the cross

9.8 Ising Model

Let us now discuss an application of the Metropolis algorithm to a well-known model, used by the German physicist E. Ising to explain some observed behaviours in the magnetization of materials.[4]

The binary image formed by the 4×4 pixel[5] matrix of Fig. 9.12 represents a two-dimensional portion of a crystalline solid, where an atom, with its intrinsic magnetic moment (due to spin), is located on every pixel: we associate spin value -1 to the dark colour and spin value $+1$ to the light colour.

Depending on the temperature, the interaction existing between the nuclear spins at the microscopic level defines the behaviour of the material at the macroscopic level, determining its ferromagnetic or antiferromagnetic properties. In Fig. 9.12 the simplest model of microscopic interaction is shown, in which the atom in the middle of the cross only interacts with its horizontal and vertical nearest neighbours.

Under this approximation, the Ising model defines the energy of a ferromagnetic material with $n \times n$ atoms as:

$$H(\boldsymbol{x}) = -\beta \sum_{i,j;i\sim j} x_i x_j, \quad \beta > 0\,,$$

where $\boldsymbol{x} = (x_1, \ldots, x_{n^2})$, x_i is the spin of the i-th atom and the sum is over the nearest neighbours ($i \sim j$).

[4] This model has important applications also in fields completely different from physics, since it well describes the evolution of systems in which there are changes of state as a result of interactions between individuals. For example, it is used to study the social impact of new ideas and the dynamics of opinion in complex societies [KH96] and to predict the behaviour of financial markets [Voi03].

[5] The term pixel comes from the contraction of the words *picture element* and indicates the smallest homogeneous unit constituting an artificial image.

Without an external magnetic field, a probability for each configuration can be determined depending on the energy and on the temperature T as:

$$p(\boldsymbol{x}) = \frac{\exp\left\{-\frac{H(\boldsymbol{x})}{T}\right\}}{\sum_x \exp\left\{-\frac{H(\boldsymbol{x})}{T}\right\}} \equiv \frac{h(\boldsymbol{x})}{Z} .$$

The formula shows that low-energy configurations, that is, those with neighbouring atoms with the same spin, have a higher probability to be reached.

In 1944, L.Onsager [Ons44] developed the exact analytical treatment of the Ising model in two dimensions, with the calculation of the expected number of atoms with spin 1 at a given temperature, a quantity needed to determine the total magnetic moment of the material. However, the more realistic three-dimensional model has not been solved yet, and we need to resort to simulation, which we describe in the two-dimensional case for simplicity.

In this situation, the admissible configurations of all spins are 2^{n^2}, so it is not possible, except for very small n, to calculate all $p(\boldsymbol{x})$ values and carry on the direct simulation as in Sect 8.4. We therefore use the Metropolis algorithm with an auxiliary distribution that, at each step, randomly selects an atom and proposes its spin change. It is quite easy to realize that the formula for this distribution is as follows:

$$q(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{n^2} ,$$

if $\boldsymbol{x}$ and $\boldsymbol{y}$ differ in the spin of an atom only, whereas $q(\boldsymbol{x}, \boldsymbol{y}) = 0$ in all the other cases. To easily calculate the acceptance ratio, we can notice that the energy of the system can be written as:

$$H(\boldsymbol{x}) = \beta(n^-(\boldsymbol{x}) - n^+(\boldsymbol{x})) ,$$

where n^+ and n^- indicate the number of nearest neighbour atom pairs with concordant and discordant sign, respectively. Then, since $q(\boldsymbol{x}, \boldsymbol{y}) = q(\boldsymbol{y}, \boldsymbol{x})$, the acceptance ratio becomes simply:

$$\exp\{\beta[(n^+(\boldsymbol{y}) - n^+(\boldsymbol{x})) - (n^-(\boldsymbol{y}) - n^-(\boldsymbol{x}))]/T\} .$$

Now, if the auxiliary distribution has chosen the i-th atom, the differences in the exponent depend exclusively on the signs of the nearest neighbours of that atom. By denoting with $n^+(x_i)$ and with $n^-(x_i)$ the number of nearest neighbours with the same sign of x_i and with opposite sign, respectively, and taking into account that $n^-(x_i) = 4 - n^+(x_i)$, the acceptance ratio becomes:

$$\exp\{2\beta[n^+(y_i) - n^+(x_i)]/T\} = \exp\{2\beta[4 - 2n^+(x_i)]/T\} .$$

Therefore, once an initial configuration has been selected (e.g. by randomly choosing the spin of each atom), the algorithm checks the spin orientations both of the atom involved in the change and of its nearest neighbours. If $g(\boldsymbol{x}^{(i)})$ is the number of atoms with spin 1 at the i-th step of the algorithm, the next term $g(\boldsymbol{x}^{(i+1)})$ used for the calculation of the mean (9.36) is unchanged, if the chosen atom does not change spin or, otherwise, is easily obtained by adding 1 or -1. From an algorithmic point of view, the index of the atom to be changed is obtained by generating a uniform variate ξ_1 in (0, 1) and selecting the smallest integer i that exceeds $\xi_1 n^2$. Given i, if a second uniform variate ξ_2 is less than the acceptance probability

$$\alpha(\boldsymbol{x}, \boldsymbol{y}) = \min\left\{1, \exp\{2\beta[4 - 2n^{+}(x_i)]/T\}\right\}$$

a spin change occurs. This algorithm is implemented in our code MCising.

We performed a simulation of the Ising model with $N = 100\,000$ iterations, $\beta = 0.3$ and $T = 1$, starting from the configuration where all atoms have spin -1 in a lattice $n \times n$ with $n = 20$. Our goal was to determine the expected number of atoms with spin 1, so $g(\boldsymbol{x})$ will give this value for the configuration $\boldsymbol{x}$. In Fig. 9.13 the dashed line shows the time evolution of the sequence $\{g(\boldsymbol{x}^{(i)})\}$. The sample running mean $\langle g(\boldsymbol{x})\rangle$ calculated as a function of the number of iterations is displayed with a continuous line. Some main features can be easily noted:

(1) After a few thousand iterations, the sequence $\{g(\boldsymbol{x}^{(i)})\}$ stabilizes and begins to oscillate around its presumed expected value.
(2) The plot of the sample mean instead tends to converge to a constant value, which, according to Theorem 9.1, is $\langle g(\boldsymbol{X})\rangle$.

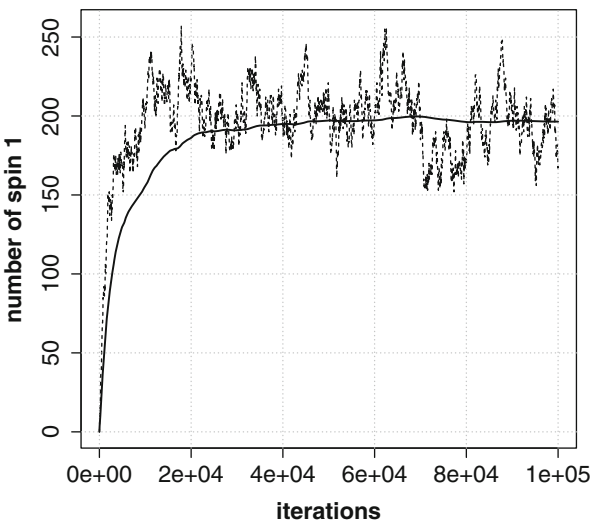


Fig. 9.13 Result of the Metropolis simulation for the Ising model: the number of atom with spin 1 (dashed line) and the sample mean of this parameter (solid line) are shown as a function of the number of iterations

(3) Since the simulation starts from very low probability values that are quickly abandoned by the system, a distortion in the estimate of the mean is present, since the configuration with all spin -1 atoms will never be spontaneously reached during any finite length simulation. This situation gives the initial states a greater weight than it should. We have then to discard an initial number of iterations (e.g. the first 10,000) to reach a high probability zone and start to accumulate data from this point to compute $\langle g(\boldsymbol{x})\rangle$.
(4) The amount of data to be collected can be evaluated using the plot of the sample mean. The simulation can be stopped when the mean oscillations have an amplitude lower than a certain threshold (which is subjective and depends on the aimed precision).
(5) The group size b for the *batch means* method must be increased until the autocorrelation of the series $\{\langle g(\boldsymbol{x})^{(i)}\}_{i\geq 1}$ becomes negligible. As suggested at the end of the previous section, we set b equal to the square root of the sample size (for the details see again [FJ10]). This in turn may require increasing the number of iterations in order not to have a too small number of groups. Our example has precisely these characteristics.

Considering Fig. 9.13, if the first 10,000 iterations are discarded and the other 90,000 are retained, the interval estimate (9.38), with $b = 300$ (i.e. 300 groups), is:

$$200.9 \pm 2 \times 1.1 = (198.7, 203.1)\,,$$

which was obtained by applying our `Batchmeans` routine to the sample sequence of the number of spin 1 atoms.

We conclude with an important warning. We suggest you to perform, using our `MCising` code, additional simulations with temperatures T lower than 1. You will find that, when the temperature gives $\beta/T > 0.44$, the plot of $g(\boldsymbol{x})$ will move fairly quickly to one of the two modes of its distribution (i.e. $g(\boldsymbol{x}) = 0$ or $g(\boldsymbol{x}) = 400$), without being able to evolve from one to the other. In this type of situations, although Theorem 9.1 still holds, it is practically impossible to carry out the number of iterations necessary to visit the state space regions of greatest probability; hence the sample mean of the plot is by no means a good estimate of the true one.

9.9 Definite Integral Calculation

The numerical computation of the value of a definite integral, one of the best known and most widespread applications of the MC methods, is also a typical example of the use of simulation techniques for problems that, at first sight, would seem not to allow a statistical approach.

As we will see shortly, with the MC methods it is convenient to solve multi-dimensional integrals, where the other numerical methods present some relevant

application problems. In the following, however, to compact and simplify the notation, we will consider only the single-valued functions, bearing in mind that the whole discussion can be very easily extended to the case of functions of many variables.

There are two different fundamental approaches that can be used to calculate the definite integral:

$$I = \int_a^b f(x)\,\mathrm{d}x \tag{9.39}$$

using random numbers.

The first method, called hit or miss, is based on the geometric interpretation of the value of a definite integral as a measure of the area under $f(x)$ and within the integration interval $[a, b]$ (dashed area of Fig. 9.14). By exploiting, from a different point of view, the same properties applied in the rejection method, we can in fact determine I by multiplying the area of a rectangle enclosing $f(x)$ by the probability that any point P inside the rectangle is also inside the area under $f(x)$ (hatched area of Fig. 9.14). Recalling relation (8.41), we have:

$$p = \frac{\text{hatched area}}{\text{area of the rectangle ABCD}} = \frac{I}{h \cdot (b-a)}\,, \tag{9.40}$$

so that:

$$I = pA\,, \quad \text{with } A = h \cdot (b-a)\,. \tag{9.41}$$

If we randomly generate N points (x_1, y_1), (x_2, y_2), $\ldots$, (x_N, y_N) uniformly distributed inside the rectangle ABCD and count the number N_S of "successes", i.e. the number of times in which $y_i \leq f(x_i)$, an approximate evaluation of p is

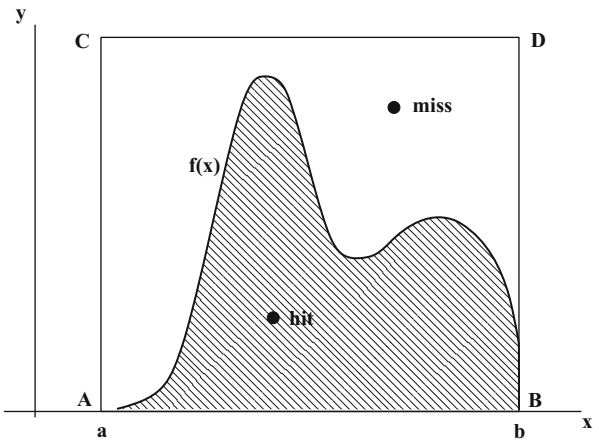


Fig. 9.14 Graphic representation of the hit or miss method: the integral (9.39) is evaluated using the proportion of random points uniformly distributed inside the rectangle ABCD that also "hit" the dashed zone. The rectangle has basis $(b-a)$ and height $h \geq$ the maximum value of $f(x)$

obtained using the ratio N_S/N (see Fig. 9.14). The estimate of the integral (9.39) then becomes:

$$I = pA \simeq p_N A = A\frac{N_S}{N} \equiv I_N^{HM} \,. \tag{9.42}$$

This value should be considered as the realization of a new "Hit or Miss"statistical variable $\boldsymbol{I_N^{HM}}$. To derive its error, it is sufficient to observe that N_S follows a binomial distribution with mean $Np = NI/A$ and variance $Np(1-p)$; we therefore get:

$$\mathrm{Var}[\boldsymbol{I_N^{HM}}] = \frac{A^2}{N^2}\mathrm{Var}[\boldsymbol{N_S}] = \frac{I(A-I)}{N} \simeq \frac{I_N^{HM}(A - I_N^{HM})}{N} \,. \tag{9.43}$$

The second MC integration method, known as crude Monte Carlo, considers instead x as a uniformly distributed random variable within the integration interval $[a, b]$. Recalling Definition (2.68) of the expected value of a function of random variable, we can write the identity:

$$I = (b-a)\int_a^b \frac{1}{(b-a)} f(x)\,\mathrm{d}x = \langle f(x)\rangle\,(b-a) \,, \tag{9.44}$$

with $\langle f(x)\rangle$ equal to the mean of the values assumed by the integrand function on $[a, b]$.

One of the ways to roughly evaluate $\langle f(x)\rangle$ is to calculate the average of N values $f(x_1), f(x_2), \ldots, f(x_N)$, with $x_1, x_2, \ldots, x_N$ randomly and uniformly sampled within $[a, b]$:

$$I \simeq \frac{(b-a)}{N}\sum_{i=1}^{N} f(x_i) \equiv I_N^M \,. \tag{9.45}$$

The variance of the random variable $\boldsymbol{I_N^M}$ is easily obtained from the properties of the variance operator defined in Sect 2.9:

$$\mathrm{Var}[\boldsymbol{I_N^M}] \equiv \sigma^2_{I_N^M} = \frac{(b-a)^2}{N^2}\mathrm{Var}\Big[\sum_{i=1}^{N} f(X_i)\Big] = \frac{(b-a)^2}{N}\mathrm{Var}[f(X_i)] \,. \tag{9.46}$$

Since $f(X_i)$ can be considered as a function of the random variable X_i, taking into account Eqs. (9.44) and(2.49), one gets:

$$\begin{aligned}\mathrm{Var}[\boldsymbol{I_N^M}] &= \frac{1}{N}\left[(b-a)\int_a^b f^2(x)\,\mathrm{d}x - \left(\int_a^b f(x)\,\mathrm{d}x\right)^2\right] \\ &\simeq \frac{(b-a)^2}{N(N-1)}\left[\sum_{i=1}^{N} f^2(x_i) - \frac{1}{N}\left(\sum_{i=1}^{N} f(x_i)\right)^2\right] .\end{aligned} \tag{9.47}$$

The term within square brackets is a measure of how much $f(x)$ differs from its mean value in the integration region, so $\mathrm{Var}[\boldsymbol{I}_N^{\boldsymbol{M}}]$ strongly depends on the codomain of $f(x)$.

Exercise 9.2
Compute, using MC, the integral:

$$I = \int_0^{\frac{\pi}{2}} \sqrt{\sin x}\, \mathrm{d}x \ , \tag{9.48}$$

Answer Despite the apparent simplicity, this integral is not analytically solvable. The exact solution, obtainable only numerically, is:

$$I = \int_0^{\frac{\pi}{2}} \sqrt{\sin x}\, \mathrm{d}x = 1.19814\ldots \tag{9.49}$$

We apply the two MC integration methods previously described using our `MCinteg` routine.

(a) Crude MC method of Eqs. (9.45, 9.47)
By generating 1000 random points, we obtained $I = 1.199 \pm 1.2 \cdot 10^{-2}$, as given in the first row of Table 9.3. As we have already noted previously, if you want to obtain a very precise solution (e.g. within a few per thousand), you need to generate a large number of points, given the low convergence speed of the MC result to I.

(b) Hit or miss method of Eqs. (9.42, 9.43)
In this case we have: $h = 1$, $(b - a) = \pi/2$, and, with 1 000 random points, we obtained $I = 1.202 \pm 2.1 \cdot 10^{-2}$, a result with an error about two times the previous one.

In the MC framework, one defines as efficiency η of an algorithm the quantity:

$$\eta = 1/(t_N \mathrm{var}[S_N]) \ , \tag{9.50}$$

in which S_N is the integral value given by the algorithm and t_N the time needed for its computation. In our example, since t_N of the two procedures is about the same, with the hit or miss method, we have decreased the efficiency of the numerical integration programme by about $four$ times.

This difference obviously depends on the particular integral considered, but it is easy to demonstrate (see, e.g. [Jam80]) that the hit or miss method always gives a less precise result than the crude MC. This fact can be intuitively understood by observing that, in the first method, for each generated point x_i, the value 1 is added with probability $f(x_i)$, instead of adding the corresponding value $f(x_i)$. In this way, an estimate is used instead of the exact value and, therefore, an additional error is introduced.

Table 9.3 Comparison between the estimates of the integral (9.48) obtained by the generation of 1000 random points with our `MCinteg` routine. The different algorithms are explained in the text

Algorithm	I_N	σ_{I_N}
Crude MC	1.199	$1.2 \cdot 10^{-2}$
"Hit or miss"	1.202	$2.1 \cdot 10^{-2}$
Importance sampling	1.1987	$2.5 \cdot 10^{-3}$
Stratified sampling		
(a) Proportional	1.1981	$1.0 \cdot 10^{-3}$
(b) Optimal	1.19785	$6.8 \cdot 10^{-4}$

9.10 Importance Sampling

The method of Eqs. (9.45, 9.47) is called "crude"since, for a given number of samplings N, a better precision can be achieved with more sophisticated techniques. This result can be obtained, according to Eq. (8.6), by reducing the variance σ_T of the distribution of the simulated variable T_i.

This operation becomes relatively easy in the integration as σ_T coincides with the standard deviation of the integrand function (see Eq. 9.46), whose manipulation is generally quite simple.

For example, while discussing Eq. (9.47), we noted that the greater the variation of $f(x)$ in the integration region, the greater will be the error on the result that, conversely, becomes more precise when the generated values of $f(x)$ are not too dissimilar to each other. Then, the MC estimate of I can be made more precise when a positive integrable function $g(x)$ is found such that $g(x) \simeq f(x)$, which allows us to rewrite I as:

$$I = \int_a^b f(x)\,\mathrm{d}x = \int_a^b \frac{f(x)}{g(x)} g(x)\,\mathrm{d}x = \int_a^b \frac{f(x)}{g(x)}\,\mathrm{d}G(x)\,, \tag{9.51}$$

with:

$$G(x) = \int g(x)\,\mathrm{d}x\,. \tag{9.52}$$

If we assume that $g(x)$ is also normalized within $[a, b]$, it is easy to conclude that Eq. (9.51) represents nothing more than the mean of the random function $f(X)/g(X)$, where X is a random variable with density $g(x)$, according to Eq. (2.68). We then can write I as:

$$I = \left\langle \frac{f(X)}{g(X)} \right\rangle\,. \tag{9.53}$$

If we consider a random sample $X_1, X_2, \ldots, X_N$ from the density $g(x)$ we obtain:

$$I \simeq \frac{1}{N} \sum_{i=1}^{N} \frac{f(X_i)}{g(X_i)} = I_N^{IS} , \quad X \sim g(x) . \tag{9.54}$$

Instead of uniformly generating x to integrate $f(x)$, a random variable distributed as $g(x)$ is generated to integrate $f(x)/g(x)$, thus giving more weight to the more "important"parts of $f(x)$ (hence the name of importance sampling given to this algorithm). The final variance now depends on the ratio $f(x)/g(x)$, and, recalling Eq. (9.47), it is given by:

$$\text{var}[\boldsymbol{I_N^{IS}}] = \frac{1}{N}\left[\int_a^b \frac{f^2(x)}{g^2(x)}\,\mathrm{d}G(x) - I^2\right] = \frac{1}{N}\left[\int_a^b \frac{f^2(x)}{g(x)}\,\mathrm{d}x - I^2\right], \tag{9.55}$$

or, in an approximated way, as in Eq. (9.47):

$$\text{var}[\boldsymbol{I_N^{IS}}] \simeq \frac{1}{N(N-1)}\left\{\sum_{i=1}^{N} \frac{f^2(x_i)}{g^2(x_i)} - \frac{1}{N}\left[\sum_{i=1}^{N} \frac{f(x_i)}{g(x_i)}\right]^2\right\} . \tag{9.56}$$

We recall that in both previous equations, x is sampled from $g(x)$.

From Eq. (9.55) we immediately notice how the variance becomes zero if $g(x) = f(x)/I$. Unfortunately, in real problems, we will never be able to apply this equation since it is necessary to know I, which is exactly the solution of the problem. However, Eq. (9.55) quantitatively demonstrates that $g(x)$ must be chosen as much as possible similar to $f(x)$. This choice ensures that the ratio between the two functions varies inside a limited range within the integration region, thus maximizing the gain in precision.

9.11 Stratified Sampling

The idea behind the stratified sampling, a well-known technique used in statistics, is similar to the one we have just described: a greater number of points are concentrated in those areas that are more important for the calculation. The difference is now that, instead of changing the integrand, the integration region is divided into different subintervals; then points are sampled uniformly but with different densities, depending on the particular considered interval.

Thus, the interval $[a, b]$ is divided into k segments defined by the points $a = \alpha_0 < \alpha_1 < \alpha_2 < \ldots < \alpha_k = b$. The width and the sample size of the j-th subinterval are denoted by $\Delta_j = (\alpha_j - \alpha_{j-1})$ and N_j, respectively. For the well-known integral properties, we can also write:

$$I = \int_a^b f(x)\,\mathrm{d}x = \sum_{j=1}^{k} \int_{\alpha_{j-1}}^{\alpha_j} f(x)\,\mathrm{d}x = \sum_{j=1}^{k} I_j \ . \tag{9.57}$$

In stratified sampling, each interval I_j which appears in the right term of the previous equation is approximated with the crude method:

$$I_j = \int_{\alpha_{j-1}}^{\alpha_j} f(x)\,\mathrm{d}x \simeq \frac{\Delta_j}{N_j} \sum_{i=1}^{N_j} f(\alpha_{j-1} + \Delta_j \xi_{ij}) = \frac{\Delta_j}{N_j} \sum_{i=1}^{N_j} f(x_{ij}) \ , \tag{9.58}$$

where x_{ij} is the i-th point sampled from the uniform distribution within the j-th subinterval). The I estimate then becomes:

$$\boldsymbol{I} = \sum_{j=1}^{k} \Delta_j \left\langle f_j(X) \right\rangle \simeq \sum_{j=1}^{k} \sum_{i=1}^{N_j} \frac{\Delta_j}{N_j} f(x_{ij}) = I_N^{CS} \ , \tag{9.59}$$

where $\left\langle f_j(X) \right\rangle$ is the mean value of $f(X)$ in the j-th subinterval. The global variance, which is nothing more than the sum of the variances of the single I_j taken separately, is given by Eq. (9.47):

$$\mathrm{var}[\boldsymbol{I}_N^{CS}] = \sum_{j=1}^{k} \frac{\Delta_j^2}{N_j} \sigma_j^2 = \sum_{j=1}^{k} \frac{1}{N_j} \left\{ \Delta_j \int_{\alpha_{j-1}}^{\alpha_j} f^2(x)\,\mathrm{d}x - \left[\int_{\alpha_{j-1}}^{\alpha_j} f(x)\,\mathrm{d}x \right]^2 \right\} \tag{9.60}$$

(σ_j^2 is the variance of $f(x)$ in the j-th subinterval), or, in an approximated way:

$$\mathrm{var}[\boldsymbol{I}_N^{CS}] \simeq \sum_{j=1}^{k} \frac{\Delta_j^2}{N_j(N_j-1)} \left\{ \sum_{i=1}^{N_j} f^2(x_{ij}) - \frac{1}{N_j} \left[\sum_{i=1}^{N_j} f(x_{ij}) \right]^2 \right\} . \tag{9.61}$$

The error on the final result now depends not only on the behaviour of $f(x)$ but also on the way in which the integration domain is divided and on how the points are distributed within each subinterval. Once the subintervals Δ_j are fixed, with a simple but rather involved procedure, one can derive from Eq. (9.60) that the better choice for N_j is given by the rule [Coc77]:

$$N_j = \frac{N \Delta_j \sigma_j}{\sum_{j=1}^{k} \Delta_j \sigma_j} \ . \tag{9.62}$$

As intuitively expected, this equation prescribes to concentrate the random generation in the largest subintervals and in those with the greatest variations. Even in this case, however, this result is not directly applicable, as the σ_j values are obviously unknown a priori. To use the previous formula, a short preliminary test is usually performed to obtain a fairly correct estimate of σ_j, from which to derive an appropriate value for N_j. In doing this, a reasonable compromise must be made between the required calculation time and the increase in precision that can be obtained on the final result.

When this procedure is too long or complicated, it is possible to demonstrate (see again [Coc77]) that the best way to proceed is to generate a number of points proportional to the length of each subinterval:

$$N_j = N \frac{\Delta_j}{(b-a)} \,. \tag{9.63}$$

This property can be intuitively understood if we observe that with the stratified proportional sampling, the uniformity of generation of random points is improved compared to the crude method, thus reducing the statistical fluctuations due to a relevant increase of their density in specific zones of the integration interval.

The subinterval lengths can instead be optimized only when the points N_j are chosen on the basis of the Eq. (9.62), and the integration domain is simultaneously divided into a large number of subintervals (see again [Coc77]), conditions that are not always satisfied in practice. Otherwise, there are no strict prescriptions; usually, to simplify the programmes, the subintervals Δ_j are all taken with the same length:

$$\Delta_j = \frac{(b-a)}{k} \qquad \forall\, j = 1, \ldots, k \,. \tag{9.64}$$

Exercise 9.3
Calculate integral (9.48) with the importance and stratified sampling techniques.

Answer (a) Importance sampling
Since, in a neighbourhood of zero, the function $\sin x$ can be expanded as:

$$\sin x \simeq x - \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots \tag{9.65}$$

(continued)

Exercise 9.3 (continued)

we can try to approximate the integrand function ($\sqrt{\sin x}$) with $h(x) = \sqrt{x}$, as shown in Fig. 9.15. Hence, we choose this form for $g(x)$ that, after normalization, becomes:

$$g(x) = \frac{3}{\pi}\sqrt{\frac{2}{\pi}}\sqrt{x}\,, \tag{9.66}$$

giving the cumulative function:

$$G(x) = \int_0^x g(x)\,\mathrm{d}x = \left(\frac{2}{\pi}x\right)^{3/2}. \tag{9.67}$$

Equation (8.12) results in:

$$x = \frac{\pi}{2}(\xi)^{2/3}\,. \tag{9.68}$$

The x variates generated with this formula are used to calculate the integral according to Eq. (9.54). This algorithm is performed by our `MCinteg` routine.

The result with $N = 1000$, reported in the third row of Table 9.3, clearly shows an improvement in precision of about an order of magnitude compared to the crude method. Since the computing time is roughly the same for the 2 algorithms, the gain in efficiency is nearly around 100 ! However, to successfully apply this method, $g(x)$ must be easy to sample from (recall that its cumulative must be known), even when the behaviour of $f(x)$ is complicated. Moreover, in the multidimensional case, to minimize the computation time, it is definitely preferable that $G(x_1, x_2, \ldots, x_n)$ is a separable function:

$$G(x_1, x_2, \ldots, x_n) = g(x_1) \cdot g(x_2) \cdot \ldots \cdot g(x_n)\,.$$

The classes of functions satisfying all these conditions are a few, and essentially: trigonometric, exponential, low order polynomials and some combinations of them.

(b) Stratified sampling

We apply this algorithm using the uniform stratified sampling technique. We then consider equal subintervals and generate the same number of points in each of them. By dividing the integration domain into 20

(continued)

Exercise 9.3 (continued)

subintervals and generating 1 000 points, we obtained the value reported in the fourth row of Table 9.3. Compared to the crude method, we note an improvement in the standard deviation of a factor of $\simeq$ 12, which, even taking into account an additional calculation time of about 30%, corresponds to an increase in efficiency of more than *two* orders of magnitude.

With 10,000 points and 100 layers, we obtained a very accurate result:

$$I = 1.198154 \pm 0.000023\,.$$

With the same total number of points and subintervals, the precision can be further increased if the number of points in each subinterval is determined with the optimal method of Eq. (9.62)), as shown in the last row of Table 9.3. This algorithm is present in our `MCinteg` and `MCintopt` routines; the last one applies the optimized stratified sampling to an input function.

In addition to the methods presented here, there are several other variance reduction techniques; those interested can consult [Kah56, KW86, Rub81, Rip86].

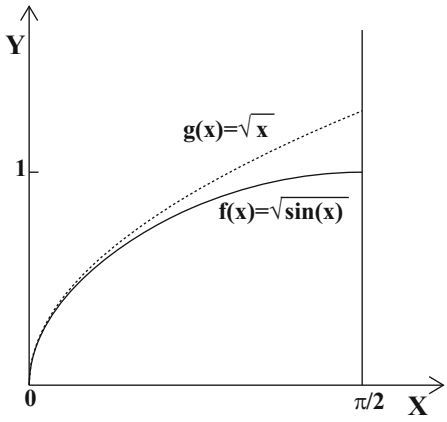


Fig. 9.15 Comparison between the functions $f(x) = \sqrt{\sin x}$ and $h(x) = \sqrt{x}$ in the interval $[0, \pi/2]$

9.12 Multidimensional Integrals

Up to now, we have shown how it is possible to improve, even significantly, the precision on the MC estimate of the value of a definite integral by using variance reduction techniques. But, despite this progress, MC methods, for one-dimensional functions, are always less efficient than the standard numerical approximation procedures, which converge as N^{-k} with $k \geq 2$.

However, this situation changes when we consider the multidimensional case, as the error of the MC methods is independent of the dimensionality d of the integral, while the precision of the other numerical techniques varies as $N^{-k/d}$, with $k \geq 2$ (a process requiring N points in one dimension will then need to get the same precision, N^2 points in two dimensions, N^3 points in three dimensions and so on). Furthermore, all the other methods assume a smooth polynomial behaviour of the integrand function, and, therefore, their use is questionable in case of discontinuities. For all these reasons, MC methods start to be competitive for five-dimensional integrals, becoming the only ones actually applicable in ten or more dimensions.

In case you have to compute for the first time the multidimensional integral:

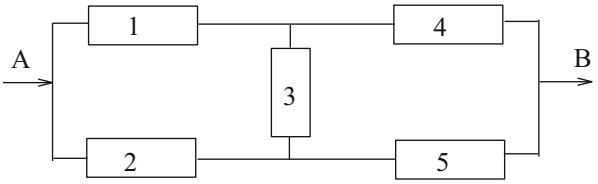
$$I = \int_{\Omega} f(x_1, \ldots, x_d)\, \mathrm{d}x_1 \ldots \mathrm{d}x_d \tag{9.69}$$

with MC methods, we suggest you to try to use some existent software. There are several reliable codes available on the market, which exploit the variance reduction techniques described before (see, e.g. [PFTW92]). In these codes, to make the algorithms simpler and more reliable, the integration is carried out on a domain having independent integration limits (a hypercube or a hyper-rectangle) obtained by performing an appropriate change of coordinates. If this transformation turns out to be too complex, it is often preferable to geometrically consider the integral (9.69) as the volume of a solid W in the space with $d + 1$ dimensions $(x_1, \ldots, x_d, y)$ and apply the hit or miss method introduced in Sect. 9.9.

9.13 Problems

9.1 Often the maximum deviation is used in quality controls: a characteristic parameter X, considered as a normal random variable, is chosen, and the difference between the maximum and minimum values of X found in a control lot of n elements is examined. Knowing that the average production deviation of X is $\sigma = 0.5$, write a simulation code to determine the value $\Delta = x_{\max} - x_{\min}$ above which to discard the production with $CL = 99\%$ for a control batch of $n = 100$ elements.

9.2 A device is formed by five components, according to the scheme shown in the figure:



The system operates if the "workflow"goes from A to B, that is, if at least one of paths (1,3,5), (2,3,4), (1,4) e (2,5) is working. Determine, using a simulation, the average operating time of the device, knowing that the failure time of each component is a negative exponential random variable with a mean value of $\langle t_1 \rangle$, $\langle t_2 \rangle$, $\langle t_4 \rangle$, $\langle t_5 \rangle = 3$ days and $\langle t_3 \rangle = 5$ days, respectively.

9.3 Modify the integration routine `MCinteg` (used in the Exercises 9.2 and 9.3) to calculate the integral of the standard Gaussian:

$$(1/\sqrt{2\pi}) \int_0^x \exp(-t^2/2)\ \mathrm{d}t\ .$$

Compare the results with those of Table E.1 for various x values. Exclude the importance sampling technique. Try also to use our routine `MCintopt`.

9.4 Using all the MC methods discussed in the text, calculate, possibly modifying the `MCinteg` routine, the integral:

$$I = \int_0^1 \log(1+x)\,\mathrm{d}x\ .$$

9.5 Calculate with the MC methods the integral:

$$\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} (x_1^2 + x_2^2 + x_3^2)\ \mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3\ .$$

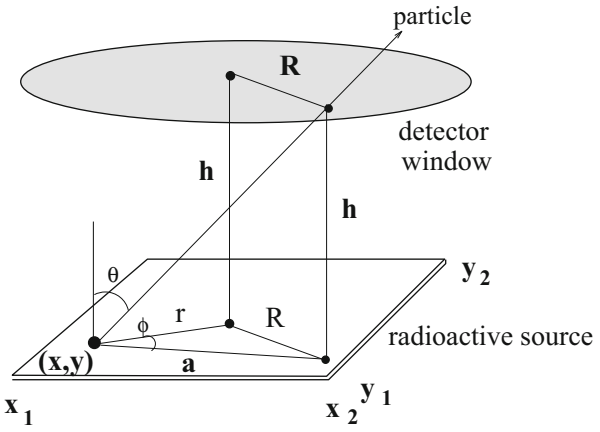
9.6 Compute, with the hit or miss method, the area of the ellipse:

$$\frac{x^2}{2} + y^2 = 1\ .$$

9.7 Implement a Metropolis algorithm to estimate the mean and variance of a standard Gaussian distribution, using a simulation of length N. Set $U(-a, a)$ as an

auxiliary distribution. Then, extract a sample of size N from a standard Gaussian with the `rnorm` routine of R. Compare the trajectories of the estimators as N increases with $a = 2$. Repeat the experiments with different values of a: is the speed of convergence of the trajectories of the mean and variance estimators influenced by a? Is it reasonable to use $a > 3$?

9.8 A detector has an entrance window of radius $R_d = 3\,\text{cm}$, and a plane radioactive source, with sides $(x_2 - x_1) = 4\,\text{cm}$ and $(y_2 - y_1) = 6\,\text{cm}$, is placed at a distance of $h = 5\,\text{cm}$ from it (see figure). The source emits particles isotropically with angles (θ, ϕ) from evenly distributed points (x, y). Calculate the geometric efficiency of the detector (i.e. the probability for an emitted particle to enter the detector) using the configuration proposed in the figure.



9.9 The Von Mises distribution has density:

$$p(x) = \frac{1}{2\pi I_0(c)} e^{c\,\cos(x)}, \quad -\pi \le x \le \pi\ ,$$

where $I_0(c)$ is the zero order modified Bessel function of the first type. Use the Metropolis algorithm to generate a sample from this distribution without calculating $I_0(c)$.

9.10 Generate various samples from the Ising model with $\beta = 0.3$ and $T = 1$, increasing step by step the dimension $n \times n$ of the lattice with the number N of iterations fixed at 100,000. On the basis of the plot of the number of atoms with spin $= 1$, what is, in your opinion, the maximum value of n for which the algorithm is working correctly?

Chapter 10
Statistical Inference and Likelihood

In speaking of the most probable consequence, we must remember that in reality the probability of transition to states of higher entropy is so enormous in comparison with that of any appreciable decrease in entropy that in practice the latter can never be observed in nature.

L.D. Landau, E.M. Lifchitz and L.P. Pitaevskij,
"STATISTICAL PHYSICS".

10.1 Introduction

In Chaps. 6 and 7, we introduced estimation theory and hypothesis testing in the context of elementary statistics: from a data sample, an estimate of a statistical parameter (mean, probability, correlation coefficient, etc.) with its error and the related confidence interval is determined using a data sample. Afterwards, if necessary, the compatibility of this estimate with a model, generally called the null hypothesis H_0, is verified by means of χ^2 or other tests.

In this scheme, which can be defined "static", the model to be checked is given a priori and is not modified by the information coming from the collected sample.

However, a more "dynamic" and efficient approach can be adopted, which consists in determining *the most likely model of the parent population* on the basis of the collected data. In this case, the estimation intervals (which, as we know, may depend on the population model assumed) are then modified accordingly, before performing the test between data and model. The scheme is that of Fig. 10.1, where the bold arrows indicate the differences with respect to the static model considered so far.

The methods that determine the most likely population model from a set of data by improving the parameter estimation are known as *best fit methods*. In practice, the optimization of the model from the data is usually achieved assuming the family of the density function (binomial, Gaussian, Poissonian, uniform or other) to be known and considering the characteristic parameters of the distribution such as mean, variance or existence limits, *as free parameters*.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3_10

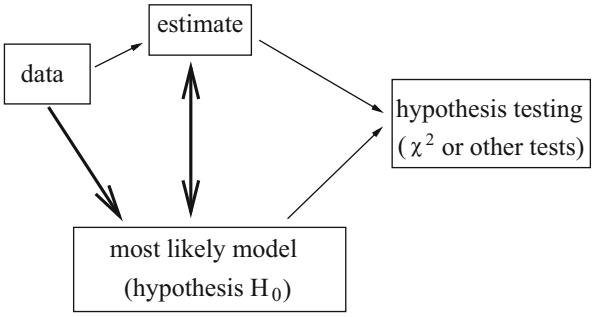


Fig. 10.1 Parameter estimation and hypothesis testing. The bold arrows denote the steps of the model optimization

In this chapter we will use the following notation: $\boldsymbol{X}$ is a vector of random variables, each of them generate a sample, according to the Definition 2.12. Therefore, we will refer to:

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_m) \tag{10.1}$$

as the observed values of $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)$ in a trial, and we will write that:

$$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{mi})$$

are the occurrences of $\boldsymbol{X}$ in the i-th trial.

We introduce also a new notation, distinguishing between the variable $\boldsymbol{X}$ and a n-dimensional sample:

$$\underline{\boldsymbol{X}} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)\ . \tag{10.2}$$

The values assumed by the sample after the experiment (i.e. after n trials) are:

$$\underline{\boldsymbol{x}} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)\ . \tag{10.3}$$

Hence, we consider a probability space $(S, \mathcal{F}, P_\theta)$ depending on one or more parameters, according to Eq. (6.1). The density to be optimized is therefore of the type $p(\boldsymbol{x}; \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ is a p-dimensional parameter. To optimize the density then means to determine the values of the $\boldsymbol{\theta}$ parameters which better fit to the collected data, having a priori fixed a functional form.

Let us start with a single random variable X, and let $\underline{x} = (x_1, x_2, \ldots, x_n)$ be the observed values in n independent trials. If $p(x; \boldsymbol{\theta})$ is the density of X (depending on a set of parameters $\boldsymbol{\theta}$), we can apply the law of compound probabilities to the case

of n independent trials carried out on the same variable, and, recalling Theorem 4.1, we can associate the observation with the probability density:

$$L(\boldsymbol{\theta};\, \underline{x}) = p(x_1;\boldsymbol{\theta})\, p(x_2;\boldsymbol{\theta}) \cdots p(x_n;\boldsymbol{\theta}) \equiv \prod_{i=1}^{n} p(x_i;\boldsymbol{\theta})\ . \tag{10.4}$$

The name given to this product, considered as a function of $\boldsymbol{\theta}$, is *likelihood function*. For any fixed $\boldsymbol{\theta}$, it represents, apart from the differential factors to be integrated, the probability to obtain the values $\underline{x}$.

For m variables, the likelihood function is generalized in an obvious way through Eqs. (10.1)–(10.3):

$$\begin{aligned} L(\boldsymbol{\theta};\, \underline{\boldsymbol{x}}) = {} & p(x_{11}, x_{21}, \ldots, x_{m1};\boldsymbol{\theta})\; p(x_{12}, x_{22}, \ldots, x_{m2};\boldsymbol{\theta}) \ldots \\ & \times p(x_{1n}, x_{2n}, \ldots, x_{mn};\boldsymbol{\theta}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i;\boldsymbol{\theta})\ , \end{aligned} \tag{10.5}$$

where the product is extended to all n experimental values obtained for each of the m variables $\boldsymbol{X}$. The general definition of the likelihood function also includes the case of non-independent trials; we will take this possibility into account in Sects. 12.7 and 12.8, when considering experimental data affected by systematic errors. As we will see, the mathematical properties of the likelihood function of interest in statistics are the same as its logarithm. It is then possible to eliminate the product in Eqs. (10.4), and (10.5) by defining the new function:

$$\mathcal{L} = -\ln\left(L(\boldsymbol{\theta};\, \underline{\boldsymbol{x}})\right) = -\sum_{i=1}^{n} \ln\left(p(\boldsymbol{x}_i;\boldsymbol{\theta})\right)\ , \tag{10.6}$$

where the minus sign in front of the logarithm should be noticed. If this convention is adopted, a maximum of L corresponds to a minimum of $\mathcal{L}$. In the following, to simplify calculations, we often use, instead of the L function, its negative logarithm $\mathcal{L}$.

10.2 Maximum Likelihood (ML) Method

The maximum likelihood(ML) method for estimating the $\boldsymbol{\theta}$ parameters was introduced by R.A. Fisher in 1912. It can be stated as follows:

Definition 10.1 (Maximum Likelihood Method (ML)) Given a set of observed values $\underline{\boldsymbol{x}} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ from a random sample $\underline{\boldsymbol{X}} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n)$ with p.d.f. $p(\boldsymbol{x};\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter varying in an set Θ, the maximum likelihood

estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is the maximum point (if it exists) of the likelihood function (10.5). Shortly:

$$\max_{\Theta}\left[L(\boldsymbol{\theta};\,\underline{\boldsymbol{x}})\right]=\max_{\Theta}\left[\prod_{i=1}^{n}p(\boldsymbol{x}_i;\boldsymbol{\theta})\right]=L(\hat{\boldsymbol{\theta}};\,\underline{\boldsymbol{x}})\,. \tag{10.7}$$

Alternatively, the principle requires minimizing the logarithmic function $\mathcal{L}$ of Eq. (10.6). In this case, the minimum point of the function is easily obtained by solving the likelihood equations with respect to $\boldsymbol{\theta}$:

$$\frac{\partial\mathcal{L}}{\partial\theta_k}=\sum_{i=1}^{n}\left[\frac{1}{p(\boldsymbol{x}_i;\boldsymbol{\theta})}\,\frac{\partial p(\boldsymbol{x}_i;\boldsymbol{\theta})}{\partial\theta_k}\right]=0\,,\quad(k=1,2,\ldots,p)\,. \tag{10.8}$$

Three points have to be emphasized to clarify the ML method:

- *Before the trial*, the likelihood function $L(\boldsymbol{\theta};\,\underline{\boldsymbol{X}})$ is proportional to the p.d.f. of $(\boldsymbol{X}_1,\boldsymbol{X}_2,\ldots\boldsymbol{X}_n)$. In general, there is proportionality and not coincidence because, in the maximization of the likelihood function, constant factors not affecting the maximum of $\hat{\boldsymbol{\theta}}$ are sometimes omitted;
- *After the trial*, during the likelihood minimization, the occurrences $\underline{\boldsymbol{x}}$ of the variables $\underline{\boldsymbol{X}}$ are used. The quantities $\underline{\boldsymbol{x}}$ are then regarded as fixed;
- *After the trial* the likelihood function (or its negative logarithm) only depends on the values $\boldsymbol{\theta}$, which now are the variables with respect to which to maximize (or minimize). The maximum (or minimum) point $\hat{\boldsymbol{\theta}}$ is the ML parameter estimate obtained from the data.

Now we apply the principle to some simple cases.

Exercise 10.1
In n independent trials, x successes have been obtained. Find the ML estimate of the probability p.

Answer We assign the binomial density (2.29) to the population:

$$b(n,\,p;\,x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\,.$$

Since here we have only one observed value, the likelihood function coincides with the binomial function, which must be maximized with respect to p, keeping fixed the parameters x and n. Since the factorial term does not contain

(continued)

Exercise 10.1 (continued)
p, we can neglect it during the maximization procedure. For simplicity, we minimize the logarithmic likelihood (10.6), which now becomes:

$$\mathcal{L} = -x \ln(p) - (n - x) \ln(1 - p) \, .$$

To find the minimum, we set the derivative to zero:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p} = -\frac{x}{p} + \frac{n - x}{1 - p} = 0 \, ,$$

that gives:

$$\hat{p} = \frac{x}{n} = f \, , \tag{10.9}$$

with the notation of Eq. (10.7). It is also easy to prove, from the second derivative, that this is the absolute minimum of the function.

The result shows that the ML probability estimate *is nothing else than the observed frequency*. At the beginning of the book, we *postulated* the existence of the probability and noted the possible convergence of the frequency to it, as indicated in Eq. (1.3). Alternatively, we could have taken the ML method as a starting point and deduce Eq. (1.3) from this. In fact, Eq. (10.9) is a special case of this principle: all fundamental statistical estimators can be deduced starting from the ML method. In the following we will soon see other examples of this principle. The next two exercises show a likelihood function in the form of a density product.

Exercise 10.2
Two experiments give x_1 successes over n_1 trials and x_2 successes over n_2 trials, respectively. Find the ML estimate of p.

Answer Operating as in the previous exercise, we can write, apart from inessential constant factors, the ML function as the product of two binomials having the same probability p:

$$L = p^{x_1} \, p^{x_2} \, (1 - p)^{n_1 - x_1} \, (1 - p)^{n_2 - x_2} \, .$$

Using logarithms, we obtain:

$$\mathcal{L} = -(x_1 + x_2) \ln(p) - (n_1 - x_1 + n_2 - x_2) \ln(1 - p) \, ,$$

(continued)

Exercise 10.2 (continued)
and hence:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p} = -\frac{x_1 + x_2}{p} + \frac{(n_1 + n_2) - x_1 - x_2}{1 - p} = 0 \implies \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} .$$

The ML estimate is nothing else than the sum of the successes over the sum of the trials.

Exercise 10.3
Given n variates x_i from a one-dimensional Gaussian, find the ML estimate of mean and variance.

Answer Since we are dealing with n measurements, the likelihood function is the product of n Gaussians with the same μ and σ parameters:

$$L(\mu, \sigma) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\, e^{-\frac{1}{2\sigma^2}\sum_i (x_i - \mu)^2} .$$

The logarithmic likelihood (10.6) is then:

$$\mathcal{L}(\mu, \sigma) = \frac{n}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 ,$$

which, setting the derivatives to zero, provides the required estimates:

$$\frac{\partial \mathcal{L}}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \implies \hat{\mu} = \sum_{i=1}^{n}\frac{x_i}{n} \equiv m$$

$$\frac{\partial \mathcal{L}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \implies \hat{\sigma^2} = \sum_{i=1}^{n}\frac{(x_i - m)^2}{n} .$$

In the variance formula, the true mean μ has been replaced with the estimate $\hat{\mu} = m$.

Notice that the ML estimate of the mean coincides with the usual sample mean m, whereas the ML estimate of σ^2 gives the estimator (6.59), which is unbiased only asymptotically.

When the likelihood function is differentiable with respect to θ, the ML method reduces to a differential analysis problem. However, the terms containing θ can be

discrete, or, even if continuous, they could not admit continuous derivatives. It is therefore necessary to resort to finite difference methods or to problem-dependent considerations. The following exercise is an example of this type of issues.

Exercise 10.4
The sampling of a variable from the uniform density (3.79):

$$u(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b \,, \\ 0 & \text{for } x < a \,, x > b \end{cases}$$

gives the n variates:

$$\underline{x} = x_1 < x_2 \cdots < x_{n-1} < x_n \,.$$

Find the ML estimate of a and b.

Answer Since the values have been sorted in increasing order, the condition:

$$a \le x_1 \,, \qquad b \ge x_n$$

must apply. The likelihood function (10.5) becomes:

$$L(a, b; \underline{x}) = \frac{1}{(b-a)^n} \quad a \le x_1 \,, \;\; b \ge x_n \,.$$

In this case, even if we have continuous parameters, the maximum of the function cannot be found by differentiation due to its discontinuities. Since this specific likelihood is maximal when the denominator is minimal, the ML estimate of a and b coincides with the smallest and largest the observed values, respectively:

$$\hat{a} = x_1 \,, \qquad \hat{b} = x_n \,.$$

It is important to note that the remaining part of the observed values:

$$x_2, \; x_3, \; \ldots \; x_{n-1}$$

does not provide any information about the estimate of a and b. Since the extremes x_1 and x_n contain all the information necessary for the estimation, they are said to be a *sufficient statistic* for (a, b).

This concept will be developed in more detail in the next section.

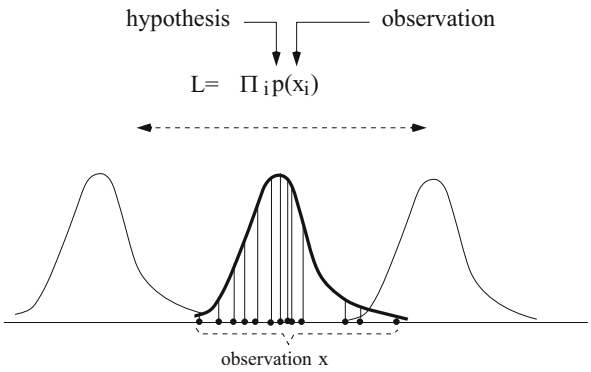


Fig. 10.2 Intuitive justification of the maximum likelihood method: the best density reproducing the data is the one in bold, which maximizes the product of the ordinates

We have shown what the ML method consists in and the results it provides. However, we believe that it is useful to suggest an intuitive argument that explains *why the method works*, that is, why it gives reasonable parameter estimates. Look at Fig. 10.2, where the maximization of the likelihood function is sketched as a shift of the $p(x, \theta)$ density along the abscissa axis (here θ is a location parameter of the distribution). The observed values are on the abscissa axis, represented by bold points; they will concentrate in one or more regions, with some values dispersed in the other areas. The likelihood function is obtained by the *product of the ordinates of the observed data calculated through the density function*. As can be seen from the figure, the maximum likelihood is obtained when the density is "best adjusted" to the data. The ML parameter estimate corresponds to this maximum.

Now is the time to make a little effort of abstraction and study the fundamental results of the estimator theory. After this step, developed in the next two sections, the theory of estimation will appear clearer, and probably even more interesting.

10.3 Estimator Properties

In this section we consider the θ parameter as a scalar, but Eqs. (10.2), (10.3), and (10.4) hold without modifications also in the multidimensional case. To begin with, let us go back to the definition of estimator, which we briefly mentioned in Sect. 2.11.

Definition 10.2 (Point Estimator) Given a sample $\underline{\boldsymbol{X}}$ of size n from an m-dimensional random variable $\boldsymbol{X}$ with density $p(\boldsymbol{X}; \theta)$, with $\theta \in \Theta$, a point estimator (or estimator) of the parameter θ is a statistic (see Definition 2.13):

$$T_n \equiv t_n(\underline{\boldsymbol{X}}) \,, \tag{10.10}$$

with values in Θ.

The estimator is then defined as a function $T : S \to \Theta$ that maps from the sample space S to the parameter space Θ, used to estimate θ. As we know, the mean M and the variance S^2 of a sample are estimators of the parameters μ and σ^2. The conventions we will use for estimators are as follows: T_n (or T) is a random variable from Eq. (10.10), $t_n(\cdot)$ is the associated functional form and t_n (or t) is an occurrence of T_n (or T) after a trial or experiment.

A reasonable estimator should get closer and closer to the true value of the parameter as the number of observations increases. In this respect, a useful property is consistency:

Definition 10.3 (Consistent Estimator) An estimator T_n of the parameter θ is called consistent if it converges in probability towards θ according to Eq. (2.73):

$$\lim_{n\to\infty} P\{|T_n - \theta| < \epsilon\} = 1, \quad \forall\, \epsilon > 0\,. \tag{10.11}$$

As we can see, the consistency of the estimator requires only the convergence in probability. However, for the cases considered in this text, the almost certain convergence of Eq. (2.74) also holds. Using the expected value of an estimator (see Sect. 2.11 for this important concept), from Tchebychev's inequality (3.92), it is easy to demonstrate that a sufficient condition for Eq. (10.11) to hold is:

$$\lim_{n\to\infty} \langle T_n \rangle = \theta\,, \tag{10.12}$$

$$\lim_{n\to\infty} \mathrm{Var}[T_n] = \lim_{n\to\infty} \left(\left\langle T_n^2 \right\rangle - \langle T_n \rangle^2 \right) = 0\,. \tag{10.13}$$

If n remains finite, it is reasonable to expect that also the true mean of the distribution of the estimators T_n coincides with θ. However, this condition is not always satisfied: an example that we have already discussed several times is that of the incorrect sample variance (6.59). Therefore, the following definition is necessary:

Definition 10.4 (Unbiased Estimator) An estimator T_n of a parameter θ is unbiased when:

$$\langle T_n \rangle = \theta, \quad \forall\, n\,. \tag{10.14}$$

Otherwise, the estimator is called biased and one has:

$$\langle T_n \rangle = \theta + b_n\,, \tag{10.15}$$

where b_n is called systematic effect or *bias*. In general the bias depends on θ. When Eq. (10.14) does not hold, but Eq. (10.12) remains valid, then:

$$\lim_{n\to\infty} b_n = 0\,,$$

and the estimator is called asymptotically correct.

The variance (6.59) is then a consistent, biased and asymptotically unbiased estimator.

Besides consistency and unbiasedness, the third important property of an estimator is efficiency.

Definition 10.5 (Most Efficient Estimator) Given two unbiased estimators T_n and Q_n of the same parameter θ, T_n is more efficient than Q_n if the relation:

$$\text{Var}[T_n] < \text{Var}[Q_n], \quad \forall\, \theta \in \Theta \tag{10.16}$$

holds.

Clearly, having to choose between two estimators, all other conditions being equal, the more efficient one is preferable, because it allows us to obtain smaller confidence intervals for θ estimation.

Another important feature related to statistic is sufficiency, introduced by R.A. Fisher in 1925. We report the simplest formulation of this property, which makes use of the likelihood function:

Definition 10.6 (Sufficient Statistic) The statistic S_n is called sufficient if the likelihood function (10.5) can be factorized into the product of two functions h and g such that:

$$L(\theta;\, \underline{\boldsymbol{x}}) = g\left(s_n(\underline{\boldsymbol{x}}), \theta\right)\, h(\underline{\boldsymbol{x}})\,, \tag{10.17}$$

where $h(\underline{\boldsymbol{x}})$ does not depend on θ.

For a multidimensional parameter, the function g is written as:

$$g\left(s_n(\underline{\boldsymbol{x}}), q_n(\underline{\boldsymbol{x}}), \ldots, \boldsymbol{\theta}\right)$$

and one says that the statistics S_n, Q_n, ... are jointly sufficient for $\boldsymbol{\theta}$.

In some texts, Eq. (10.17) is denominated as *factorization theorem*. In fact, the definition of sufficiency is sometimes defined in very general terms, and subsequently Eq. (10.17) is shown to be a necessary and sufficient condition for the validity of this property. In this text we instead adopt the factorization theorem as *a definition* of sufficiency.

In practice, the sufficient statistic contains all the information about the parameter to be estimated. Indeed, when deriving Eq. (10.17) with respect to θ to obtain the maximum likelihood, the function $h(\underline{\boldsymbol{x}})$ plays the role of a simple constant and is therefore irrelevant in the estimation of θ. It is also clear that, if S is a sufficient statistic, a statistic $Q = f(S)$, where f is an invertible function of S, is also sufficient, since the likelihood function can be written under the form:

$$L(\theta;\, \underline{\boldsymbol{x}}) = g(f^{-1}(q),\, \theta)\, h(\underline{\boldsymbol{x}})\,. \tag{10.18}$$

A suitable example of sufficient statistic is given by the extreme values of a sample drawn from the uniform density, as discussed in Exercise 10.4. Instead, an example of jointly sufficient statistics is given by the mean and variance estimators for Gaussian samples:

$$T_n = \sum_i X_i^2 \ , \qquad Q_n = \sum_i X_i \ .$$

Indeed, from Exercise 10.3 one has:

$$L(\mu, \sigma;\ \underline{x}) = \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\, e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2} = \\ = \frac{1}{(\sqrt{2\pi}\,\sigma)^n}\, e^{-\frac{1}{2\sigma^2}\left(\sum_i x_i^2 - 2\mu\sum_i x_i + \mu^2 \sum_i\right)} \ ,$$

from which:

$$L(\mu, \sigma;\ \underline{x}) = \frac{1}{\sigma^n}\, e^{-\frac{1}{2\sigma^2}\left(t_n(\underline{x}) - 2\mu q_n(\underline{x}) + n\mu^2\right)}\, h \equiv g\left(t_n(\underline{x}), q_n(\underline{x}), \mu, \sigma\right) h \ ,$$

with h being a constant. This formal treatment corresponds to the well-known practical fact that to estimate the mean and variance of a sample, it is enough to calculate the mean of the squares and the square of the mean.

10.4 Theorems on Estimators

In this section we have gathered all the important results of the theory of ML estimators and the Cramér-Rao lower bound theorem, also valid for other estimators.

For simplicity, we will consider populations with probability density $p(\boldsymbol{x}; \theta)$ depending on a scalar parameter θ. However, the formulae remain valid even in the case of a vector of parameters, if θ is replaced with the vector $\boldsymbol{\theta}$ and the partial derivative $\partial/\partial\theta$ with $\partial/\partial\theta_k$ with respect to any k-th element of the vector $\boldsymbol{\theta}$. Multi-parameter generalization will only be discussed in cases where this procedure is not entirely obvious.

The first theorem links the ML estimators to the sufficient statistic, showing that the sufficient statistic is the best way to summarize the experimental information about θ.

Theorem 10.1 (Sufficient Statistics) *If $S_n = s_n(\underline{\boldsymbol{X}})$ is a sufficient statistic for θ, the ML estimator $\hat{\theta}$, if it exists, is always a function of S_n.*

Proof From Eq. (10.17) it results that $L(\theta;\ \underline{\boldsymbol{x}})$ has the maximum $\hat{\theta}$ at the point where $g\left(s_n(\underline{\boldsymbol{x}}), \theta\right)$ has its maximum and this point obviously depends on $\underline{\boldsymbol{x}}$ through s_n only. The theorem is easily extended to an set of jointly sufficient statistics. □

Theorem 10.2 (Reparameterizations) *If $\eta \in H$ is a parameter depending on another parameter θ through a one-to-one function $g : \Theta \to H$*

$$\eta = g(\theta) \,, \tag{10.19}$$

the ML estimate of η is given by:

$$\hat{\eta} = g(\hat{\theta}) \,, \tag{10.20}$$

where $\hat{\theta}$ is the ML estimate of θ.

Proof We first prove that η and θ have the same likelihood function. In fact, since g is invertible, one has:

$$L_\theta(\theta;\, \underline{\boldsymbol{x}}) = L_\theta(g^{-1}(\eta);\, \underline{\boldsymbol{x}}) \equiv L_\eta(\eta;\, \underline{\boldsymbol{x}}) \,, \tag{10.21}$$

where we have highlighted the equality between two different functional forms, L_θ and L_η. For instance, if $\eta = \ln\theta$ and θ is the mean of a Gaussian, one has:

$$L_\eta \propto \exp\left[-\frac{1}{2}\frac{(x - e^\eta)^2}{\sigma^2}\right] .$$

We also note that this transformation does not have the complications seen with random variables, where the Jacobian determinants are necessary to change the differentials, because what is transformed here are the *parameters*, not the *variables*.

Since, by definition:

$$L_\theta(\hat{\theta};\, \underline{\boldsymbol{x}}) \geq L_\theta(\theta;\, \underline{\boldsymbol{x}}) \quad \text{for any } \theta \,,$$

and Θ is in one-to-one correspondence with H through g, by applying Eq. (10.21) to both members of the inequality, one has:

$$L_\eta(g(\hat{\theta});\, \underline{\boldsymbol{x}}) \geq L_\eta(\eta;\, \underline{\boldsymbol{x}}) \quad \text{for any } \eta \,,$$

and hence:

$$\hat{\eta} = g(\hat{\theta}) \,.$$

□

This theorem is very useful: for example, all the quantities that are function of the sample mean can be considered as a result of ML estimates.

We will now assume that the so-called regularity properties are satisfied:

- θ belongs to Θ.
- The function $p(x;\theta)$ is a p.d.f. for any $\theta \in \Theta$.
- The set $\{x : \ p(x;\theta) > 0\}$, which is the p.d.f. *support*, does not depend on θ.
- If $\theta_1 \neq \theta_2$, then there exists at least a set $B \in \mathcal{F}$ for which $\int_B p(x;\theta_1)\,\mathrm{d}x \neq \int_B p(x;\theta_2)\,\mathrm{d}x$.
- $p(x;\theta)$ is differentiable for any $\theta \in \Theta$.
- The operations of sum and integration in x and of differentiation by θ can be exchanged.

In order to avoid confusions when studying the likelihood function, it is always necessary to check whether analysis is performed by considering $L(\boldsymbol{\theta};\ \underline{\boldsymbol{X}})$ as a *random function* of $\underline{\boldsymbol{X}}$ with $\boldsymbol{\theta}$ fixed or as a *function* of the parameter $\boldsymbol{\theta}$ for a particular observation $\underline{\boldsymbol{x}}$. Therefore, we recommend to pay attention to the upper or lower case notation to easily understand the following discussions. For example, averages of the type:

$$\left\langle \left(\frac{\partial}{\partial\theta} \ln L \right)^2 \right\rangle ,$$

refer to functions of random variables as $\ln L(\boldsymbol{\theta};\ \underline{\boldsymbol{X}})$ at fixed $\boldsymbol{\theta}$ values, whereas the ML method applies *to the sample of observed values* $\underline{\boldsymbol{x}}$ and consider the likelihood as a function of $\boldsymbol{\theta}$.

The most important theorem on the estimator variance was formulated by the statisticians Cramér and Rao in 1944–45, who established a lower bound for it. This theorem can be easily understood if some fundamental relations are kept in mind.

The first one exploits the fact that, if the regularity conditions are valid, the density function is normalized independently of θ:

$$\int p(\boldsymbol{x};\theta)\ \mathrm{d}\boldsymbol{x} = 1 ,$$

and hence:

$$\int \frac{\partial p(\boldsymbol{x};\theta)}{\partial\theta}\ \mathrm{d}\boldsymbol{x} = \frac{\partial}{\partial\theta} \int p(\boldsymbol{x};\theta)\ \mathrm{d}\boldsymbol{x} = 0 . \tag{10.22}$$

From this relation one can also show that:

$$\left\langle \frac{\partial}{\partial\theta} \ln p(\boldsymbol{X};\theta) \right\rangle = 0 . \tag{10.23}$$

Indeed:

$$\int \frac{\partial p(\boldsymbol{x};\theta)}{\partial\theta}\,\mathrm{d}\boldsymbol{x} = \int \frac{1}{p(\boldsymbol{x};\theta)}\,\frac{\partial p(\boldsymbol{x};\theta)}{\partial\theta}\,p(\boldsymbol{x};\theta)\,\mathrm{d}\boldsymbol{x}$$
$$= \int \frac{\partial \ln p(\boldsymbol{x};\theta)}{\partial\theta}\,p(\boldsymbol{x};\theta)\,\mathrm{d}\boldsymbol{x} \equiv \left\langle \frac{\partial}{\partial\theta}\ln p(\boldsymbol{X};\theta)\right\rangle = 0\,.$$

Obviously, also the following relation is valid:

$$\frac{\partial}{\partial\theta}\int \frac{\partial p(\boldsymbol{x};\theta)}{\partial\theta}\,\mathrm{d}\boldsymbol{x} = \int \frac{\partial^2 p(\boldsymbol{x};\theta)}{\partial\theta^2}\,\mathrm{d}\boldsymbol{x}$$
$$= \int \frac{1}{p(\boldsymbol{x};\theta)}\,\frac{\partial^2 p(\boldsymbol{x};\theta)}{\partial\theta^2}\,p(\boldsymbol{x};\theta)\,\mathrm{d}\boldsymbol{x} = \left\langle \frac{1}{p}\,\frac{\partial^2 p}{\partial\theta^2}\right\rangle = 0\,. \tag{10.24}$$

Equation (10.23) shows that the derivative with respect to θ of the logarithm of a density, the so-called score function, is always a function of $\boldsymbol{X}$ with zero mean. The variance of the score function is known as *Fisher information* and, for $\boldsymbol{\theta}$ present in the p.d.f. of $\boldsymbol{X}$, is usually denoted as $I(\theta)$. Recalling Eq. (2.67), it is given by:

$$\mathrm{Var}\left[\frac{\partial}{\partial\theta}\ln p(\boldsymbol{X};\theta)\right] = \left\langle \left(\frac{\partial}{\partial\theta}\ln p(\boldsymbol{X};\theta) - \left\langle \frac{\partial}{\partial\theta}\ln p(\boldsymbol{X};\theta)\right\rangle\right)^2\right\rangle$$
$$= \left\langle \left(\frac{\partial}{\partial\theta}\ln p(\boldsymbol{X};\theta)\right)^2\right\rangle \equiv I(\theta)\,. \tag{10.25}$$

Notice that the remarkable relation:

$$I(\theta) = \left\langle \left(\frac{\partial}{\partial\theta}\ln p(\boldsymbol{X};\theta)\right)^2\right\rangle = -\left\langle \frac{\partial^2}{\partial\theta^2}\ln p(\boldsymbol{X};\theta)\right\rangle\,, \tag{10.26}$$

is valid, since:

$$\left\langle \frac{\partial^2}{\partial\theta^2}\ln p\right\rangle = \left\langle \frac{\partial}{\partial\theta}\frac{\partial \ln p}{\partial\theta}\right\rangle = \left\langle -\frac{1}{p^2}\left(\frac{\partial p}{\partial\theta}\right)^2 + \frac{1}{p}\frac{\partial^2 p}{\partial\theta^2}\right\rangle = -\left\langle \left(\frac{\partial \ln p}{\partial\theta}\right)^2\right\rangle\,,$$

where Eq. (10.24) has been used.

It should be kept in mind that the fundamental relations (10.22)–(10.26) *are valid for any density* satisfying the regularity conditions summarized above. We recommend to do some exercises and check them for some of the known densities, such as binomial and Gaussian. It is important to notice that these relations also hold for the likelihood function, which, as we know, is the p.d.f. of a sample of size n of one or more random variables (apart from a constant factor). For example,

Fisher information about θ contained in the likelihood is related to that contained in $p(\boldsymbol{x_i}; \theta)$ by the crucial relation, correspondent to Eq. (10.25):

$$\left\langle \left(\frac{\partial}{\partial \theta} \ln L \right)^2 \right\rangle = \left\langle \left(\frac{\partial}{\partial \theta} \sum_i \ln p(\boldsymbol{X}_i; \theta) \right)^2 \right\rangle = n \left\langle \left(\frac{\partial}{\partial \theta} \ln p \right)^2 \right\rangle = n I(\theta) \,, \tag{10.27}$$

where Eqs. (5.81), (10.5), and (10.25) have been used together with the fact that the sample consists of n independent occurrences of $\boldsymbol{X}$.

A last useful equation applies to any (unbiased or biased) estimator T_n of θ:

$$\left\langle T_n \frac{\partial \ln L(\theta;\, \underline{\boldsymbol{X}})}{\partial \theta} \right\rangle = -\left\langle T_n \frac{\partial \mathcal{L}(\theta;\, \underline{\boldsymbol{X}})}{\partial \theta} \right\rangle = \frac{\partial \tau(\theta)}{\partial \theta} \,, \tag{10.28}$$

where $\tau(\theta) = \langle T_n \rangle$ and $\tau(\theta)$ are assumed differentiable in θ. Notice that $\tau(\theta) = \theta$ for an unbiased estimator and that, in this case, the last member of Eq. (10.28) is equal to 1. The previous formula can be easily derived from the equivalence between the density function of the sample and the likelihood function and from the regularity conditions, since:

$$\begin{aligned} \frac{\partial \tau(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} \langle T_n \rangle \\ &= \int T_n \frac{\partial L(\theta;\, \underline{\boldsymbol{x}})}{\partial \theta} \, \mathrm{d}\underline{\boldsymbol{x}} \\ &= \int T_n \frac{1}{L(\theta;\, \underline{\boldsymbol{x}})} \frac{\partial L(\theta;\, \underline{\boldsymbol{x}})}{\partial \theta} L(\theta;\, \underline{\boldsymbol{x}}) \, \mathrm{d}\underline{\boldsymbol{x}} = \int T_n \frac{\partial \ln L(\theta;\, \underline{\boldsymbol{x}})}{\partial \theta} L(\theta;\, \underline{\boldsymbol{x}}) \, \mathrm{d}\underline{\boldsymbol{x}} \\ &= \left\langle T_n \frac{\partial}{\partial \theta} \ln L(\theta;\, \underline{\boldsymbol{X}}) \right\rangle . \end{aligned}$$

We can now prove the

Theorem 10.3 (Cramér-Rao Bound) *Let $p(\boldsymbol{x}; \theta)$ be the p.d.f. of $\boldsymbol{X}$ and let T_n be an estimator of θ with finite variance based on the sample $\underline{\boldsymbol{X}}$. If $\tau(\theta) = \langle T_n \rangle$ is differentiable and the information $I(\theta)$ (10.25) of the density p remains finite for any θ, the estimator variance can never be less than $\tau'(\theta)^2/[nI(\theta)]$:*

$$\mathrm{Var}[T_n] \geq \frac{\tau'(\theta)^2}{n \left\langle \left(\frac{\partial}{\partial \theta} \ln p(\boldsymbol{X}; \theta) \right)^2 \right\rangle} = \frac{\tau'(\theta)^2}{nI(\theta)} \,. \tag{10.29}$$

Proof From Eqs. (10.23), and (10.28) it results:

$$\left\langle T_n \frac{\partial}{\partial \theta} \ln L \right\rangle = \left\langle (T_n - \theta) \frac{\partial}{\partial \theta} \ln L \right\rangle = \tau'(\theta) \,. \tag{10.30}$$

By squaring this expression and applying the Cauchy-Schwarz inequality (4.29), one obtains:

$$\left\langle (T_n - \theta)^2 \right\rangle \left\langle \left(\frac{\partial}{\partial \theta} \ln L \right)^2 \right\rangle \geq \left\langle (T_n - \theta) \left(\frac{\partial}{\partial \theta} \ln L \right) \right\rangle^2 = \tau'(\theta)^2 \,. \tag{10.31}$$

From Eq. (10.27) one then obtains:

$$\left\langle (T_n - \theta)^2 \right\rangle \equiv \mathrm{Var}[T_n] \geq \frac{\tau'(\theta)^2}{nI(\theta)} \,.$$

□

If T_n is an unbiased estimator of θ, the Cramér-Rao lower bound becomes $1/[nI(\theta)]$. The Cramér-Rao theorem allows to evaluate in a precise way the estimator efficiency through the definition of its lower bound. Indeed, the ideally correct estimator, based on Definition 10.5, is the one with minimum variance, that is, the estimator which satisfies the condition:

$$\mathrm{Var}[T_n] = \left\langle \left(\frac{\partial \ln L}{\partial \theta} \right)^2 \right\rangle^{-1} = -\left\langle \frac{\partial^2 \ln L}{\partial \theta^2} \right\rangle^{-1} = \frac{1}{nI(\theta)} \,. \tag{10.32}$$

This estimator, among the unbiased ones, is considered as the most efficient or as the *best estimator*. It is also obvious to define the efficiency of a generic correct estimator as:

$$\varepsilon(T_n) = [\mathrm{Var}[T_n]\, nI(\theta)]^{-1} \,. \tag{10.33}$$

For the best correct estimator, the condition $\varepsilon(T_n) = 1$ holds. An estimator that is not the most efficient is also said to be *inadmissible*.

The variance of the most efficient estimator, that is, *the optimal confidence interval of* θ, is small when the information is large. This explains the word *information* given to I.

For a p-dimensional parameter $\boldsymbol{\theta}$, estimated by $T_n(\underline{\boldsymbol{X}})$ in an unbiased way, the generalization of Eq. (10.32) provides the variance and covariance matrix of T_n. Its (i, j) elements are given by :

$$(nI_{ij})^{-1} = \mathrm{Cov}[T_i, T_j] = \left\langle (T_i - \theta_i)(T_j - \theta_j) \right\rangle = -\left\langle \frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j} \right\rangle^{-1} \,, \tag{10.34}$$

where T_i, $i = 1, \ldots, p$, is the i-th component of T_n.

Exercise 10.5
Find the information on the probability contained in the binomial distribution and that on the mean for the Poisson and Gauss distributions. Comment on the results obtained.

Answer Denoting by $b(x;\, p)$, $p(x;\, \mu)$ and $g(x;\, \mu, \sigma)$ the binomial, Poissonian and Gaussian p.d.f.s, respectively, one easily derives from Eqs. (2.29), (3.14), and (3.28):

$$\begin{aligned}
\ln b(X;\, p) &= \ln n! - \ln(n - X)! - \ln X! + X \ln p + (n - X) \ln(1 - p) \,, \\
\ln p(X;\, \mu) &= X \ln \mu - \ln X! - \mu \,, \\
\ln g(X;\, \mu, \sigma) &= \ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right) - \frac{1}{2} \left(\frac{X - \mu}{\sigma} \right)^2 .
\end{aligned}$$

Notice that these functions are random variables, because they are function of X (in capital letter). By differentiating with respect to the parameters of interest one obtains:

$$\begin{aligned}
\frac{\partial}{\partial p} \ln b(X;\, p) &= \frac{X}{p} - \frac{n - X}{1 - p} = \frac{X - np}{p(1 - p)} \,, \\
\frac{\partial}{\partial \mu} \ln p(X;\, \mu) &= \frac{X}{\mu} - 1 = \frac{X - \mu}{\mu} \,, \\
\frac{\partial}{\partial \mu} \ln g(X;\, \mu, \sigma) &= -\frac{X - \mu}{\sigma} \left(-\frac{1}{\sigma} \right) = \frac{X - \mu}{\sigma^2} \,.
\end{aligned} \tag{10.35}$$

All these derivatives have null mean, according to Eq. (10.23), since the difference $(X - \mu)$ appears in their numerators.

The information can be now calculated by applying Eqs. (10.26)–(10.35), by using the square of the first or the second derivative, whichever is more convenient:

$$\begin{aligned}
I(p) &= \frac{1}{p^2(1 - p)^2} \left\langle (X - np)^2 \right\rangle = \frac{np(1 - p)}{p^2(1 - p)^2} = \frac{n}{p(1 - p)} \,, \\
I(\mu) &= \frac{1}{\mu^2} \left\langle (X - \mu)^2 \right\rangle = \frac{\sigma^2}{\mu^2} = \frac{1}{\mu} = \frac{1}{\sigma^2} \,, \\
I(\mu) &= \frac{1}{\sigma^4} \left\langle (X - \mu)^2 \right\rangle = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2} \,,
\end{aligned} \tag{10.36}$$

(continued)

Exercise 10.5 (continued)
where the property $\sigma^2 = \langle (X-\mu)^2 \rangle$ has been applied to the explicit form of the variances of these densities (see for example Table 3.1).

What considerations do these findings now suggest?

First, the information is proportional to the inverse of the variance: a "narrow" density, with little dispersion around the mean, will have high information, as is intuitive.

Furthermore, if we introduce the frequency estimator:

$$T_1 \equiv F = \frac{X}{n} ,$$

which estimates the probability based on the number of successes X and the sample mean estimator:

$$T_n \equiv M = \sum_i \frac{X_i}{n} ,$$

which evaluated the mean μ from n variates of X, we see that, for these estimators, the Cramér-Rao bound coincides with the statistical uncertainty deduced in Chap. 6:

$$\mathrm{Var}[F] = \frac{1}{I(p)} = \frac{p(1-p)}{n} ,$$

$$\mathrm{Var}[M] = \frac{1}{nI(\mu)} = \frac{\sigma^2}{n} .$$

For large samples, this error is evaluated under the approximations $p \simeq f$ and $\sigma^2 \simeq s^2$, which provide the well-known estimation intervals (6.33) and (6.50).

We deduce that the frequency is the best estimator of the probability that appears as a parameter in the binomial density and that the sample mean is the best estimator of the mean of the Poissonian and Gaussian distributions.

Given all these premises, we can now introduce the pivotal ML theorem, the one that assigns to the method the fundamental role in parameter estimation.

Theorem 10.4 (About the Most Efficient Estimator) *If T_n is an unbiased estimator of $\tau(\theta)$ with minimum variance (i.e. the best estimator), it coincides with the ML estimator, if it exists:*

$$T_n = \tau(\hat{\theta}) .$$

Proof Since the Cramer-Rao lower bound is valid for the best estimator, we can write, from Eq. (10.29):

$$\left\langle [T_n - \tau(\theta)]^2 \right\rangle \frac{nI(\theta)}{[\tau'(\theta)]^2} = 1 \,. \tag{10.37}$$

This relation is satisfied if and only if:

$$\frac{\partial \ln L}{\partial \theta} = \frac{nI(\theta)}{\tau'(\theta)} [T_n - \tau(\theta)] \,. \tag{10.38}$$

In fact, Eq. (10.37) is easily obtained if we square Eq. (10.38), take the expectation value and use Eqs. (10.27).

On the contrary, if Eq. (10.37) holds, then the Cauchy-Schwarz inequality (10.31) becomes a strict equality, so $\partial \ln L/\partial\theta = c\,[T_n - \tau(\theta)]$ for a given constant c. Taking into account also Eq. (10.30) one can write:

$$\frac{nI(\theta)}{\tau'(\theta)} = \frac{1}{\tau'(\theta)} \left\langle \left(\frac{\partial \ln L}{\partial \theta} \right)^2 \right\rangle = c \left\langle [T_n - \tau(\theta)] \frac{1}{\tau'(\theta)} \frac{\partial \ln L}{\partial \theta} \right\rangle = c \,,$$

from which, since $c = nI(\theta)/\tau'(\theta)$, Eq. (10.38) follows.

If now $\boldsymbol{x}$ is fixed and θ is variable, the ML estimate is obtained setting Eq. (10.38) to zero:

$$\frac{\partial \ln L}{\partial \theta} = \frac{nI(\hat{\theta})}{[\tau'(\hat{\theta})]^2} [T_n - \tau(\hat{\theta})] = 0 \,, \qquad \text{from which} \qquad T_n = \tau(\hat{\theta}) \,.$$

□

We now investigate the asymptotic properties of ML estimators, starting from consistency. The rigorous proof of its subsistence can be found in [Cra51, Azz96, Jam08]; here we will present a simple heuristic argument. Equations (10.23) and (10.26) shows that, for the law of large numbers:

$$\frac{1}{n} \frac{\partial \ln L}{\partial \theta} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \ln p(\boldsymbol{x}_i; \theta)}{\partial \theta} \,,$$

converges in probability (and also almost surely) to:

$$\left\langle \frac{\partial \ln p(\boldsymbol{X}; \theta)}{\partial \theta} \right\rangle = 0 \,,$$

and that:

$$\frac{1}{n}\,\frac{\partial^2 \ln L}{\partial \theta^2} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \ln p(\boldsymbol{x}_i;\theta)}{\partial \theta^2}$$

converges in probability (and also almost surely) to:

$$\left\langle \frac{\partial^2 \ln p(\boldsymbol{X};\theta)}{\partial \theta^2} \right\rangle = -I(\theta) < 0\,,$$

if θ is the true parameter value. This shows that, as n increases, the first derivative of the likelihood function calculated at θ tends in probability to zero and that the second derivative is negative in probability. We are therefore led to think that the distance between θ and the maximum point $\hat{\theta}$ of the log-likelihood tends in probability to zero, which corresponds precisely to the consistency of the ML estimator.

Let us now consider a series of experiments, in each of which we obtain a value of $\boldsymbol{x}$ and maximize the likelihood function with respect to θ. The set of the $\hat{\theta}_i$ estimates thus obtained forms a sample *of the random variable* $\hat{\Theta}$. If we perform an infinite series of experiments, we will get the true distribution of $\hat{\Theta}$. Will it be Gaussian? The asymptotic normality theorem ensures that, if the sample size n is large enough, the answer is affirmative. We give only a hint of the proof, avoiding in particular to precisely consider the negligible terms (according to the convergence in probability) in the Taylor series expansions.

Considering, for simplicity, a single-valued random variable X, we develop the derivative of the logarithm of L around the true value θ:

$$\left.\frac{\partial \ln L}{\partial \theta}\right|_{\hat{\theta}} = \left.\frac{\partial \ln L}{\partial \theta}\right|_{\theta} + \left.\frac{\partial^2 \ln L}{\partial \theta^2}\right|_{\theta}(\hat{\theta}-\theta) + \frac{1}{2}\left.\frac{\partial^3 \ln L}{\partial \theta^3}\right|_{\theta}(\hat{\theta}-\theta)^2 + \ldots \qquad (10.39)$$

If n is the sample size, we can write:

$$\begin{aligned}\frac{1}{n}\left.\frac{\partial \ln L}{\partial \theta}\right|_{\hat{\theta}} &= \frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial \ln p(x_i;\theta)}{\partial \theta}\right|_{\theta} + \frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^2 \ln p(x_i;\theta)}{\partial \theta^2}\right|_{\theta}(\hat{\theta}-\theta) \\ &\quad + \frac{1}{2n}\sum_{i=1}^{n}\left.\frac{\partial^3 \ln p(x_i;\theta)}{\partial \theta^3}\right|_{\theta}(\hat{\theta}-\theta)^2 + \ldots \qquad (10.40)\end{aligned}$$

Since the hypothesis of the consistency of the ML estimator is valid, for n large enough, $(\hat{\theta}-\theta)$ will become small, and, if the average values of the derivatives remain bounded (regularity condition), the terms higher than the first order can

be neglected in Eq. (10.40). Moreover, since for $\theta = \hat{\theta}$ the first derivative of $\ln L$ vanishes, Eq. (10.40) becomes:

$$\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial \ln p(x_i;\theta)}{\partial\theta}\right|_{\theta} \simeq -\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^2 \ln p(x_i;\theta)}{\partial\theta^2}\right|_{\theta}(\hat{\theta}-\theta)\ .$$

The previous formula can be rewritten as follows:

$$\frac{\sum_{i=1}^{n}\left.\frac{\partial \ln p(x_i;\theta)}{\partial\theta}\right|_{\theta}}{\sqrt{nI(\theta)}} \Bigg/ \frac{-\frac{1}{n}\sum_{i=1}^{n}\left.\frac{\partial^2 \ln p(x_i;\theta)}{\partial\theta^2}\right|_{\theta}}{I(\theta)} \simeq \sqrt{nI(\theta)}\,(\hat{\theta}-\theta)\ .$$

From the Central Limit Theorem and Eqs. (10.23) and (10.27), the numerator of the first member converges in distribution to a standard Gaussian, while, by the law of large numbers and Eq. (10.26), the denominator almost certainly converges to 1. It can then be shown that this implies the convergence in distribution to a Gaussian for the whole ratio at the first member. Therefore, this also applies to the second member, which we consider identical except for negligible terms. We then conclude that, for large n, we have approximately:

$$(\hat{\theta}-\theta) \sim N\left(0, \frac{1}{nI(\theta)}\right)\ . \tag{10.41}$$

We can finally state the important

Theorem 10.5 (Asymptotic Normality) *If the regularity conditions of Sect. 10.4 hold, the ML estimators are asymptotically normal with an expected value equal to the true value of the parameter and have asymptotic efficiency equal to 1.*

Proof Equation (10.41) shows that the estimator $\hat{\theta}$, for large n, is normally distributed with mean θ and variance given by the Cramér-Rao bound (10.29). □

In practice, the values of n at which the distribution of $\hat{\theta}$ can be approximated with a Gaussian depend both on the sample parent population $p(x;\theta)$ and on the estimator type. For the sample mean, as we already know, the normality is reached quite fast ($n > 10$). For other estimators, such as the sample variance, the convergence to normality is much slower.

All previous arguments show that the ML method provides consistent, asymptotically correct (with a distortion factor $O(1/n)$, as in the Exercise 10.3) and asymptotically normal estimators with variance given by the Cramér-Rao bound. Estimators of this type are called BAN (*Best Asymptotically Normal*).

These properties make maximum likelihood the most used method in statistics for the point estimation of parameters when the density $p(\boldsymbol{x};\boldsymbol{\theta})$ is known a priori.

10.5 Confidence Intervals

The point estimation of the parameters through the maximization of the likelihood also provides the elements to carry out the interval estimations. Indeed:

(a) We know, from Theorem 10.5, that $(\hat{\theta} - \theta)$ is asymptotically normal with null mean and variance $1/[nI(\theta)]$.
(b) During the proof of Theorem 10.5, we have seen that $(\partial \ln L/\partial\theta)$ has zero mean, variance $nI(\theta)$ and nearly normal distribution for large n.

These two methods give practically the same results for the interval estimation. If the estimated information $I(\hat{\theta})$ is used instead of the expected one (which correspond, in elementary statistics, to the plug-in rule $s \simeq \sigma$), usually the confidence intervals already presented in Chap. 6 are found. The distortion of the interval introduced by this approximation is studied in detail in [Jam08] and is of the order of $1/n$. You can go deeper into these aspects by solving Problem 10.8.

Let us now look in detail at a third method for the determination of confidence intervals, which is of fundamental importance in multidimensional cases. We anticipate the result, which is:

(c) The variable $2[\ln L(\hat{\boldsymbol{\theta}}) - \ln L(\boldsymbol{\theta})]$ is asymptotically distributed as $\chi^2(p)$, where p is the size of $\boldsymbol{\theta}$.

For simplicity, we consider an asymptotic approximation of $2[\ln L(\hat{\theta}) - \ln L(\theta)]$ in the one-dimensional case and expand, up to the second order, the negative logarithm of $L(\theta)$ around $\hat{\theta}$, where θ is the true value of the parameter. Neglecting, as before, the higher-order terms according to the convergence in probability, we have:

$$\begin{aligned}\mathcal{L}(\theta) &\simeq \mathcal{L}(\hat{\theta}) + \mathcal{L}'(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2}\,\mathcal{L}''(\hat{\theta})(\theta - \hat{\theta})^2 \\ &\simeq \mathcal{L}(\hat{\theta}) + \frac{n}{2}\,\frac{\mathcal{L}''(\theta)}{n}(\theta - \hat{\theta})^2 \simeq \mathcal{L}(\hat{\theta}) - \frac{nI(\theta)}{2}\,(\theta - \hat{\theta})^2\,.\end{aligned}$$

The error term in the first row, which is $o(\theta - \hat{\theta})^2$, can be neglected, thanks to the consistency of $\hat{\theta}$. This justifies also the exchange of $\hat{\theta}$ with θ in the argument of $\mathcal{L}''$; finally, for the last step, the law of large numbers was used. Therefore, we can write:

$$2[\ln L(\hat{\theta}) - \ln L(\theta)] \simeq nI(\theta)\,(\hat{\theta} - \theta)^2\,. \tag{10.42}$$

We now perform a reparametrization using an invertible function $\eta(\theta)$ such that we have $nI_\eta(\eta) = 1$; the function $\eta(\cdot)$ which fulfills this requirement is any primitive of $\sqrt{nI(\theta)}$. Equation (10.42), reformulated as a function of η becomes:

$$\ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta) \simeq \frac{1}{2}\,(\hat{\eta} - \eta)^2\,. \tag{10.43}$$

From Theorem 10.5 and taking into account the reparametrization, we know that approximately $\hat{\eta} \sim N(\eta, 1)$. Therefore, the confidence intervals for η of half-width equal to one or two standard deviations are respectively $\hat{\eta} \pm 1$ and $\hat{\eta} \pm 2$, with an (approximate) confidence level of 68.3% and 95.4%. Thanks to Eq. (10.43), the extremes of the corresponding confidence intervals are calculated, with respect to η, with the equations:

$$\ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta) = 0.5 \,,$$
$$\ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta) = 2 \,.$$

The same result can also be obtained by noting that $(\hat{\eta} - \eta)^2 \sim \chi^2(1)$, since $(\hat{\eta} - \eta)$ is distributed as a standard Gaussian. Then, if $1 - \alpha = CL$ is the confidence level, the extremes of the corresponding interval will be given by the η values satisfying the equation:

$$2\left(\ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta)\right) = \chi^2_\alpha(1) \,, \tag{10.44}$$

where $\chi^2_\alpha(1)$ is the α quantile of the χ^2 distribution with one degree of freedom. If $\alpha = 0.683$, then $\chi^2_\alpha(1) = 1.00$, which corresponds to solve the equation $[\ln L_\eta(\hat{\eta}) - \ln L_\eta(\eta)] = 0.5$ for η.

We have then found a reparameterization that produces symmetric intervals around $\hat{\eta}$ with Gaussian confidence levels. But now the important point comes: it is not necessary to explicitly perform the reparameterization, *it is enough to know that it exists*. In fact, thanks to Theorem 10.2, we know that *numerically* the likelihoods L_θ and L_η are equal; then one can just use the original likelihood L_θ and find the values θ_1 and θ_2 for which:

$$2\Delta[\ln L] \equiv 2\left(\ln L(\hat{\theta}) - \ln L(\theta)\right) = \chi^2_\alpha(1) \,, \tag{10.45}$$

to obtain the interval estimate for θ:

$$\theta_1 \le \theta \le \theta_2 \,, \qquad CL = 1 - \alpha \,. \tag{10.46}$$

For instance, to determine the confidence intervals with $CL = 68.3\%$ and $CL = 95.4\%$, the extremes which give $2\Delta[\ln L] = 1$ or $2\Delta[\ln L] = 4$ must be found. An application of this method is shown in Problem 10.8.

If the likelihood function is sufficiently regular, it is therefore possible to determine both the confidence intervals and their corresponding confidence levels. They are Gaussian but *do not correspond to Gaussian-like intervals*, since in general they are asymmetric with respect to the point estimate $\hat{\theta}$. The intervals that are more similar to the Gaussian ones are those of the transformed parameter η of Eq. (10.19), whose existence is warranted by Theorem 10.2 and which have been used to determine the variation of the likelihood as a function of the confidence levels. All these remarks are schematized in Fig. 10.3. Moreover, the consistency

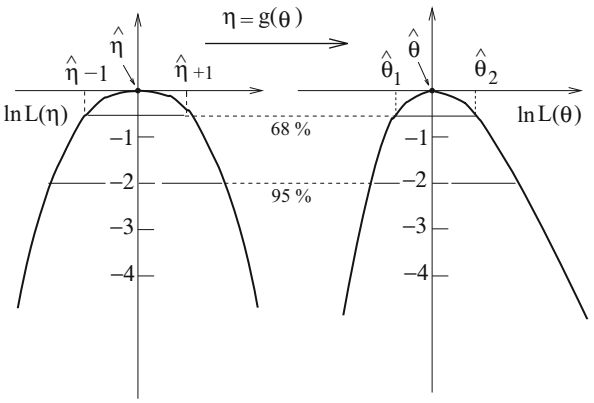


Fig. 10.3 Determination of the confidence intervals for the ML estimation of a one-dimensional parameter. The probability levels indicated in the figure reflect the fact that $-\ln L \simeq \chi^2/2$ and that the $\chi^2(1)$ quantile takes the values 1 and 4 for cumulative probabilities of 68.3% and 95.4%, respectively

of $\hat{\theta}$ ensures that, as n increases, the estimation interval (10.46) also tends to be symmetrical around θ. For a p-dimensional parameter $\boldsymbol{\theta}$, the boundary of the confidence set is found using the condition equivalent to Eq. (10.45):

$$2\Delta[\ln L] \equiv 2\left(\ln L(\hat{\boldsymbol{\theta}}) - \ln L(\boldsymbol{\theta})\right) = \chi^2_\alpha(p)\ , \tag{10.47}$$

where χ^2_α is the quantile of the assigned CL and the asymptotic distribution of $2\Delta[\ln L]$ is $\chi^2(p)$. The values χ^2_α, as a function of the degrees of freedom, can be read in Table E.4. For example, from this table we see that, with two degrees of freedom, the regions enclosed by the contours $2\Delta[\ln L] \simeq 2.4$ and $2\Delta[\ln L] \simeq 6.0$ correspond to $CL \simeq 68\%$ and $CL \simeq 95\%$, respectively. Similarly, we find intervals by solving $2\Delta[\ln L] \simeq 1$ and $2\Delta[\ln L] \simeq 4$ in the one-dimensional case.

Equation (10.47) requires to explore the χ^2 hypersurfaces, and its application is then often difficult. In practice, almost always, the p one-dimensional confidence intervals, each of level α, are determined numerically by varying one parameter θ_i at a time (according to a grid of values) and maximizing the likelihood with respect to the other parameters for each value of θ_i. The interval $(\theta_{i1}, \theta_{i2})$ is obtained by solving, with respect to θ_i, an equation similar to Eq. (10.45):

$$2[\ln L(\hat{\theta}) - \ln L(\theta_i, \hat{\theta}(\theta_i))] = \chi^2_\alpha(1)\ . \tag{10.48}$$

Here $L(\theta_i, \hat{\theta}(\theta_i))$ is the maximized likelihood with respect to all the other components of $\boldsymbol{\theta}$ having fixed θ_i, and $L(\hat{\theta})$ is the best fit value obtained by maximizing all free parameters. This procedure is justified considering the following identity:

$$\ln L(\hat{\theta}) - \ln L(\theta) = [\ln L(\hat{\theta}) - \ln L(\theta_i, \hat{\theta}(\theta_i))] + [\ln L(\theta_i, \hat{\theta}(\theta_i)) - \ln L(\theta)]\ .$$

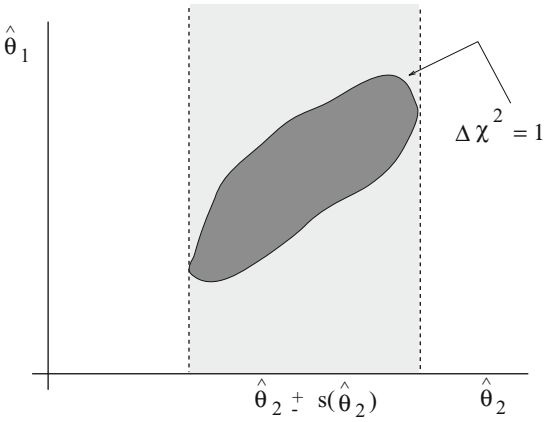


Fig. 10.4 Forms assumed by the confidence regions given by Eqs. (10.47) (darker region) and (10.45) (lighter band). The dark region is the random region which contains the true value of the pair (θ_1, θ_2) with probability 39.3%; the light band, which is the projection of the dark region on the abscissa axis, is the occurrence of the random interval containing with a 68.3% probability the true value of θ_2 (or of θ_1, if the other axis is considered)

The first member has the asymptotic distribution $\chi^2(p)$, while the second addendum of the second member has the asymptotic distribution $\chi^2(p-1)$, being in fact $2\Delta[\ln L]$ when θ_i is known. This suggests, by analogy with the χ^2 Theorem 3.4 of additivity discussed in Appendix C, that the first term on the second member has an asymptotic distribution $\chi^2(1)$. The errors found with Eqs. (10.47) and (10.48) have the meaning shown in Fig. 4.8 and in Fig. 10.4 for the two-parameter case. The outline of the darker region corresponds to Eq. (10.47) for a variation of a χ^2 unit: according to Table E.4, with two degrees of freedom, this contour has a confidence level of almost 40% (the exact value is 39.3%, as seen from Eq. (4.83)). Instead Eq. (10.48) corresponds to the Gaussian confidence level in one dimension for θ_2, shown by the light-coloured region of the figure. The errors usually provided by the minimizing programmes, if the boundaries of the χ^2 regions are not explicitly required, are calculated with Eq. (10.48) and refer to a $CL = 0.68$ for each single parameter, regardless of the others.

10.6 Least Squares Method and Maximum Likelihood

Let us consider the observation of n independent Gaussian variables, coming from n different Gaussian distributions. In this case, the likelihood function is:

$$L(\boldsymbol{\theta};\, \underline{x}) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\sigma_i(\boldsymbol{\theta})} \exp\left(-\frac{1}{2} \frac{(x_i - \mu_i(\boldsymbol{\theta}))^2}{\sigma_i^2(\boldsymbol{\theta})} \right) \right], \tag{10.49}$$

where, in general, the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ depend, in turn, on a multidimensional parameter $\boldsymbol{\theta}$. Note that the previous formula generalizes Eq. (10.4), where the n measurements came from the same population. As a matter of fact, the n populations are now different, although they all have the same Gaussian density. Since, also in

this case, the likelihood represents the probability density of the sample, we can apply the same approach developed up to this point. The ML then allows to evaluate the parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\mu}$ through the maximization of Eq. (10.49) or the minimization of its negative logarithm:

$$\mathcal{L} \equiv -\ln L(\boldsymbol{\theta};\, \underline{x}) = -\sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\sigma_i(\boldsymbol{\theta})}\right) + \frac{1}{2}\sum_{i=1}^{n} \frac{(x_i - \mu_i(\boldsymbol{\theta}))^2}{\sigma_i^2(\boldsymbol{\theta})} \,. \tag{10.50}$$

Since:

$$-\sum_{i=1}^{n} \ln\left(\frac{1}{\sqrt{2\pi}\sigma_i(\boldsymbol{\theta})}\right) = \frac{1}{2}\sum_{i=1}^{n} \ln(2\pi) + \frac{1}{2}\sum_{i=1}^{n} \ln(\sigma_i^2(\boldsymbol{\theta})) \,,$$

and both the first constant term and the 1/2 multiplicative factor are inessential, Eq. (10.50) is equivalent to the minimum search of the function:

$$\mathcal{L} \equiv -\ln L(\boldsymbol{\theta};\, \underline{x}) = \sum_{i=1}^{n} \ln(\sigma_i^2(\boldsymbol{\theta})) + \sum_{i=1}^{n} \frac{(x_i - \mu_i(\boldsymbol{\theta}))^2}{\sigma_i^2(\boldsymbol{\theta})} \,. \tag{10.51}$$

If we assume the standard deviations σ_i to be known (or approximated with the s_i estimated in previous experiments), the first term of the Eq. (10.50) is constant, and the ML method is reduced to the search for the minimum of the χ^2 function:

$$\chi^2 = \sum_{i=1}^{n} \frac{\big(x_i - \mu_i(\boldsymbol{\theta})\big)^2}{s_i^2} \,, \tag{10.52}$$

by setting to zero the derivatives:

$$\frac{\partial \chi^2\big(\boldsymbol{x}, \boldsymbol{\mu}(\hat{\theta}_j)\big)}{\partial \theta_j} = 0 \,, \quad (j = 1, 2, \ldots, k) \,. \tag{10.53}$$

Equation (10.53) represents the least squares (LS) method, which is discussed in detail in Chap. 11. Here, this procedure turns out to be a consequence of the ML approach *when the data come from populations having a Gaussian density of known variance and unknown mean to be determined*.

The important feature of the LS method is to require, for its application, only the knowledge of the expected value and of the variance of the observed variables. Moreover, after the minimization, it is always possible to calculate the final χ^2 value using the parameter best fit values at the minimum $\boldsymbol{\mu}(\hat{\boldsymbol{\theta}})$:

$$\hat{\chi}^2 = \sum_{i=1}^{n} \frac{\big(x_i - \mu_i(\hat{\boldsymbol{\theta}})\big)^2}{s_i^2} \,, \tag{10.54}$$

and it is then also possible to perform the χ^2 test, according to the procedure discussed in Sects. 7.5 and 7.6.

Although the parameters $\hat{\theta}_i$ are estimates and not the true values, under certain assumptions, which will be presented in Chap. 11, when the sample size n is large, the variable (10.54) actually tends to the χ^2 density with $(n-p)$ degrees of freedom, where p is, as usual, the size of $\boldsymbol{\theta}$ [SW89]. Using this value, it is possible to verify whether the functional forms assumed to calculate the true means μ_i are compatible with the data. It is important to note that, while the mathematical part of minimization or maximization can always be performed, hypothesis testing with χ^2 is only meaningful when the involved variables are Gaussian.

One of the most important applications of the LS method is the study of histograms, as shown in the next section.

10.7 Best Fit of Densities to Data and Histograms

Here we resume and complete the topics we have already presented in Sects. 6.14 and 7.5, regarding the estimate of the parent population of the data.

Given a sample of raw data x_i, $i = 1, 2, \ldots, n$, the most direct way to fit a density $p(x; \boldsymbol{\theta})$ to the data sample, when an appropriate code is available, is to minimize the function:

$$\mathcal{L}(\boldsymbol{\theta}) = -2 \sum_{i=1}^{n} \ln \left(p(x_i, \boldsymbol{\theta}) \right) . \tag{10.55}$$

The R routines `optim` and `mle` minimize a user-supplied function `fn` with sophisticated algorithms. Our `FitLike` routine manages the calls to `optim` and gives the output results.

For example, the instruction to fit a set of 1000 simulated Gaussian data with $\mu = 70$ and $\sigma = 10$, contained in a `gauss` vector, to a Gaussian distribution are the following:

```
>gauss<-rnorm{1000,mean=70,sd=10)
>f<-function(par,x){(0.399/par[2])*exp(-0.5*((x-par[1])/par[2])^2)}
>FitLike(x=gauss,parf=c(65,11),fun=f)
```

This code, given the initial conditions contained in `parf`, returns the values $\hat{\mu} = 69.3 \pm 0.2$ and $\hat{\sigma} = 10.1 \pm 0.2$. Errors are evaluated with numerical algorithms based on Eq. (10.48). This method is very efficient in estimating the parameters, but does not allow the user, after the minimization, to easily perform goodness of fit tests. It should therefore be used only when the functional form of the density is certainly defined and when the original raw experimental data are available.

A different procedure is possible with histogrammed data. In this case, the random variable defined in Eq. (10.51) is the number of events I_i in the i-th

histogram bin of width Δ_i (i.e. the occurrence $I_i = n_i$), whereas μ_i is the expected (theoretical) number of events. From Eq. (6.99), μ_i is given by:

$$\mu_i(\boldsymbol{\theta}) = N \int_{\Delta_i} p(x;\boldsymbol{\theta})\,\mathrm{d}x \simeq N p(x_{0i};\boldsymbol{\theta})\Delta_i \equiv N p_i(\boldsymbol{\theta})\ , \tag{10.56}$$

where the assumed density $p_i(x_{0i},\boldsymbol{\theta})$ is calculated in the bin midpoint x_{0i} and N is the sample size. Here the variable x of Eq. (10.56) represents the support (spectrum) of X, that is the histogram abscissa.

The likelihood function is proportional to the multinomial probability (4.89) of having n_i events in the i-th bin over a total of k bins. Neglecting the factors independent of θ, one has:

$$L(\boldsymbol{\theta};\,\underline{n}) = \prod_{i=1}^{k} [p_i(\boldsymbol{\theta})]^{n_i}\ , \tag{10.57}$$

$$\mathcal{L} = -\ln L(\boldsymbol{\theta};\,\underline{n}) = -\sum_{i=1}^{k} n_i \ln[p_i(\boldsymbol{\theta})]\ . \tag{10.58}$$

To find the ML estimate of the p-dimensional parameter $\boldsymbol{\theta}$, we can maximize Eq. (10.57) or minimize Eq. (10.58). Usually minimization codes are used, and logarithmic likelihoods are then minimized.

It is interesting to verify that the minimization of Eq. (10.58) implies the least squares method, which is the most commonly used algorithm in these cases. In fact, we can differentiate Eq. (10.58) with respect to the j-th component of $\boldsymbol{\theta}$, θ_j. After a sign change, one gets:

$$\sum_{i=1}^{k} \frac{n_i}{p_i(\boldsymbol{\theta})} \frac{\partial p_i(\boldsymbol{\theta})}{\partial \theta_j} = \sum_{i=1}^{k} \frac{n_i - N p_i(\boldsymbol{\theta})}{p_i(\boldsymbol{\theta})} \frac{\partial p_i(\boldsymbol{\theta})}{\partial \theta_j}\ .$$

The equality holds because, taking into account the property $\sum_i p_i(\boldsymbol{\theta}) = 1$ and the regularity condition (10.22), the sum of partial derivatives vanishes.

It is easy to realize that the second member of the equality coincides, apart from a multiplicative constant, with the partial derivative of:

$$\chi^2 = \sum_i \frac{(n_i - N p_i(\boldsymbol{\theta}))^2}{N p_i(\boldsymbol{\theta})}\ , \tag{10.59}$$

when *the denominator is regarded as a constant*.

The best fit parameters $\hat{\boldsymbol{\theta}}$ can therefore be found by minimizing Eq. (10.59) with respect to the numerator. It is easily recognized that this procedure is nothing more than an application of the least squares method represented by Eq. (10.54). Since, in this approximation, the χ^2 denominator is kept constant, the modified χ^2 is

sometimes used in the minimization process:

$$\chi^2 = \sum_i \frac{(n_i - N p_i(\boldsymbol{\theta}))^2}{n_i} , \tag{10.60}$$

where the statistical errors have been estimated with the approximation $\sigma_i^2 = N p_i \simeq n_i$. Therefore, to estimate $\boldsymbol{\theta}$ we can use both Eqs. (10.58) and (10.60). As discussed in Sect. 7.4, the two methods give equivalent results, if the sample is large.

However, Eq. (10.58) is the most general, since it is also appropriate even for small samples. Instead Eq. (10.60) is, in principle, valid only when all the bins have at least a few dozen events. In practice, one is often less restrictive, and a sample that has more than five events per channel is considered large enough. If this is not the case, adjacent channels with few events can be grouped into a single one before the χ^2 minimization. As an advantage, unlike Eq. (10.55), once this procedure has been completed, one can proceed to the χ^2 test through Eq. (10.59). The number of degrees of freedom is equal to $(\nu - p)$, where ν is k or $(k - 1)$, depending on whether the sample size N is variable or constant, respectively.

This procedure fully implements the diagram of Fig. 10.1: a population model of density $p(\boldsymbol{x}; \boldsymbol{\theta})$ is devised and from the data the most likely density, given by the best fit parameter $\hat{\boldsymbol{\theta}}$ with its error, is evaluated; then, the χ^2 test is performed to verify the hypothesis on the functional form chosen for the density. Now let us see in detail this procedure by examining some cases already discussed in the sections about basic statistics.

Exercise 10.6

Perform the best fit of the histogram (6.97) obtained with a computer simulation of 1000 variates from a Gaussian of true parameters $\mu = 70$, $\sigma = 10$:

x_i	n_i	x_i	n_i
37.5	1	72.5	207
42.5	4	77.5	153
47.5	16	82.5	101
52.5	44	87.5	42
57.5	81	92.5	7
62.5	152	97.5	6
67.5	186		

Answer Assuming that the histogram comes from a Gaussian of unknown parameters (as would happen for a real, non-simulated data set), we perform

(continued)

Exercise 10.6 (continued)
the best fit of the data to a Gaussian distribution. Equation (10.56) then becomes:

$$N\, p_i(\boldsymbol{\theta}) \equiv N\, p_i(\mu,\sigma) = N\, p_i(x_i;\mu,\sigma)\, \Delta_i$$
$$= 1000 \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right] \cdot 5\,, \qquad (10.61)$$

since $N = 1000$ and $\Delta_i = 5$.

To apply the two different best fit procedures described before, we use our `Gaussfit` routine, which calls both the R minimization function `mle` and our routine `Nlinfit`, that you can find on our web site. The description of the code and of the statistical methods used in the estimation and determination of errors on the parameters can be found in the routine comment lines.

This code, using Eq. (10.60), performs the minimization of a user-defined function, given as χ^2 dependent on p parameters. If requested, the likelihood function $-2\ln L$ of Eq. (10.58) is used. At the end of the minimization, the final χ^2 value is calculated with Eq. (10.59), and the user can proceed to the χ^2 test.

Since to use the function χ^2 (both in the minimization and in the test phase) an event content at least > 5 per bin is required, in the application of Eqs. (10.60) and (10.59), the χ^2 is calculated by grouping the first two channels. For example, Eq. (10.59) becomes ($n_1 = 1, n_2 = 4$):

$$\chi^2 = \frac{(n_1+n_2-Np_1(\mu,\sigma)-Np_2(\mu,\sigma))^2}{Np_1(\mu,\sigma)+Np_2(\mu,\sigma)} + \frac{(n_3-Np_3(\mu,\sigma))^2}{Np_3(\mu,\sigma)} + \dots\,.$$

The results are reported in Table (10.62), which shows the formula used for the minimization, the best fit parameter value with error, the final $\hat{\chi}^2$ value obtained from Eq. (10.59) and the observed significance SL (p-value) of the test, obtained from Eq. (7.33) for the one-tailed test, provided by Table E.3 with $(\nu - k)$ degrees of freedom. Since N is constant and the first two bins have been combined together, $\nu = (13-1-1) = 11$. Moreover, considering that two parameters have been minimized, $k = 2$ and hence $(\nu - k) = 9$.

Equation	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\chi}^2$	SL
(10.58)	70.09 ± 0.31	9.75 ± 0.22	9.18	$\simeq 42\%$
(10.60)	69.95 ± 0.34	9.62 ± 0.25	10.74	$\simeq 29\%$

(10.62)

(continued)

Exercise 10.6 (continued)
This table shows that the two methods used give equivalent results, even if the experimental sample was not particularly large. The estimate with the exact formula (10.58) is associated with a higher p-value, because a more accurate estimate of the parameters results in a better fit.

Now let us go back to an old acquaintance of ours, the ten-coin experiment, that we discussed for the last time in Exercise 7.7.

Exercise 10.7
Find the probability p to obtain head in a single coin tossing through a best fit to the binomial density using the data of Table 2.2 (reported again here for convenience):

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	0	0	5	13	12	25	24	14	6	1	0

Answer In this case, the parent population has binomial density (2.29), and Eq. (10.56) can be written as:

$$N\, p_i(\boldsymbol{\theta}) \equiv N\, p_i(p) = N\, b(x_i;\, 10,\, p)$$
$$= N\, \frac{10!}{x_i!(10 - x_i)!}\, p^{x_i}(1-p)^{10-x_i}\,, \tag{10.63}$$

where $N = 100$ is the number of trials and the Δ_i interval is missing because now the variable is discrete. The probability p is the unknown parameter to be determined through the best fit procedure.

Similarly to the previous exercise, in the χ^2 calculation, the first and the last three bins have to be grouped ($n_1 = 0, n_2 = 1, n_3 = 5$) and ($n_9 = 6, n_{10} = 1, n_{11} = 0$) to have a number of events ≥ 5. The numerators of the χ^2 function then become:

$$[0 + 0 + 5 - Nb(10,\, p;\, 0) - Nb(10,\, p;\, 1) - Nb(10,\, p;\, 2)]^2\,,$$
$$[13 - Nb(10,\, p;\, 3)]^2\,, \ldots\,.$$

Therefore, the histogram bins are 7 and the degrees of freedom are $\nu = 7 - 1 = 6$, since the total number N of trials is fixed. A further degree of freedom

(continued)

Exercise 10.7 (continued)
is lost because the probability is determined by data. The actual number of degrees of freedom is then $(\nu - k) = 5$.

The results obtained with our code `Coinfit` are reported in Table (10.64): as in the previous exercise, we have minimized both the logarithm of the negative likelihood (10.58) and the modified χ^2 (10.60), and have performed the χ^2 test with Eq. (10.59).

Equation	$\hat{p}$	$\hat{\chi}^2$	*SL*
(10.58)	0.521 ± 0.016	3.79	$\simeq 58\%$
(10.60)	0.528 ± 0.014	4.17	$\simeq 53\%$

(10.64)

Again, the two methods provide similar results. However, in the case of small samples as the present one, we recommend, as a general rule, to use Eq. (10.58).

The last time we considered the ten-coin experiment in elementary statistics, Exercise 7.7, we estimated the probability directly from the data, based on the experimental result of 521 heads out of 1000 tosses. From (6.33) we then got:

$$p \in 0.521 \pm \sqrt{\frac{0.521(1-0.521)}{1000}} = 0.521 \pm 0.016 \,.$$

In this case, the optimization of the parameter with Eq. (10.58) has led to the same result obtained from the observed relative frequency of heads since both point estimates coincide with the ML estimate. Moreover, the sample size is large, so that no difference appears between the two different approximation methods.

The experimental data and the best fit curves of the last two exercises are reported in Fig. 10.5.

10.8 Weighted Mean

An important application of the least squares method consists in finding the mean of Gaussian variables all having the same true mean μ but different variances.

This is a common case in laboratory activities, where it is often necessary to combine the results of measurements of the same quantity (same true mean) carried out with different devices (different measurement errors).

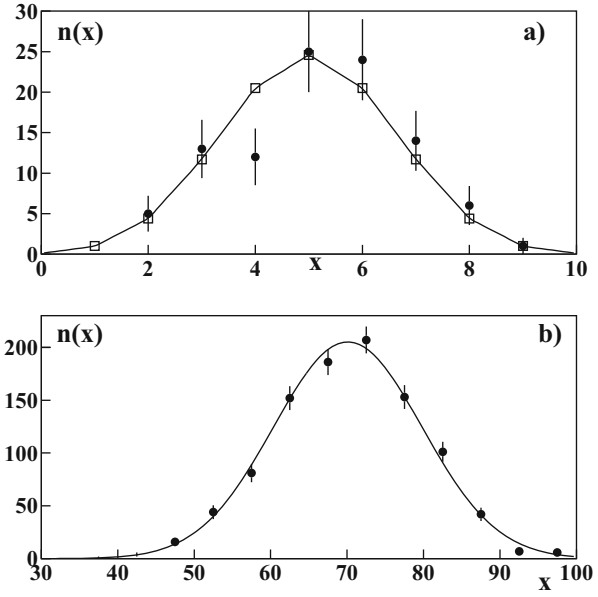


Fig. 10.5 Experimental data with error bars and best fit curves (**a**) for a binomial density (Exercise 10.7) and (**b**) Gaussian (Exercise 10.6). To facilitate the comparison, discrete points (empty squares) of the binomial of Fig. (**a**) have been joined by segments

As in the previous paragraph, we obtain the result starting from the maximum likelihood estimate under the hypothesis of Gaussian observations and then verifying that, more generally, this is derivable as a consequence of the least squares method.

To exemplify the problem, we first propose an interesting question. Suppose you have a sample of n independent observations with the same mean μ and standard deviation σ; now group them into two samples of k and $n-k$ observations, respectively. Since the relation $\mu = (\mu_1 + \mu_2)/2$ holds for true averages, intuitively we are led to think that the average of the n observations should be equivalent to half the sum of the two partial averages, $m = (m_1 + m_2)/2$.

However, this conclusion is wrong. Indeed, as it can be easily verified:

$$\frac{1}{n}\sum_{i=1}^{n} x_i \neq \frac{1}{2}\left[\frac{1}{k}\sum_{i=1}^{k} x_i + \frac{1}{n-k}\sum_{i=k+1}^{n} x_i\right].$$

The explanation of this apparent paradox is subtle and conceptually important: it is wrong to average the two partial means, because, if the two subsamples have a different number of events, *they have different variances* σ^2/k *and* $\sigma^2/(n-k)$. We can then think of "weighting" the two averages in order to assign a greater

importance to those giving a more precise estimate. If we define the inverse of the variance as a weight, then $p_1 = (\sigma^2/k)^{-1}$ and $p_2 = [\sigma^2/(n-k)]^{-1}$, and one has:

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{p_1 + p_2}\left[p_1 \frac{1}{k}\sum_{i=1}^{k} x_i + p_2 \frac{1}{n-k}\sum_{i=k+1}^{n} x_i\right] .$$

This is the right solution, as we will now show in a general way, deriving the *weighted average formula*.

We specialize Eq. (10.50) to the case of n independent Gaussian observations divided into k subgroups of size n_i each ($\sum_{i=1}^{k} n_i = n$), all having mean μ but different variance σ_i^2 for each subgroup. The negative logarithm of the likelihood function is:

$$\mathcal{L} \equiv -\ln L(\mu, \boldsymbol{\sigma};\, \underline{x}) = -\sum_{i=1}^{k} n_i \ln\left(\frac{1}{\sqrt{2\pi}\sigma_i}\right) + \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\frac{(x_{ij}-\mu)^2}{\sigma_i^2} , \tag{10.65}$$

where x_{ij} is the j-th observation in the i-th subgroup. If we assume σ_i as known, the first term of this expression does not depend on any parameter and can be neglected in the minimum search of $\mathcal{L}$, which becomes dependent on μ only. If m_i is the mean of the i-th subgroup, the minimum condition is then given by:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\mu} = \frac{1}{2}\frac{\mathrm{d}\chi^2}{\mathrm{d}\mu} = -\sum_i \sum_j \frac{x_{ij}-\mu}{\sigma_i^2} = -\sum_i \frac{n_i m_i}{\sigma_i^2} + \mu \sum_i \frac{n_i}{\sigma_i^2} = 0 , \tag{10.66}$$

that is:

$$\hat{\mu} \equiv m = \frac{\sum_i (n_i m_i)/\sigma_i^2}{\sum_i n_i/\sigma_i^2} , \tag{10.67}$$

which is the well-known weighted average formula. This formula gives the data "center of mass", by weighting each term by:

$$p_i = \frac{n_i}{\sigma_i^2} . \tag{10.68}$$

If all the data come from the same population, then $\sigma_i = \sigma$, and Eq. (10.67) becomes the usual sample mean formula:

$$\hat{\mu} \equiv m = \frac{(1/\sigma^2)\sum_i n_i m_i}{(1/\sigma^2)\sum_i n_i} = \sum_i \sum_j \frac{x_{ij}}{n} .$$

To determine the statistical error of the weighted average, the transformation law (5.74) for independent variables must be applied to Eq. (10.67), as in the case of the sample mean:

$$\mathrm{Var}\left[\frac{\sum_i n_i M_i/\sigma_i^2}{\sum_i n_i/\sigma_i^2}\right] = \left(\frac{1}{\sum_i n_i/\sigma_i^2}\right)^2 \sum_i \frac{n_i^2\,\mathrm{Var}[M_i]}{\sigma_i^4} \,.$$

Since $\mathrm{Var}[M_i] = \sigma_i^2/n_i$, one obtains:

$$\sigma_{\hat{\mu}}^2 = \left(\frac{1}{\sum_i n_i/\sigma_i^2}\right)^2 \sum_i \frac{n_i}{\sigma_i^2} = \frac{1}{\sum_i n_i/\sigma_i^2} \,. \tag{10.69}$$

The weighted average interval estimation at one standard deviation is then given by:

$$\mu \in \frac{\sum_{i=1}^{k} m_i p_i}{\sum_{i=1}^{k} p_i} \pm \sqrt{\frac{1}{\sum_{i=1}^{k} p_i}} \,, \qquad p_i = \frac{n_i}{\sigma_i^2} \,. \tag{10.70}$$

If $\sigma_i = \sigma$, this equation transforms into Eq. (6.50). The confidence levels to be assigned to the interval are Gaussian (i.e. they follow the 3σ law), because they refer to linearly combined Gaussian variables. Also for non-Gaussian variables, the Central Limit Theorem ensures that normality will be reached for n greater than about ten. In practice, often the variances are unknown and are estimated from data by setting $\sigma_i \simeq s_i$. It is possible to show that, also in this case, the confidence levels to be assigned to the interval (10.70) are Gaussian when n is large. The proof exploits the consistency of s_i's as estimators of the σ_i's, the Central Limit Theorem for the convergence in distribution of M_i, and the independence between observations. A necessary hypothesis for the demonstration is that the weight of each subgroup of observations does not become negligible with respect to the others, which is defined by requiring that n_i/n tends to a constant for $n \to \infty$.

We can now capitalize on our knowledge of estimator theory by asking whether the weighted mean is an efficient estimator. Since it is an ML estimator, on the basis of Theorem 10.4, we deduce that either it is the most efficient estimator or there is no optimal Cramér-Rao estimator for the weighted sums of data. Applying Eq. (10.27) to Eq. (10.66), it is immediate to see that the weighted average is the most efficient estimator, because it satisfies the Cramér-Rao limit of Theorem 10.3:

$$\sum_i n_i I_i(\mu) = \left\langle -\frac{\mathrm{d}^2 \ln L}{\mathrm{d}\mu^2} \right\rangle = \left\langle -\frac{\mathrm{d}}{\mathrm{d}\mu} \sum_i \sum_j \frac{(x_{ij} - \mu)}{\sigma_i^2} \right\rangle = \sum_i p_i \,, \tag{10.71}$$

which is just the inverse of variance of Eq. (10.70) (here $I_i(\mu)$ indicates the Fisher information for μ belonging to the distribution of X_{ij}).

In R, the `weighted.mean(x, w)` routine calculates the weighted average of the data of a `x` vector of weights `ww`, but we have not found an R code for the error calculation. We have then implemented this possibility in our `MeanEst` routine, already described in Sect. 6.9, with the call `MeanEst(x,sigma=sx)`, where `sx` is the vector of the true or estimated standard deviations of `x`. If the vector `sigma` is absent, the non-weighted mean is performed.

Exercise 10.8
Using the computer routine `random`, 20 variates $0 \leq x_i \leq 1$ have been extracted from the uniform density. They are reported in the following table:

0.198	0.530	0.005	0.147
0.898	0.445	0.573	0.943
0.127	0.870	0.859	0.608
0.605	0.729	0.160	0.555
0.202	0.313	0.782	0.112

Compute the mean of the whole sample, the weighted mean of the first 15 and of the last 5 data (the first three columns and the last column of the table), and compare the results.

Answer The mean and the standard deviation of the three samples are given, with obvious notation, by:

$$m_{20} = 0.485\,, \qquad s_{20} = 0.302$$
$$m_{15} = 0.489\,, \qquad s_{15} = 0.298$$
$$m_5 = 0.473\,, \qquad s_5 = 0.347\,.$$

These data, according to Eq. (3.82), are the variates of a uniform distribution with $\mu = 0.5$ and $\sigma = 1/\sqrt{12} = 0.289$.

By applying Eq. (6.50) to the three samples, one gets:

$$m_{20} \in 0.485 \pm \frac{0.302}{\sqrt{20}} = 0.485 \pm 0.068\,,$$
$$m_{15} \in 0.489 \pm \frac{0.298}{\sqrt{15}} = 0.489 \pm 0.077\,,$$
$$m_5 \in 0.473 \pm \frac{0.347}{\sqrt{5}} = 0.47 \pm 0.15\,.$$

(continued)

Exercise 10.8 (continued)
The weighted mean of the two partial averages, which are independent since they come from samples without common data, is given by Eq. (10.70):

$$\mu \in \frac{0.489 \cdot 168.7 + 0.473 \cdot 41.6}{168.7 + 41.6} \pm \frac{1}{\sqrt{168.7 + 41.6}} = 0.486 \pm 0.069 \,,$$

where $p_{15} = 15/s_{15}^2 = 168.7$ and $p_5 = 5/s_5^2 = 41.6$ are the weights. This result can be obtained with the code:

```
> MeanEst(x=c(0.489,0.470),sigma=c(0.077,0.150))
```

As you can see, the result is practically identical to the total mean m_{20}. If we did not know the weighted average formula, we could have applied a different estimator, namely, the one given by the arithmetic mean of the two partial averages with the relative error obtained from Eqs. (5.67) and (5.74):

$$\begin{aligned}\mu \in & \frac{m_{15} + m_5}{2} \pm \frac{1}{2}\sqrt{\frac{s_{15}^2}{15} + \frac{s_5^2}{5}} \\ = & \frac{0.489 + 0.473}{2} \pm 0.5\sqrt{0.077^2 + 0.155^2} = 0.481 \pm 0.086 \,.\end{aligned}$$

Is this estimator acceptable?

If we denote with M_p and M_s the weighted mean and the arithmetic mean, we can verify that these estimators are not biased. In fact, by applying Eq. (10.14), we get:

$$\begin{aligned}\langle M_p \rangle &= \left\langle \frac{\sum_i M_i\, p_i}{\sum_i p_i} \right\rangle = \frac{\sum_i \langle M_i \rangle\, p_i}{\sum_i p_i} = \left\langle \mu \frac{\sum_i p_i}{\sum_i p_i} \right\rangle = \mu \\ \langle M_s \rangle &= \left\langle \frac{\sum_i M_i}{n} \right\rangle = \sum_i \frac{\langle M_i \rangle}{n} = \mu \frac{n}{n} = \mu \,.\end{aligned}$$

The crucial difference is that M_p, being an ML estimator, is the most efficient. In fact, it can be proved that the statistical error of M_s is about 20% greater than that of M_p. The estimator M_s is therefore not acceptable. That is, M_s correctly estimates an interval that contains the true mean (in a frequentist sense), but the width of this interval is greater than that of M_p.

10.9 Test of Hypotheses

In this and in the next paragraphs, we complete the important topic of hypothesis testing, which we introduced for the first time at a somewhat intuitive level in Exercises 3.13–3.17 and in Sect. 7.1, without specifying exactly the alternatives against which the null hypothesis was tested.

After having defined the likelihood function, we are now able to address the topic with greater precision, considering a null hypothesis against an alternative hypothesis and introducing an optimality criterion for the choice between the two hypotheses. However, this subject is much broader, because it is possible to deal with cases in which both the null and the alternative hypotheses are actually sets of hypotheses. Here we will limit ourselves to give the basic ideas, which however will already allow us to define a series of methods that are applicable in a simple and direct way to many concrete cases.

From now on we will assume that the likelihood function $L(\theta; x)$ is not simply proportional but coincides with the considered density function. If the observation consists of only one measurement, then $L(\theta; x) = p(x; \theta)$, where $p(\cdot)$ is the p.d.f. of the random variable X. Usually the hypothesis is tested by checking if the value of an estimator of θ (often a function of a sufficient statistic for θ) belongs to a "critical set". The considered likelihood function considered will then be the probability density of the estimator sample distribution.

Let x be an observation, and consider two hypotheses H_0 (the main one, called null hypothesis) against H_1, which represents a possible alternative to H_0. If their exclusive effect is to have a different value of the parameters in the density function, we can construct, with obvious notation, two likelihood functions: $L(\theta_0; x)$ when $H_0 : \theta = \theta_0$ and $L(\theta_1; x)$ when $H_1 : \theta = \theta_1$. In order not to burden the notation, we will consider here only the one-dimensional case, but all the conclusions we will draw also apply to an observation $\boldsymbol{x}$ and/or a set of parameters $\boldsymbol{\theta}$.

For the testing of hypotheses, the situation is that of Fig. 10.6, and the terminology is that of Tables 10.1 and 10.2, where some new terms appear, besides those already known.

If $p(x|H_0)$ is the density of the estimator corresponding to the null hypothesis, H_0 is accepted if $t_{\alpha/2} \leq x \leq t_{1-\alpha/2}$, or is rejected if x is in the critical region, defined by $x < t_\alpha/2$, $x > t_{1-\alpha/2}$. The area subtended by the critical region is the significance level SL, corresponding to the probability of making a mistake by rejecting the H_0 hypothesis when it is true (type I error). In the case of a one-tailed test, in which the hypothesis H_1 corresponds to a single distribution, to the left (or to the right) of $p(x|H_0)$, the quantities $t_{\alpha/2}$ and $t_{1-\alpha/2}$ are replaced only by one quantile t_α (or $t_{1-\alpha}$).

If, on the other hand, H_0 is wrong and the right hypothesis is given by the density $p(x|H_1)$, the tail area, indicated as β in Fig. 10.6, corresponds to the probability of rejecting the correct hypothesis, because in this case the null hypothesis H_0 is accepted (type II error). The area $\pi = 1 - \beta$ is called *the power of the test* and corresponds to the probability of discarding H_0 when H_1 is true. The density

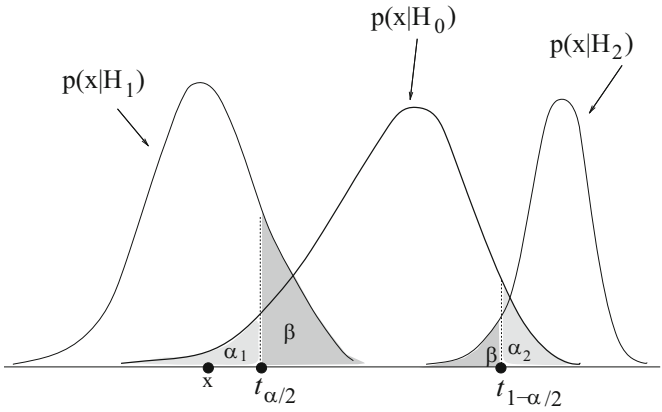


Fig. 10.6 Graphic representation of the quantities involved in the test between two hypotheses. The quantiles t refer to the $p(x|H_0)$ distribution

Table 10.1 The language of statistical tests

Term	Meaning
Null hypothesis H_0	Reference model
Alternative hypothesis H_1	Alternative model
Type I error	To reject H_0 when it is true
Type II error	To accept H_0 when H_1 is true
Significance level SL	Probability of type I error
β area	Probability of type II error
Test level α	A priori fixed value of SL
Critical or rejection region	Interval ($x < t_\alpha/2$ or $x > t_{1-\alpha/2}$) of Fig. 10.6
Power of the test $\pi = 1 - \beta$	Probability to reject H_0 when H_1 is true
More powerful test	For a given α, the test with the highest power
One-tailed test	One tail only, to the left or to the right
Two-tailed test	Two tails, to the left and to the right

$p(x|H_2)$ of Fig. 10.6 indicates the symmetric situation when the maximum of the density relative to the alternative hypothesis H_1 lies to the right of the maximum of the null hypothesis.

We introduced for the first time the power of the test in Sect. 7.7, as a mean fraction of null hypotheses correctly discarded when there are no models for the alternative hypotheses. Here, the power $1 - \beta$ depends on the alternative hypothesis to be examined.

Table 10.2 Testing between two hypotheses: terminology and corresponding probability levels

True hypothesis	Decision	
	H_0	H_1
H_0	Correct decision	Type I error
	$1-\alpha$	α
H_1	Type II error	Correct decision
	β	$1-\beta$

If a significance level α, related to two tail values $t_{\alpha/2}$ and $t_{1-\alpha/2}$ is fixed a priori (test level), the null hypothesis H_0 is accepted when $x \in [t_{\alpha/2}, t_{1-\alpha/2}]$ with probability equal to:

$$P\{t_{\alpha/2} \leq X \leq t_{1-\alpha/2}|H_0\} \equiv P\{X \in A|H_0\} = \int_{t_{\alpha/2}}^{t_{1-\alpha/2}} L(\theta_0; x)\,\mathrm{d}x = 1-\alpha\ , \tag{10.72}$$

where A is the acceptance interval $[t_{\alpha/2}, t_{1-\alpha/2}]$ for the one-dimensional case and a subset of the spectrum of $\boldsymbol{X}$ in the multidimensional case.

The power of the test is the probability to reject H_0 when H_1 is true, that is, the probability to obtain results into the critical region when H_1 is true:

$$1 - P\{X \in A|H_1\} = \int_{-\infty}^{t_{\alpha/2}} L(\theta_1; x)\,\mathrm{d}x + \int_{t_{1-\alpha/2}}^{+\infty} L(\theta_1; x)\,\mathrm{d}x = 1-\beta\ , \tag{10.73}$$

where β is the type II error probability, that is, to accept H_0 when H_1 is true:

$$P\{X \in A|H_1\} = \int_{t_{\alpha/2}}^{t_{1-\alpha/2}} L(\theta_1; x)\,\mathrm{d}x = \beta\ . \tag{10.74}$$

For an ideal test, where the densities corresponding to H_0 and H_1 have disjoint support, $\beta = 0$ and the power $1-\beta = 1$ is maximum.

These are the definitions related to hypothesis testing in the more general case of a two-tailed test. For the one-tailed test the rejection region has the form $(-\infty, c)$ or $(c, +\infty)$.

10.10 One- or Two-Sample Tests

Before examining the optimality criterion for choosing between two hypotheses, let us familiarize ourselves with the concept of power in the case of the tests on the mean already seen in Sect. 7.2. Suppose we have a sample mean M calculated from a Gaussian sample of n events and we want to verify the compatibility with one of two theoretical means μ_0 (hypothesis H_0) and μ_1 (alternative hypothesis H_1) of

two populations having the same variance σ^2. In addition to the test level, we also want to check its power, choosing n to have an assigned value of $1-\beta$. Referring to Fig. 10.6, if we suppose $\mu_1 < \mu_0$, we can write the probabilities:

$$P\left\{\frac{M-\mu_0}{\sigma/\sqrt{n}} \le -|t_\alpha| \,\middle|\, H_0\right\} = P\left\{M \le \mu_0 - |t_\alpha|\frac{\sigma}{\sqrt{n}} \,\middle|\, H_0\right\} = \alpha \,. \tag{10.75}$$

This critical region must have probability of $1-\beta$ under H_1, that is,

$$P\left\{M \le \mu_0 - |t_\alpha|\frac{\sigma}{\sqrt{n}} \,\middle|\, H_1\right\} = 1-\beta \,. \tag{10.76}$$

This equation is valid if and only if:

$$\mu_0 - |t_\alpha|\frac{\sigma}{\sqrt{n}} = \mu_1 + t_{1-\beta}\frac{\sigma}{\sqrt{n}} = \mu_1 + |t_\beta|\frac{\sigma}{\sqrt{n}} \,, \tag{10.77}$$

so that the required minimum n is given by:

$$n = \left[\sigma\frac{|t_\alpha| + |t_\beta|}{|\mu_1 - \mu_0|}\right]^2 \,. \tag{10.78}$$

If $\mu_1 > \mu_0$, Eqs. (10.75) and (10.76) become:

$$P\left\{\frac{M-\mu_0}{\sigma/\sqrt{n}} \ge |t_\alpha| \,\middle|\, H_0\right\} = P\left\{M \ge \mu_0 + |t_\alpha|\frac{\sigma}{\sqrt{n}} \,\middle|\, H_0\right\} = \alpha \,, \tag{10.79}$$

$$P\left\{M \ge \mu_0 + |t_\alpha|\frac{\sigma}{\sqrt{n}} \,\middle|\, H_1\right\} = 1-\beta \,, \tag{10.80}$$

giving the condition:

$$\mu_0 + |t_\alpha|\frac{\sigma}{\sqrt{n}} = \mu_1 - |t_\beta|\frac{\sigma}{\sqrt{n}} \,, \tag{10.81}$$

which again leads to Eq. (10.78).

In the case of a two-tailed test, one proceeds as above, taking both cases $\mu_1 < \mu_0$ and $\mu_0 < \mu_1$ into consideration, but with the quantile $\alpha/2$ instead of α. One then gets Eq. (10.78) again, with $t_{\alpha/2}$ replacing t_α.

If one substitutes σ with the sample standard deviation s, the Student quantiles must be used. Since these quantiles depend on n, a closed-form solution as in the Gaussian case is no longer possible, and one must iterate over n until the solution is reached. This calculation is performed by the R routine `power.t.test`, which requires the difference $\mu_1 - \mu_2$ as input, an assumption on the value of σ and on

those of α, $1-\beta$ and n. Giving four of these five values as inputs, the routine calculates the missing one. An example of use is given by the following exercise.

Exercise 10.9
It is assumed as a null hypothesis H_0 that a variable X is Gaussian with mean $\mu_0 = 10$ and standard deviation $\sigma = 10$. Find the optimal sample size n and the critical region to accept H_0 with $\alpha = 0.05$ and power $1-\beta = 0.95$ against the alternative hypothesis H_1 of a Gaussian with the same $\sigma = 10$ and mean $\mu_1 = 20$. Consider also the test with $\sigma \simeq s$.

Answer We are in the case of the one-tailed test; hence we use the Gaussian quantile $|t_\alpha| = |t_{0.95}|$. We then obtain, from Eq. (10.78):

$$n = \left[\frac{10 \cdot (1.645 + 1.645)}{20 - 10}\right]^2 = 10.8 \,.$$

The critical value m is obtained from Eq. (10.81):

$$m = 10 + 1.645\frac{10}{\sqrt{10.8}} = 20 - 1.645\frac{10}{\sqrt{10.8}} \simeq 15 \,.$$

The required test must then sample $n = 11$ values of X, calculate the sample mean M, accept H_0 if $\{M \leq 15\}$ and accept H_1 if $\{M > 15\}$. If we assume that $\sigma \simeq s = 10$, we can use R with the call:

```
power.t.test(delta=10,sd=10,sig.level=0.05,power=0.95,alt='one',
             type='one') ,
```

to obtain, as a result, $n = 12.3$. Therefore:

$$m = 10 + 1.78\frac{10}{\sqrt{12.3}} = 15.1 \,,$$

where 1.78 is the Student quantile $t_{0.95}$ with 12 degrees of freedom. The call `delta`$= \mu_1 - \mu_0$, `sig.level`$= \alpha$, `alt` refers to one-tailed test and `type` to the single sample case. In summary, we obtain $n \simeq 12$, $m = 15$.

In the two-sample problem, the question is whether they come from populations with the same mean. In this case, it is necessary to find the smallest common dimension n of these two samples in order to distinguish between the two means, with the predefined test levels. To solve this problem it is enough to replace M by $M_1 - M_0$, μ_0 by zero, and μ_1 by $\mu_1 - \mu_0$ in Eqs. (10.75)–(10.78). Moreover, the uncertainty on the difference between the two means $\sigma\sqrt{1/n + 1/n} = \sqrt{2}\sigma/\sqrt{n}$ must substitute the error $\sigma/\sqrt{n}$ in the denominator. The result $2n$ is immediately obtained for the number of events, where n is from Eq. (10.78). In the two-sample

case, the rule is therefore to double the result of the one sample case. If $\sigma \simeq s$, the Student's density calculation gives a value slightly smaller than the one-sample doubled value. If we reconsider Exercise 10.9, in the Gaussian case, we obtain $n = 21.6$ and $n = 15$, whereas with the Student's density with the call to `power.t.test` with `alt='two'`, we obtain $n = 22.4$ and $m = 15.1$ for the Student quantile $t_{0.95} = 1.71$ with 23 degrees of freedom.

The same type of test is often used with frequencies. Using the Gaussian approximation, if p_0 and p_1 are the true probabilities under H_0 and H_1, respectively, and a frequency f is measured, from Eq. (3.6) and when $p_1 < p_0$, Eq. (10.77) becomes:

$$p_0 - |t_\alpha| \sqrt{\frac{p_0(1-p_0)}{n}} = p_1 + |t_\beta| \sqrt{\frac{p_1(1-p_1)}{n}} \,, \tag{10.82}$$

whereas, when $p_1 > p_0$, it is necessary to exchange the signs on both sides of the equation.

Since we are using the Gaussian approximation, these formulae are typically used for $n > 5$. The minimum n value is obtained by repeating exactly the procedure that led to Eq. (10.78):

$$n = \left[\frac{|t_\alpha| \sqrt{p_0(1-p_0)} + |t_\beta| \sqrt{p_1(1-p_1)}}{|p_1 - p_0|} \right]^2 . \tag{10.83}$$

In the two-sample case, the procedure for the left-tailed test is based on the distribution of the difference between the measured frequencies $f_1 - f_0$, similar to that just seen for the two Gaussian samples, except that now the variance of $f_1 - f_0$ depends on the values assumed for p_1 and p_0. Under H_0, when $p_1 = p_0 = p$, the variance of $f_1 - f_0$ is $2p(1-p)/n$ and usually $p = (p_1 + p_0)/2$; under H_1, when $p_1 < p_0$, the variance of $f_1 - f_0$ is $[p_1(1-p_1) + p_0(1-p_0)]/n$. Under the Gaussian approximation, we use Eq. (10.82), replace the standard deviations at the first and second member with those just calculated and substitute the first p_0 at the first member with zero and the first p_1 at the second member with $(p_1 - p_0)$. In the end, we get the following result:

$$n = \left[\frac{|t_\alpha| \sqrt{2p(1-p)} + |t_\beta| \sqrt{p_0(1-p_0) + p_1(1-p_1)}}{|p_1 - p_0|} \right]^2 , \tag{10.84}$$

also valid for the one-tailed right test. We recall again that for the two-tailed test, the quantile t_α must be replaced with $t_{\alpha/2}$ in Eqs. (10.83) and (10.84).

In R, the test on two binomial samples (proportions) is done by the `power.prop.test` routine, which takes as input the values p_0, p_1, α, $1 - \beta$, n and gives in output the value that is not entered as input. This routine does not consider the one-sample case of Eq. (10.83).

Finally, we recall that the R library `pwr` includes many routines that calculate various one- and two-sample cases for the currently used statistical variables, also including the cases of samples with different size.

Exercise 10.10
A null hypothesis H_0 with $p_0 = 0.01$ is assumed. Find the optimal sample size and the critical region to accept H_0 with $\alpha = 0.05$ and power $1-\beta = 0.80$ against the alternative hypothesis H_1 that $p_1 = 0.02$. Carry out the one and two-sample test.

Answer Using Eq. (10.83) for the one-tailed test, one obtains:

$$n = \left[\frac{1.645\sqrt{0.099} + 0.842\sqrt{0.140}}{0.01}\right]^2 = 793 \,.$$

From Eq. (10.82) the limit of the critical region is given by $f = 0.01 + 1.645\sqrt{0.01 \cdot 0.99/793} = 0.02 - 0.842\sqrt{0.02 \cdot 0.98/793} = 0.0158$, corresponding to a number of successes $x = 0.0158 \cdot 793 = 12.5 \simeq 13$. If $x \leq 13$ H_0 is accepted; otherwise H_1 is chosen.

For the two-sample case, from Eq. (10.84), one obtains $n = 1845$. The call `power.prop.test(p1=0.01,p2=0.02,sig=0.05,power=0.8,alt='one')`, provides the value $n = 1826$, corresponding to $f = 0.0138$ and $x = 25.2 \simeq 25$. The R routine contains continuity corrections that are absent in Eq. (10.84).

10.11 Most Powerful Tests

Suppose you are testing a filter for the identification of email spams. This filter analyses mails and assigns a score. After the analysis of correct emails and spams performed by the filter, two histograms of the scores similar to distributions of the type shown in Fig. 10.6 are obtained, where on the x axis, there are the evaluated scores and on the y axis the number of emails with that given score. If the score is higher for spams, we will have two distributions like $p(x|H_0)$ for good mails and $p(x|H_2)$ for spams. If the two distributions are disjoint, selecting good emails would be trivial. If instead the two distributions partially overlap, as in Fig. 10.6, we should devise a selection criterion, that is to say, to evaluate the score above which an email is discarded. The first parameter to take into consideration is the type I error α_2: in this case we have to fix it as small as possible, to avoid discarding good mails. Then, we will set an upper score limit, allowing for some spams to enter the system. At this point the β error determines the percentage of spams accepted and the power

$1-\beta$ the percentage of those correctly rejected. In other types of problems, as in production quality controls, where it is more important to avoid a "dirty" sample than to miss good events, a larger value of α must be chosen to decrease that of β.

This example clearly shows the standard procedure used to perform statistical tests. At first, the level of significance α appropriate for the problem is fixed a priori; then an estimator $T_n = t_n(\underline{X})$ is chosen (the mean, the variance, the chi-square, etc.), the sample distribution of the estimator is found, and its quantile values t_α or $t_{1-\alpha}$ are determined. If α remains fixed and the estimator is changed, the power of the test associated to the alternative hypothesis H_1 *turns out to depend on the chosen estimator*. Now let us ask ourselves the fundamental question: given an observation of a variable X, which is the most powerful test, that is, the one that has the *best critical region*?

The question may in some cases have considerable practical interest, because statistical tests often have a high economic and/or management cost (as an example, think of quality control tests in industry). It is therefore important to choose the most powerful test, for which the type II error is minimal and, consequently, the decision criterion is the most reliable. An answer to this issue is given by the Neyman-Pearson(NP) theorem:

Theorem 10.6 (Neyman-Pearson) *Let θ be a parameter of a likelihood function and consider the null hypothesis H_0 and the alternative one H_1:*

$$H_0 : \theta = \theta_0 \,, \qquad H_1 : \theta = \theta_1 \,.$$

The likelihood ratio:

$$R(X) = \frac{L(\theta_0;\, X)}{L(\theta_1;\, X)} \,, \tag{10.85}$$

takes large values if H_0 is true and small values if H_1 is true. The most powerful test among all those of level α is given by:

$$\textit{reject} \quad H_0 \quad \textit{if} \quad \left\{ R(X) = \frac{L(\theta_0;\, X)}{L(\theta_1;\, X)} \le r_\alpha \right\} \,, \tag{10.86}$$

where r_α is the $R(X)$ value of Fig. 10.7, such that $P\{R(X) \le r_\alpha | H_0\} = \alpha$.

Proof Let Ω be the H_0 rejection region (of α level) for the NP test and Ω' be the rejection region (of α' level) for any other test. The theorem holds if it is proved that the probability of the type II error for the same α level is minimal for the NP test:

$$\alpha = \alpha' \quad \Longrightarrow \quad \beta' > \beta \,. \tag{10.87}$$

By assumption, one has:

$$\alpha = \int_\Omega L(\theta_0;\, x)\, \mathrm{d}x = \alpha' = \int_{\Omega'} L(\theta_0;\, x)\, \mathrm{d}x \,.$$

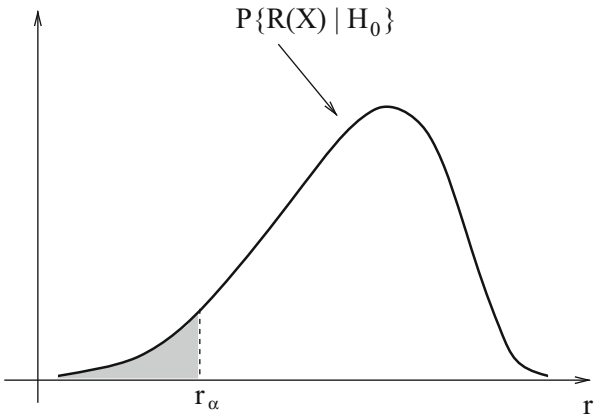


Fig. 10.7 Using the R likelihood ratio, the null hypothesis is rejected if the value of the ratio is lower than the limit r_α, which determines the significance level given by the shaded area

If $\beta' \neq \beta$ one can write (see Fig. 10.6):

$$\begin{aligned}\beta' - \beta \equiv \Delta\beta &= 1 - \int_{\Omega'} L(\theta_1; x)\,\mathrm{d}x - \left[1 - \int_{\Omega} L(\theta_1; x)\,\mathrm{d}x\right] \\ &= \int_{\Omega} L(\theta_1; x)\,\mathrm{d}x - \int_{\Omega'} L(\theta_1; x)\,\mathrm{d}x\ . \end{aligned} \tag{10.88}$$

If $x \in \Omega$, Eq. (10.86) is always true:

$$\frac{L(\theta_0; x)}{L(\theta_1; x)} \leq r_\alpha \implies L(\theta_1; x) \geq \frac{1}{r_\alpha} L(\theta_0; x)\ ,$$

whereas, if $x \in (\Omega' - \Omega)$:

$$\frac{L(\theta_0, x)}{L(\theta_1; x)} > r_\alpha \implies L(\theta_1; x) < \frac{1}{r_\alpha} L(\theta_0; x)\ , \quad x \notin \Omega\ .$$

By replacing these last two relations in Eq. (10.88) and taking into account that the integrations on $\Omega \cap \Omega'$ do not contribute, Eq. (10.87) is obtained:

$$\Delta\beta > \frac{1}{r_\alpha}\left[\int_{\Omega} L(\theta_0; x)\,\mathrm{d}x - \int_{\Omega'} L(\theta_0; x)\,\mathrm{d}x\right] = \frac{1}{r_\alpha}[\alpha - \alpha'] = 0\ .$$

This proves the theorem. □

We now apply this theorem to test the hypothesis of two different means $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$ with samples extracted from Gaussian populations of equal variance σ^2. Under these conditions, the ratio (10.85) becomes:

$$\frac{L_0}{L_1} = \exp\left\{\frac{1}{2\sigma^2}\Big[\sum (x_i - \mu_1)^2 - \sum (x_i - \mu_0)^2\Big]\right\} \le r_\alpha \ ,$$

and hence, using logarithms:

$$2(\mu_0 - \mu_1)\sum_i x_i \le 2\sigma^2 \ln r_\alpha + n(\mu_0^2 - \mu_1^2) \ . \tag{10.89}$$

Dividing by $2n(\mu_0 - \mu_1)$, if $(\mu_0 - \mu_1) > 0$ one obtains:

$$m_n = \frac{1}{n}\sum_i x_i \le \frac{\sigma^2}{n(\mu_0 - \mu_1)} \ln r_\alpha + \frac{\mu_0 + \mu_1}{2} = m_0 \ . \tag{10.90}$$

In this case the Neyman-Pearson test with the explicit calculation of r_α is equivalent to the test on the sample mean: H_0 is discarded if $\{M_n < m_0\}$. The value m_0 is determined by the type I error under H_0:

$$\alpha = P\{M_n < m_0\} = P\left\{\frac{M_n - \mu_0}{\sigma/\sqrt{n}} < \frac{m_0 - \mu_0}{\sigma/\sqrt{n}}\right\} \ .$$

This condition holds when m_0 satisfies the equality:

$$\frac{m_0 - \mu_0}{\sigma/\sqrt{n}} = t_\alpha \quad \Longleftrightarrow \quad m_0 = \mu_0 + \frac{\sigma}{\sqrt{n}}\, t_\alpha \ ,$$

where t_α is the α quantile of the standard Gaussian. If $\mu_0 < \mu_1$, it is necessary to change the sign of the inequality (10.89), and H_0 is discarded if $\{M_n > m_0 = \mu_0 + (\sigma/\sqrt{n})\, t_{1-\alpha}\}$.

It is also easy to verify that the test on the sample mean is the most powerful even for exponential and Poisson distributions.

10.12 Test Functions

Neyman-Pearson's test requires the knowledge of the density $p(r)$ of the likelihood ratio R. In principle, it is always possible to get this by using the techniques developed in Chap. 5 for functions of random variables, but sometimes these calculations turn out to be difficult and laborious.

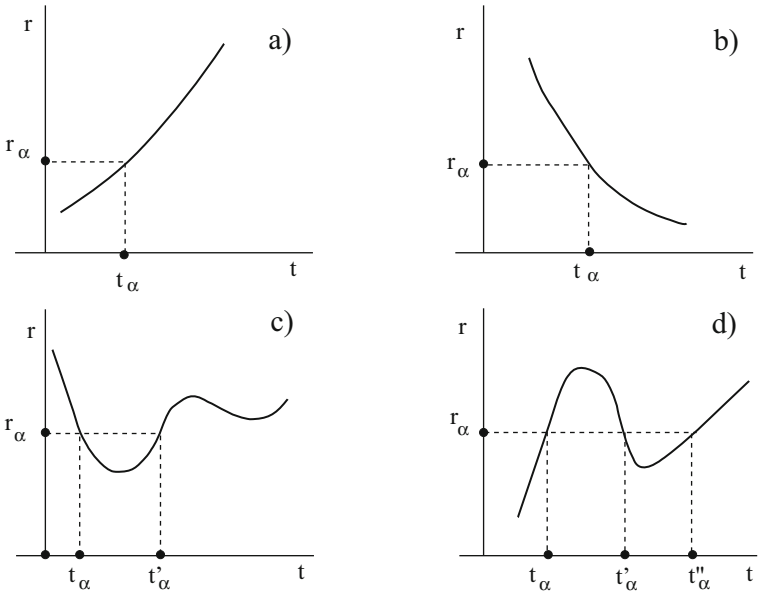


Fig. 10.8 When the likelihood ratio can be expressed as a function of another statistic, the interval $R \leq r_\alpha$ relative to R can correspond to an interval $T \leq t_\alpha$ (**a**), $T \geq t_\alpha$ (**b**), $t_\alpha \leq T \leq t'_\alpha$ (**c**), $T \leq t_\alpha$, $t'_\alpha \leq T \leq t''_\alpha$ (**d**), depending on the function $R = \psi(T)$

Fortunately, as we just saw in Eq. (10.90), the problem is simplified drastically if R can be written under the form:

$$R = \psi(T) \,, \tag{10.91}$$

where T is a statistic with a known distribution.

In fact, a test using the T operator determines the limits t_α corresponding to the chosen significance level; using Eq. (10.91), one can therefore determine the limit r_α and apply Theorem 10.6. However, this does not necessarily have to be done in practice: if Eq. (10.91) holds, the $R(X)$ *test can be replaced by a* $T(X)$ *test*, which also satisfies the maximum power property. The functional relation between R and T given by Eq. (10.91) can turn the one-tailed test $R(x) \leq r_\alpha$ (see Eq. (10.86)) into a one-tailed test to the right or the left, a two-tailed test, or a test with a more complicated critical region, as shown in Fig. 10.8. For example, from Eq. (10.90) the quantity R is a function of the sample mean. It follows that the mean estimator $T = \sum X_i/n$ *is the most powerful test function of the mean* for samples extracted from Gaussian populations. This function maximizes $1 - \beta$ once the α test level has been chosen.

Exercise 10.11
A company wins a tender for the supply of an electronic component stating that the percentage of defective parts at the origin is less than 1%. Find the limit below which the number of defective pieces found in a control batch of 1000 pieces must remain to be confident, at a level $\alpha = 5\%$, that the firm's statement is correct. Also find the power of this test with respect to the alternative hypotheses of a defect probability of 2% and 3%, and verify if the test has maximum power.

Answer If the defect rate is 1%, the number of discards on a batch of 1000 pieces is a binomial variable with an expected value (true average) equal to 10. Since the probability is small (0.01) and the mean value is large, it is possible approximate the binomial using a Poisson distribution. We are brought back to the case of Exercise 10.10, with the difference that here n is fixed and β unknown. Since the problem requires a significance level of 5%, from Table E.1 and from Eqs. (3.43) and (3.44), we see that the one-tailed test:

$$P\{T \geq t_{1-\alpha}\} = 1 - \Phi(t_{1-\alpha}) = 0.05\,,$$

is satisfied for a value of the standard variable $T = t_{1-\alpha} \simeq 1.645$. If S is the number of discards, from Eq. (3.37) we have:

$$\frac{S-\mu}{\sigma} = \frac{S-10}{\sqrt{10}} \geq 1.645\,, \quad \text{and hence:} \quad S \geq 15.2\,.$$

Since the considered variable is discrete, there is no critical value that coincides with the assigned SL value. At this point we could randomize the test by setting in Eq. (7.4) $\alpha_L = 0.05$ and:

$$SL_2 = P\{S \geq 15\} = 1 - \Phi\left(\frac{15-10}{\sqrt{10}}\right) = 0.057\,,$$
$$SL_1 = P\{S \geq 16\} = 1 - \Phi\left(\frac{16-10}{\sqrt{10}}\right) = 0.029\,,$$

where Table E.1 has been used. As shown in Problem 10.15, the randomized test accepts the batch one time over four when $S = 16$. Here we prefer to proceed in a simpler (even if approximate) way, adopting as a decision rule the acceptance of the batch if $S \leq 15$ and its rejection if more than 15 defective pieces are found.

The power of the test is given by the probability $1 - \beta$ to discard H_0 (1% defect rate) when the alternative hypothesis H_1 is true. We recall again that

(continued)

Exercise 10.11 (continued)
β is the area of Fig. 10.6, which means the probability of the type II error. If H_1 refers to a defect rate of 2%, from Eq. (3.43), and using the R statistic, we obtain:

$$1-\beta = 1 - P\{S < 16; H_1\} = P\{S \geq 16; H_1\} \tag{10.92}$$
$$= 1 - \Phi\left(\frac{16-20}{\sqrt{20}}\right) = \texttt{1 - pnorm((16 - 20)/sqrt(20))} \simeq 0.81 \,.$$

The power of the test is about 80%, and the probability to accept H_0 when H_1 is true is about 20%.

The same calculation when H_1 refers to a 3% defect rate, gives:

$$1-\beta = 1 - \Phi\left(\frac{16-30}{\sqrt{30}}\right) \simeq 0.995 \,. \tag{10.93}$$

Therefore, the hypotheses H_0 : 1% and H_1 : 3% give distributions having basically a disjoint support. In practice, the type II error occurs with a nonnegligible probability only for the alternative hypothesis of a defect rate of about 2%.

Due to Eq. (10.90), this is the most powerful test. We also verify the results by considering the Poisson distribution (3.14) to be the likelihood function approximating the binomial distribution. When the alternative hypothesis H_1 assumes a defect rate of 2%, we write the likelihood ratio (10.85) as:

$$R(S) = \frac{10^S\, e^{-10}}{20^S\, e^{-20}} = e^{10}\left(\frac{1}{2}\right)^S \equiv \psi(S) \,, \tag{10.94}$$

which shows that Eq. (10.91) is valid. Since the exponential factor is constant, the link between R and S decreases monotonically, as in Fig. 10.8b. The inverse function $S = \psi^{-1}(R)$ is:

$$S = \frac{10 - \ln R}{\ln 2} \,. \tag{10.95}$$

We then determine an interval $S \geq 15.5$ (intermediate value between the discrete limit values found on S) and an interval on R approximately equal to

$$R \simeq r_\alpha \leq e^{10}(0.5)^{15.5} = 0.47 \,,$$

to reject the null hypothesis at the chosen test level.

The test has then the maximum power among all possible tests at the 5% level.

Up to now we have considered alternative hypotheses of the type $H_1 : \theta = \theta_1$, called *punctual or simple*. Let us now consider the case of alternative hypotheses called *composite*, which include a set of values, such as $H_1 : \theta > \theta_1$. In these cases we need to explore the power of the test for the whole set of parameter values of the alternative hypothesis, in order to find the *uniformly most powerful test*. In these cases an extension of Eq. (10.85) is usually used, given by the generalized likelihood ratio:

$$R = \frac{L(\theta_0; X)}{\max_\theta L(\theta; X)} . \tag{10.96}$$

This equation represents the ratio between the likelihood calculated at the parameter optimal value and the value $L(\theta, X)$ calculated at the maximum. We do not address this rather complex topic, but we limit ourselves to observing that the generalized likelihood ratio test usually performs well in this context, because likelihood is always a function of a sufficient statistic of the problem.

In all cases (actually not many) where the power can be explicitly calculated, one can still determine for which alternative hypotheses a chosen test is sufficiently powerful by examining the *power function*, defined as a function of the parameters of the alternative hypothesis as:

$$\pi(\theta) = 1 - \beta(\theta) = 1 - P\{X \in A; H_1\} , \tag{10.97}$$

where A is the acceptance region of H_0, that is, the complementary set of the critical region.

Exercise 10.12
Find the power function for the case of Exercise 10.11.

Answer The decision criterion obtained consisted in the rejection of the hypothesis H_0 for a number of defects $S \geq 16$ on a batch of 1000 elements. Using the Gauss approximation (10.92) of the binomial density, the power function (10.97) results:

$$\pi(\mu) = 1 - \Phi\left(\frac{16 - \mu}{\sqrt{\mu}}\right) , \tag{10.98}$$

(continued)

Exercise 10.12 (continued)
This curve can be obtained with the R instructions:

```
> mu<- seq(0,50,by=0.2)
> pow<- 1-pnorm((16-mu)/sqrt(mu))
> plot(mu,pow,type='l')
> grid()
```

where `mu` is the expected value of the number of defective pieces. The function is shown in Fig. 10.9. It shows that the support of the sample distribution under H_0 (defect rate of 1%) becomes practically disjoint from the one under H_1 when the alternative percentage of defects is larger than 3%. We arrived more intuitively to the same conclusion in Exercise 10.11.

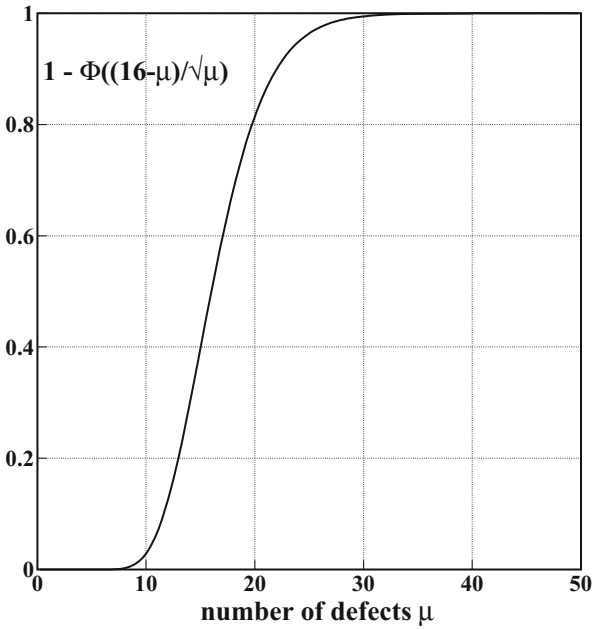


Fig. 10.9 Power function for the case of the Exercise 10.11

10.13 Sequential Tests

In addition to the transformation (10.91), the difficulty in finding the distribution of the Neyman-Pearson variable R can be overcome also in another very elegant, but approximate, way.

Consider two hypotheses H_0 and H_1 with the corresponding likelihood functions $L(\theta_0; x) \equiv L_0(x)$ and $L(\theta_1; x) \equiv L_1(x)$. The ratio:

$$R \equiv R(X) = \frac{L_0(X)}{L_1(X)} \equiv \frac{L_0}{L_1} \tag{10.99}$$

tends to assume large values if H_0 is true, small values if H_1 is true. We can then define the test of the two hypotheses as follows:

$$\begin{array}{lll} \text{accept } H_0 \text{ if} & (L_0/L_1) \geq r_H & \Longrightarrow L_0 \geq r_H L_1 \\ \text{accept } H_1 \text{ if} & (L_0/L_1) \leq r_h & \Longrightarrow L_0 \leq r_h L_1 \\ \text{no decision if} & r_h < (L_0/L_1) < r_H & \end{array} \tag{10.100}$$

The areas corresponding to the probabilities of type I and type II errors α and β are shown in Fig. 10.10.

For a given observation of dimension N, the spectrum of the variable R is then divided into three disjoint regions, R_0, R_1 and R_c, corresponding respectively to the decision to accept H_0, to accept H_1 and to not make any decision.

The interval R_0 then corresponds to the condition $R > r_H$, whereas R_1 is the interval $R < r_h$. From Fig. 10.10 and Eqs. (10.100), it is clear that the probabilities

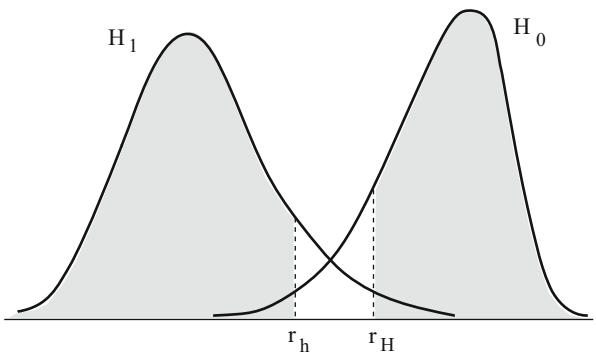


Fig. 10.10 In hypothesis testing, when using the likelihood ratio r, H_0 is accepted if $R \geq r_H$; H_1 is accepted if $R \leq r_h$; no decision is taken if $r_h < R < r_H$

to assume correct decisions are given by :

$$1-\alpha = \int_{R_0} R(x)L_0(x)\,\mathrm{d}x \geq r_H \int_{R_0} \frac{L_0(x)}{L_1(x)} L_1(x)\,\mathrm{d}x$$
$$= r_H \int_{R_0} R(x)L_1(x)\,\mathrm{d}x = r_H\,\beta\,, \tag{10.101}$$
$$1-\beta = \int_{R_1} R(x)L_1(x)\,\mathrm{d}x \geq \frac{1}{r_h} \int_{R_1} \frac{L_0(x)}{L_1(x)} L_0(x)\,\mathrm{d}x$$
$$= \frac{1}{r_h} \int_{R_1} R(x)L_0(x)\,\mathrm{d}x = \frac{\alpha}{r_h}\,, \tag{10.102}$$

where, as usual, α and β are the type I and type II error probabilities. From these inequalities, the lower limit of r_h and the upper limit of r_H can be immediately obtained as:

$$\frac{\alpha}{1-\beta} \leq r_h\,, \qquad r_H \leq \frac{1-\alpha}{\beta}\,. \tag{10.103}$$

In conclusion, Eq. (10.100) can be described as follows: in a test of H_0 against H_1, when α and β are fixed a priori, if:

$$R \leq \frac{\alpha}{1-\beta} \leq r_h \quad \text{accept } H_1\,, \tag{10.104}$$
$$\frac{\alpha}{1-\beta} < R < \frac{1-\alpha}{\beta} \qquad \text{no decision is taken}\,, \tag{10.105}$$
$$R \geq \frac{1-\alpha}{\beta} \geq r_H \quad \text{accept } H_0\,. \tag{10.106}$$

This test is approximate but *is independent of the knowledge of the probability distribution of the likelihood ratio (10.85)*.

Which type I and II errors α' and β' are actually associated with the test resulting from Eqs. (10.104)–(10.106)? Indicating the new threshold values of the approximate test with r'_H and r'_h, a partial answer to this question can be given because one always has $\alpha' + \beta' \leq \alpha + \beta$. Indeed, from Eqs. (10.101)–(10.103) one easily obtains:

$$1-\alpha' \geq \beta'\, r'_H = \beta'\,\frac{1-\alpha}{\beta}\,,$$

$$1-\beta' \geq \frac{\alpha'}{r'_h} = \alpha'\,\frac{1-\beta}{\alpha}\,,$$

so that:

$$\frac{\alpha'}{1-\beta'} \leq \frac{\alpha}{1-\beta}\,, \qquad \frac{1-\alpha}{\beta} \leq \frac{1-\alpha'}{\beta'}\,,$$

and hence:

$$\alpha'(1-\beta)+\beta'(1-\alpha) \leq \alpha(1-\beta')+\beta(1-\alpha') \Longrightarrow \alpha'+\beta' \leq \alpha+\beta\,. \qquad (10.107)$$

Exercise 10.13
Perform the approximate likelihood ratio test on data from Exercise 10.11, assuming a type I error of 5% and type II error of 20%.

Answer In Exercise 10.11 we studied the probability function of the sum variable S, connected to the Neyman-Pearson variable R by Eq. (10.95). We found that the hypothesis H_0 of a 1% defect rate was discarded, at a level of significance α of 5%, for a number of defective pieces greater than 15 on a batch of 1000 pieces. We had also found a probability β of about 20% for the type II error when the hypothesis H_1 has a defect rate of 2%.

Also in this case the levels $\alpha = 0.05$ and $\beta = 0.20$ are requested. From Eqs. (10.104)–(10.106) one obtains:

$$\begin{array}{lll} \text{accept} & H_0 \text{ if} & R \geq 4.75 \\ \text{no decision} & \text{if} & 0.0625 < R < 4.75 \\ \text{accept} & H_1 \text{ if} & R \leq 0.0625 \end{array}$$

From Eq. (10.95) it results that to the values corresponding respectively to $r_H = 4.75$ and $r_h = 0.0625$ are:

$$s_h = 18.4\,, \qquad s_H = 12.2\,.$$

The hypothesis (H_0) of a defect rate less than 1% should be accepted if the number of defective pieces remains below 12, whereas when this value exceeds 18, the hypothesis H_1 of a defect rate greater than 2% must be preferred. Values between 12 and 18 relate to an estimated defect rate between 1% and 2% and therefore represent an area of uncertainty where it is advisable not to make any decision.

The fact that the limits in Eqs. (10.104)–(10.106) do not depend on the sample size N suggests an application of the likelihood ratio method to be carried out iteratively, as N increases, during an experiment. The test is performed until when

the results remain in the zone of indecision. More precisely, the sample size is increased if:

$$r_h < R_N < r_H, \quad N \geq 1\,, \tag{10.108}$$

and the test is stopped as soon as Eq. (10.108) is not satisfied. The limits of the indecision interval are to be specified according to the level of the test and the type II error.

Notice that both the sample size N and the final likelihood ratio value

$$R_N = \frac{L_0(\theta_0; X_N)}{L_1(\theta_1; X_N)}\,, \tag{10.109}$$

have to be considered as random variables when the indecision zone is left. These tests are called *sequential*.

If we assume that we have determined r_h and r_H so that the probabilities of type I and II errors are α and β, Eq. (10.103) is valid as well. Hence, in practice, a sequential test of approximate level will have as an area of indecision:

$$\frac{\alpha}{1-\beta} < R_N < \frac{1-\alpha}{\beta}, \quad N \geq 1\,. \tag{10.110}$$

Also in this case, Eq. (10.107) determines the actual levels of the test: $\alpha' + \beta' \leq \alpha + \beta$.

Although the sequential test is approximate, it requires on average a smaller sample size compared to tests with fixed N to reach the same power, thus giving a considerable saving in sampling time, even of 50%. Intuitively, this happens because there are often sequences of favourable (or unfavourable) cases that quickly give values of R_N outside the indecision interval (10.110), allowing to immediately choose the correct hypothesis (H_0 or H_1).

We omit here the formal demonstration of this very useful property; it can be found, under very general conditions, in [MGB73]. Instead, in the next exercise, we will present a different approach based on simulation techniques.

Exercise 10.14

Apply the sequential test with variable N, to the data of Exercise 10.11, assuming $\alpha = 0.05$, $\beta = 0.20$ and a number of defective pieces of 1% (H_0) against the alternative hypothesis of 2% (H_1). Estimate, with simulation methods, the distributions of the number of defective pieces S_N and of the sampled pieces N.

Answer If N is variable, Eq. (10.94) can be written as:

(continued)

Exercise 10.14 (continued)

$$R_N = \left(\frac{1}{2}\right)^{S_N} e^{N/100} ,$$

where S_N is the number of defective pieces. Passing to logarithms one has:

$$\ln R_N = -S_N \ln 2 + \frac{N}{100} \simeq -\frac{S_N}{1.44} + \frac{N}{100} ,$$

and Eq. (10.110) becomes:

$$1.44 \ln\left(\frac{\alpha}{1-\beta}\right) - \frac{1.44}{100}N < -S_N < -\frac{1.44}{100}N + 1.44 \ln\left(\frac{1-\alpha}{\beta}\right) .$$

Changing the signs and inserting the values $\alpha = 0.05$ and $\beta = 0.20$, the indecision interval, where the sampling continues, becomes:

$$-2.25 + 0.0144\, N < S_N < 3.99 + 0.0144\, N . \tag{10.111}$$

H_0 is accepted for S_N values to the left of this interval, whereas, for values to the right, the alternative H_1 is chosen.

The two discrete random variables S_N and N appearing in Eq. (10.111) have a distribution difficult to be determined with analytical methods. However, we can simulate them, proceeding as follows: consider a number $0 \leq \xi \leq 1$ supplied by the `rndm` routine and increment N by one; if $\xi \leq 0.01$ (hypothesis H_0), S_N is also increased by one; if S_N and N satisfy Eq. (10.111), a further step is needed, if $S_N \geq 3.99 + 0.0144\, N$ then H_1 is chosen; if $S_N \leq -2.25 + 0.0144\, N$, H_0 is kept. The same test can be simulated under the alternative hypothesis H_1, increasing S_N if $\xi \leq 0.02$. The results, obtained by repeating the computer test for 100,000 times with our code `Sequen` are displayed in Fig. 10.11. The simulation shows, in Fig. 10.11a and b, that the average number of pieces sampled is 451 ± 1 under H_0 and 570 ± 1 under H_1. This number is about half of that required for the fixed sample test with the same values of $\alpha = 0.05$ and $\beta = 0.20$, which is approximately 800, as shown in Exercise 10.10. It can also be noticed that N has a distribution with an exponential-like tail on the right, with a small (but not negligible) number of tests reaching larger values than the sample size needed in the fixed N test. The density of S_N below H_0, shown in Fig. 10.11c, is of exponential type, with an average value of about five pieces. In the top part of Fig. 10.11a and b, an estimate of the "experimental" values of the levels $1-\alpha'$ and $1-\beta'$ is reported, as the fraction of tests where H_0 was accepted when true (Fig. 10.11a) and the same for H_1 (Fig. 10.11b). From the data the values of $\alpha' \in (3.8 \pm 0.1)\%$ and $\beta' \in (19.3 \pm 0.1)\%$ are obtained (check, as an exercise, the values of the

(continued)

Exercise 10.14 (continued)
statistical errors). They do not coincide with the a priori fixed values because the test deals with discrete variables. The sum $\alpha' + \beta' \in (23.1 \pm 0.1)\%$ is less than the value $\alpha + \beta = 25\%$, according to Eq. (10.107). In Fig. 10.11d the values in the plane (N, S_N) obtained in 100,000 tests under the two hypotheses are displayed. The points are arranged just above or below the two boundary lines of Eq. (10.111). This curve allows the graphical determination of the indecision interval; for example, when $N = 500$, if $S_N < 5$ one chooses H_0, if $S_N > 11$ one chooses H_1. Notice also that the discrete structure of the problem is clearly visible in each histogram.

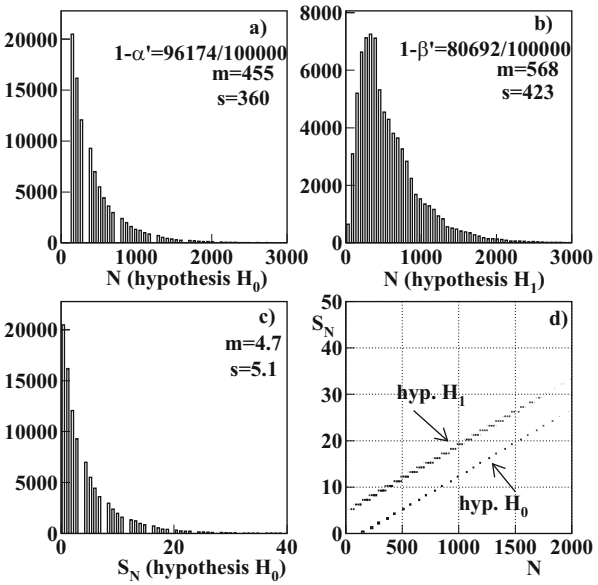


Fig. 10.11 Results of the simulation of 100,000 sequential tests of Exercise 10.14. Histogram of the number N of pieces to be sampled if the defect probability is 1% (hypothesis H_0) (**a**); the same if this probability is 2% (hypothesis H_1) (**b**); histogram of the number of discards S_N under H_0 (**c**); points in the plane (N, S_N) corresponding to the acceptance of the two hypotheses (**d**)

10.14 Problems

10.1 An urn contains black and white marbles in the proportion of 2 : 1 or 1 : 2 (it is not known in favour of which colour). In the case of four extractions with replacement, find the ML estimate of the proportion as a function of the possible results of extracted black marbles ($x = 0, 1, 2, 3, 4$).

10.2 Using the method of the previous problem, find the ML estimate of the probability p as a function of the successes x obtained in $n = 3$ attempts, considering the *nine* possible values: $p = 0.1, 0.2, 0.3, \ldots, 0.9$.

10.3 If $\hat{\mu}$ and $\hat{\sigma}$ are ML estimates of μ and σ, find the ML estimate of the quantile $F(x_\alpha) = \alpha$.

10.4 Calculate the ML estimate of λ using n observed split times t_i from a population having exponential density $\lambda \exp[-\lambda t]$.

10.5 A method used to estimate the number N of elements in a finite population is to take a random sample of n elements, mark them and re-enter them into the population. A second sample of n elements is then drawn, and the number x of marked elements is determined. If in an experiment we fix $n = 150$ and find $x = 37$, calculate the ML estimate of N, (a) exactly and (b) under the approximation $N \gg n$.
Hint: use the hypergeometric law (1.33).
What methods could be used to roughly determine the interval estimate of N, i.e. $N \in \hat{N} \pm \sigma[\hat{N}]$?

10.6 If X is a Bernoulli variable with p.d.f. $b(x) = p^x(1-p)^{1-x}$, verify whether the statistics $S = X_1 + X_2$ and $P = X_1 X_2$ are sufficient.

10.7 The sum $w = \sum_i x_i^2$ is obtained from a sample of size n. Knowing that the variable has the normal distribution $N(0, \sigma^2)$, with a known zero mean and variance to be determined, apply the methods (a) and (b) of Sect. 10.5 to determine the confidence interval of σ^2.

10.8 A sampling from a normal population $N(\mu, \sigma^2)$ with $\mu = 0$ has given the values $x_i = 1.499, 5.087, 0.983, 2.289, 1.045, -1.886$. Find the ML point and interval estimate of the variance σ^2, with $CL = 95.4\%$.
Then, calculate another estimate using Eq. (6.76). Which method is preferable?

10.9 From 100 simulated values of $X \sim N(0, \sigma^2)$, a sum $w = \sum_i x_i^2 = 1044$ is obtained. Estimate the variance with the methods used in the two previous problems.

10.10 Two samples of ball bearing diameters (in cm), produced by the same machine in two different weeks, have number of events, mean and standard

deviation, respectively, equal to $n_1 = 100$, $m_1 = 2.08$, $s_1 = 0.16$ and $n_2 = 200$, $m_2 = 2.05$, $s_2 = 0.15$. Calculate both the ML and the interval estimate of the mean diameter value.

10.11 A radioactivity monitor detects the presence or not of nuclear particles without counting their number. From a measurement of 50 homogeneous random samples with the same amount of substance and within the same unit time interval, 45 of them tested positive by the monitor. Find the ML estimate of the average number of particles emitted by that amount of substance per unit time.

10.12 The recording of $N = 1000$ arrival times provided the histogram:

t	2	4	6	8	10	12	14	16	18	20
$n(t)$	472	276	132	51	36	12	11	7	1	2

which shows that between 0 and 2 s, there were 472 split times, between 2 and 4 s 276 split times and so on. Perform the best fit analysis using the exponential density $e(t) = \lambda \exp(-\lambda t)$ by determining the ML estimate of λ. Verify the validity of the hypothesis with the χ^2 test.

10.13 Under H_0, a binary variable X assumes values between 0 and 1 with probability $1 - \varepsilon$ and ε, respectively: $P\{X = 0, 1; H_0\} = (1 - \varepsilon), \varepsilon$. Let $0 < \varepsilon \ll 1$ be a small positive number. Consider the alternative hypothesis $P\{X = 0, 1; H_1\} = \varepsilon, (1 - \varepsilon)$. Calculate significance level and the power of the following test: accept H_0 if in an experiment $\{X = 0\}$, whereas H_1 is accepted if $\{X = 1\}$.

10.14 In the situation of the previous problem, consider the result of two independent trials and the following test: H_0 is accepted if $\{X_1 = 0, X_2 = 0\}$; H_1 is chosen if $\{X_1 = 1, X_2 = 1\}$; a coin is tossed to randomly choose between H_0 and H_1 when $\{X_1 + X_2 = 1\}$. Find the significance level and the power of this test, and compare the result with that of the previous problem.

10.15 How is the randomized test performed in the case of Exercise 10.12?

10.16 A car company uses suspensions withstanding an average of 100 h in an extreme fatigue test (H_0 hypothesis). A supplier offers a new type of suspension claiming an average of 110 h (alternative hypothesis H_1). Knowing that the lifetime distribution is negative exponential, design a test on the new type of suspension. (a) Keep the sample size fixed, use the normal approximation for the mean and make sure that the probability of being wrong if the new suspensions are the same as the old ones is 1% and the probability of accepting good suspensions is 95%. (b) Now solve the problem with a sequential test and evaluate, with a simulation code, the distributions of the average times and of the number of pieces required for the test.

Should the fixed-size sample or sequential test be used? What strategy would you adopt?

10.17 Find the power function for the Problem 10.16. *Hint*: compare Eq. (3.57) with Eq. (3.67).

10.18 We want to test the null hypothesis that the probability of heads in one toss of a given coin is $p = 0.5$ against the alternative hypothesis $p = 0.3$. What is the number n of tosses needed to decide between these two hypotheses, assuming a test level of 10% and a power of 90%, i.e. $\alpha = 0.10$ and $\beta = 0.10$? Use the Gaussian approximation.

10.19 With the same values of α and β as in the previous problem, find the number of successes x as a function of the number n of tosses that allows to choose between the two hypotheses with a sequential test. Determine, by flipping a coin or with a simulation code, the average number of flips n needed to decide between the two hypotheses. Compare the result with that of the previous problem.

10.20 If X has density $p(x) = (1 + \theta)x^{\theta}$, $0 \leq x \leq 1$, find, with a sample of $n = 100$ events, the best critical region for testing $\theta = \theta_0 = 1$ against $\theta = \theta_1 = 2$ at a level $\alpha = 0.05$. Determine the test power and the power function.

Chapter 11
Least Squares

Of all the principles which can be proposed for that purpose, I think there is none more general, more exact, and more easy of application, than that of which we have made use in the preceding researches, and which consists of rendering the sum of the squares of the errors a minimum.

Adrian Legendre, "NEW METHODS FOR DETERMINATION OF THE ORBITS OF COMETS".

11.1 Introduction

In Sect. 10.6 we showed that the least squares (LS) method originates from the maximum likelihood principle when the variables are Gaussian.

Historically, however, things have developed differently. Indeed, while maximum likelihood was introduced by Fisher in the early 1900s, the least squares method was first applied by the French mathematician Legendre in 1803, as indicated in the epigraph of this chapter. Later Laplace, in his famous treatise "Thèorie Analitique des Probabilitès" of 1812, showed that the LS method produces unbiased estimates even in the case of non-Gaussian variables. The decisive step for the correct collocation of the LS method in statistics was then taken by Gauss in 1821, with the proof that the LS estimators are unbiased and efficient (minimal variance), when the observations are linear functions of the parameters to be determined. This fundamental theorem was extended and better formalized by Markov in 1912 and is now known as the Gauss-Markov theorem.

On the basis of these results, we can state that, when the densities of the sample populations are not a priori known, the least squares method can be used as an alternative to the maximum likelihood method when the expected values of the observations can be expressed as linear combinations of the parameters. The LS principle is therefore not only a consequence but also *a complement to the* ML method, in respect of the class of unbiased estimators with minimum variance. If the expected values are a nonlinear combination of the parameters, the LS method is still applicable, but the resulting estimators could be biased and not of minimal variance. Later, we will also briefly comment this case. The LS method, as

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3_11

discussed in Sect. 10.6, consists in finding the minimum of a χ^2-type formula, such as Eq. (10.52). Generalizing, we can define the method as:

Definition 11.1 (Least Squares (LS) Method) Let $y_1, y_2, \ldots, y_n$ be the observed values of the random variable Y (called "dependent or response variable") such as $\langle Y_i \rangle = \mu_i(x_i, \boldsymbol{\theta}) \equiv \mu_i(\boldsymbol{\theta})$. Concerning the function $\mu_i(x_i, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$ is a vector of parameters with dimension $(p + 1)$, and $x_1, \ldots, x_n$ are n observed values of a variable X called "predictor or independent variable". The variances $\mathrm{Var}[Y_i] = \sigma_i^2$ are known. The least squares estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is the one that minimizes the quantity:

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{[y_i - \mu_i(\boldsymbol{\theta})]^2}{\sigma_i^2} \,. \tag{11.1}$$

The predictor can also be a vector $\boldsymbol{X}$. In this case $\mu_i(\boldsymbol{x}_i, \boldsymbol{\theta})$ denotes the ith observed value of $\boldsymbol{X}$.

From the definition it is understood that the functions $\mu_i(\boldsymbol{\theta})$, as $\boldsymbol{\theta}$ varies, constitute a class of models that relate the (random or deterministic) variable X with $\langle Y \rangle$. Since this relation allows us to evaluate $\langle Y_i \rangle$ without having observed y_i, X is called the predictor, while Y is called the response. The minimization of Eq. (11.1) with respect to $\boldsymbol{\theta}$ then identifies the best model within this class. Unlike the ML method, we note that here it is not necessary to know the distribution of the variables to estimate $\boldsymbol{\theta}$, but only their expected value and variance. Another important aspect of the method is that the variances σ_i^2 do not depend on $\boldsymbol{\theta}$.

After the θ minimization, it is possible to test the model, if we are not sure of its validity. If the variances σ_i^2 are known, the χ^2 test can be performed as explained in Sects. 7.5 and 7.6.

To avoid misunderstandings, we remind you that:

- The minimization procedure is purely mathematical and not statistical and consists in finding the minimum of the " χ^2-type" quantity (11.1).
- The quantity:

$$(\chi^2)_{min} \equiv \chi^2(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^{n} \frac{\left[y_i - \mu_i(\hat{\boldsymbol{\theta}})\right]^2}{\sigma_i^2} \tag{11.2}$$

does follow the $\chi^2(n-p-1)$ *distribution if* Y_i *are independent Gaussian random variables with expected value* $\mu_i(\boldsymbol{\theta})$, *which is a linear function of* $\boldsymbol{\theta}$.
After this clarification, we think that to always denote as χ^2 the quantity to be minimized, as we will do in the following, even if it may *not* follow the $\chi^2(n - p - 1)$ distribution, does not lead to confusion.[1]

[1] The degrees of freedom are now $(n - p - 1)$ and not $(n - p)$ because in Definition 11.1 the size of θ is $(p + 1)$.

If the goodness of fit test does not reject the null hypothesis or if one is a priori sure of the adopted model, it may be useful to give confidence intervals for $\boldsymbol{\theta}$ or for $\mu(\boldsymbol{\theta})$. We will see how to carry out this operation in the linear Gaussian case. The confidence intervals for $\mu(\boldsymbol{\theta})$ are then used to predict $\langle Y_i \rangle$ at values of x_i for which no experiments have been performed.

In Sect. 10.7 we have already discussed an example, derived from the ML principle, where the LS method is used to fit density functions to histograms. In this chapter we will instead describe the most important and common applications of the LS method, i.e. the search for functional forms of $\mu_i(\boldsymbol{\theta})$. Our goal is twofold: to explain a method (although not necessarily always the best) to perform and verify the fit that can be used in every situation, and enable you to choose the right algorithm and to use the minimization codes in a statistically correct way for any type of problem.

We will not describe in detail how these algorithms work, because this is essentially a topic of numerical computation, rather than of statistics. For more details you can refer to the texts [BR92] and [PFTW92] or directly to the specific user manuals of the minimization codes, such as [Jam92]. In our web pages [RPP], you will also find the `Linfit` and `Nlinfit` routines to perform linear and nonlinear least squares fits and further technical-numerical information.

Let us now begin by discussing two types of dependency between variables that may occur and which will be the focus of most of this chapter.

11.2 No Errors on Predictors

A very frequent statistical situation is given by a random variable Y written as:

$$Y = f(x) + Z, \quad f(x) \equiv f(x, \boldsymbol{\theta}) \,, \tag{11.3}$$

where $f(x)$ is a known function (apart from the $\boldsymbol{\theta}$ parameter) of the non-random predictor x.

Note that one can always set $\langle Y \rangle = f(x)$ and $\langle Z \rangle = 0$, because if $\langle Z \rangle = z_0$, it would be enough to redefine f as $f(x) + z_0$ and Eq. (11.3) would hold again. This trick allows us to represent Y as the sum of a deterministic component (its expected value), which carries the essential part of the relationship with the predictor x and of a random part, which represents the fluctuations around the mean.

In the case of a set of fixed and measurable (without error) x_i values, a repeated sampling of Y for each value x_i provides a graph similar to that of Fig. 11.1a, with fixed x_i and Y_i fluctuating around their average values. If only one measurement is performed and the standard deviation of the Z_i is known, the plot is drawn as in Fig. 11.1b, with the *error bars* equal to the standard deviations σ_i of Z_i, and centred on the measured values y_i.

The goal of the least squares method is to determine the *functional form* $f(x)$ which links $\langle Y \rangle$ to the deterministic variable x. In other words, we need to

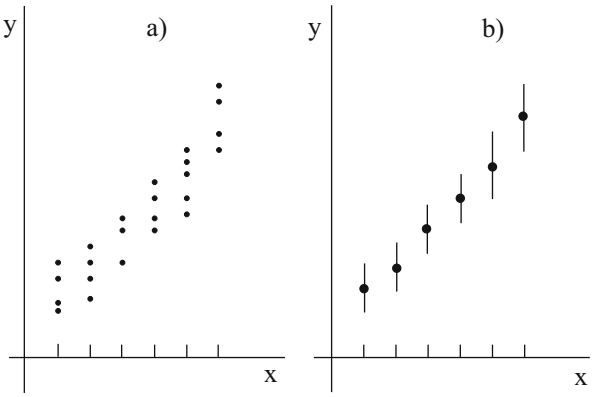


Fig. 11.1 Sampling of a random variable y as a function of a non-random variable x; repeated samplings (**a**) and standard representation in the case of a single sampling (**b**)

determine, for each x, the mean $f(x, \hat{\boldsymbol{\theta}})$ as the *curve with respect to which the fluctuations of Y are random*. Here "random" means that the dependence between $\langle Y \rangle$ and x is completely described by f, the remainder being a pure error term. If in Eq. (11.1) $\mu_i(\boldsymbol{\theta}) = f(x_i, \boldsymbol{\theta})$, the χ^2 is given by:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_i^2} , \tag{11.4}$$

where the variance of Y appears in the denominator. Based on Eq. (11.3), it is given by:

$$\sigma_i^2 \equiv \mathrm{Var}[Y_i] = \mathrm{Var}[Z_i] .$$

For example, for a linear relation, one has: $f(x, \boldsymbol{\theta}) = \theta_1 + \theta_2 x$.

The search for a density starting from a histogram, obtained minimizing Eq. (10.60), is an instance of the non-random predictor case. In fact, the function $f(x_i, \boldsymbol{\theta}) = Np(x_i, \boldsymbol{\theta})$ coincides with the expected value of $Y_i = I_i$, the variables x are the spectrum values in the discrete case or the midpoint of the spectrum within the predetermined bin Δx in the continuous case, and there is no error on x. The statistical error is only on Y and is given by the fluctuations in the number of events falling in the different histogram bins during sampling.

Another very frequent situation comes from the measurements of physical quantities, when one (the x variable) is measured with negligible error and basically under the experimenter full control, while the other (the Y variable) is evaluated with lower precision. In this case the function $y = f(x, \boldsymbol{\theta})$ represents the physical law between the two variables.

Let us now see a second type of dependence between variables defined by the relations:

$$Y = f(X) + Z, \quad f(X) \equiv f(X, \boldsymbol{\theta}) , \tag{11.5}$$

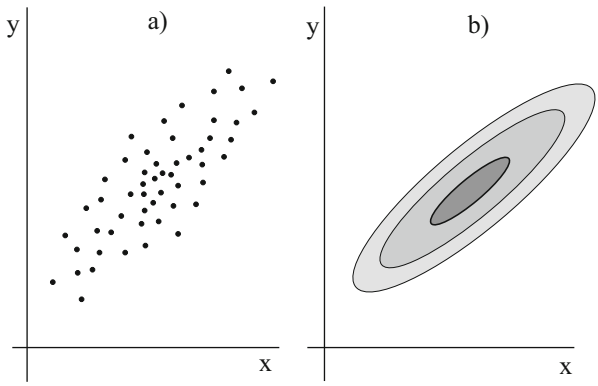


Fig. 11.2 Sampling of pairs of correlated random variables (**a**) and representation of the corresponding density (**b**)

where X and Z are random variables and $\langle Z \rangle = 0$. Unlike Eq. (11.3), X is now a random variable, and between X and Y, there is a *statistical correlation*, in general different from zero. In a repeated sampling with n independent trials, we observe as many values of the random vector (X_i, Y_i), which we indicate with the pairs (x_i, y_i), $i = 1, \ldots, n$. If we plot them on a graph, we obtain a point cloud like in Fig. 11.2a. The joint density of (X, Y), whose functional form is given by Eq. (4.3), can be represented as in Fig. 11.2b, indicating the values of its integral with different shades of grey in the different regions of the plane.

As in Eq. (11.3), we assume that the argument of f is observed exactly. If we are interested not so much in the random mechanism that generates X as in the relation between X and Y, we can go back to Eq. (11.4), set a value x and use the distribution of Y conditional on $\{X = x\}$ to calculate the χ^2 function as:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\mathrm{Var}[Y_i|x_i]} \,. \tag{11.6}$$

Here the uncertainties σ_i^2 have been replaced with the conditional variances $\mathrm{Var}\,[Y_i|x_i]$, and the expected values of Y_i have been replaced with the corresponding conditional expected values of $f(x_i, \boldsymbol{\theta}) = \langle Y_i|x_i \rangle$. Therefore, keeping $x_1, \ldots, x_n$ fixed, we obtain $\hat{\boldsymbol{\theta}}$ by minimizing Eq. (11.6). According to the model, $\langle Y|x \rangle$ is given by $f(x)$ independently of any (X, Z) distribution. Instead, the latter distribution influences $\mathrm{Var}\,[Y|x]$, since:

$$\mathrm{Var}[Y|x] = \mathrm{Var}[f(X)|x] + \mathrm{Var}[Z|x] = \mathrm{Var}[Z|x] \,.$$

If X and Z are independent, then:

$$\mathrm{Var}[Y|x] = \mathrm{Var}[Z|x] = \mathrm{Var}[Z], \quad \text{and} \quad \mathrm{Var}[Y] = \mathrm{Var}[f(X)] + \mathrm{Var}[Z] \,. \tag{11.7}$$

The denominator of Eq. (11.6) becomes constant, whereas if there is a dependence between X and Z the condition $\mathrm{Var}[Z|x] = \mathrm{Var}[Z]$ is no longer valid.

The functional dependence described by Eq. (11.5) is very common, and the case of Table 6.4 is one of them. In fact, the chest perimeter of a certain soldier will be related not to the average height value, but *just to the height of that same soldier*. We can also think of a financial study that tries to correlate the trend of the Milan stock exchange index with that of New York: the Milan index will be correlated to the *current value* of the New York index, with the superposition of other random variables containing (local) sources of variation.

Equations (11.3) and (11.5) are formally the same, but the meaning of the function f is different: in the first case, x is a non-random variable, and $f(x)$ represents an analytic functional dependence between x and $\langle Y \rangle$, while in the second case, $f(X)$ induces the statistical correlation between them. Then, referring to Eq. (11.5), we can call $f(X)$ as the *correlation function* between X and Y.

To better clarify this concept, when X and Z are independent, $\mathrm{Var}[Z] = \mathrm{Var}[Z|x]$, and from Eq. (11.7), we have:

$$\sigma_y^2 \equiv \mathrm{Var}[Y] = \mathrm{Var}[f(X)] + \mathrm{Var}[Z|x] \,. \tag{11.8}$$

Since $\mathrm{Var}[Y|x] = \mathrm{Var}[Z|x]$, Eq. (11.8) can be rewritten as:

$$\mathrm{Var}[Y|x] = \sigma_y^2 \left(1 - \frac{\mathrm{Var}[f(X)]}{\sigma_y^2}\right) \,. \tag{11.9}$$

This equation should be compared to Eq. (4.55), which represents the conditional variance of Y given $\{X = x\}$ when the p.d.f. is the bivariate Gaussian. For convenience we rewrite this equation here:

$$\mathrm{Var}[Y|x] = \sigma_y^2 (1 - \rho^2) \,.$$

By analogy between this expression and Eq. (11.9), we are then led to define a more general form of the correlation coefficient between X and Y:

$$\rho = \pm \frac{\sigma[f(X)]}{\sigma[Y]} \,, \tag{11.10}$$

where the sign is that of the covariance between X and Y, $\mathrm{Cov}[X, Y]$. Notice that, from Eqs. (11.7) and (11.10), it follows that $\rho = \pm 1$ when $\mathrm{Var}[Z] = 0$ and that $\rho = 0$ when $f(X)$ is constant. It is also easy to show that, if $f(X) = a + bX$, the correlation coefficient takes the familiar form (see Problem 11.1):

$$\rho = \pm \frac{\sigma[f(X)]}{\sigma[Y]} = \frac{\sigma_{xy}}{\sigma_x \, \sigma_y} \,. \tag{11.11}$$

We recall that, in Chap. 4, we have demonstrated that the correlation function between two jointly Gaussian random variables can only be linear and that $\langle Y|x\rangle$ follows the regression line (see Eq. 4.54):

$$f(x) = \langle Y|x\rangle = \mu_y + \rho\,\frac{\sigma_y}{\sigma_x}(x - \mu_x)\ .$$

This situation appears here as a special case of Eq. (11.5), valid for any function $f(X)$, where the correlation coefficient is defined more generally according to Eq. (11.10). Also in this case, the Cauchy-Schwarz inequality (4.28) assures that the property $-1 \leq \rho \leq 1$ remains valid.

11.3 Errors in Predictors

A particular but important case regarding the study of a functional dependence between $\langle Y\rangle$ and $\langle X\rangle$ occurs when:

$$X = x_0 + X_R\ , \quad Y = f(x_0) + Z\ , \quad f(x_0) \equiv f(x_0, \boldsymbol{\theta})\ , \tag{11.12}$$

where X_R and Z are independent with a null expected value. The relation that determines Y is identical to Eq. (11.3), but here x_0 is subject to statistical fluctuations, that is, the predictor x_0 is unknown, and only the random variable X is observed.

A repeated sampling of (X, Y), in correspondence with n true values $\boldsymbol{x}_0^\dagger = (x_{01}, \ldots, x_{0n})$, gives rise to a plot as that of Fig. 11.3a, with the variables x_i and y_i fluctuating around their mean values x_{0i} and $f(x_{0i})$. If a single measurement is performed and the standard deviations of X_{Ri} and Z_i are known, the graph is drawn as in Fig. 11.3b, with the standard deviations of X_{Ri} and Z_i drawn as error bars centred on the measured values x_i and y_i. This is often the case of

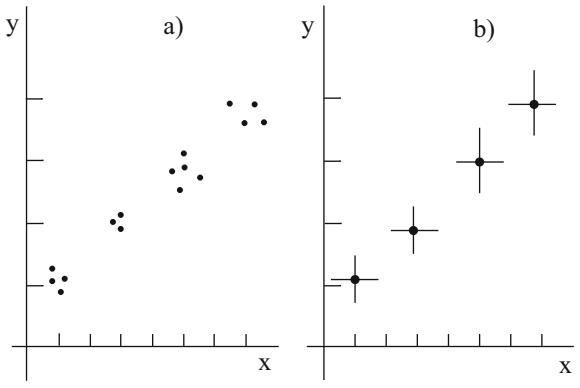


Fig. 11.3 Sampling of a random variable Y as a function of the mean value of another random variable X; repeated samplings (**a**) and standard representation in the case of an experiment with a single sampling only (**b**)

laboratory measurements, when two quantities linked by a deterministic physical law are measured with relevant uncertainties. Here too, the task of the least squares method is to find the physical law as a *best-fit curve*, reducing the effects of the uncertainties introduced by the measurement.

From Definition 11.1, the χ^2 to be minimized results in this case:

$$\chi^2(\boldsymbol{\theta}, \boldsymbol{x}_0) = \sum_i \frac{(x_i - x_{0i})^2}{\sigma_{x_i}^2} + \sum_i \frac{[y_i - f(x_{0i}, \boldsymbol{\theta})]^2}{\sigma_{y_i}^2} \,. \tag{11.13}$$

Now (x_i, y_i) are the observed pairs of points, and the crucial aspect of this equation is that, according to Eq. (11.12), f is a function of the unknown mean variables x_0, not of the measured ones. The free parameters to be estimated are then $(n + p + 1)$, i.e. the sum of the dimensions of $\boldsymbol{x}_0$ and of $\boldsymbol{\theta}$.

The minimization of Eq. (11.13) is not difficult if one has a *nonlinear* χ^2 minimizing program which accepts the equation of the function to be minimized. This is the case of our `Linemq` routine that we use to solve Problems 11.2 and 11.3. However, the doubled number of experimental values and the high number $(n+p+1)$ of free parameters can cause problems when dealing with large samples. To overcome these problems, we approximate the minimum point of Eq. (11.13) by setting $x_{0i} = x_i$ in the first sum, which thus is always zero, and by evaluating the errors σ_{y_i} with a first-order Taylor expansion of $f(X)$ around x_0:

$$f(X) = f(x_0) + (X - x_0) f'(x_0) + o(X - x_0)^2 \,. \tag{11.14}$$

If $\mathrm{Var}[X_R]$ is small enough to omit the terms of order higher than the first (even if random), we can apply the variance operator to Eq. (11.14) to obtain:

$$\mathrm{Var}[f(X)] \simeq f'^2(x_0)\, \sigma_x^2 \,.$$

Since $f'^2(x_0)$ depends on the unknown x_0 value, we replace it with the observed value x, to obtain the so-called effective variance σ_E^2:

$$\mathrm{Var}[Y - f(X)] = \mathrm{Var}[Y] + \mathrm{Var}[f(X)] \simeq \sigma_y^2 + f'^2(x)\, \sigma_x^2 \equiv \sigma_E^2 \,, \tag{11.15}$$

where the assumption of the independence between X_R and Z has been used.

Using the differences $(y_i - f(x_i, \boldsymbol{\theta}))$ in the χ^2 numerator and the effective variance in the denominator, one obtains the simple result:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{Ei}^2} = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{y_i}^2 + f'^2(x_i, \boldsymbol{\theta}) \sigma_{x_i}^2} \,, \tag{11.16}$$

where only the measured values (x_i, y_i) appear together with their variances taken into account through the effective variance σ^2_{Ei}. In Eq. (11.16) the free parameters are again only the $(p+1)$ parameters $\boldsymbol{\theta}$.

If the values of $f'^2(x_i, \boldsymbol{\theta})$ are unknown, Eq. (11.16) does not follow Definition 11.1, and the χ^2 minimization gives always rise to nonlinear equations due to the derivative of f in the denominator, even when f is a linear function of the $\boldsymbol{\theta}$ parameters. However, if one has only linear minimization programs or intends to reduce the computation time, an iterative method can be used, starting at the beginning with $\sigma^2_{Ei} = \sigma^2_{y_i}$ and setting $\sigma^2_{Ei} = \sigma^2_{y_i} + f'^2(x_i, \boldsymbol{\theta}^{(k-1)})\sigma^2_{x_i}$ in the k-th cycle, using the estimates $\boldsymbol{\theta}^{(k-1)}$ of the previous iteration. In this way, the denominator becomes independent of the parameters being minimized in the current cycle, and the linearity is restored [Ore82]. However, it can be shown that this procedure gives inconsistent estimators (even if closer to the true parameter values than the methods that ignore the errors on X [Lyb84]), so that a nonlinear minimization code to solve Eq. (11.16) should be used. The results obtained with this algorithm are practically coincident with those given by the general formula (11.13). This fact can be verified with our `FitLineBoth` routine, which performs the minimization of Eq. (11.16). In the case of a straight line, if the data x, y, s_x, s_y have been stored in the vectors `x, y, sx, sy`, respectively, the calling instructions are the following (for more details you can see the comments inside the routine):

```
>fun<-function(par,x){return(par[1]+par[2]*x)}
>dfun<-function{par,x}{return(par[2])}  # derivative of fun
>FitLineBoth(x,y,sx,sy,par=c(0.5,0.5),fun=fun,dfun=dfun)
```

The effective variance has an intuitive geometric interpretation, shown in Fig. 11.4: the term $f'^2\sigma^2_x$, which is added to the initial variance σ^2_y, is just the effect of the fluctuations in x projected on the Y axis using the tangent of the angle α (i.e. the derivative $f'(x)$).

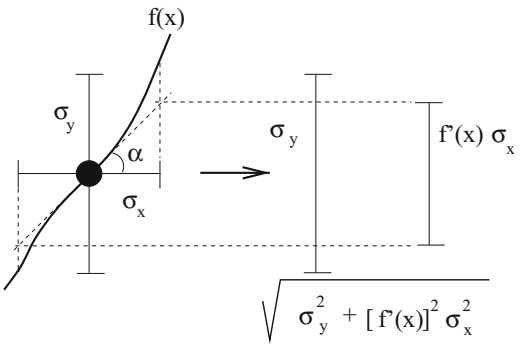


Fig. 11.4 Geometric interpretation of the effective variance

If the variables X_i and Y_i have a non-zero covariance $\mathrm{Cov}[X_i, Y_i]$, using Eq. (5.65), we can also further generalize Eq. (11.16) as follows:

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_{y_i}^2 + f'^2(x_i, \boldsymbol{\theta})\sigma_{x_i}^2 - 2f'(x_i, \boldsymbol{\theta})\,\sigma_{x_i y_i}} \,. \tag{11.17}$$

11.4 Least Squares Regression Lines: Unweighted Case

In this section we examine a particularly simple case of the relation (11.3), which includes also Eq. (11.5), considering X and Z independent and conditioned to the observed values of X. Therefore we will always write x in lowercase.

When the function connecting $\langle Y|x\rangle$ to x is a straight line, that is, $f(x, \boldsymbol{\theta}) = a + bx$, with $\boldsymbol{\theta} = (a, b)$, the relation between x and Y becomes:

$$Y = a + bx + Z \,, \tag{11.18}$$

where we assume also that $\mathrm{Var}[Y] = \mathrm{Var}[Z] = \mathrm{Var}[Z|x] = \sigma_z^2$ independently of x. For example, this happens when the distribution of (X, Y) is a bivariate Gaussian (see Eq. (4.55)).

Moreover, we assume σ_z^2 to be unknown, since the opposite case can be considered as a particular situation of weighted linear least squares, that we will describe in Sect. 11.6.

After the observation of an experimental random sample $(x_1, y_1), \ldots, (x_n, y_n)$, the χ^2 to be minimized can be written as:

$$\chi^2(a, b) = \frac{1}{\sigma_z^2} \sum_i (y_i - a - bx_i)^2 \,. \tag{11.19}$$

The minimization of this function does not depend on σ_z^2 and requires setting to zero the partial derivatives:

$$\begin{aligned} \frac{\partial \chi^2}{\partial a} &= -2\left[\sum_i y_i - n\,a - b\sum_i x_i\right] = 0 \,, \\ \frac{\partial \chi^2}{\partial b} &= -2\left[\sum_i x_i y_i - a\,\sum_i x_i - b\sum_i x_i^2\right] = 0 \,. \end{aligned} \tag{11.20}$$

To simplify the notation we define:

$$S_x = \sum_{i=1}^{n} x_i \,, \quad S_y = \sum_{i=1}^{n} y_i \,, \quad S_{xx} = \sum_{i=1}^{b} x_i^2 \,, \quad S_{xy} = \sum_{i=1}^{n} x_i y_i \,, \tag{11.21}$$

so that Eq. (11.20) generates the system of equations:

$$a\,n + b\,S_x = S_y \tag{11.22}$$

$$a\,S_x + b\,S_{xx} = S_{xy}\ , \tag{11.23}$$

or, in matrix form:

$$\begin{pmatrix} n & S_x \\ S_x & S_{xx} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}\ . \tag{11.24}$$

The parameters a and b represent the unknowns to be determined, while the sums (11.21) are known numerical coefficients. The least squares estimates of the parameters are immediately obtained from the matrix analysis methods:

$$\hat{a} = \frac{1}{D}\left[S_{xx}S_y - S_x S_{xy}\right]\ , \tag{11.25}$$

$$\hat{b} = \frac{1}{D}\left[nS_{xy} - S_x S_y\right]\ , \tag{11.26}$$

where:

$$D = n\,S_{xx} - S_x^2\ , \tag{11.27}$$

is the determinant of the system (11.24).

The parameters thus evaluated are marked with the symbol "^" to indicate that they are estimates, not the true values of a and b.

Using Eq. (4.23), after some algebra, Eq. (11.26) can be rewritten as:

$$\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{s_x s_y}\frac{s_y}{s_x} = r\frac{s_y}{s_x}\ , \tag{11.28}$$

where r is the estimate given by Eq. (6.117) of the linear correlation coefficient ρ of Eq. (11.11). Moreover, from Eq. (11.22), we obtain $\hat{a} = \langle y\rangle - \hat{b}\,\langle x\rangle$. Given these coefficients, the estimated least squares line $\hat{y}(x) = \hat{a} + \hat{b}x$ can be considered as an estimate of the regression line (4.54), also discussed at the end of Sect. 11.2:

$$\hat{y}(x) = \hat{a} + \hat{b}\,x = \langle y\rangle + r\frac{s_y}{s_x}(x - \langle x\rangle)\ . \tag{11.29}$$

So we conclude that *the regression line represents the locus of the points of the conditional distribution averages for Gaussian variables or the least squares line for non-Gaussian variables*. The error on the parameter estimates, which will be indicated as usual with the symbol s, is obtained by applying Eq. (5.65) on the variance of functions of random variables. This allows us to evaluate the propagation

effects of the fluctuations of the variables Y (i.e. of the random part) on the parameters $\hat{a}$ and $\hat{b}$ using Eqs. (11.25 and 11.26):

$$\sigma_f^2 = \sigma_z^2 \sum_i \left(\frac{\partial f}{\partial y_i} \right)^2 . \tag{11.30}$$

The covariance $\mathrm{Cov}[Y_i, Y_j]$ now does not appear because the independence of the observations Y_i is assumed. Taking into account that:

$$\frac{\partial S_y}{\partial y_i} = 1 \quad \text{and} \quad \frac{\partial S_{xy}}{\partial y_i} = x_i ,$$

one obtains:

$$\frac{\partial \hat{a}}{\partial y_i} = \frac{1}{D} \left[S_{xx} - S_x \, x_i \right] , \quad \frac{\partial \hat{b}}{\partial y_i} = \frac{1}{D} \left[n \, x_i - S_x \right] . \tag{11.31}$$

Applying Eq. (11.30) and having in mind Eq. (11.27), it easy to deduce that:

$$\mathrm{Var}[\hat{a}] = \frac{\sigma_z^2}{D^2} \sum_i \left([S_{xx} - S_x x_i]^2 \right) = \sigma_z^2 \, \frac{S_{xx}}{D} , \tag{11.32}$$

$$\mathrm{Var}[\hat{b}] = \frac{\sigma_z^2}{D^2} \sum_i \left([n x_i - S_x]^2 \right) = \sigma_z^2 \, \frac{n}{D} . \tag{11.33}$$

These are exact formulae of the estimator variances, because $\hat{a}$ and $\hat{b}$ depend linearly on y_i. We may wonder whether these errors behave like $1/\sqrt{n}$, as the error on the mean estimate of a random sample. Having denoted by $\bar{s}_x^2$ the sample variance of $(x_1, \ldots, x_n)$ with denominator n instead of $n - 1$:

$$\bar{s}_x^2 = \frac{S_{xx}}{n} - \frac{S_x^2}{n^2} , \tag{11.34}$$

with some little algebra, one finds indeed that:

$$\mathrm{Var}[\hat{a}] = \frac{\sigma_z^2}{n} \left(1 + \frac{\langle x \rangle^2}{\bar{s}_x^2} \right) , \quad \mathrm{Var}[\hat{b}] = \frac{\sigma_z^2}{n} \, \frac{1}{\bar{s}_x^2} . \tag{11.35}$$

If $(x_1, \ldots, x_n)$ behaves as a random sample, $\bar{s}_x^2$ and $\langle x \rangle^2$ almost certainly converge to constant values, and therefore the errors converge to zero as $1/\sqrt{n}$. We have already noticed several times how this behavior is common to many statistical estimators.

When σ_z^2 is unknown, Eqs. (11.35) does not provide an estimation of the uncertainties on the parameters. However, the problem can be solved using the experimental data. If we rewrite Eq. (11.18) as $Z = Y-(a+bx)$, since $\sigma_z^2 = \mathrm{Var}[Z]$, its estimate can be obtained from the sample variance of the residuals of the least squares line:

$$\hat{z}_i = y_i - (\hat{a} + \hat{b}x_i) = y_i - \hat{y}_i \ , \tag{11.36}$$

with $\langle \hat{z} \rangle = 0$ because, from Eq. (11.29), $\langle y \rangle = \langle \hat{y} \rangle$. Therefore, the sample variance of the residuals is:

$$s_z^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{z}_i^2 = \frac{1}{n-2} \sum_{i=1}^{n} \left[y_i - \hat{a} - \hat{b}x_i \right]^2 \ , \tag{11.37}$$

where the variance estimator is unbiased due to the $(n-2)$ denominator. The estimates of $\hat{a}$ and $\hat{b}$ are then:

$$s^2(\hat{a}) = \frac{s_z^2}{n} \left(1 + \frac{\langle x \rangle^2}{\bar{s}_x^2} \right) \ , \tag{11.38}$$

$$s^2(\hat{b}) = \frac{s_z^2}{n} \frac{1}{\bar{s}_x^2} \ . \tag{11.39}$$

Now we have all the elements needed to compute the confidence intervals for the parameters of the regression line, when the Y_i are assumed to be independent Gaussians with mean $(a + bx_i)$ and variance σ_z^2. Let us start from the fact that if we knew that $b = 0$, the least squares problem would coincide with that of the sample mean estimate $(y_1, \ldots, y_n)$. This has been solved in Sect. 6.11, showing that M (i.e. $\hat{a}$) and S^2/σ^2 are independent and follow a Gaussian and a reduced χ^2 (with $n-1$ degrees of freedom) distribution, respectively. So $\sqrt{n}(M-\mu)/S$ has Student's distribution t with $n-1$ degrees of freedom. A similar result still holds when we also have to estimate b, since both $\hat{a}$ and $\hat{b}$ are linear combinations of independent Gaussian variables, while $s^2(\hat{a})/\sigma^2(\hat{a})$ and $s^2(\hat{b})/\sigma^2(\hat{b})$ have a reduced χ^2 distribution with $n-2$ degrees of freedom. Therefore, both have a $(\hat{a}-a)/s(\hat{a})$ and $(\hat{b}-b)/s(\hat{b})$ Student's t-distribution with $n-2$ degrees of freedom, respectively. Using the corresponding $t_{1-\alpha/2}$ quantiles, the confidence intervals for a and b with CL$= 1-\alpha$ are then given by:

$$\hat{a} \pm t_{1-\alpha/2}\, s(\hat{a}) \ , \qquad \hat{b} \pm t_{1-\alpha/2}\, s(\hat{b}) \ . \tag{11.40}$$

Once the model is estimated, it can be used to evaluate $\langle Y|x \rangle$ at a point x also not included among the experimental points $(x_1, \ldots, x_n)$ or to predict the value we would observe for Y in correspondence of x.

To predict $\langle Y|x\rangle$ it is natural to use the estimate $\hat{y}(x) = \hat{a}+\hat{b}(x)$. Its error depends on the covariance between $\hat{a}$ and $\hat{b}$. In fact, applying Eqs. (5.65) and (11.32, 11.33), we obtain:

$$s^2(\hat{y}(x)) = \left(\frac{\partial \hat{y}}{\partial \hat{a}}\right)^2 s^2(\hat{a}) + \left(\frac{\partial \hat{y}}{\partial \hat{b}}\right)^2 s^2(\hat{b}) + 2\frac{\partial \hat{y}}{\partial \hat{a}}\,\frac{\partial \hat{y}}{\partial \hat{b}}\, s(\hat{a},\hat{b}) \,, \tag{11.41}$$

where $s(\hat{a},\hat{b})$ is the covariance estimate. To estimate this parameter, with the usual notation $\Delta X = X - \langle X\rangle$, we apply Eq. (5.83) and expand to the first order $\hat{a}$ and $\hat{b}$ in the variables y_i, to first calculate the covariance $\mathrm{Cov}[\hat{a},\hat{b}]$:

$$\mathrm{Cov}[\hat{a},\hat{b}] = \left\langle \Delta\hat{a}\,\Delta\hat{b}\right\rangle = \sum_i \frac{\partial \hat{a}}{\partial y_i}\,\frac{\partial \hat{b}}{\partial y_i}\left\langle (\Delta Y_i)^2\right\rangle = \sum_i \frac{\partial \hat{a}}{\partial y_i}\,\frac{\partial \hat{b}}{\partial y_i}\,\mathrm{Var}[Y_i] \,,$$

where the mean value is not applied to the derivatives because they are constant in the measured values y_i. Recalling Eq. (11.31) and with the notation of Eq. (11.21), one finally obtains:

$$\begin{aligned}
\mathrm{Cov}[\hat{a},\hat{b}] &= \frac{\sigma_z^2}{D^2}\sum_i (S_{xx} - S_x x_i)\,(n x_i - S_x) \\
&= \frac{\sigma_z^2}{D^2}\left[n S_{xx} S_x - n S_{xx} S_x - n S_x S_{xx} + S_x^2 S_x\right] \\
&= -\sigma_z^2\, S_x\,\frac{[n S_{xx} - S_x^2]}{D^2} = -\sigma_z^2\,\frac{S_x}{D} \,,
\end{aligned} \tag{11.42}$$

and then, using Eq. (11.34), one gets the estimate:

$$s(\hat{a},\hat{b}) = -s_z^2\,\frac{S_x}{D} = -\frac{s_z^2}{n}\,\frac{\langle x\rangle}{\bar{s}_x^2} \,.$$

Substituting this value into Eq. (11.41), and using Eqs. (11.38, 11.39), we finally obtain:

$$s^2(\hat{y}) = \frac{s_z^2}{n}\left[1 + \frac{(x - \langle x\rangle)^2}{\bar{s}_x^2}\right] . \tag{11.43}$$

Since $\hat{y}(x)$ is a linear combination of independent Gaussian variables, it is again possible to prove that

$$\frac{\hat{y}(x) - (a + bx)}{s(\hat{y}(x))}$$

is Student's variable with $n-2$ degrees of freedom. Hence, a confidence interval for $y(x) = a + bx$ with CL$= 1-\alpha$ is given by:

$$y(x) \in \hat{y}(x) \pm t_{1-\alpha/2} s(\hat{y}(x)) \,. \tag{11.44}$$

Equation (11.43) is quite instructive, because it clearly shows the risk inherent in extrapolations: the expected error increases roughly as $(x - \langle x \rangle)$ as x moves away from the "center of gravity" of the experimental points, represented by $\langle x \rangle$. This is true even assuming that outside the experimental spectrum the relationship between x and $\langle Y|x \rangle$ is still linear, which we cannot affirm or deny only based on the observed sample.

If linearity is not maintained, Eq. (11.43) is not even an appropriate estimate of variance. This remark is confirmed by evaluating the intervals (11.44) as a function of x and using $\hat{y}(x)$ as ordinate. The result is a band around the least squares line as shown in Fig. 11.5. Remember, however, that this *is not* a simultaneous confidence set with CL $= 1-\alpha$, but only the combination of different univariate intervals.

A new observation Y for a given value $X = x$ still has the expected value $a+bx$; we will therefore use $\hat{y}(x)$ as an estimator of Y. However, the variance of the random part of Y must also be taken into account in the prediction, as well as that of $\hat{y}(x)$. Therefore, the error associated with the estimate is now given by:

$$\mathrm{Var}\,[\hat{y}(x) + Z|x] = \mathrm{Var}\,[\hat{y}(x)] + \mathrm{Var}\,[Z|x] = \mathrm{Var}\,[\hat{y}(x)] + \sigma_z^2 \,, \tag{11.45}$$

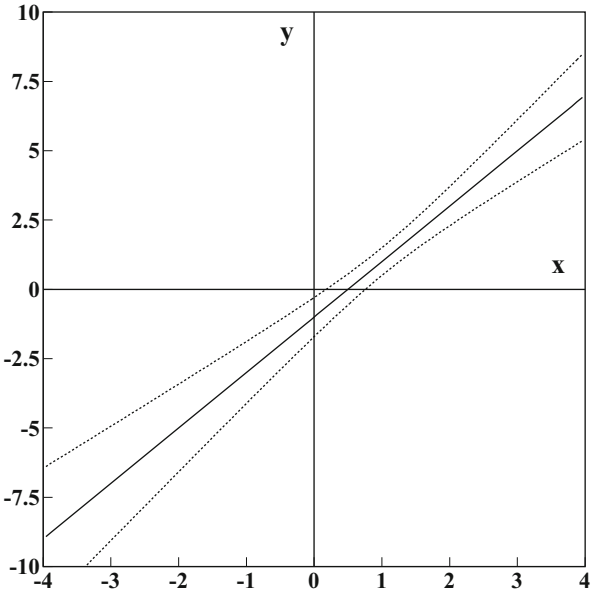


Fig. 11.5 Confidence belt $\hat{y} \pm t_{1-\alpha/2} s(\hat{y})$ of the least squares line. $s(\hat{y})$ is estimated with Eq. (11.43) for $\langle x \rangle = \langle y \rangle = 1$, $s_x = 1$ and $s/n = 0.5$

where the variances are added together because the new random part of Y is independent of the fluctuations of $\hat{a}$ and $\hat{b}$.

Substituting the estimator obtained in Eqs. (11.45) and (11.43) and replacing σ_z^2 by s_z^2, we obtain the error as:

$$s^2(\hat{a}+\hat{b}x+z) = s_z^2 + s^2(\hat{y}) = s_z^2\left[1+\frac{1}{n}\left(1+\frac{(x-\langle x\rangle)^2}{\bar{s}_x^2}\right)\right]. \tag{11.46}$$

In summary, the prediction interval of Y for a fixed x value and for a given confidence level $CL = 1-\alpha$ becomes:

$$Y(x) \in \hat{a}+\hat{b}x \pm t_{1-\alpha/2}\, s_z \sqrt{1+\frac{1}{n}\left(1+\frac{(x-\langle x\rangle)^2}{\bar{s}_x^2}\right)}, \tag{11.47}$$

where, as for the $a+bx$ estimate, $t_{1-\alpha/2}$ is Student's quantile with $n-2$ degrees of freedom.

Exercise 11.1
Consider again Exercise 6.10, and recalculate the interval estimate for the thoracic perimeter of a 170-cm-tall soldier.

Answer The solution previously found did not take into account the uncertainty due to the use of sample means, variances and correlation coefficient instead of the true ones. Now, if we recall the result of Exercise 6.10:

$$t \in m(t|s) \pm s(t|s) = 86.9 \pm 4.3 \ \text{cm},$$

and apply Eq. (11.47), we can include this uncertainty to get the correct estimate:

$$\begin{aligned} t \ &\in \ 86.9 \pm 4.3\sqrt{1+\frac{1}{1665}\left(1+\frac{(170-163.7)^2}{5.79^2}\right)} \\ &= 86.9 \pm 4.3 \cdot 1.0006 \simeq 86.9 \pm 4.3\,. \end{aligned}$$

The previous result, even after the correction, remains virtually unchanged. The confidence levels are calculated with Student's density with $(1665-2)$ degrees of freedom, which can be safely considered as Gaussian. Therefore, 68% of 170-cm-high soldiers has the thoracic perimeter between 82.6 and 91.2 cm. The confidence interval (11.44) containing the true mean (in the frequentist sense) with $CL \simeq 68\%$ for $x = 170$ cm is:

$$86.9 \pm 0.1\ cm\,.$$

11.5 Unweighted Linear Least Squares

Here we generalize the discussion of the previous section to the multidimensional case.

In the most general linear minimization problem, we have p predictors collected inside a vector $\boldsymbol{x}$. The equation corresponding to (11.18) can then be written as:

$$Y = \mu(\boldsymbol{x}, \boldsymbol{\theta}) + Z = f(\boldsymbol{x}, \boldsymbol{\theta}) + Z = \theta_0 + \sum_{j=1}^{p} \theta_j x_j + Z \,, \tag{11.48}$$

where x_j is the jth element of $\boldsymbol{x}$. As in Sect. 11.4, we set $\langle Z|\boldsymbol{x}\rangle = 0$ and $\mathrm{Var}[Z|\boldsymbol{x}] = \sigma_z^2$. Therefore:

$$\langle Y|\boldsymbol{x}\rangle = \theta_0 + \sum_{j=1}^{p} \theta_j x_j \,, \quad \mathrm{Var}[Y|\boldsymbol{x}] = \sigma_z^2 \,.$$

The components of $\boldsymbol{x}$ can be different variables or functions of the same variable, or a combination of the two, depending on the problem. For example, we could have a response which depends on a single predictor through a polynomial of degree p:

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 + \sum_{j=1}^{p} \theta_j x^j \,, \tag{11.49}$$

with $\boldsymbol{x}^\dagger = (x, x^2, \ldots, x^p)$ or, more generally:

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = \theta_0 + \sum_{j=1}^{p} \theta_j f_j(x) \,, \tag{11.50}$$

with $\boldsymbol{x}^\dagger = (f_1(x), f_2(x), \ldots, f_p(x))$. If $p = 1$ in Eq. (11.49), we are again in the particular case of the least squares line. The goal is still to get an estimate of $\boldsymbol{\theta}$ from a random sample $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$, with $(p + 1) < n$.

In a multidimensional problem, it is convenient to switch to matrix notation. To facilitate the reading, we briefly recall some matrix calculus rules considering two generic matrices A and B. Denoting with $\dagger$ the matrix transposition, that is, the exchange between rows and columns, the following properties hold:

$$(AB)^\dagger = B^\dagger A^\dagger \,, \quad (A + B)^\dagger = A^\dagger + B^\dagger \quad (A^\dagger)^\dagger = A \,, \tag{11.51}$$

if the matrix dimensions are compatible. If A is a square matrix, its inverse A^{-1} has the property $AA^{-1} = I$, where I is the identity or unit matrix. If the inverse matrix exists, A is said to have rank equal to the number of its rows or "full rank". If B is

also an invertible square matrix, the following properties hold:

$$(AB)^{-1} = B^{-1}A^{-1}\,, \quad A^{-1}A = I\,, \quad (A^{-1})^{-1} = A\,. \tag{11.52}$$

If A is a matrix with m rows and k columns ($k \leq m$) and of rank k, the $k \times k$ matrix:

$$A^\dagger A = (A^\dagger A)^\dagger \tag{11.53}$$

is a symmetric positive definite matrix. This property means that, for any vector $\boldsymbol{c}$ different from zero and with compatible dimensions, $\boldsymbol{c}^\dagger A^\dagger A \boldsymbol{c} > 0$. In addition to these formulae, it is also worthwhile to read again Sect. 4.5.

Let us now go back to our original problem. Denoting with x_{ij} the value of the jth element of $\boldsymbol{x}_i$, we define the predictor matrix as:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}\,. \tag{11.54}$$

We then denote with $\boldsymbol{y}$ the column vector of the responses $(y_1, \dots, y_n)$ and by $\boldsymbol{\theta}$ the column vector of the parameters $(\theta_0, \theta_1, \dots, \theta_p)$. The column filled with unit values in the X matrix takes into account the constant part of the model, which should always be included, unless one is sure that $\theta_0 = 0$. In this way we can rewrite Eq. (11.48) with the compact matrix equation:

$$\boldsymbol{Y} = X\,\boldsymbol{\theta} + \boldsymbol{Z}\,, \tag{11.55}$$

where $\boldsymbol{Z}$ is the column vector $(Z_1, \dots, Z_n)$ of uncorrelated random variables with zero mean and constant marginal variances $\mathrm{Var}[Z_i] = \sigma_z^2$.

The χ^2 to be minimized for the estimate of $\boldsymbol{\theta}$ can be obtained from Eq. (11.1):

$$\chi^2(\boldsymbol{\theta}) = \frac{1}{\sigma_z^2} \sum_{i=1}^{n} [y_i - \theta_0 - \sum_{j=1}^{p} \theta_j x_{ij})]^2\,, \tag{11.56}$$

where the point of minimum depends on the numerator only. This relation, written in a matrix form, becomes:

$$\sigma_z^2 \chi^2 = (X\boldsymbol{\theta} - \boldsymbol{y})^\dagger\,(X\boldsymbol{\theta} - \boldsymbol{y})\,. \tag{11.57}$$

The condition for the minimum of Eq. (11.56):

$$\frac{\partial \chi^2}{\partial \theta_k} = 0\,, \quad k = 0, \dots, p$$

leads to the so-called normal equations:

$$\sum_{i=1}^{n}\left[\left(y_i-\theta_0-\sum_{j=1}^{p}\theta_j x_{ij}\right)x_{ik}\right]=0\,,\quad k=0,1,\ldots,p\,, \tag{11.58}$$

where the equality $x_{i0}=1$ must be used when $k=0$. The minimum given by the normal equations (11.58) can be written in the compact form:

$$X^{\dagger}\,\boldsymbol{y}=(X^{\dagger}X)\,\boldsymbol{\theta}\quad\text{or}\quad\boldsymbol{\beta}=\boldsymbol{\alpha}\,\boldsymbol{\theta}\,, \tag{11.59}$$

where the matrices $\boldsymbol{\beta}=X^{\dagger}\,\boldsymbol{y}$ and $\boldsymbol{\alpha}=X^{\dagger}X$ have been introduced following a rather common notation [BR92, PFTW92]. This equation can be solved for the unknown parameters $\boldsymbol{\theta}$, by inverting the $(p+1)$ rank matrix $X^{\dagger}X$, which gives the LS parameter estimates. This result finally represents the solution of the linear unweighted least squares with multiple predictors (also known as multiple regression problem):

$$\hat{\boldsymbol{\theta}}=(X^{\dagger}X)^{-1}\,X^{\dagger}\,\boldsymbol{y}\equiv\boldsymbol{\alpha}^{-1}\,\boldsymbol{\beta}\,. \tag{11.60}$$

These fundamental equations are encoded within the R library by the `lm` function, which is used in our `Linfit` routine. When $\langle Y|x\rangle=a+bx$, it is easy to verify that normal Eq. (11.59) just corresponds to Eq. (11.24).

We now calculate the errors on these estimates using the matrix formalism and bearing in mind that Eq. (11.60) is a particular case of Eq. (5.75). Then, we can write Eq. (11.60) as:

$$\hat{\theta}_j=f_j(y_1,\ldots,y_n)\,,\quad j=0,\ldots,p\,,$$

where any function f_j is a linear combination of the elements of $\boldsymbol{y}$. We apply the variance transformation of Eq. (5.77), where in this operation the jith element of the transport matrix, given by $\partial f_j/\partial y_i$, is nothing else than the jith element of the matrix $(X^{\dagger}X)^{-1}X^{\dagger}$. Applying this result to Eq. (11.60), the covariance matrix of $\hat{\boldsymbol{\theta}}$ is obtained as:

$$\begin{aligned}V(\hat{\boldsymbol{\theta}})&=(X^{\dagger}X)^{-1}\,X^{\dagger}\,V(\boldsymbol{Y})\,X\,[(X^{\dagger}X)^{-1}]^{\dagger}\\&=(X^{\dagger}X)^{-1}\,X^{\dagger}\,\sigma_z^2 I\,X\,[(X^{\dagger}X)^{-1}]^{\dagger}\\&=\sigma_z^2\,[(X^{\dagger}X)^{-1}]^{\dagger}=\sigma_z^2\,(X^{\dagger}X)^{-1}\,,\end{aligned} \tag{11.61}$$

where Eqs. (11.51 and 11.52) have been used, the symmetry of $(X^{\dagger}X)$ has been considered together with the fact that $V(\boldsymbol{Y})=\sigma_z^2 I$ is a diagonal matrix with all the elements equal to σ_z^2. This important relation shows that *the inverse of the matrix* $\boldsymbol{\alpha}=X^{\dagger}X$ *contains all information about the errors of the parameter estimates.*

Indeed this matrix, called *error matrix*, is a square matrix of dimension $(p+1)\times(p+1)$, given by the number of free parameters, and is symmetric and positive definite. The diagonal elements are the variances of the LS estimates of the parameters $\hat{\boldsymbol{\theta}}$, and the non-diagonal elements represent the covariance between all pairs of estimates $(\hat{\theta}_i, \hat{\theta}_j)$.

An instructive verification of this statement can be done by applying Eq. (11.61) to the two-dimensional case (least squares line); in fact, we obtain:

$$V(\hat{a}, \hat{b}) = \frac{\sigma_z^2}{n S_{xx} - S_x^2} \begin{pmatrix} S_{xx} & -S_x \\ -S_x & n \end{pmatrix} ,$$

in agreement with Eqs. (11.24, 11.32, 11.33, and 11.42).

The evaluation of $\langle Y|\boldsymbol{x}\rangle$ at the generic point $\boldsymbol{x}$ is performed, as in Sect. 11.4, via the linear transformation $\hat{y}(\boldsymbol{x}) = \boldsymbol{x}^\dagger \hat{\boldsymbol{\theta}}$, where we can consider $\boldsymbol{x}^\dagger$ as a row of the X matrix or a new set of predictor values. Applying again Eqs. (5.75) and (5.77) to this new particular transformation, we immediately obtain the variance of $\hat{y}(\boldsymbol{x})$:

$$\mathrm{Var}[\hat{y}(\boldsymbol{x})] = \boldsymbol{x}^\dagger V(\hat{\boldsymbol{\theta}}) \boldsymbol{x} , \tag{11.62}$$

which is the generalization of Eq. (11.41).

The correct error estimate σ_z can be obtained analogously to Eq. (11.37):

$$s_z^2 = \frac{(X\hat{\boldsymbol{\theta}} - \boldsymbol{y})^\dagger (X\hat{\boldsymbol{\theta}} - \boldsymbol{y})}{n - p - 1} . \tag{11.63}$$

The degrees of freedom at the denominator are still given by the number of points n minus the number $(p+1)$ of estimated parameters.

If the Y_i are independent and Gaussian-distributed variables, we can obtain, for each θ_j, confidence intervals that are similar to those given by Eq. (11.40):

$$\theta_j \in \hat{\theta}_j \pm t_{1-\alpha/2}\, s(\hat{\theta}_j) , \tag{11.64}$$

where $s(\hat{\theta}_j)$ can be derived from Eq. (11.61) by replacing σ_z with s_z:

$$s(\hat{\theta}_j) = s_z \sqrt{(X^\dagger X)^{-1}_{jj}} . \tag{11.65}$$

The confidence interval of $y(\boldsymbol{x}) = \langle Y|\boldsymbol{x}\rangle$ with $CL = 1-\alpha$ assumes a form analogous to Eq. (11.44):

$$y(\boldsymbol{x}) \in \hat{y}(\boldsymbol{x}) \pm t_{1-\alpha/2}\, s_z \sqrt{\boldsymbol{x}^\dagger (X^\dagger X)^{-1} \boldsymbol{x}} , \tag{11.66}$$

where $t_{1-\alpha/2}$ is Student's quantile with $(n-p-1)$ degrees of freedom. The prediction interval for Y (similar to that of Eq. (11.47)) at a given value $\boldsymbol{x}$ is identical

to that of Eq. (11.66) after adding 1 to the term under square root:

$$Y(\boldsymbol{x}) \in \hat{y}(\boldsymbol{x}) \pm t_{1-\alpha/2}\, s_z \sqrt{1 + \boldsymbol{x}^\dagger (X^\dagger X)^{-1} \boldsymbol{x}}\ . \tag{11.67}$$

Often the fitting procedure is complicated and difficult to interpret due to the correlations between the LS estimates of the parameters, whose values change from one fit to another if we increase p in Eq. (11.48) by adding additional predictors to the ones used in the previous fit. To obtain uncorrelated estimates, the $(X^\dagger X)^{-1}$ matrix must be diagonalized using sophisticated matrix calculus techniques or orthogonalizing the X matrix to satisfy the condition:

$$\sum_i x_{ik} x_{il} = 0 \quad \text{if } k \neq l\ . \tag{11.68}$$

In this way $X^\dagger X$ becomes diagonal, and also the error matrix, which is its inverse, is diagonal. Then, the covariances are all zero and the parameters are uncorrelated. Although important in practice, we will not discuss diagonalization methods here, since they are quite laborious. Interested readers can easily find them in texts devoted to numerical computation techniques, such as [PFTW92].

Exercise 11.2
Write the normal equations for the quadratic function:

$$Y = f(X; a, b, c) = a + bX + cX^2\ .$$

Answer Using the notation of Eqs. (11.21) and (11.59) becomes:

$$\begin{pmatrix} S_1 & S_x & S_{x^2} \\ S_x & S_{x^2} & S_{x^3} \\ S_{x^2} & S_{x^3} & S_{x^4} \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \\ S_{x^2 y} \end{pmatrix}\ .$$

The comparison of these equations with Eq. (11.24) immediately suggests the generalization of the normal equations to a polynomial of any degree.

11.6 Weighted Linear Least Squares

In Sects. 11.4 and 11.5, we assumed the different observations of the response variable as uncorrelated and with constant variance, even if unknown. A first

deviation from this hypothesis is to assume non-constant variances, i.e.:

$$V \equiv V(\boldsymbol{Y}) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} , \tag{11.69}$$

that is, $\text{Var}[Y_i] = \sigma_i^2 \ \forall i$ and $\text{Cov}[Y_i, Y_j] = 0$, for $i \neq j$. More generally, the covariances may also be non-zero, with a non-diagonal V matrix.

When all the elements of V are unknown, the estimation problem would be even more difficult. Therefore, here we analyse the situations where all σ_i are known, showing also that a solution is also feasible when the ratios σ_i/σ_j are known, i.e when we can quantify how much the ith response is more (or less) variable than the jth one. In this last case, we can write the covariance matrix as $V = \sigma_z^2 W^{-1}$, where:

$$W^{-1} = \begin{pmatrix} \frac{1}{w_1} & 0 & \dots & 0 \\ 0 & \frac{1}{w_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{w_n} \end{pmatrix} , \tag{11.70}$$

and σ_z is a common error scale factor to be estimated from the data (see Eq. (11.79)).

The *weights* $(w_1, \dots, w_n)$ are all known; then $\sigma_i^2 = \sigma_z^2/w_i$ and also the ratios σ_i^2/σ_j^2 are known and equal to w_j/w_i.

When all σ_i^2 are known (absolute weights), $V = W^{-1}$ and the the weight matrix can be formally written as in the previous case with:

$$w_i = 1/\sigma_i^2 \, . \tag{11.71}$$

The χ^2 to be minimized for the multiple regression with $\boldsymbol{x}_i^\dagger = (1, x_{i1}, \dots, x_{ip})$ now becomes:

$$\begin{aligned} \chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{[y_i - \boldsymbol{x}_i^\dagger \boldsymbol{\theta}]^2}{\sigma_i^2} &= \sum_{i=1}^{n} \left[\sqrt{w_i}\ y_i - \sqrt{w_i}\ \boldsymbol{x}_i^\dagger \boldsymbol{\theta} \right]^2 \\ &= \sum_{i=1}^{n} \left[\tilde{y}_i - \tilde{\boldsymbol{x}}_i^\dagger \boldsymbol{\theta} \right]^2 \, . \end{aligned} \tag{11.72}$$

Instead, in the case of relative weights, the χ^2 to be minimized takes the form:

$$\chi^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{[y_i - \boldsymbol{x}_i^\dagger \boldsymbol{\theta}]^2}{\sigma_i^2} = \frac{1}{\sigma_z^2} \sum_{i=1}^{n} \left[\sqrt{w_i}\, y_i - \sqrt{w_i}\, \boldsymbol{x}_i^\dagger \boldsymbol{\theta} \right]^2$$
$$= \frac{1}{\sigma_z^2} \sum_{i=1}^{n} \left[\tilde{y}_i - \tilde{\boldsymbol{x}}_i^\dagger \boldsymbol{\theta} \right]^2 . \tag{11.73}$$

In both situations, we set $\tilde{y}_i = \sqrt{w_i}\, y_i$ and $\tilde{\boldsymbol{x}}_i = \sqrt{w_i}\, \boldsymbol{x}_i$. Passing to matrix notation, we denote by $W^{\frac{1}{2}}$ the diagonal matrix of the weight square roots and set $\tilde{\boldsymbol{y}} = W^{\frac{1}{2}} \boldsymbol{y}$ and $\tilde{X} = W^{\frac{1}{2}} X$. With the transformed variables, the linear model we are considering is:

$$\tilde{\boldsymbol{Y}} = \tilde{X} \boldsymbol{\theta} + \boldsymbol{Z} , \tag{11.74}$$

where $\boldsymbol{Z}$ is the same of model (11.55). It immediately turns out that minimizing the χ^2 of Eq. (11.72) is equivalent to minimize:

$$\chi^2 = (\tilde{X} \boldsymbol{\theta} - \tilde{\boldsymbol{y}})^\dagger \, (\tilde{X} \boldsymbol{\theta} - \tilde{\boldsymbol{y}}) . \tag{11.75}$$

This χ^2 has the same form given in Eq. (11.57), so that, taking into account that $W^{\frac{1}{2}} W^{\frac{1}{2}} = W$, the symmetry of W and Eqs. (11.51), we have:

$$\hat{\boldsymbol{\theta}} = (\tilde{X}^\dagger \tilde{X})^{-1} \, \tilde{X}^\dagger \, \tilde{\boldsymbol{y}} = (X^\dagger \, W \, X)^{-1} \, X^\dagger \; W \boldsymbol{y} \equiv \boldsymbol{\alpha}^{-1} \, \boldsymbol{\beta} , \tag{11.76}$$

where the matrices defined in Eq. (11.59) now become $\boldsymbol{\alpha} = (X^\dagger \, W \, X)$ and $\boldsymbol{\beta} = X^\dagger \; W \boldsymbol{y}$.

The errors of the estimates are easily evaluated from Eq. (11.61):

$$V(\hat{\boldsymbol{\theta}}) = (\tilde{X}^\dagger \tilde{X})^{-1} = (X^\dagger \, W \, X)^{-1} \;\; \text{absolute weights;} \tag{11.77}$$
$$V(\hat{\boldsymbol{\theta}}) = \sigma_z^2 \, (\tilde{X}^\dagger \tilde{X})^{-1} = \sigma_z^2 \, (X^\dagger \, W \, X)^{-1} \;\; \text{relative weights} . \tag{11.78}$$

In this last case, recalling Eq. (11.63), the estimate of σ_z^2 becomes:

$$s_z^2 = \frac{(\tilde{X} \, \hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{y}})^\dagger \, (\tilde{X} \, \hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{y}})}{n - p - 1} = \frac{(X \, \hat{\boldsymbol{\theta}} - \boldsymbol{y})^\dagger \, W \, (X \, \hat{\boldsymbol{\theta}} - \boldsymbol{y})}{n - p - 1} . \tag{11.79}$$

If the responses Y_i are independent Gaussians, the confidence intervals for the θ_j can be easily evaluated by applying Eq. (11.64), with the error of Eq. (11.65) replaced by:

$$\sqrt{(X^\dagger \, W \, X)^{-1}_{jj}} = \sqrt{\boldsymbol{\alpha}^{-1}_{jj}} \quad \text{absolute weights} ; \tag{11.80}$$

$$s_z\sqrt{(X^\dagger\, W\, X)^{-1}_{jj}} = s_z\sqrt{\boldsymbol{\alpha}^{-1}_{jj}} \quad \text{relative weights}\,. \tag{11.81}$$

All the output results of the `Linfit` routine are obtained with the weighted least squares Eqs. (11.72–11.79).

The "prediction" $\tilde{y}(\boldsymbol{x})$ for a given $\boldsymbol{x}$ value is generally not of much interest; therefore it is not worth applying Eq. (11.66) to $\tilde{y}$ directly. We just remark that $\hat{y}(\boldsymbol{x}) = \boldsymbol{x}^\dagger\,\hat{\boldsymbol{\theta}}$ still holds for the untransformed response variable and thus the variance estimate of $\hat{y}(\boldsymbol{x})$, in the case of absolute weights, is $\boldsymbol{x}^\dagger(X^\dagger W X)^{-1}\boldsymbol{x}$, as in Eq. (11.62). With relative weights we have instead $(s_z^2)\boldsymbol{x}^\dagger(X^\dagger W X)^{-1}\boldsymbol{x}$. The confidence interval with $CL = 1-\alpha$ is then as that of Eq. (11.66):

$$y(\boldsymbol{x}) \in \hat{y}(\boldsymbol{x}) \pm t_{1-\alpha/2}\,(s_z)\,\sqrt{\boldsymbol{x}^\dagger(X^\dagger W X)^{-1}\boldsymbol{x}}\,, \tag{11.82}$$

where the term (s_z) is included only in the case of relative weights. The prediction interval for $Y(\boldsymbol{x})$ is obtained from this equation by adding to the estimate of $\hat{y}(\boldsymbol{x})$ that of the random part of the model for Y for a given $\boldsymbol{x}$:

$$Y(\boldsymbol{x}) = \boldsymbol{x}^\dagger\boldsymbol{\theta} + \frac{z}{\sqrt{w}}\,,$$

that is, $\mathrm{Var}[z/\sqrt{w}] = 1/w$ for absolute weights and $\mathrm{Var}[z/\sqrt{w}] = \sigma_z^2/w$ or its estimate s_z^2/w for relative ones. Finally, as in Eq. (11.67):

$$Y(\boldsymbol{x}) \in \hat{y}(\boldsymbol{x}) \pm t_{1-\alpha/2}\,(s_z)\,\sqrt{\frac{1}{w} + \boldsymbol{x}^\dagger(X^\dagger W X)^{-1}\boldsymbol{x}}\,, \tag{11.83}$$

where again the term within brackets (s_z) must be included only for relative weights.

The previous discussion is valid if W is a diagonal matrix. To be more general, we examine the transformation $\tilde{\boldsymbol{y}} = W^{\frac{1}{2}}\boldsymbol{y}$ together with the obvious factorization of V, that is, $V = \sigma_z^2 W^{-1} = \sigma_z^2 W^{-\frac{1}{2}}\, W^{-\frac{1}{2}}$. If V is non-diagonal and we set $V = \sigma_z^2 W^{-1}$, where W is a known matrix, we can obtain a similar factorization by applying Eq. (4.68): $W^{-1} = H\,H^\dagger$. Here H plays the same role of $W^{-\frac{1}{2}}$. Based on the results of Sect. 4.5, we realize that the transformation $\tilde{\boldsymbol{Y}} = H^{-1}\boldsymbol{Y}$ makes $\tilde{\boldsymbol{Y}}$ a Gaussian vector with covariance matrix equal to $\sigma_z^2 I$ and vector of the means $\tilde{X}\boldsymbol{\theta} = H^{-1}X\boldsymbol{\theta}$, bringing us back to the hypotheses used in the previous paragraph to obtain all the confidence intervals. Therefore the results from Eqs. (11.75) to (11.82) hold, without modifications, for any covariance matrix $\sigma_z^2 W^{-1}$. Also Eq. (11.83) continues to hold by replacing $1/w$ with the ith element on the diagonal of W^{-1}, if we consider an experimental point $\boldsymbol{x}_i$, or with a new coefficient if not.

11.7 Properties of Least Squares Estimates

We now demonstrate the three fundamental theorems (including that of Gauss-Markov) which are the basis of the linear least squares estimation. If you do not appreciate mathematics, you can only read the theorems statements (and their consequences!) and move on to the next section.

Theorem 11.1 (On Correct Estimates) *In the case of linear dependence on the parameters, the least squares (LS) estimates are unbiased.*

Proof Applying the operator of the mean to Eq. (11.60) and recalling Eq. (11.55), one immediately gets:

$$\left\langle \hat{\boldsymbol{\theta}} \right\rangle = (X^\dagger X)^{-1} X^\dagger \, \langle \boldsymbol{Y} \rangle = (X^\dagger X)^{-1} X^\dagger X \, \boldsymbol{\theta} = \boldsymbol{\theta} \ ,$$

according to Eq. (10.14). □

Theorem 11.2 (Gauss-Markov) *With reference to the model of Eq. (11.55), the LS estimator has minimal variance (i.e. is the most efficient) among all the unbiased and linear estimators of* $\boldsymbol{\theta}$.

Proof We must show that, if $\boldsymbol{\theta}^*$ is an estimated unbiased parameter of a linear model $\boldsymbol{Y}$, one has:

$$a^\dagger V(\hat{\boldsymbol{\theta}}) a \leq a^\dagger V(\boldsymbol{\theta}^*) a \ , \tag{11.84}$$

where a is any vector of dimension $(p+1)$. In particular, if a contains all zeros and value 1 in the ith position, Eq. (11.84) includes also the cases:

$$V(\hat{\theta}_{ii}) \leq V(\theta^*_{ii}) \ , \quad i = 0, 1, \ldots, p \ . \tag{11.85}$$

We therefore consider a generic unbiased estimate of the parameters $\boldsymbol{\theta}$ which is linear in $\boldsymbol{Y}$:

$$\boldsymbol{\theta}^* = U \boldsymbol{Y} \ .$$

For the least squares estimators, $U = (X^\dagger X)^{-1} X^\dagger$. Since an unbiased estimate has been assumed, from Eq. (11.55) one has:

$$\langle \boldsymbol{\theta}^* \rangle = U \, \langle \boldsymbol{Y} \rangle = U X \boldsymbol{\theta} = \boldsymbol{\theta} \ ,$$

and hence:

$$UX = I \ , \quad (UX)^\dagger = I \ . \tag{11.86}$$

It is crucial to note that this property does not imply $U = X^{-1}$, because U and X are not square matrices. Based on Eq. (11.61), we have:

$$V(\boldsymbol{\theta}^*) = \sigma_z^2 U\, U^\dagger \,.$$

The following identity is also valid:

$$U\, U^\dagger = C + (U - CX^\dagger)\,(U - CX^\dagger)^\dagger \,, \tag{11.87}$$

where C is the LS error matrix divided by σ_z^2:

$$C = (X^\dagger X)^{-1} \,.$$

This relation can be easily verified by developing the right term of the previous equation using Eq. (11.86) and because C, V and their inverse are symmetric matrices coincident with their transpose:

$$\begin{aligned} &C + UU^\dagger - CX^\dagger U^\dagger - UXC^\dagger + CX^\dagger XC^\dagger \\ &= C + UU^\dagger - CX^\dagger U^\dagger - UXC^\dagger + C(X^\dagger X)C \\ &= C + UU^\dagger - C - C + C(X^\dagger X)C \\ &= C + UU^\dagger - C - C + C = UU^\dagger \,. \end{aligned}$$

From Eq. (11.61) we can write Eq. (11.87) as:

$$V(\boldsymbol{\theta}^*) = V(\hat{\boldsymbol{\theta}}) + \sigma_z^2 (U - CX^\dagger)(U - CX^\dagger)^\dagger \,,$$

which shows that the covariance matrix $\boldsymbol{\theta}^*$ is equal to the covariance matrix of the LS estimates $\hat{\boldsymbol{\theta}}$ plus a symmetric positive definite matrix written in the form of $HH^\dagger$, as in Eq. (4.68). This proves the theorem.

Clearly, the equality $V(\boldsymbol{\theta}^*) = V(\hat{\boldsymbol{\theta}})$ occurs when $U = CX^\dagger = (X^\dagger X)^{-1} X^\dagger$, as it is easy to verify.

It is important to note that this theorem does not imply any assumptions about the population distributions or about the size of the sample $\boldsymbol{y}$. The only requirement is that the average values $\langle \boldsymbol{Y} \rangle$ are linear functions of the parameters. □

For the weighted least squares of Sect. 11.6, the theorems just proved continue to hold by replacing X with $\tilde{X}$ and $\boldsymbol{Y}$ with $\tilde{\boldsymbol{Y}}$.

We have seen that, in the case of Gaussian variables and linear models, the LS estimates also provide Gaussian intervals for the parameters $\boldsymbol{\theta}$ for $\langle Y|\boldsymbol{x} \rangle$ and prediction intervals for Y itself. Moreover, it is possible to perform a χ^2 test of the fitted model using the statistic $\chi^2(\hat{\boldsymbol{\theta}})$ of Eq. (11.2).

The $\chi^2(\hat{\boldsymbol{\theta}})$ variate has a $\chi^2(n-p-1)$ distribution, and the model fit is usually considered unsatisfactory at a level α if:

$$SL = P\left\{Q_R \geq \frac{\chi^2(\hat{\boldsymbol{\theta}})}{n-p-1}\right\} \leq \alpha \,, \tag{11.88}$$

where Q_R follows the reduced χ^2 distribution with $(n-p-1)$ degrees of freedom.

Less used, but safer, is the two-tailed test of Eq. (7.34). Following the discussion in Chap. 6, we know that this test also protects against the use of models with too many parameters, which tend to interpolate the data, and then produces too small χ^2 values (overparametrization).

Under suitable conditions an important theorem [SW89] allows us to extend these properties even to nonlinear models and non-Gaussian variables.

The following discussion applies to χ^2 functions that follow Definition 11.1, hence excluding the effective variance case. The models considered in the theorem are in fact of the type:

$$Y_i = \mu_i(\boldsymbol{x}_i, \boldsymbol{\theta}) + Y_{Ri}, \quad i = 1, \ldots, n \,, \tag{11.89}$$

where $\boldsymbol{x}_i$ are fixed and Y_{Ri} are iid variables with null mean and known variances σ_i^2. Definition 11.1 follows Eq. (11.89), because the assumption $\langle Y_i | \boldsymbol{x}_i \rangle = \mu_i(\boldsymbol{\theta})$ coincides with Eq. (11.89) and includes the cases of both the non-random and random predictors conditional on $\boldsymbol{x}$. We assume that $\mu_i(\boldsymbol{\theta}) = f(\boldsymbol{x}_i, \boldsymbol{\theta})$ for all i and use the notations:

$$\boldsymbol{\mu}(\boldsymbol{\theta}) = (\mu_1(\boldsymbol{\theta}), \ldots, \mu_n(\boldsymbol{\theta}))^\dagger \,,$$

$$F_{ij} = \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \theta_j} \,,$$

where F_{ij} are the elements of an F matrix of dimension $n \times (p+1)$ that generalizes Eq. (11.54) introduced in the linear LS case.

Theorem 11.3 (Least Squares Estimates) *Given the model (11.89) with iid variables $Y_{R1}, \ldots, Y_{Rn}$ with zero mean and variances $\sigma_1^2, \ldots, \sigma_n^2$, one approximately has:*

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \sim N_p(\mathbf{0}, \Sigma^{-1}) \,, \quad \Sigma = F^\dagger W F \,, \tag{11.90}$$

where W is a diagonal matrix with $1/\sigma_i^2$ at the position (i,i) and $\hat{\boldsymbol{\theta}}$ is the LS estimator (see [SW89], Sect. 12.2).

If, in addition, Y_{Ri} are Gaussian, $\chi^2(\hat{\boldsymbol{\theta}})$ is the variate of a random variable asymptotically distributed as $\chi^2(\nu)$, where $\nu = (n-p-1)$, with $(p+1)$ corresponding to the $\boldsymbol{\theta}$ dimension (see [SW89], Theorem 2.1).

Table 11.1 Properties of the LS estimators *in the case of known errors*. The symbol (∗) indicates that the property is valid under the conditions of Theorem 11.3. The efficiency refers to the Cramér-Rao lower bound and is always maximal in the first (with finite sample size) and in the second row, where the estimates are from the ML method, while in the third row, it is instead limited, due to the Gauss-Markov theorem, to the correct and linear estimators

Problem type			Properties	
Gaussian data?	Linear model?	Efficient estimator?	Gaussian estimates?	χ^2 test possible?
YES	YES	YES	YES	YES
YES	NO	YES	YES (∗)	YES (∗)
NO	YES	YES	YES (∗)	NO
NO	NO	UNCERTAIN	YES (∗)	NO

Then in practice we can apply Eqs. (11.63–11.67) to the nonlinear case by replacing everywhere $\boldsymbol{x}^\dagger \hat{\boldsymbol{\theta}}$ with $f(\boldsymbol{x}, \hat{\boldsymbol{\theta}})$ and every row $\boldsymbol{x}_i^\dagger$ of X with the vector of derivatives:

$$\left.\frac{\partial f(\boldsymbol{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\dagger}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

when X is not multiplied by $\boldsymbol{\theta}$.

Theorem 11.3 partially justifies the widespread habit of applying the 3σ law to the estimation intervals provided by nonlinear LS algorithms and of performing the χ^2 test to check the best-fit solution (see also Table 11.1). However, in important cases when any doubts arise, we advise you to simulate the LS procedure with artificial data to directly verify the distribution of the estimated parameters and of the χ^2 values using the methods of Sect. 8.10.

11.8 Model Testing and Search for Functional Forms

The results we have presented so far are valid if the model assumed for the expected response value is correct.

Indeed, in these situations, the functional form of the model is known, and hence the final χ^2 value can be used to readjust the errors of the experimental points and of the obtained estimates. In this case, it obviously makes no sense to perform the χ^2 test.

If, on the other hand, one is not sure of the functional form, and, for example, a model selection is performed by adding or removing parameters from polynomials or other empirical functions, the model validity can be judged with the χ^2 test only if the absolute errors are known. If the assumed model is correct and the errors are

Gaussian, we get:

$$\sum_{i=1}^{n} \frac{[Y_i - \boldsymbol{x}_i^{\dagger} \hat{\boldsymbol{\theta}}]^2}{\sigma_i^2} \sim \chi^2(n - p - 1) \,,$$

and the fit quality can be controlled with the two-tailed test described in the previous section.

As we will show below, in addition to the χ^2 test, it is also possible to perform the F test and/or the residual trend analysis. The latter two methods can also be used when only the relative error weights are known, as they are not affected by the value of a common scale factor. In the general linear model problem, when the functional form is not known a priori, one starts by estimating a given hypothetical function belonging to the class defined by Eq. (11.48). Consider, for example, a useful subclass of regression models where each of the p predictors is a function of the same variable x (see also Eqs. (11.49 and 11.50)):

$$f(x, \boldsymbol{\theta}) = \theta_0 + \sum_{k=1}^{p} \theta_k f_k(x) \,, \quad f(x, \boldsymbol{\theta}) = \theta_0 + \sum_{k=1}^{p} \theta_k x^k \,. \tag{11.91}$$

As usual, this function has to be fitted to n experimental points. In the following it is necessary to pay attention to distinguish the cases where the absolute error is known or not. The latter situation corresponds to have an unknown σ_z value in Eq. (11.73).

Recall that the problem consists in finding the curve around which *the fluctuations of the points are random*, that is, the curve representing the functional dependence between $\boldsymbol{x}$ and the average value of Y. This curve *must not pass exactly through points* (interpolation) and therefore must have fewer parameters than the number of experimental points; for this reason it is also called regression curve (although historically this term was introduced in a different context).

The initial choice of a particular function within the subclass (11.91) is made on the basis of available information on the problem and, whenever possible, of the graphic patterns of the (x_i, y_i) pairs, as in Figs.11.1 and 11.2.

The first test to be performed after the χ^2 minimization is to check that the residuals:

$$\hat{z}_i = y_i - \hat{y}(\boldsymbol{x}_i) = y_i - \hat{y}_i \tag{11.92}$$

behave just like random fluctuations since they are, in fact, an estimate of the random fluctuations z_i. A plot of the $(\hat{y}_i, \hat{z}_i)$ pairs is very informative in this sense, since the residuals always have zero mean. It can be shown that, if the model includes the constant term θ_0, the sample correlation between the predicted values $\hat{y}_i$ and the residuals $\hat{z}_i$ vanishes and there is no linear relation between them.

If the assumptions made on $f(x, \boldsymbol{\theta})$ and on the mean (zero) and variance (constant) of the random fluctuations z_i are compatible with the data, we should

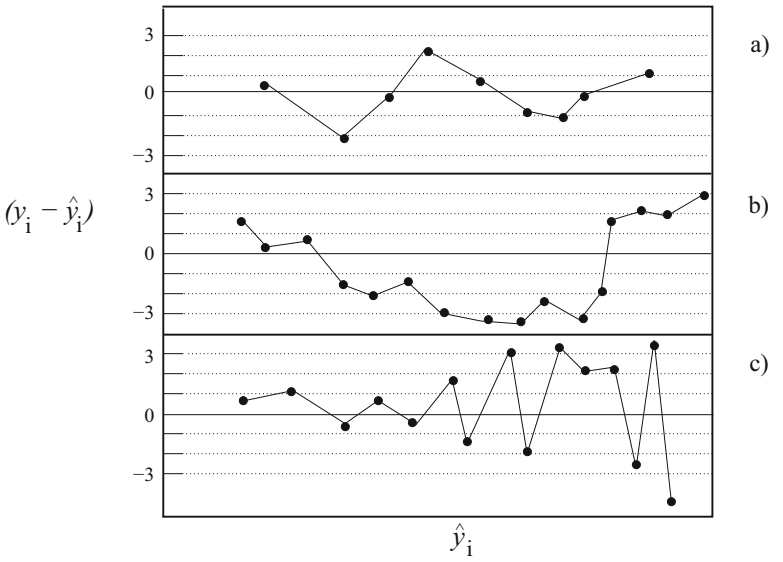


Fig. 11.6 Typical residual graphs after a best fit: good fit (**a**), wrong functional form (**b**) and heteroskedasticity (see text) (**c**)

therefore observe in the plot a cloud of points enclosed within a zero-centred band of constant width.[2]

In particular, if the residuals are also Gaussian, the band will be roughly symmetrical around zero. In the weighed case, the residuals for each predictor $\boldsymbol{x}_i$ must be standardized (i.e. divided by the error):

$$t(x_i) = \frac{\hat{z}_i}{\sigma_i} = \frac{y_i - \hat{y}(\boldsymbol{x}_i)}{\sigma_i} \,. \tag{11.93}$$

The trend must appear random, and the values outside the band $|t_i| < 3$ must be rare, in agreement with the 3σ law (note that this is a rough check, since the variance of $\hat{Z}_i$ is the one given in Footnote 2, with $\tilde{X}$ instead of X). If σ_z^2 is unknown, the residuals must still be weighed, and in the graph, the pairs $(\hat{y}(\boldsymbol{x}_i), \sqrt{w_i}\hat{z}_i)$ are represented. Figure 11.6 shows three possible cases. In case (a) the fit is satisfactory: in case (b) the points show a trend violating the 3σ law, due to a fit with an inadequate functional form; and case (c) instead occurs when the errors of the higher values of y have been underestimated. Frequently this happens when a non-weighted fit is performed (with absolute errors kept constant) to data that instead have a constant

[2] Note that $\mathrm{Var}[\hat{Z}_i] = \sigma_z^2(1 - l_{ii})$, where $0 \le l_{ii} \le 1$ is the element of place ii of the matrix $X(X^\dagger X)^{-1}X^\dagger$. It may happen that, for certain matrices X, some l_{ii} are significantly different from most of the others.

relative error. The absolute error then increases with y, and the residual graph is as in Fig. 11.6c; statisticians name this behaviour as *heteroskedasticity*.

Several models can produce a visually correct residual plot. In this case, iterative methods are used, obtained by adding or removing predictors and by comparing pairs of consecutive models from this sequence. In the case of the polynomial of Eq. (11.91), we could increase its degree by one unit at a time, minimize the χ^2 and decide when to stop. The method is based on the fit of the models and on the answer to the question: "when specific predictors $f_k(x)$ are removed, is the worsening of the fit statistically significant or not?" To respond, let us start by examining Fig. 11.7, which shows the vectors involved in our least squares problem in the $\mathbb{R}^n$ space. To understand its meaning, we note that the estimate of the response for each row of X, that is, the vector of the quantities $\hat{y}(\boldsymbol{x}_i)$, corresponds to:

$$\hat{\boldsymbol{y}} = X\,\hat{\boldsymbol{\theta}} = X(X^\dagger X)^{-1}X^\dagger \boldsymbol{y} = P\boldsymbol{y}\ . \tag{11.94}$$

It can be demonstrated that the P matrix simply implements orthogonal projection of $\boldsymbol{y}$ on the vector space generated by the columns of X, i.e. $\hat{\boldsymbol{\theta}}$ is the parameter vector defining the linear combination of the columns of X that is closest to $\boldsymbol{y}$. Let us now consider the X_0 matrix, obtained from X by removing a certain number of columns and keeping only p_0 predictors (hence X_0 has dimension $n \times (p_0 + 1)$). The estimate of the response with this reduced model will of course be $\hat{\boldsymbol{y}}_0 = P_0\boldsymbol{y}$, with $P_0 = X_0(X_0^\dagger X_0)^{-1}X_0^\dagger$, i.e. the projection matrix on the column space of X_0. This is equivalent to setting $p - p_0$ parameters inside the vector $\boldsymbol{\theta}$ to zero.

By applying the Pythagorean theorem to the dashed right-angled triangle of Fig. 11.7, we obtain an important relation involving the regression residuals of both the full and the reduced models. Recalling that $\sum_i \hat{z}_i^2 = \|\hat{\boldsymbol{z}}\|^2 = \sum_i (y_i - \hat{y}(x_i))^2$

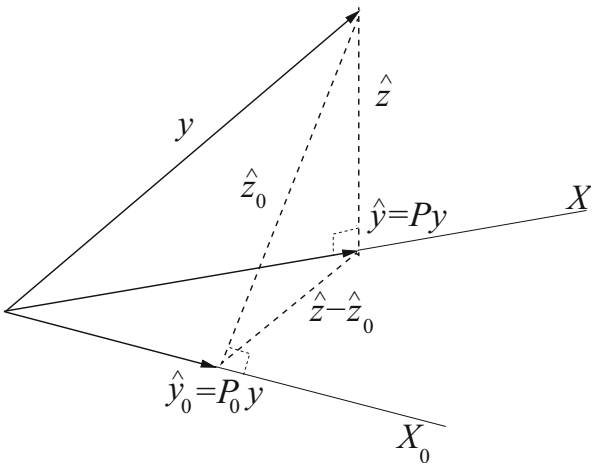


Fig. 11.7 Vector representation in the $\mathbb{R}^n$ space. Regression with all predictors (X matrix) and with a subset of predictors (X_0 matrix). The response estimate $\hat{\boldsymbol{y}}(X) \equiv \hat{\boldsymbol{y}}$ at each row of X is the orthogonal projection of $\boldsymbol{y}$ on the linear space generated from X columns. A similar argument applies to $\hat{\boldsymbol{y}}_0$. It is therefore immediate to derive the vectors of the residuals of the two regressions and the relation between them

and $\|\hat{z}_0\|^2 = \sum_i (y_i - \hat{y}_0(x_i))^2$, we have:

$$\|\hat{z}_0\|^2 = \|\hat{z} - \hat{z}_0\|^2 + \|\hat{z}\|^2 \,. \tag{11.95}$$

If $y(x)$ is given by the first of Eq. (11.91) and the functions $f_k(x)$ satisfy the orthogonality property (11.68), by explicitly writing the residuals, it can easily be shown that Eq. (11.95) is equivalent to the condition:

$$\sum_i [y_i - \hat{y}(x_i)][\hat{y}(x_i) - \hat{y}_0(x_i)] = \sum_{k=p_0+1}^{p} \sum_i [y_i - \hat{y}(x_i)]\, \hat{\theta}_k f_k(x_i) = 0 \,, \tag{11.96}$$

which can be verified with the normal equations (11.58). From Eq. (11.36) we can easily deduce that the elements of the vectors of residuals are linear combinations of the responses y_i. If the response is Gaussian, the two norms at the second member are relative to orthogonal (and then uncorrelated) vectors, which are also independent being a linear combination of Gaussian vectors. Moreover, if the reduced model with p_0 predictors holds, from Cochran's Theorem 4.5, we have:

$$\frac{1}{\sigma_z^2}\|\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_0\|^2 \sim \chi^2(p - p_0) \,, \quad \frac{1}{\sigma_z^2}\|\hat{\mathbf{Z}}\|^2 \sim \chi^2(n - p - 1) \,. \tag{11.97}$$

These results on the independence and distribution of squared norms give us the answer to the question we asked ourselves a little while ago since, *if the reduced model is valid*, then:

$$\frac{\|\hat{z} - \hat{z}_0\|^2/(p - p_0)}{\|\hat{z}\|^2/(n - p - 1)} = \frac{\left(\|\hat{z}_0\|^2 - \|\hat{z}\|^2\right)/(p - p_0)}{\|\hat{z}\|^2/(n - p - 1)} \tag{11.98}$$

follows the Snedecor F density with $(p - p_0, n - p - 1)$ degrees of freedom. This density has been defined by Eq. (5.46) as the distribution of the ratio between two independent reduced χ^2 variables and has been used in Exercise 7.11. Often Eq. (11.98) is written with a different notation:

$$F = \frac{(RSS_0 - RSS)/(p - p_0)}{RSS/(n - p - 1)} \,, \tag{11.99}$$

with $RSS_0 = \|\hat{z_0}\|^2$ and where $RSS = \|\hat{z}\|^2$ is the acronym of *residual sum of squares*. The statistics (11.99) quantitatively measures the worsening of the reduced model fit with respect to the full model by means of the RSS increase.

We note that the F test, being a ratio between χ^2 variables, can be performed also when only the relative errors are known. A significantly large value of F indicates that the predictors removed from the complete model were important. For this reason we reject at the α level the hypothesis that the corresponding coefficients

are null if:

$$F > F_{1-\alpha}(p - p_0, n - p - 1) \; . \tag{11.100}$$

A particular case is when the reduced model is obtained by removing only a single predictor, for instance, by setting $\theta_p = 0$ into one of Eq. (11.91), so that $p_0 = p-1$. The $\boldsymbol{F}$ variable, under $H_0 : \theta_p = 0$, has an $F(1, n - p - 1)$ distribution, and it is easy to recognize that the equality $F_{1-\alpha}(1, n - p - 1) = t^2_{1-\alpha/2}(n - p - 1)$ holds, where t is Student's t percentile with $(n - p - 1)$ degrees of freedom. Indeed, F is also the square of:

$$T = \frac{|\hat{\theta}_p|}{s(\hat{\theta}_p)} \; , \tag{11.101}$$

where $s(\hat{\theta}_p)$ is given by Eq. (11.63) and T has Student's t-distribution with $(n - p - 1)$ degrees of freedom.

All this can be verified by reading the discussion about the confidence intervals for the least squares line parameters of Eq. (11.40), where $p = 1, a = \theta_0$ and $b = \theta_1$. Having this in mind, we will reject H_0 and accept the new $(p_0 + 1)$th parameter if T exceeds $t_{1-\alpha/2}(n - p - 1)$. In other words, a large value of T indicates a small relative error and therefore the importance of the parameter. Recalling the problem of choosing the degree of the polynomial in Eq. (11.91), we then could, for example, start from the model with only θ_0 and verify the hypothesis $H_0 : \theta_1 = 0$. If this is rejected, we will add the second degree term and test the hypothesis $H_0 : \theta_2 = 0$ and so on. If k is the index of the first test that does not reject H_0, we will set $p = k - 1$. This is just one of the possible procedures. If by chance the second degree term is not significant, but the third degree term is significant, with this procedure we would not notice it.

It is important to point out that all what we have described so far are necessary but not sufficient conditions to ensure that the functional form has been correctly found. In other words, there are many functional forms (not just the "true"one) that satisfy the best fit criteria listed above. This is a basic ambiguity, *not resolvable only statistically*, which one must always be aware of. To reduce this uncertainty, it is therefore essential to introduce, inside the model, all the a priori known information about the problem and try to reduce errors as much as possible.

We exemplify these concepts by simulating a realistic case. Starting from the polynomial:

$$y = \theta_0 + \theta_2 x^2 + \theta_3 x^3 = 3 + 5x^2 - 0.5x^3 \; , \tag{11.102}$$

and applying the techniques of Chap. 8, we generated simulated (artificial) Gaussian data having a rather large relative standard deviation of $\pm 15\%$:

$$y_i = (1 + 0.15\, g_i)(3 + 5x_i^2 - 0.5x_i^3) \; ,$$

Table 11.2 Results of the best fits to the data of Eq. (11.103) with polynomials of the type $y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$ assuming known relative errors. The estimate s_z is given by Eq. (11.79), whereas RSS is the weighted sum of squared residuals, that is, the numerator of Eq. (11.79)

	FIT1	FIT2	FIT3	FIT4	FIT5
θ_0	-5.1 ± 2.7	-15.1 ± 3.0	-6.1 ± 5.1		1.8 ± 1.1
θ_1	12.8 ± 1.4	23.2 ± 2.8	10.6 ± 6.7	2.7 ± 1.4	
θ_2		-1.5 ± 0.4	2.8 ± 2.1	5.1 ± 0.8	6.0 ± 0.5
θ_3			-0.4 ± 0.2	-0.6 ± 0.1	-0.6 ± 0.1
RSS	0.380	0.097	0.049	0.066	0.079
s_z	0.252	0.139	0.111	0.115	0.126

where g_i is a standard normal variate. The data are reported in Eq. (11.103):

$$\begin{array}{lllllllll} x & 1.0 & 2.0 & 3.0 & 4.0 & 5.0 & 6.0 & 7.0 & 8.0 \\ y & 6.9 & 22.4 & 40.8 & 64.4 & 60.0 & 81.0 & 78.0 & 70.5 \\ \sigma & 1.1 & 2.9 & 5.2 & 7.7 & 9.8 & 11.3 & 11.5 & 10.1 \end{array} , \tag{11.103}$$

where $\sigma = \sigma_z/\sqrt{w} = 0.15/\sqrt{w}$ and $1/\sqrt{w} = 3 + 5x^2 - 0.5x^3$. The different coefficients of the different regression polynomials obtained from the fits of these data are shown in Table 11.2, together with the values of RSS. They have been calculated with our code `FitPolin` which uses the `Linfit` routine. By storing the vectors x, y of Eq. (11.103) and the weights w in `x, y, w`, the instructions for the polynomial FIT3 of Table 11.2 are the following:[3]

```
>class(fitfun<- y~x+I(x^2)+I(x^3))
>FitPolin(x=x,y=y,dy=1/sqrt(w),fitfun=fitfun,ww='REL')
```

With the options `ww='REL'` and `dy=1/sqrt(w)`, we assume for the moment that only the relative weights w are known, while the overall scale factor σ_z is not. Therefore the confidence intervals of the parameters are given by Eq. (11.64) with the error given by Eq. (11.81). However, it should be recalled that the errors on the parameters and the estimate s_z of σ_z are reliable only if we use a functional form of the model not far from the true model, so that $s_z\sqrt{(X^\dagger W X)^{-1}_{jj}} \simeq \sigma_z\sqrt{(X^\dagger W X)^{-1}_{jj}}$. Moreover, as we will briefly explain in Sect. 11.10, the χ^2 test has not to be performed. We can only compare the fit results between two models using the F test of Eq. (11.99), where the weighted sum of squares of the residuals must be used: $RSS = (\boldsymbol{y} - X\hat{\boldsymbol{\theta}})^\dagger W(\boldsymbol{y} - X\hat{\boldsymbol{\theta}})$. This comparison is also useful to discard over-parameterized models.

In our example, the fit with the straight-line model:

$$y(x) = \theta_0 + \theta_1 x \ ,$$

[3] The use of the objects of the `formula` class, as `fitfun`, is described in the R on line manual.

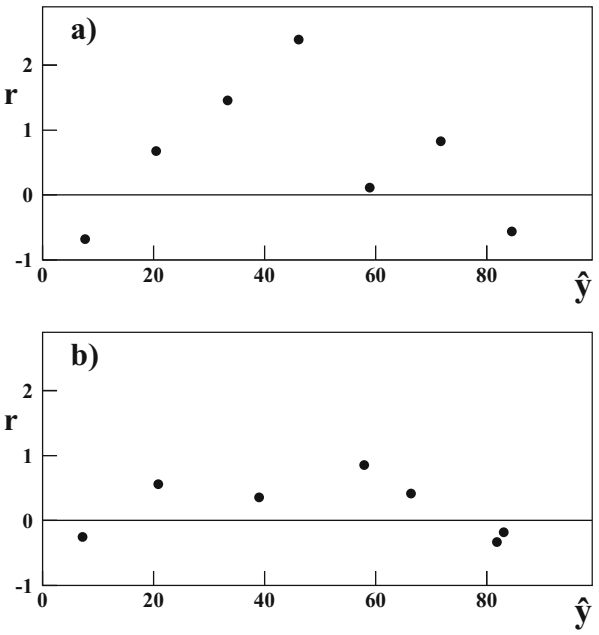


Fig. 11.8 Residuals (11.92) of the regression for (**a**) FIT1 and (**b**) FIT5. Compare these values with those shown in Fig. 11.6

denoted as FIT1 in Table 11.2, fails the check with the residual plot: indeed, Fig. 11.8 has a behaviour similar to that of Fig. 11.6b.

We then consider a quadratic polynomial:

$$y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 \,,$$

obtaining the FIT2 result. The F test ratio (11.98) between FIT2 and FIT1 gives:

$$F(1,5) = \frac{0.380 - 0.097}{0.097} 5 = 14.6 \,.$$

From Table E.5 at 5% and 1% levels, one obtains:

$$14.6 > F_{0.95}(1,5) = 6.61 \,, \quad 14.6 < F_{0.99}(1,5) = 16.3 \,.$$

The test shows that the null hypothesis (i.e. the uselessness of FIT2) can be rejected at the 5% level, but not at the 1% level. If we consider that the solution is very close to the 1% level, it is legitimate to consider the three-parameter solution of FIT2 to be significant.

We could stop at this point, but let us see what happens with FIT3, in which a cubic polynomial with four parameters is used:

$$y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \,.$$

The F test now gives $3.9 < F_{0.99}(1, 4) = 21.20$, showing that the fourth parameter is useless. Moreover, all four parameters of FIT3 have a small T value in Eq. (11.101) and are therefore compatible with zero. All this shows that four free parameters are redundant. In FIT4 and FIT5 the fit is attempted with cubic curves but with the suppression of the parameters θ_0 and θ_1, respectively. The F test with the comparison of FIT4 and FIT5 versus FIT3 again indicates that the four free parameters of FIT3 are too many compared to the three of FIT4 and FIT5, showing that the resulting fits are practically equivalent to FIT2. Finally, in the last row of Table 11.2, the different estimates s_z of the common multiplicative scale error σ_z are reported. They have been obtained from the `FitPolin` routine using Eq. (11.79). Apart from the clearly wrong case of FIT1, all the other fits give similar values.

Let us now consider the case of known absolute errors σ_i. Our code `FitPolin`, called with the options `dy=sy` and `ww='ABS'`, where `sy` is the vector containing the last row of Eq. (11.103), gives the output results reported in Table 11.3. The parameter errors, given now by Eq. (11.80), are different, and the final value of χ^2/ν can be used to perform also the χ^2 test. The observed significance level SL, reported in the last line of Tabble 11.3, confirms the previous conclusions.

The results FIT2, FIT4 and FIT5 of Table 11.2 are then all equally compatible, although we know (but only because the data have been simulated by Eq. (11.102)) that the solution closest to the true one is given by FIT5.

The correct conclusion is therefore the following: we have proved that, starting from a polynomial of first degree and progressively adding higher degree terms, the eight values of Table 11.103, affected by a a relevant $\pm 15\%$ relative error, are compatible with a quadratic or cubic functional form, having no more than three parameters. This class of solutions includes the true function (11.102). The use of orthogonal functions with the property (11.68) greatly optimizes the minimization procedure, because, due to their independence, the addition of new parameters does not change, the fitted values of the previous ones. However, this procedure generally does not help resolve ambiguities. The situation is effectively summarized in Fig. 11.9, where we immediately see that, in the case of absolute errors, second- and third-degree polynomials are statistically compatible with the data within the statistical fluctuations.

Table 11.3 Best-fit results with polynomials of the type $y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$ for the data of Eq. (11.103) for the case of known absolute errors

	FIT1	FIT2	FIT3	FIT4	FIT5
θ_0	-5.1 ± 1.6	-15.1 ± 3.2	-6.1 ± 7.0		1.8 ± 1.3
θ_1	12.8 ± 0.9	23.2 ± 3.1	10.6 ± 9.1	2.7 ± 1.8	
θ_2		-1.5 ± 0.4	2.8 ± 2.9	5.1 ± 1.0	6.0 ± 0.6
θ_3			-0.4 ± 0.2	-0.6 ± 0.1	-0.6 ± 0.1
χ^2/ν	2.8	0.9	0.5	0.6	0.7
SL	2%	98%	60%	58%	72%

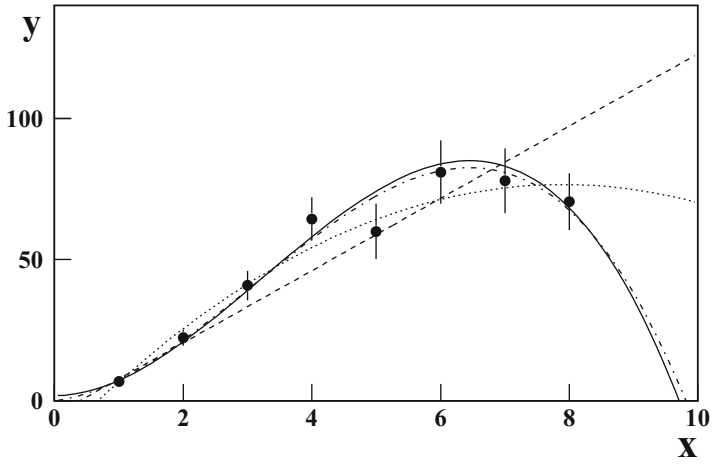


Fig. 11.9 Experimental points from Eq. (11.103) and and results of the polynomial fits given in Table 11.2. Dashed line, FIT1; dotted line, FIT2; point-dashed line, FIT4; and full line, FIT5. Apart from FIT1, all the other polynomials have three free parameters

The figure also shows that the *extrapolation* of the curves outside the data range is extremely dangerous: the correct extrapolated value, for $x = 10$, is that of FIT5, which is the full curve; completely different results are obtained with the other curves, which however well represent the data within the measurement interval.

11.9 Search for Correlations

The search for functional forms, which we have just described, mainly investigates the analytical relation (11.3) between x and y described in Sect. 11.2. On the contrary, when also X is a random variable, as in the case of Eq. (11.5), and we examine the link (11.11) between $f(X)$ and the correlation coefficient, the problem is usually called *search for correlations*. Now, according to Eqs. (11.6–11.7), the variance at the denominator of the χ^2 variable is a constant representing the response fluctuations. We are therefore in the unweighted fit case.

Suppose to have n occurrences of the pairs of variables (X, Y). If we denote by:

$$\hat{y}_i \equiv \hat{y}(x_i) = f(x_i, \hat{\boldsymbol{\theta}}) \tag{11.104}$$

the estimate of the Y mean values at any given x, the decomposition of Eq. (11.95) becomes in this case a decomposition of the sample deviance of Y. Therefore,

Eq. (2.40) can be written as:

$$\underbrace{\sum_{i=1}^{n}(y_i - \langle y\rangle)^2}_{(RSS_0)} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \langle y\rangle)^2}_{(RSS_0 - RSS)} + \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{(RSS)} . \tag{11.105}$$

In words, this decomposition corresponds to:

$$\text{total sum of squares} = \text{explained sum of squares} + \text{residual sum of squares} . \tag{11.106}$$

This equality is the simplest form of *analysis of variance*, which we have extensively described in Sect. 7.9, and has a very interesting interpretation: the dispersion of y_i around $\langle y\rangle$ is decomposed into two uncorrelated parts: the first one is the one identified by the linear regression model with p predictors (i.e. by $f(X)$ in Eq. (11.5)), and the second one is the residual, that is, the sum of squares of the deviations, denoted by Z in Eq. (11.5), that the model cannot explain. If the model interpolated the data, we would have $y_i - \hat{y}_i = 0$ for each i, a statistically unacceptable result.

From Eq. (11.60) we deduce also $\hat{\theta}_0 = \langle y\rangle = \hat{y}_{0i}$ and hence $\hat{z}_{0i} = y_i - \langle y\rangle$. Therefore:

$$\|\hat{\boldsymbol{z}}_0\|^2 = \sum_{i=1}^{n}(y_i - \langle y\rangle)^2 = (n-1)s_y^2 .$$

Moreover, from Fig. 11.7, we see that $\|\hat{\boldsymbol{z}} - \hat{\boldsymbol{z}}_0\| = \|\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_0\|$. By dividing the explained sum of squares of Eq. (11.106) by the total sum of squares, we obtain the *coefficient of determination*:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \langle y\rangle)^2}{\sum_{i=1}^{n}(y_i - \langle y\rangle)^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \langle y\rangle)^2} = 1 - \frac{RSS}{(n-1)s_y^2} . \tag{11.107}$$

Recalling Eq. (11.105), we see that $0 \leq R^2 \leq 1$. The upper limit is reached when the model interpolates the data while $R^2 = 0$ only if $\hat{\theta}_1 = \hat{\theta}_2 = \ldots = \hat{\theta}_p = 0$, a condition that occurs if y does not depend linearly on any of the model predictors, so that it has "residual" fluctuations only. In practice, the zero value is never reached exactly, but sometimes very small values can occur.

From Fig. 11.7, it can be shown that R^2 is the square of the sample correlation coefficient between y_i and $\hat{y}_i$. This parameter is called *multiple correlation coefficient* and can be interpreted as a measure of the sample correlation between y_i and the rows $\boldsymbol{x}_i$ of the X matrix, since each $\hat{y}_i$ is the prediction of Y at $\boldsymbol{x}_i$. In the specific case of the least squares line (when $p = 1$), substituting Eq. (11.29) into the numerator

of Eq. (11.107), one obtains the relation:

$$\sum_{i=1}^{n}(\hat{y}_i - \langle y\rangle)^2 = r^2\frac{s_y^2}{s_x^2}\sum_{i=1}^{n}(x_i - \langle x\rangle)^2 = r^2(n-1)s_y^2 \,,$$

which shows that $R^2 = r^2$ is the square of the linear correlation coefficient between the values x_i and y_i.

To complete the analysis of the decomposition (11.105), we mention the coefficient of determination corrected for degrees of freedom, which is given in the output of many least squares estimation codes, as in the case of our `Linfit` routine. It is defined by:

$$R_a^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2/(n-p-1)}{\sum_{i=1}^{n}(y_i - \langle y\rangle)^2/(n-1)} = 1 - \frac{RSS/(n-p-1)}{s_y^2} \,.$$

The search for correlations aims *to find the function* $f(x)$ *maximizing* R^2, that is, the function that maximizes the explained sum of squares and minimizes the residual one. We outline the procedure for functional forms of the type (11.91), keeping in mind that it still holds for any multiple linear regression model:

(a) Given a function $f(x, \boldsymbol{\theta})$ which is linear in the parameters, the LS estimate $\hat{\boldsymbol{\theta}}$ is evaluated by minimizing the quantity:

$$\sigma_z^2\chi^2(\boldsymbol{\theta}) = \sum_i [y_i - f(x_i, \boldsymbol{\theta})]^2 \,. \tag{11.108}$$

(b) Once the estimate has been obtained, the explained sum of squares is calculated together with the coefficient of determination R^2.

(c) Among all the functional forms hypothesized in point (a), the one that has the maximum R^2 is selected. If there are two or more almost equivalent solutions, the one having the smaller number of parameters is chosen (the result does not change if $RSS = \sigma_z^2\chi^2(\hat{\boldsymbol{\theta}})$ is minimized instead of maximizing R^2).

(d) The trend of the residual plot of the chosen solution is checked; a random behaviour indicates the absence of structures clearly due to functional dependencies not seized by $f(x, \hat{\boldsymbol{\theta}})$.

Let us apply these rules to a data set simulated with the algorithm:

$$\begin{aligned} x &= x_0 + x_R = 10 + 2\,g_1 \\ f(x) &= 2 + x^2 \\ y &= f(x) + Z = 2 + x^2 + 5\,g_2 \,, \end{aligned} \tag{11.109}$$

where g_1 and g_2 are standard Gaussian variates. X is then a normal variable with $\mu_x = 10, \sigma_x = 2$, whereas Y is not normal because the correlation $f(X)$ is nonlinear; however, the fluctuations around the correlation function are Gaussian with $\sigma_z =$

5. Twenty (x_i, y_i) pairs obtained using the previous algorithm are reported in the following table (11.110):

x	6.28	6.62	7.10	7.46	7.54	8.11	8.62	8.95	9.00	9.62
y	40.2	47.3	47.4	48.7	52.2	68.3	79.8	88.3	86.4	97.7
x	9.92	10.10	10.11	10.25	10.34	11.09	12.23	12.46	13.31	13.82
y	98.7	102.2	101.7	111.2	105.4	118.5	154.3	164.0	182.5	191.1

(11.110)

The results obtained with our routine `FitPolin` by minimizing the χ^2:

$$\chi^2 = \sum_i (y_i - \theta_0 - \theta_1 x_i - \theta_2 x_i^2 - \theta_3 x_i^3)^2$$

and considering polynomials up to the third degree are reported in Table 11.4 and Fig. 11.10. The parameter errors are calculated with the `Linfit` routine using Eq. (11.79).

As an example, assuming to have loaded the vectors x and y of Eq. (11.110) in `xx,yy`, the code instructions for the FIT2 polynomial of Table 11.4 are:

```
>class(fitfun<-yy~x+I(x^2))
>FitPolin(x=xx,y=yy,fitfun=fitfun,ww='NO')
```

where the parameter `ww` specifies that the errors on the variables are not given. If we now apply rules (a)–(d) just discussed, FIT3 appears as the best solution:

$$f(x) = -1.02 + 1.03\, x^2 \ .$$

This solution maximizes the explained variance to 99.1% and minimizes the Gaussian fluctuations around the regression curve to $s_z^2 = 4.4$. The FIT2 solution gives the same results, but with one additional parameter and too large parameter errors (i.e. with too small values of the T statistic of Eq. (11.101)).

The most reasonable conclusion is therefore that the data have a parabolic correlation function $f(x) \simeq x^2$, with a small or even negligible constant term

Table 11.4 Best-fit results with polynomials of the type $y(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$ for the data of Eq. (11.110)

	FIT1	FIT2	FIT3	FIT4
θ_0	-98.7 ± 6.6	-12.2 ± 20.2	-1.02 ± 2.45	32.4 ± 2.8
θ_1	20.5 ± 0.7	2.3 ± 4.2		
θ_2		0.91 ± 0.21	1.03 ± 0.02	
$s_z = \sqrt{\frac{\chi^2}{n-p-1}}$	6.3	4.4	4.4	6.9
R^2	98.1 %	99.1 %	99.1 %	97.7 %

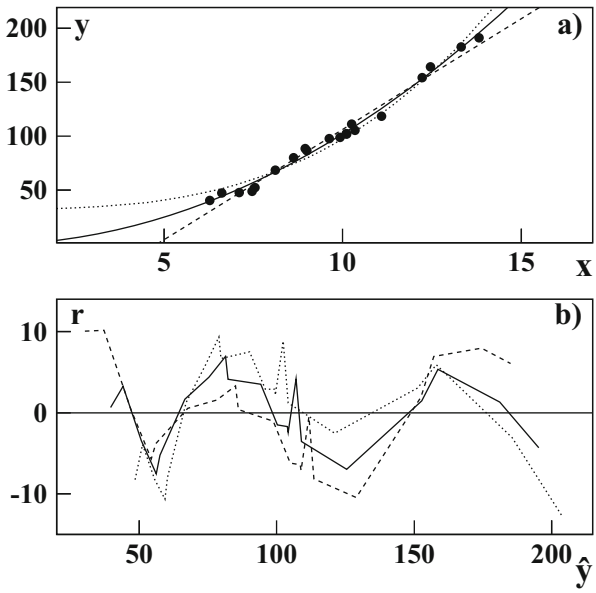


Fig. 11.10 (**a**) Experimental points of Eq. (11.110) and fitted polynomials from Table 11.4. Dashed line, FIT1; full line, FIT3; dotted line, FIT4. The FIT2 solution gives graphically indistinguishable results from those of FIT3. (**b**) Residuals (11.92) corresponding to the three solutions of Fig. (**a**)

compared to the mean of the values of y, since the error on the parameter is large ($\theta_0 \in -1.02 \pm 2.45$). The absolute fluctuations around this curve are $\simeq 4.4$ (in s_z units). As you can see, this conclusion is quite close to the "truth"represented by Eq. (11.109). Also the residual plot, shown in Fig. 11.10b, demonstrates that the FIT3 solution has a more regular residual fluctuations than the others.

Finally, we note that if we progressively increased the number of parameters, at some point we would certainly find even larger values of R^2 and even smaller values of s_z^2. In the most extreme situation, with 20 parameters, we would go through all the points exactly, getting $R^2 = 1$ and $s_z^2 = 0$! However, these correlation studies must be performed with functions far from the interpolation limit and for which the fluctuations of the data must appear as random. The functional forms to be tested, such as the maximum degree of polynomials, must therefore be determined a priori on the basis of the available information about the problem, to reach a compromise between the best possible fit and a model with the minimum number of free parameters.

To conclude this section, we report the connection between the F statistic used to verify the null hypothesis $H_0 : \theta_1 = \ldots = \theta_p = 0$ and the R^2 parameter. Comparing Eqs. (11.105) and (11.107), we get:

$$R^2 = \frac{RSS_0 - RSS}{RSS_0} = \frac{(RSS_0 - RSS)/RSS}{(RSS_0 - RSS)/RSS + 1} = \frac{pF}{pF + (n - p - 1)} ,$$

where $\boldsymbol{F} \sim F(p, n-p-1)$. Since R^2 is an increasing function of F, the choice of the functional form having the highest values of R^2 is equivalent to select the solution that, when tested against H_0, rejects it with the smallest p-value.

11.10 Fit Strategies

At the end of this long discussion on model fitting and testing, we try to give you some further guidance on the practical procedures to follow.

We start by recalling that in the previous sections we have always considered the case of Gaussian errors.

If the errors are reliable but the observed variables are not Gaussian, high χ^2 values are usually obtained because often the points are more scattered than expected from normal deviates. In this case, χ^2 values corresponding to significance levels which are small from a purely statistical point of view (even of per thousand levels or less) are sometimes accepted. This choice, i.e. to associate the high χ^2 value to the data non-Gaussianity rather than to the model function, could be sometime justifiable from a practical point of view.

An approximate but immediate view of the result quality can also be seen by eye before carrying out the χ^2 test: if the experimental points "touch"the curve $y = f(x, \hat{\boldsymbol{\theta}})$ within one, two or three error bars according to the 3σ law, then the χ^2 will probably be acceptable. Figure 11.11 exemplifies the situation; remember that, by convention, the error bar is always equal to $\pm\sigma$.

Most minimization codes, if the errors are not specified, tacitly assume that they are all the same and that $\sigma_z = 1$, without warning the users of ... the risks they are taking. If we can actually assume $\sigma_i^2 = \sigma_z^2$ for each i without knowing σ_z^2, we will then have to provide its estimate s_z through Eq. (11.63) to be able to calculate the errors on the fit results. This process is known as error readjustment or rescaling. We recall that this procedure is valid only if there are no doubts both about the error constancy and on the functional model $\mu_i(\boldsymbol{x_i}, \boldsymbol{\theta}) = \boldsymbol{x}_i^{\dagger}\boldsymbol{\theta}$ used. Furthermore, it makes

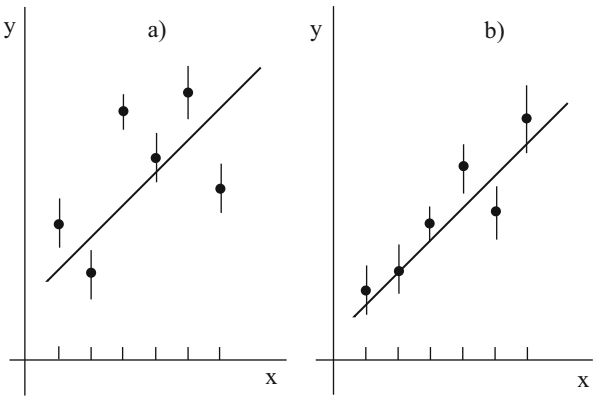


Fig. 11.11 Bad fit (**a**) and a good fit (**b**); in the second case, about 68% of points "touch" the regression curve within one error bar

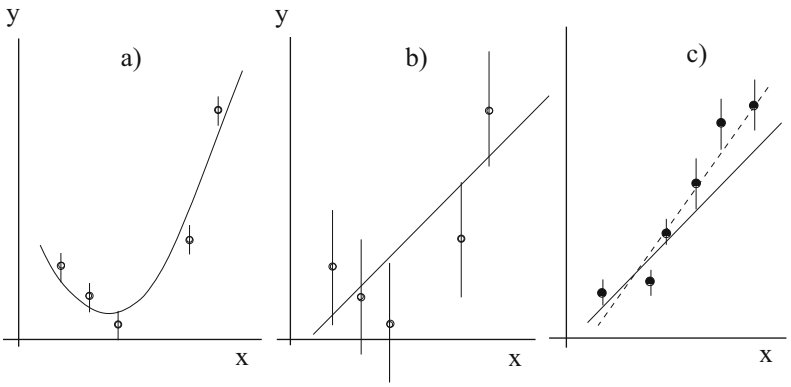


Fig. 11.12 When data fluctuate around a parabola with constant errors as in (**a**), a fit assuming a straight-line model and using the error rescaling will find a good χ^2 but with an incorrect error overestimation as in (**b**). In (**c**) a linear weighted fit of points with different errors is shown, giving as the result the full line. The solution of an unweighted fit, which assumes equal errors, assigns the same relevance to all points, thus giving an incorrect result, given by the dashed line

no sense to perform the χ^2 test at the end of the fit, because, after replacing σ_z^2 with s_z^2 in the $\chi^2(\hat{\boldsymbol{\theta}})$ formula, we will always get $\chi^2_{\sigma_z=s_z}(\hat{\boldsymbol{\theta}}) = (n - p - 1)$ (a constant!) for any functional model. Without these precautions, we could mistakenly evaluate as correct a fit that is not, since the expected value $(n - p - 1)$ of $\chi^2(n - p - 1)$ is well below the critical value. The problem is visualized in Fig. 11.12, where we assume the data to be parabolic and with a constant error, as in Fig. 11.12a. A straight-line fit with error adjustment will give the result in Fig. 11.12b, where the final χ^2 is obviously good, but only because errors larger than the real ones have been estimated. The correct result is only obtained by representing $f(x, \boldsymbol{\theta})$ as a parabola $\theta_0+\theta_1 x+\theta_2 x^2$, which, of course, presupposes the a priori knowledge of the functional form of the model. The χ^2 test to check the validity of the model function *must therefore be performed only when the absolute errors are known.* However, it is possible, using the F test of Eq. (11.99), to check on the elements of linear models even when only the relative error ratios are known.

Finally, another frequent gross mistake is to provide minimization codes only with data without errors even in the case of variable errors; most of the time this procedure gives wrong results, as shown in Fig. 11.12c.

11.11 Nonlinear Least Squares

As mentioned in the previous section, the model function to be used in the χ^2 minimization can be nonlinear in the parameters. An example of this is given in Exercises 10.6 and 10.7, where the model introduced in the function to be minimized

(11.2) was represented by the Gaussian and binomial densities, respectively. We solved these exercises, using our routine `Nlinfit` for nonlinear minimization, and the results have been also examined. In this section we intend to briefly describe the algorithms used in this class of codes to evaluate the minimum of the function $\chi^2(\boldsymbol{\theta})$ in a p-dimensional space, without giving too many numerical calculation details, which are extensively described in other texts such as [BR92] or [PFTW92] and also mentioned in our web pages [RPP].

The most efficient algorithms are based on the negative gradient method, since the opposite of the gradient is a vector that always points towards the minimum. Therefore, proceeding by successive steps, they reach a region where the second derivative is positive, until when the function starts to increase again. Around the minimum the χ^2 function is expanded in parabolic approximation:

$$\chi^2 \simeq \chi_0^2 + \sum_j \frac{\partial \chi^2}{\partial \theta_j}\,\theta_j + \frac{1}{2}\sum_{jk} \frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k}\,\theta_j \theta_k \ . \tag{11.111}$$

To fix ideas, let us assume that the χ^2 function is of the type of Eq. (11.6):

$$\chi^2(\boldsymbol{\theta}) = \sum_i \frac{[y_i - f(x_i, \boldsymbol{\theta})]^2}{\sigma_i^2} \equiv \sum_i H_i^2 \ ,$$

where the index i refers to the measured points; the second derivatives then take the form:

$$\begin{aligned} \frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k} &= \frac{\partial}{\partial \theta_j}\frac{\partial}{\partial \theta_k}\sum_i H_i^2 = \frac{\partial}{\partial \theta_j}\sum_i 2\,H_i \frac{\partial H_i}{\partial \theta_k} \\ &= 2\sum_i \frac{\partial H_i}{\partial \theta_j}\frac{\partial H_i}{\partial \theta_k} + 2\sum_i H_i \frac{\partial^2 H_i}{\partial \theta_j \partial \theta_k} \ . \end{aligned} \tag{11.112}$$

Neglecting the second term, the formula:

$$\frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k} \simeq 2\sum_i \frac{\partial H_i}{\partial \theta_j}\frac{\partial H_i}{\partial \theta_k} = 2\sum_i \frac{1}{\sigma_i^2}\frac{\partial f_i}{\partial \theta_j}\frac{\partial f_i}{\partial \theta_k} \ . \tag{11.113}$$

is obtained. Experience has shown that this approximation has many advantages: the algorithm is faster, as accurate as the algorithms that use higher-order expansions, and the *matrix of second derivatives is always positive definite* [Jam08, Jam92].

It is also important to note that setting to zero the second derivatives of the function:

$$H_i(\boldsymbol{\theta}) = \frac{y_i - f(x_i, \boldsymbol{\theta})}{\sigma_i}$$

with respect to the θ parameters, implies that the second derivatives of the function $f(x, \boldsymbol{\theta})$ vanish. This means a linear first-order Taylor expansion of f around the starting point $\boldsymbol{\theta}^*$:

$$f(x, \boldsymbol{\theta}) = f(x, \boldsymbol{\theta}^*) + \sum_k \left. \frac{\partial f}{\partial \theta_k} \right|_{\theta_k^*} \Delta\theta_k \ , \tag{11.114}$$

where $\Delta\theta = (\theta - \theta^*)$. In this case Eq. (11.111) is an exact relation, since the χ^2 derivatives of higher order than the second are zero. If we now keep in mind the formulae described in Sect. 11.5 and operate, in Eqs. (11.55–11.58), the substitutions:

$$\theta_k \to \Delta\theta_k = \theta_k - \theta_k^* \ , \quad y_i \to y_i - f(x_i, \boldsymbol{\theta}^*) \ , \quad x_{ik} \to \left. \frac{\partial f(x_i)}{\partial \theta_k} \right|_{\theta_k^*} \ ,$$

we see that the minimization equations are, step by step, identical to Eqs. (11.57–11.60).

The X matrix of Eq. (11.54) now contains the p first derivatives of $f(x, \boldsymbol{\theta})$ calculated at the point $\boldsymbol{\theta}^*$ and at the n experimental points x_i. The matrix $(X^\dagger W X)$ of Eq. (11.59), which in this case is called curvature matrix, contains the χ^2 second derivatives under the form of products between the first derivatives of f and the weights $1/\sigma_i^2$. Once the minimum with respect to the $\Delta\theta_k$ values is found, the procedure restarts from the new point θ_k. At the minimum point, the errors on the parameters are obtained through the error matrix (11.62), as in the linear case. For more details, you can examine the `Nlinfit` routine and at its comment lines.

When the χ^2 to be minimized is a likelihood function, from Eq. (10.45) it results:

$$\chi^2(\boldsymbol{\theta}) = -2 \ln L(\boldsymbol{\theta}) + C \equiv -2 \sum_i \ln f(x_i, \boldsymbol{\theta}) + C \ . \tag{11.115}$$

The repetition of the previous calculation shows that, after the linearization of the function f, the second derivative of χ^2 is:

$$\frac{\partial^2 \chi^2}{\partial \theta_j \partial \theta_k} \simeq 2 \sum_i \frac{1}{f_i^2} \frac{\partial f_i}{\partial \theta_j} \frac{\partial f_i}{\partial \theta_k} \ . \tag{11.116}$$

Also in this case, it can be shown that the second derivative matrix is always positive definite [Jam08, Jam92].

11.12 Problems

11.1 Find the linear correlation coefficient (4.31) when $f(X) = a + bX$ and X and Z are independent.

11.2 Two quantities X and Y, linked by a linear dependence, are measured with an instrument having an accuracy of 10%:

x	10	20	30	40	50	60	70
y	21.4	38.8	52.2	88.1	99.5	120.4	158.3

Determine the regression line using the routines `Linemq` and `FitPolin`, assuming the accuracy to be an uniformly distributed interval $[x-0.05\,x, x+0.05\,x]$ (and similarly for y) that contains the true value with $CL = 100\%$. Calculate the χ^2 value and comment the result.

11.3 Two measured quantities X and Y have Gaussian distributed errors:

x	10	20	30	40	50	60	70
s_x	0.3	0.6	0.9	1.2	1.4	1.7	2.0
y	20.5	40.0	63.6	86.7	104.3	123.3	144.7
s_y	0.6	1.1	1.9	2.5	3.1	3.5	4.1

Determine the regression line with the routines `Linemq` and `FitPolin`. Calculate the χ^2 value and comment the result.

11.4 The vertex problem: using the least squares method, determine the common vertex (x_0, y_0) of a set of straight lines $y = a_i + b_i x$. (Hint: consider the equation of a pencil of straight lines passing through a point: $(y - y_0)/(x - x_0) = b$).

11.5 A Gaussian variable Y, measured as a function of X, provided the values:

x	1.0	1.1	1.4	1.5	1.8	2.0	2.2	2.3	2.4	2.5
y	5.16	5.96	6.29	7.41	7.31	8.26	9.15	9.51	9.96	9.03

.

Knowing that Y has a constant standard deviation, determine if a first- or second-degree polynomial relation between X and Y is statistically compatible with the data. The `FitPolin` routine can be used.

11.6 The sampling of two Gaussian variables X and Y provided the result:

x	1.271	0.697	2.568	2.400	2.879	2.465	2.472	2.039	2.277	1.392
y	6.05	3.57	13.88	13.77	15.77	13.61	13.86	11.30	12.77	6.38

.

Determine the correlation function between these variables using the `FitPolin` routine with a first- or second-degree polynomial function.

11.7 The result of five measurements y_i as a function of exact values x_i is:

x	2	4	6	8	10
y	7.9	11.9	17.0	25.5	23.8

.

The Y values have a Gaussian relative error equal to $\pm 10\%$. Determine the coefficients of the functional dependence $Y = a + bX$. Analyse the obtained result by simulating 20 000 times the fit procedure.

11.8 The following five measurements y_i are given as a function of x_i values:

x	1.85	3.77	5.74	7.56	9.66
y	8.87	13.90	17.70	22.91	23.59

.

The values of X and Y are affected by uniform relative fluctuations of $\pm 10\%$. Determine the a and b coefficients of the functional dependence $Y = a + bX$. Analyse the obtained result by simulating 20,000 times the fit procedure.

Chapter 12
Experimental Data Analysis

You see, it depended on one or two points at the very edge of the range of the data, and there's a principle that a point on the end of the range of the data -the last point- isn't very good, because, if it was, they'd have another point further along

Richard P. Feynman, "SURELY YOU'RE JOKING, MR. FEYNMAN!: ADVENTURES OF A CURIOUS CHARACTER".

12.1 Introduction

The technical and more extensive part of this chapter describes how to apply statistical and probabilistic methods to the various types of measurements and experiments that are usually carried out in a scientific laboratory.

Our main purpose is to enable the researcher or the experimenter to recognize the type of measurement he/she is carrying out and to properly evaluate its accuracy and precision. Here, we will not deal with the important problem of finding physical laws through best-fit techniques, because this crucial aspect has been extensively treated in the previous Chap. 11.

Together with the technical topics, we will also address some very important conceptual and methodological aspects directly related to the foundations of the scientific method that permitted the birth and the development of modern science. This method is based on the observation of natural phenomena, i.e. on the data collection and analysis, according to those principles and procedures that were systematically adopted for the first time by Galileo Galilei and which have then consolidated and improved over the last four centuries. Mathematical and statistical analysis of data play, in this context, a role of primary importance.

Today, disciplines such as physics and medicine are considered sciences, as they are based on experimental facts, while this is no longer true for astrology, because the latter is based on people's expectations and is totally disproved by the facts, when these are analysed, as in Exercise 3.17, with the scientific method.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2 139, https://doi.org/10.1007/978-3-031-09429-3_12

The distinction between the sciences in a broad sense, such as medicine, and the so-called hard (or exact) sciences, such as physics, is more subtle. More correctly, the distinction should be made between totally and partially quantitative sciences. An adequate definition of hard science, if we really want to use this term, could be the following:

Statement 12.1 (Hard Science) *A science is said to be hard (or exact) when it is always able to associate an error (uncertainty) with its predictions and results.*

Basically, the term "exact" doesn't imply that the results must be affected by zero or negligible error; it is instead synonymous with "quantitative", which indicates a method providing results that are predictable with certainty or *with reliable confidence levels*. As you already know, the calculation of the measurement error or uncertainty is of fundamental importance for the correct determination of the confidence levels.

From the theoretical point of view, errors are sometimes present when simplified models of the phenomenon under study are used or when calculations are carried out with approximate numerical methods; however, we will not delve here into these particular aspects.

In the following, we will then only consider the point of view that the experimental errors derive from the random or systematic fluctuations or uncertainties connected with the measurement operations. Often the most difficult and laborious part of an experiment is precisely the evaluation of errors. In this phase, the experimenter is not guided so much by theorems or precise rules but rather by experience and a series of assumptions, sometimes even a little arbitrary. However, there is an important constraint: these assumptions and rules of thumb *must always be in accordance with the fundamental laws of probability and statistics*. We also note that some of the subjective and arbitrary a priori assumptions we will make are about *the shape of the statistical distributions to be associated with the behaviour of the instruments or to the measurement operations* that are performed. For an in-depth analysis of the consequences linked to these choices, you can read [D'A99].

Finally, remember that in this chapter we will use the notation (6.16) for confidence intervals and that the meaning of the confidence levels associated with these intervals is the frequentist one widely discussed in Sect. 6.2. These are the conventions currently used in the international scientific literature for the results of laboratory experiments.

12.2 Terminology

The measurement of a physical quantity with an experimental equipment can be sketched as in Fig. 12.1. As we can see, the uncertainties characterizing the measurement can be referred to the measured physical quantity, to the instrument or to the interaction between object and instrument. There is currently no single terminology for describing these uncertainties. The ISO recommendations [fSI93]

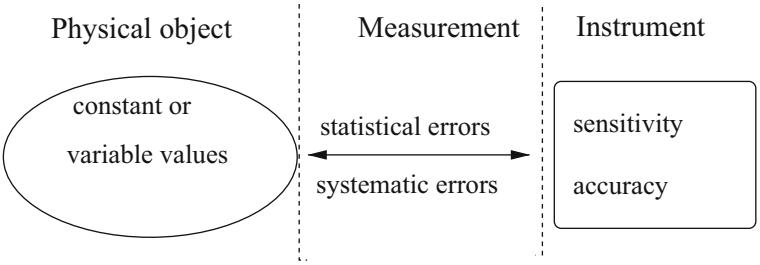


Fig. 12.1 Sketch of measurement operations

are to classify them into two types: those that can be treated only with statistical methods and those (called systematic or systematic effects) that must be treated with other methods. The current nomenclature normally uses the following terms:

$$\text{uncertainty} \begin{cases} \nearrow \text{statistical uncertainty, statistical error, random error} \\ \searrow \text{systematic uncertainty, systematic error, systematic effect} \end{cases}$$

The recommendation of [fSI93] is to use the terms statistical uncertainty and systematic effect. We believe that it is more appropriate to further distinguish between systematic effect and systematic error. The *systematic effect* must be known and corrected before the data analysis, and the *systematic error* is the uncertainty that remains after the systematic effect has been removed. For example, if data depends on atmospheric pressure, and one has daily average pressure values for a nearby location, the data can be corrected day by day with this value. One can figure out that during the day there are small variations around the average pressure values that have been used: this uncertainty must then be added to the other uncertainties of the measurement as a systematic error. We make the following choice, which we will stick to throughout the chapter:

$$\text{uncertainty} \begin{cases} \nearrow \text{statistical uncertainty or statistical error} \\ \searrow \text{systematic uncertainty or systematic error} \end{cases} \tag{12.1}$$

We will now examine in detail all the cases that may happen.

12.3 Constant and Variable Physical Quantities

When starting an experiment, first of all it is necessary to check whether the quantity to be measured is a constant or a random variable.

In the first case, if the fluctuations are observed during the experiment (different results at each measurement while keeping the experimental conditions constant), they are to be attributed to the measurement operations or to the behaviour of the

instruments used. In the second case, the observed fluctuations will also include those of the measured object. This component of the fluctuations contains the physical information about the statistical law (2.6) of the quantity being measured. The situation can be summarized in the two operational definitions:

Statement 12.2 (Constant Physical Quantity) *A physical quantity is called constant when it is a universal physical constant, i.e. a quantity that has the same value in all reference systems (Planck's constant, electron charge, rest mass of a stable particle, . . .) or a quantity which can reasonably be considered constant and stable with respect to the measurement that is being carried out.*

Statement 12.3 (Variable Physical Quantity) *A physical quantity is said to be variable when it has measurable fluctuations and variations that are intrinsic to the physical process being studied. Very often the fluctuations are purely statistical, and then the quantity is a random variable that has a specific distribution. The purpose of the measurement is precisely the determination of this distribution.*

Some examples of random physical quantities are:

- The speed of a gas molecule ($\sqrt{\chi^2}$ density with three degrees of freedom, known also as Maxwell density (see Exercise 3.10)
- The number of cosmic rays per second (approximately 100) that pass through your body as you are reading this page (Poisson law)
- The number of electrons passing through the cross section of a conductor in a given time interval
- In general, all the quantities studied in mechanics and statistical physics

12.4 Instrumental Sensitivity and Accuracy

The behaviour of an instrument is defined by two important characteristics: sensitivity (also called resolution) and accuracy.

The sensitivity denotes the smallest change in the measured variable to which the instrument responds.

Statement 12.4 (Instrumental Sensitivity) *If an instrument provides the value x in the measurement of a physical quantity x, the sensitivity interval or resolution is indicated by Δx, that is, the minimum quantity necessary to move the result of the measurement from the value x to a contiguous one. The sensitivity is defined as the ratio:*

$$S = \frac{1}{\Delta x} .$$

In an ideal instrument, with high sensitivity, $\Delta x \simeq 0$ ed $S \gg 1$. If Δy is the read-out variation for a change Δx of the measured quantity, the sensitivity can be defined as

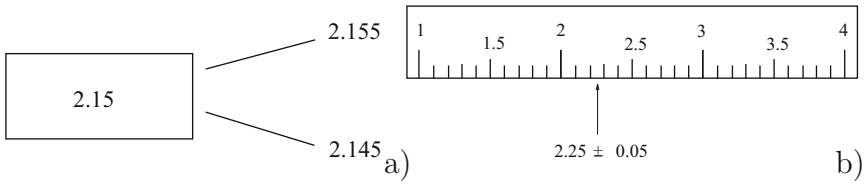


Fig. 12.2 (**a**) In digital instruments the sensitivity range becomes a rounding error. (**b**) In analogue instruments, the sensitivity interval is given by the width of the minimum read-out interval defined by the scale. The measured value is assumed as the midpoint of the interval indicated by the pointer

$S = \Delta y/\Delta x$. In digital instruments, the sensitivity interval can be clearly defined, since in this case it is nothing more than a rounding error. For example, if a well-made digital multimetre indicates a voltage of 2.15 V, the true value will be between 2.145 and 2.155 V (see Fig. 12.2a, since it is reasonable to assume that the rounding operations are carried out correctly. In this situation, the sensitivity range is $\Delta = (2.155 - 2.145)\,\mathrm{V} = 10\,\mathrm{mV}$. Since there is generally no correlation between the sensitivity range and the localization of the true value within this range, we can say that, if x is the experimental value and Δx is the sensitivity interval, the true value will be within the interval $x \pm \Delta x/2$, with uniform probability law and with a confidence level of 100%. In summary, the true value is assumed to be uniformly distributed within the range:

$$\text{true value} = x \pm \frac{\Delta x}{2}\,, \qquad (CL = 100\%)\,. \tag{12.2}$$

In the case of Fig. 12.2a, one can affirm that the true value is within the interval:

$$2.150 \pm 0.005\ \mathrm{V}\ ,$$

that the sensitivity range is 10 mV and that the sensitivity is $S = 100\,\mathrm{V}^{-1}$. This means that 1 V shifts the reading by 100 positions.

For analogue instruments, the sensitivity range has a less clear definition: a pointer (needle, small arrow, ...) moves continuously, but the thickness of the pointer itself and the distance between two adjacent marks, present on the graduated dial of the instrument, define a minimum read-out interval below which it makes no sense to proceed. Generally, the result is still reported in the form (12.2), where the measured value is assigned to the midpoint of the interval indicated by the pointer and the error is the width of the minimum interval, centred on the read value (see Fig. 12.2b). Often there is a tendency to interpolate the reading by eye and thus to reduce the error. The procedure is acceptable, but in this case, one must be aware that a subjective (finer) scale is used instead of the scale of the instrument, obtained by visually interpolating between the notches marked on the dial. In this case it is possible to associate with the interval (12.2) not the uniform distribution but the triangular one (5.35), centred on the read value and with a width equal to the

minimum "ideal" read-out interval evaluated by eye. Instead of visual interpolation, it is however better in these cases to use a digital instrument with a higher sensitivity.

In addition to sensitivity, the other fundamental parameter characterizing an instrument is accuracy. It is a non-random deviation between the measured value and the true one and usually depends on the uncertainty on the correction to be applied to remove the systematic effect. In the following, this uncertainty will be denoted by δ.

We now come to an important point: how do errors due to sensitivity combine in an instrument? If the digital multimetre sketched in Fig. 12.2a had a perfect calibration, that is, if $\delta \ll \Delta x$, the true value would certainly be located within the interval (12.2). If instead there was a calibration defect (accuracy error), let us say of 30 mV, then $\delta \gg \Delta x$ and the interval (12.2) would be meaningless. Generally, professional and well-manufactured scientific instruments that are in good operational conditions have an accuracy range always lower than or at most of the same order as the sensitivity range. These instruments are equipped with an *accuracy table*, where the rules for defining a global interval Δ are given. This table allows you to combine sensitivity and accuracy errors, for which Eq. (12.2) is valid. In this case Δ indicates a global interval:

$$\delta + \Delta x \simeq \Delta(\text{syst}) \,, \tag{12.3}$$

which is called instrumental or systematic error.

If there are sensitivity errors, it is reasonable to assume that the true value is equally likely located within the interval (12.2); if instead there is also an accuracy component, this is strictly speaking no longer true, because the calibration defect generally causes a constant and correlated deviation between the true value and the measured one. However, in the absence of more detailed information, the systematic error (12.3) is generally associated with a uniform density. At the international level, the instruments are divided into accuracy classes, defined on the basis of the relative systematic error:

Classes of accuracy					
CLASS	0.2	0.5	1	1.5	2.5
$\pm\Delta(syst)/x_{FS}$	0.2 %	0.5 %	1 %	1.5 %	2.5 %

where x_{FS} is the instrument full scale. For example, an instrument is defined to be of class 1 if its total systematic error does not exceed $\pm 1\%$ of the full scale reading.

12.5 Measurement Uncertainty

We go on with the study of the diagram of Fig. 12.1 and describe the analysis of the measurement operations.

In this process, which involves the interaction between the whole experimental apparatus (which can also include the observer) and the quantities that are being measured, two types of errors occur, statistical and systematic.

The statistical errors have been extensively discussed in Chap. 6; in laboratory measurements they occur when the stabilization of the experimental operations or the measurement operations themselves become critical due to the very high sensitivity of the experimental instruments. If you measure the length of a workshop bench with a carpenter's tape and repeat the measurement several times, you obtain always the same value, and there are no statistical errors. If, on the other hand, highly sensitive optical instruments (such as laser distance metres) are used, the superimposition of many different fluctuations (in positioning, calibration or other) means that a slightly different value is obtained at each measurement. In this case, we have a spectrum of experimental results, and the value of the bench length becomes a random variable. If the fluctuations inherent to the measurement process are numerous, linearly overlap and none of them outweigh the others, then the conditions of the Central Limit Theorem 3.1 hold, and the measurements tend to be Gaussian distributed. When only statistical errors are present, it is usually assumed that the average of the measurements should tend to the true value (mind you: this is just an assumption!). The statistical error which, as we know, is the estimate of the standard deviation of the distribution of the measures defines the *precision* of the measure.

On the contrary, as discussed above, systematic effects are instead due to incorrect operations or wrong assumptions about the physical model on which the measurement is based (e.g. describing the large oscillations of the pendulum with a linear model). Consequently, a *non-random* deviation is introduced between the measured value and the true one regardless of the number of observations made.

In principle, all possible sources of systematic effects must be eliminated *before* starting the measurement. If this is not possible, at the *end* of the measurement, but *before* carrying any statistical data analysis, systematic corrections need to be applied to measured values, using equations based on physical models or other methods. If, for example, at the end of a measurement we realize that the response of an instrument drifted due to temperature effects, we will have to quantitatively study this behaviour and elaborate equations to correct all the observed data, or we will have to repeat the measurement by thermally stabilizing the whole apparatus. However, even after all possible verifications have been made, an uncertainty about the value of these corrections may remain. This uncertainty affects all data in the same way, as in the example of the correction due to temperature. Therefore, in these cases, it is also necessary to evaluate the systematic error to be associated with the obtained results.

Sometimes, the distinction between statistical and systematic errors is not clear-cut. For example, when we read an analogue instrument and correctly try to

minimize the parallax error, we can obtain a series of different read-outs, the mean of which will be more or less coincident, for large samples, with the true value. In this case the parallax error is statistical. If, on the other hand, we always read the instrument sideways, the average of the measurements will always deviate from the true value, giving rise to a systematic error.

This previous discussion about statistical and systematic errors can be summarized with some fundamental definitions:

Statement 12.5 (Statistical Error) *It is that kind of error, due to measurement operations, which causes the result to vary according to a certain statistical distribution. The mean of this distribution (true average) is assumed to coincide with the true value of the physical quantity being measured. The standard deviation of the distribution is the measurement error. These two parameters are estimated from the mean and the standard deviation of the experimentally measured sample.*

Statement 12.6 (Precision) *The precision is determined by the statistical error of the measurement, which is given by the standard deviation estimated from the measured sample. A measurement is all the more precise the smaller the statistical error is.*

Statement 12.7 (Systematic Error) *The systematic effect causes the average of the measured values to deviate from the true value, regardless of the number of measurements that are made. The systematic error arises from the uncertainties of the corrections made to eliminate the systematic effects.*

Statement 12.8 (Accuracy) *Accuracy is determined by the systematic errors of the measurement. A measurement is all the more accurate the smaller the systematic errors are.*

Precision and accuracy can be represented in a schematic, but effective, way by representing, as in Fig. 12.3, the experiment as a target whose centre denotes the true value of the measurement. The results of the measurements can then be symbolized as shots on the target. A measurement that is neither precise nor accurate can be represented as a set of scattered points, with a centre of mass (mean) different from the true value (the target centre). In a precise, but not very accurate, measurement, experimental results are arranged around a value, which however can be very different from the true one. An accurate, but not very precise, measurement gives a set of points that are considerably dispersed, but with their average close to the true value. Finally, a precise and accurate measurement gives narrowly dispersed points that are grouped around the true value. It is evident that only an accurate measurement (precise or not) *is a good measurement*, as the mean value of the data is a correct estimate of the true mean. In this case, as we will shortly mention in Sect. 12.12, the error on the mean can be reduced by increasing the number of measures.

For a single measurement, accuracy can be also defined as the difference between the single measured value and the true one, while precision, which is given by the

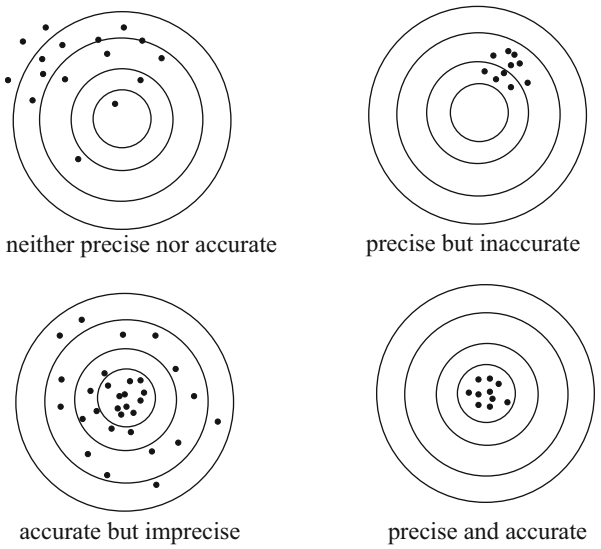


Fig. 12.3 Representation of the effects of systematic (accuracy) and statistical (precision) errors in a measurement. The true value is represented by the target centre, while the measures are represented by the points

dispersion of repeated measurements, loses its meaning in this case. Therefore, a single measurement will be accurate if, in the chosen measurement unit, it is close to the true value, not accurate if it is far from it.

Another good example to understand the difference between accuracy and precision is given by the quartz watch: if the watch is set with "the exact time", after some days it will differ slightly from this value, and we will have a precise and accurate time measurement. If, on the other hand, we set the clock 5 min ahead of the correct time, we will have a precise but not accurate measurement.

Now suppose the exact time to be unknown or, which is the same, to remove the concentric rings centred on the true value in Fig. 12.3. In this case, we are able to judge if the measurement is precise, *but not if it is accurate*; in other words, the configurations of top and bottom lines of Fig. 12.3 will look alike. Knowing how accurate a measurement is would require to already know the true value, which is the purpose of the measurement! As all experimenters know, this is the greatest difficulty encountered in laboratory measurements. It is therefore necessary to very well know the experimental apparatus and the methods of data processing that are used, in order to be reasonably certain to have a priori removed the systematic effects or to know how to evaluate them. Later on in this chapter, we will give some examples and further explore these important aspects.

12.6 Treatment of Systematic Effects

Although systematic effects can have very different characteristics, it is possible to make a fairly general treatment of them, at least for the most common types.

The first type of effect is due to the discretization operated by digital instruments, which is related to sensitivity. When δ, which denotes the accuracy error in Eq. (12.3), is negligible, the instrument sensitivity can be treated statistically, provided the standard deviation of the statistical uncertainties is much larger than Δ. In fact, we can rewrite the measured quantity as:

$$X_k = \mu + R_k + \Delta_k \equiv X'_k + \Delta_k \,, \tag{12.4}$$

where μ is the true value, R_k the random component and Δ_k the effect due to the sensitivity. We can assume $\langle R \rangle = 0$, because a value different from zero is included into the mean μ. Under these assumptions, one has $\langle X' \rangle = \mu$.

In digital instruments, Δ_k represents the distance, for the kth measurement, between $\mu + R_k$ and the closest discrete value of the instrumental scale. Unlike the calibration error, which is independent of k (i.e. of the single measurement), this error varies for each data point and can be considered as a uniform random variable, since we supposed that $\sigma \gg \Delta$. This is also the type of error that is introduced when constructing the histogram of a continuous variable, with the histogram bin smaller than the range of the data. Therefore, if Δ_k is similar to a rounding effect, the approximation $\langle \Delta \rangle \sim 0$ holds and hence:

$$\langle X \rangle = \langle X' \rangle = \mu \,, \tag{12.5}$$

so that the discretization effect of a continuous datum, typical of digital instruments and histograms, *does not alter the average of the measurements.* On the contrary, it has an effect on the dispersion of the measures. Assuming the uniform distribution of the systematic effect, we have in fact:

$$\mathrm{Var}[X] = \sigma^2 + \frac{\Delta^2}{12} \,, \tag{12.6}$$

where Δ is the step of the instrument discrete scale or the histogram bin width. To obtain the dispersion of the data without the instrumental effect, the so-called Sheppard's correction is often used:

$$\sigma^2 = \mathrm{Var}[X] - \frac{\Delta^2}{12} \,. \tag{12.7}$$

For histograms, the effect of the increased dispersion is shown in Fig. 12.4. We also recommend to solve Problem 12.13.

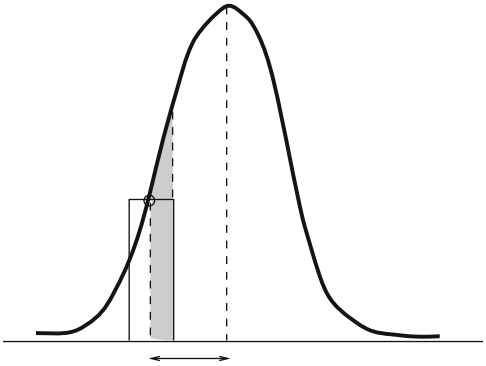


Fig. 12.4 The histogram of a continuous variable overestimates the dispersion of data when the population has a bell-shaped density as shown in the figure. In fact, the abscissa of the midpoint of the channel is attributed to all the events contained in the shaded area (which are the majority within the bin), even if it has a distance from the mean greater than the average distance of the shaded events

We now come to the second type of systematic effect, called *offset or zero-setting error*. In this case, the observed random variable must be written as:

$$X_{ik} = \mu + R_{ik} + \langle S' \rangle + S_i \equiv X'_{ik} + \langle S' \rangle + S_i \ , \tag{12.8}$$

where, as before, μ is the true mean, R is the random fluctuation and $\langle S' \rangle + S_i$ is the systematic effect, written as an average value $\langle S' \rangle$ plus a random part S_i with null mean value. The indices denote the k replicates of the ith measurement carried out with the ith instrument or by the ith laboratory. Here the systematic error S_i is the same for all the data of the same measurement or experiment and can be considered a random variable only if we consider the set of different laboratories or instruments measuring the same quantity. Before analysing the data, the systematic effect is corrected by subtracting the $\langle S' \rangle$ value (which must be known) from all the data:

$$X_{ik} - \langle S' \rangle = \mu + R_{ik} + S_i \equiv X'_{ik} + S_i \ . \tag{12.9}$$

In the following this passage will be implied, and therefore, without loss of generality, we will set $X_{ik} - \langle S' \rangle \rightarrow X_{ik}$ and $\langle S \rangle = 0$, transforming Eq. (12.8) into:

$$X_{ik} = \mu + R_{ik} + S_i \equiv X'_{ik} + S_i \ . \tag{12.10}$$

The presence of the term S_i creates a correlation among all data of the ith measurement. Since $\langle R \rangle = 0$ by construction, from Eq. (12.10) it results in $\langle X' \rangle = \langle X \rangle = \mu$; the experimental average, after the correction, is therefore a good estimator of the true mean μ. Since R_{ik} and S_i are independent, from Eq. (12.8) the

variance of this estimate is given by:

$$\mathrm{Var}[X] = \mathrm{Var}[R] + \mathrm{Var}[S] \equiv \sigma_x^2 + \sigma_{\mathrm{sys}}^2 \;\rightarrow \sigma_x^2 + \frac{\Delta^2}{12} \;, \tag{12.11}$$

where σ_x^2 is the variance of the non-systematic part. The last relation holds if the systematic errors are uniformly distributed with a total amplitude Δ, as is often the case. The validity of this formula can be verified with our routine `MCsystems`, which simulates a set of different Gaussian measurements of the same quantity μ, all carried out by different laboratories and with a uniform offset error. At the end a parameter called *pool*, i.e. the standard variable $T_p = (X_{ik} - \mu)/\,\mathrm{Var}[X]$ is calculated. The error handling is correct if T_p follows the standard Gaussian. The covariance between two measures X_1 and X_2 must be calculated considering X_1' and X_2' as independent variables but with the same systematic error. From Eq. (4.9) one has:

$$\begin{aligned}
\mathrm{Cov}[X_1, X_2] &= \langle (R_1 + S)(R_2 + S) \rangle - \langle R_1 + S \rangle \langle R_2 + S \rangle \\
&= \langle R_1 R_2 \rangle + \langle R_1 S \rangle + \langle S R_2 \rangle + \left\langle S^2 \right\rangle - \langle R_1 \rangle \langle R_2 \rangle \\
&\quad - \langle R_1 \rangle \langle S \rangle - \langle R_2 \rangle \langle S \rangle - \langle S \rangle^2 \\
&= \left\langle S^2 \right\rangle - \langle S \rangle^2 = \sigma_{\mathrm{sys}}^2 \;,
\end{aligned} \tag{12.12}$$

since R_1, R_2 and S_i are independent of each other and $\langle R_1 \rangle = \langle R_2 \rangle = 0$. The last term of the equation implies the average of S over all ith different measures, each with constant S_i. Based on Eq. (4.31), this result shows that the systematic error maximizes the correlation among measures.

We can generalize these results through Eq. (5.77). For example, if we have three statistically independent variables X_1', X_2' and X_3', with variances σ_1^2, σ_2^2 and σ_3^2, having a common systematic error σ_{sys1} and another systematic error σ_{sys2} affecting only the first two variables, the covariance matrix can be written, with obvious notation, as:

$$V(\boldsymbol{X}) = \begin{pmatrix}
\sigma_1^2 + \sigma_{\mathrm{sys1}}^2 + \sigma_{\mathrm{sys2}}^2 & \sigma_{\mathrm{sys1}}^2 + \sigma_{\mathrm{sys2}}^2 & \sigma_{\mathrm{sys1}}^2 \\
\sigma_{\mathrm{sys1}}^2 + \sigma_{\mathrm{sys2}}^2 & \sigma_2^2 + \sigma_{\mathrm{sys1}}^2 + \sigma_{\mathrm{sys2}}^2 & \sigma_{\mathrm{sys1}}^2 \\
\sigma_{\mathrm{sys1}}^2 & \sigma_{\mathrm{sys1}}^2 & \sigma_3^2 + \sigma_{\mathrm{sys1}}^2
\end{pmatrix} . \tag{12.13}$$

To calculate the variance of the sum or difference of two variables X_i, $i = 1, 2$ with parameters μ_i, σ_i and with X_1' and X_2' statistically independent, one can proceed

directly without matrix notation. In fact, from Eqs. (12.11) and (12.12), one has:

$$\begin{aligned}\mathrm{Var}[X_1 + X_2] &= \mathrm{Var}[X_1] + \mathrm{Var}[X_2] + 2\,\mathrm{Cov}[X_1, X_2] \\ &= \sigma_1^2 + \sigma_{\mathrm{sys}}^2 + \sigma_2^2 + \sigma_{\mathrm{sys}}^2 + 2\sigma_{\mathrm{sys}}^2 \\ &= \sigma_1^2 + \sigma_2^2 + 4\sigma_{\mathrm{sys}}^2 \,, \end{aligned} \tag{12.14}$$

$$\begin{aligned}\mathrm{Var}[X_1 - X_2] &= \mathrm{Var}[X_1] + \mathrm{Var}[X_2] - 2\,\mathrm{Cov}[X_1, X_2] \\ &= \sigma_1^2 + \sigma_2^2 \,. \end{aligned} \tag{12.15}$$

As is intuitive, the constant offset error increases with the sum of two variables, while it cancels out with subtraction.

The third and last case we consider is that of systematic errors known as scaling or normalization, with a scale (or multiplier) factor multiplying all the values of the ith measurement by the same constant common factor. With the same notations of Eq. (12.8), we have:

$$X_{ik} = \langle S' \rangle \, S_i (\mu + R_{ik}) \,. \tag{12.16}$$

Here the correction of the systematic effects is applied by dividing the data by $\langle S' \rangle$. The analogous of Eq. (12.10) is then:

$$X_{ik} = S_i (\mu + R_{ik}) = S_i \, X'_{ik} \,, \tag{12.17}$$

where the substitutions $X_{ik} / \langle S' \rangle \to X_{ik}$ and $\langle S \rangle = 1$, $\mathrm{Var}[S] \equiv \sigma_{\mathrm{sys}}^2$ have been done. After this correction, the systematic error affects the covariance matrix only. Since S and X' are independent, one has:

$$\langle X \rangle = \langle S X' \rangle = \langle S \rangle \langle X' \rangle = \langle X' \rangle = \mu \,, \tag{12.18}$$

from which we see that, after the correction, the data average is a good estimator of the true average. From Eqs. (5.69) and (12.16), one obtains the variance of the estimate:

$$\begin{aligned}\mathrm{Var}[X] &= \langle \mu + R_{ik} \rangle^2 \, \sigma_{\mathrm{sys}}^2 + \langle S \rangle^2 \, \sigma_x^2 \\ &= \mu^2 \sigma_{\mathrm{sys}}^2 + \sigma_x^2 + \sigma_x^2 \sigma_{\mathrm{sys}}^2 \simeq \mu^2 \sigma_{\mathrm{sys}}^2 + \sigma_x^2 \,, \end{aligned} \tag{12.19}$$

where σ_x^2 is the variance of the measure after the correction and the last term, according to Eq. (5.65), holds under linear approximation. As for Eq. (12.8), also this formula can be verified with our `MCsystemp` routine.

Always within the linear approximation, the covariance between two different measures X_1 and X_2 when the two variables X'_1 and X'_2 are independent becomes:

$$\begin{aligned}\mathrm{Cov}[X_1, X_2] &= \langle S X'_1 S_i X'_2 \rangle - \langle S X'_1 \rangle \langle S X'_2 \rangle \\ &= \left\langle S^2 \right\rangle \langle X'_1 \rangle \langle X'_2 \rangle - \langle S \rangle^2 \langle X'_1 \rangle \langle X'_2 \rangle \\ &= (\left\langle S^2 \right\rangle - \langle S \rangle^2) \langle X_1 \rangle \langle X_2 \rangle = \mu_1 \mu_2 \sigma^2_{\mathrm{sys}} \,. \end{aligned} \tag{12.20}$$

Finally, to evaluate the effect of the common scale systematic error on the product or on the ratio between X_1 and X_2, we introduce two variables $Z_1 = X_1 X_2$ and $Z = X_1/X_2$ and apply Eq. (5.77) by defining the transport matrix T and the covariance matrix $V(\boldsymbol{X})$ based on Eqs. (12.18) and (12.19) as:

$$T = \begin{pmatrix} \langle X_2 \rangle & \langle X_1 \rangle \\ \langle 1/X_2 \rangle & \langle -X_1/X_2^2 \rangle \end{pmatrix} \quad V(\boldsymbol{X}) = \begin{pmatrix} \langle X_1 \rangle^2 \sigma^2_{\mathrm{sys}} + \sigma_1^2 & \langle X_1 \rangle \langle X_2 \rangle \sigma^2_{\mathrm{sys}} \\ \langle X_1 \rangle \langle X_2 \rangle \sigma^2_{\mathrm{sys}} & \langle X_2 \rangle^2 \sigma^2_{\mathrm{sys}} + \sigma_2^2 \end{pmatrix} \tag{12.21}$$

Now the product $V(\boldsymbol{Z}) = T V(\boldsymbol{X}) T^\dagger$ must be computed. After a simple but somewhat lengthy calculation, the following matrix is obtained:

$$V(\boldsymbol{Z}) = \begin{pmatrix} \mu_2^2 \sigma_1^2 + \mu_1^2 \sigma_2^2 + 4\mu_1^2 \mu_2^2 \sigma^2_{\mathrm{sys}} & \sigma_1^2 - \sigma_2^2 \dfrac{\mu_1^2}{\mu_2^2} \\ \sigma_1^2 - \sigma_2^2 \dfrac{\mu_1^2}{\mu_2^2} & \dfrac{\sigma_1^2}{\mu_2^2} + \sigma_2^2 \dfrac{\mu_1^2}{\mu_2^4} \end{pmatrix} . \tag{12.22}$$

The diagonal elements give the variances of the product and division and the off-diagonal elements their covariance. From these results we see that the systematic error increases with the product, while the division does not contain the term σ_{sys}.

12.7 Best Fit with Offset Systematic Errors

As we have seen, the systematic error introduces a correlation between the experimental data, and we must therefore deal with a sample of non-independent measures.

In this case the ML method can still be applied, and the likelihood function for correlated variables can be obtained using the product Theorem 1.2 and generalizing Eq. (1.21). Therefore, given a set $Y_1, Y_2, \ldots Y_n$ of correlated measures,

the likelihood function $L_{\rm corr}$ can be written as:

$$L_{\rm corr}(\boldsymbol{\theta};\ \boldsymbol{y}) = p(\boldsymbol{\theta};\, y_1) \prod_{i=2}^{n} p(\boldsymbol{\theta};\, y_i, |y_{i-1}, \ldots, y_1) \ . \tag{12.23}$$

In the following we will deal, for simplicity, only with the case of Gaussian statistical errors and Gaussian systematic uncertainties. For a more general discussion, see [PS20].

To begin with, let us consider the case of offset systematic errors, for which Eqs. (12.10)–(12.12) hold. Assuming Gaussian errors, Eq. (12.23) becomes nothing else than the product of a one-dimensional Gaussian density by Gaussian marginal distributions. Recalling the considerations made in Sect. 4.4, it is easy to conclude that $L_{\rm corr}$ is simply a multivariate Gaussian function with correlated variables whose general expression is given by Eq. (4.69).

From Eq. (12.13), the covariance matrix V can be immediately obtained; along the diagonal we have the quadratic sums of the statistical and systematic errors, while the other terms represent the square of the systematic error:

$$V = \begin{pmatrix} \sigma_1^2 + \sigma_{\rm sys}^2 & \sigma_{\rm sys}^2 & \ldots & \sigma_{\rm sys}^2 \\ \sigma_{\rm sys}^2 & \sigma_2^2 + \sigma_{\rm sys}^2 & \ldots & \sigma_{\rm sys}^2 \\ \ldots & \ldots & \ldots & \ldots \\ \sigma_{\rm sys}^2 & \sigma_{\rm sys}^2 & \ldots & \sigma_n^2 + \sigma_{\rm sys}^2 \end{pmatrix} . \tag{12.24}$$

Let us denote with Δ the vector of the difference between experimental and theoretical values:

$$\Delta = \begin{pmatrix} \mu_1(\boldsymbol{\theta}) - y_1 \\ \ldots \\ \ldots \\ \mu_n(\boldsymbol{\theta}) - y_n \end{pmatrix} . \tag{12.25}$$

In matrix notation, the function to be minimized becomes:

$$\chi^2(\boldsymbol{\theta}) = \Delta^T V^{-1} \Delta \ . \tag{12.26}$$

As an example, let us consider a linear best-fit procedure, with $\mu_i(\theta_0, \theta_1) = \theta_0 + \theta_1 x_i$, applied to the points given in Eq. (12.27):

x	2.0	4.0	6.0	8.0	10.0	12.0
y	3.0	3.3	4.1	5.7	6.9	6.7 ,
$\sigma_{\rm stat}$	0.5	0.5	0.5	0.5	0.5	0.5

(12.27)

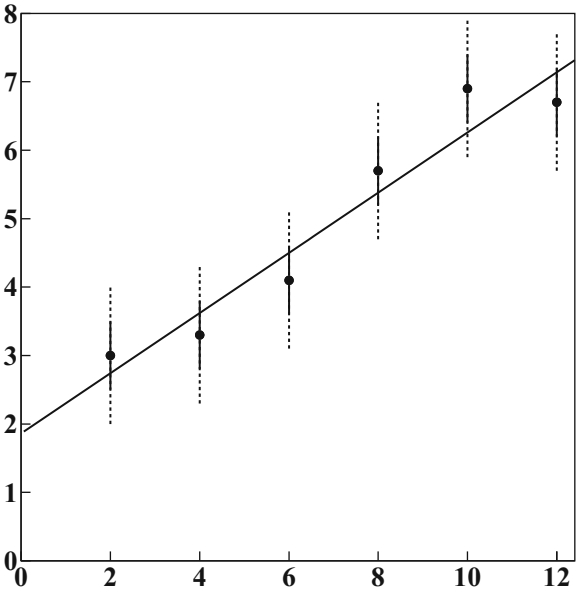


Fig. 12.5 Fit with both statistical and systematic errors. The solid line is the result of the best-fit procedure

and also shown in Fig. 12.5. The vertical bars in this figure represent, as usual, the absolute statistical errors σ_i reported in Eq. (12.27). With the dashed lines, we have instead indicated a common systematic error $\sigma_{\rm sys} = \sigma_i = 0.5$ which has been linearly added to the statistical ones.

The solution to the minimization problem can be obtained in a way similar to that of Sect. 11.4:

$$\hat{\theta}_0 = \frac{1}{D}\left[(\boldsymbol{x}^\dagger V^{-1}\boldsymbol{x})(\mathbf{1}^\dagger V^{-1}\boldsymbol{y}) - (\mathbf{1}^\dagger V^{-1}\boldsymbol{x})(\boldsymbol{x}^\dagger V^{-1}\boldsymbol{y})\right], \tag{12.28}$$

$$\hat{\theta}_1 = \frac{1}{D}\left[(\mathbf{1}^\dagger V^{-1}\mathbf{1})(\boldsymbol{x}^\dagger V^{-1}\boldsymbol{y}) - (\mathbf{1}^\dagger V^{-1}\boldsymbol{x})(\mathbf{1}^\dagger V^{-1}\boldsymbol{y})\right], \tag{12.29}$$

where $D = (\mathbf{1}^\dagger V^{-1}\mathbf{1})(\boldsymbol{x}^\dagger V^{-1}\boldsymbol{x}) - (\mathbf{1}^\dagger V^{-1}\boldsymbol{x})^2$ and $\mathbf{1}$ is a column vector (with the correct dimension) of unit elements.

If $\sigma^2_{\rm sys} = 0$ and $\sigma_i^2 = \sigma_z^2$, then $V = \sigma_z^2 I$, and we get again exactly Eqs. (11.25) and (11.26), while, if $\sigma^2_{\rm sys} = 0$, we obtain the weighted LS estimates for the regression line. In our numerical example, $\sigma^2_{\rm sys}$ is positive, but $\sigma_i = \sigma_z = 0.5$, and this implies to get the same $\hat{\theta}_0$ and $\hat{\theta}_1$ of the unweighted case. What has just been stated can be verified in the following way: if $\sigma_i = \sigma_z$ for any $i = 1, \ldots, n$ and $\sigma_{\rm sys} \geq 0$, then the element at position ij of V^{-1} can be written as $(\delta_{ij}\tau_z^2 + \tau^2_{\rm sys})$,

with $\tau_z^2 = 1/\sigma_z^2$ and:

$$\tau_{\text{sys}}^2 = -\frac{\sigma_{\text{sys}}^2}{\sigma_z^2} \frac{1}{n\sigma_z^2 + \sigma_{\text{sys}}^2} .$$

From Eqs. (12.28)–(12.29), after a little rearrangement, Eqs. (11.25)–(11.26) are again obtained. However, the error of the estimates changes: as can also be intuitively understood, the inclusion of a same constant error common to all points has no influence on the error of the slope of the fitted line, but only on the θ_0 parameter, whose estimate is then affected by a larger error than in the pure statistical case. Applying Eq. (5.73) to the estimators of Eqs. (11.25)–(11.26), that is, generalizing the error propagation formulae obtained in Sects. 11.4 and 11.5 to the case of a non-diagonal V matrix, it is easy to verify (see also [Bar89]) that the error on the parameter $\hat{\theta}_1$ remains unchanged, while the variance of $\hat{\theta}_0$ becomes:

$$\text{Var}[\hat{\theta}_0] = \frac{\sigma_z^2}{D^2} \sum_i [S_{xx} - S_x x_i]^2 + \frac{\sigma_{\text{sys}}^2}{D^2} \sum_i \sum_j [S_{xx} - S_x x_i] [S_{xx} - S_x x_j] . \tag{12.30}$$

In this equation $D = n\, S_{xx} - S_x^2$, i.e. it is equal to the sum of the variance obtained from the best-fit procedure with only statistical uncertainties and the additional source of variation due to the systematic error.

In Table 12.1 (second row), the results thus obtained have been reported. They are compared with those obtained taking into account the random errors only (first row). When systematic errors are included, from Eq. (12.30) we have that $\sigma_{\hat{\theta}_0} = \sqrt{0.47^2 + 0.5^2}$.

The equations described above can be solved with our `FitMat` routine, which provides as input to the R routine `optim` the function (12.26) to be minimized, with the possibility to have a non-diagonal covariance matrix. The results of Table 12.1 have been obtained with the instructions:

```
>xx <- c(2,4,6,8,10,12)
>y <- c(3.0,3.3,4.1,5.7,6.5.6.7)
>varmat <- matrix(rep(0.25,36),ncol=6)  # fill cov mat with 0.25
>diag(varmat) <- 0.5  # fill diagonal with 0.25+0.25
>f <- function(par,xx){par[1]+par[2]*xx}
>FitMat(xx,y,varmat,parf=c(1,0.5),fun=f)
```

where the initial values of the parameters to be fitted are contained in the vector `parf`.

Table 12.1 Best-fit results for the data of Fig. 12.5. The result considering statistical errors only (first row) is compared with that obtained with an additive constant systematic error (offset error, second row) and a constant multiplicative systematic error (scale error, third row)

Error	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{f}$
stat.	1.86 ± 0.47	0.44 ± 0.06	
stat.+sys. add. (0.5)	1.86 ± 0.69	0.44 ± 0.06	
stat.+sys. mult. (20%)	1.86 ± 0.60	0.44 ± 0.11	1.0 ± 0.2

The application of Eq. (5.73) gives exact results for the linear least squares and, in general, is valid for "small" systematic errors:

$$\frac{\sigma_{\text{sys}}^2}{\sigma_z^2 + \sigma_{\text{sys}}^2} \ll 1 \, .$$

To solve the general case of large systematic errors in a statistically correct way, we suggest to consult [HL07, PS20]. Finally, for an in-depth study on mixed linear models and for the estimation procedures with unknown σ_z^2 and σ_{sys}^2, the reference is still [Dav08].

12.8 Best Fit with Scale Systematic Errors

As we mentioned earlier, systematic uncertainties can appear not only as offset values but also as fractions or percentages of the measured values. This, for example, is the case when the number of events that are registered by a detector not having 100% efficiency has to be multiplied by a correction factor.

In this situation, described in Eqs.(12.16)–(12.19), it must be considered that S affects not only y_i but also σ_i since both these parameters have been obtained by multiplying the raw data by the same common scale factor. Furthermore, the effect of the systematic error is now intrinsically nonlinear. For this reason, as shown in [D'A94], the matrix covariance formalism under linear approximation, applied in the previous section to additive systematic errors, leads now to biased results even in the presence of not very large scale errors. This situation is, for example, described in Problem 12.18.

A way to avoid these difficulties is to introduce, in addition to $\boldsymbol{\theta}$, a further parameter a in the best-fit function, so as to have $\mu(\boldsymbol{\theta}, a)$. In the present case, we can then write $\mu(\boldsymbol{\theta}, a) = a\mu(\boldsymbol{\theta})$, since the effect of the systematic multiplicative errors is to introduce a scale constant factor to all theoretical values. The p.d.f. of

the generic variable y_i can then be written as:

$$p(y_i; \boldsymbol{\theta}, f) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\frac{(y_i - a\mu(\boldsymbol{\theta}))^2}{\sigma_i^2}\right)$$
$$= \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\frac{(fy_i - \mu(\boldsymbol{\theta}))^2}{f^2\sigma_i^2}\right), \tag{12.31}$$

where $f = 1/a$ is the factor which simultaneously multiplies both y_i and σ_i to take systematic errors into account.

Assuming f to follow the normal distribution $f \sim N(1, \sigma_{\text{sys}}^2)$, Eq. (12.23) can be written as:

$$L(\boldsymbol{\theta}, f) = \prod_i \left\{ \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{(fy_i - \mu_i(\boldsymbol{\theta}))^2}{2f^2\sigma_i^2}\right]\right\} \cdot \frac{1}{\sigma_{\text{sys}}\sqrt{2\pi}} \exp\left[-\frac{(f-1)^2}{2\sigma_{\text{sys}}^2}\right]. \tag{12.32}$$

Since the multiplicative factors in front of the exponentials have constant values, the negative logarithm of the function to be minimized becomes [D'A94]:

$$\chi^2(\boldsymbol{\theta}; \boldsymbol{f}) = -2\ln L(\boldsymbol{\theta}, f) = \sum_{i=1}^{n} \frac{(fy_i - \mu(\boldsymbol{\theta}))^2}{f^2\sigma_i^2} + \frac{(f-1)^2}{\sigma_{\text{sys}}^2}. \tag{12.33}$$

Here the standard formula (11.1) is modified by the presence of the term $(f-1)^2/(\sigma_{\text{sys}}^2)$ which takes into account the effects due to the systematic multiplicative uncertainty. By carefully considering Eq. (12.31), one sees that it has a parametrization different from the Gaussian density that would result from the application of Eq. (12.17), which would provide the relation $Y = a(\mu(\boldsymbol{\theta}) + R)$. Furthermore, the χ^2 of Eq. (12.33) will be minimized with respect to $\boldsymbol{\theta}$, a parameter, and f, which is a random variable. This procedure is allowed in the Bayesian approach. For further information, we refer to [D'A94].

Applying this procedure to the points of Fig. 12.5 and now assuming a 20% systematic error, we obtain the results reported in Table 12.1. As in the previous case, the values of $\hat{\theta}_0$ and $\hat{\theta}_1$ do not change, whereas the error on both these estimates does change because the absolute value of the systematic error now varies point by point.

The results shown in the table can be obtained from the `FitMat` routine requesting the minimization of χ^2 and providing as input the `varmat` matrix in a diagonal form and the systematic error value via the `sys` variable:

```
FitMat(xx,y,varmat,parf=c(1,0.5),type='CHIS',sys=0.2)
```

12.9 Indirect Measurements and Error Propagation

A quantity is said to be measured indirectly when it is a function $z = f(x, y, w, \ldots)$ of one or more directly measured quantities affected by uncertainties. The determination of the uncertainty on z starting from that of the measured quantities is called error propagation. In the following we will start to describe the simple case $z = f(x, y)$ which can be easily extended to any number of variables. If x and y are independent and only affected by statistical errors, Eq. (5.65) or its generalization (5.72) to n variables should be used to implement this procedure, after the substitution of the standard deviations with the measurement errors s_x and s_y:

$$s_f^2 = \left(\frac{\mathrm{d}f}{\mathrm{d}x}\right)^2 s_x^2 + \left(\frac{\mathrm{d}f}{\mathrm{d}y}\right)^2 s_y^2 \,. \tag{12.34}$$

Obviously, with n independent measures, one has:

$$s_f^2 = \sum_{i=1}^{n} \left(\frac{\mathrm{d}f}{\mathrm{d}x_i}\right)^2 s^2(x_i) \,. \tag{12.35}$$

This equation, which is the well-known error propagation law, is exact only for linear transformations. In this case, the resulting standard deviation estimate defines a Gaussian confidence interval only if all variables are Gaussian, as shown in Exercise 5.3. However, the intervals tend to be approximately Gaussian even when a nonlinear function f depends on a large number ($>5 - 10$) of random variables. Moreover, as extensively discussed in Chap. 5 and in particular in Sect. 5.4, in general Eq. (12.35) gives reliable results also in case of small relative errors.

We now come to the instrumental uncertainties, which a uniform density can be often attributed to, as shown in Eq. (12.2). In this case, for example, with two measures, an error propagation law determined from the first-order Taylor expansion and with the derivative absolute value is sometimes used:

$$\Delta_f = \left|\frac{\partial f}{\partial x}\right| \Delta_x + \left|\frac{\partial f}{\partial y}\right| \Delta_y \,. \tag{12.36}$$

The generalization for n measures obviously is:

$$\Delta_f = \sum_{i=1}^{n} \left|\frac{\partial f}{\partial x_i}\right| \Delta_i \,. \tag{12.37}$$

These formulas do not represent the standard deviation of the resulting distribution, but its total width when f is a linear function. For this reason they are generally used to estimate the upper error limit, which can be useful in the case of correlated

quantities with unknown covariances. The absolute value of the derivatives ensures that the uncertainty propagation is always calculated in the most "unfavourable" way, which corresponds to an increase in the measurement error.

If the variables are uncorrelated, then a more correct way to proceed is to apply Eqs. (12.34) and (12.35) with a uniform density, whose variance, based on Eq. (3.82), is $\Delta^2/12$. For two and n variables we have, respectively:

$$s_f^2 = \gamma \left[\left(\frac{\partial f}{\partial x} \right)^2 \Delta_x^2 + \left(\frac{\partial f}{\partial y} \right)^2 \Delta_y^2 \right] , \tag{12.38}$$

$$s_f^2 = \gamma \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 \Delta_i^2 , \tag{12.39}$$

with $\gamma = 1/12$. Therefore, to obtain the total variance, it is necessary to quadratically sum up the systematic errors and then to divide by 12. The variance has been indicated here with Latin letters, as it is still a parameter estimated from the data under the a priori uniform density assumption for the systematic effects. However, these variances should not be associated with the Gaussian density, even if the function $f(x, y)$ combines the variables linearly. In fact, while the linear combination of Gaussian errors leads to Gaussian confidence intervals, in this case the sum of two or more uniform systematic errors leads to different densities, which depend on the number of summed errors. However, from the Central Limit Theorem 3.1, we know that these densities rapidly tend to a Gaussian. The practical problem is now to evaluate the number of linearly combined measures from which it is reasonable to use the Gaussian approximation. The result we will find is quite surprising.

We solve this problem by treating in a complete way the instructive case of the sum of two systematic errors of equal value. The results of Exercise 5.2 indicates that the sum of two equal uniform variables defined in $[0, \Delta/2]$ follows the triangular density in $[0, \Delta]$:

$$p(x) = \begin{cases} \dfrac{4}{\Delta^2} x & \text{for} \quad 0 \le x \le \dfrac{\Delta}{2} \\ \dfrac{4}{\Delta^2} (\Delta - x) & \text{for} \quad \dfrac{\Delta}{2} < x \le \Delta \\ 0 & \text{otherwise,} \end{cases} \tag{12.40}$$

with parameters:

$$\mu = \frac{\Delta}{2} , \qquad \sigma^2 = \frac{\Delta^2}{24} , \qquad \sigma = \frac{\Delta}{2\sqrt{6}} . \tag{12.41}$$

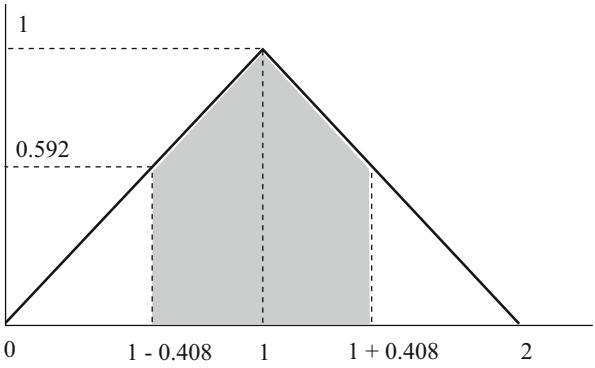


Fig. 12.6 Triangular density given by the sum of two measures affected by the same systematic error uniformly distributed in [0, 1]. The shaded area is the confidence level corresponding to the interval $\mu \pm \sigma$ and is $\simeq 65\%$

This distribution is shown in Fig. 12.6 when $\Delta/2 = 1$. If the two summed systematic errors are defined on [0, 1], Eq. (12.36) provides the value:

$$\Delta_f = \Delta_x + \Delta_y = 2\,,$$

which is just the total width of the distribution. From Eq. (12.38) we have instead:

$$s_f = \frac{1}{\sqrt{12}}\sqrt{\Delta_x^2 + \Delta_y^2} = \frac{1}{\sqrt{6}} = 0.408\,, \tag{12.42}$$

which is exactly the same value given by Eq. (12.41), since we are considering the sum of two errors. The area corresponding to the interval $|x - \mu| \leq K\sigma$, with K real number, can be directly read from Fig. 12.6 or evaluated as:

$$P\{|X - \mu| \leq K\sigma\} = \int_{\mu - K\sigma}^{\mu} x\,\mathrm{d}x + \int_{\mu}^{\mu + K\sigma} (2 - x)\,\mathrm{d}x \tag{12.43}$$

$$= \begin{cases} K\sigma(2 - K\sigma) & \text{for} \quad K\sigma \leq 1 \\ 1 & \text{per} \quad K\sigma > 1 \end{cases} = \begin{cases} 0.649 & \text{for } K = 1 \\ 0.966 & \text{for } K = 2 \\ 1 & \text{for } K = 3 \end{cases}$$

where in the last step the value of Eq. (12.42), $\sigma \simeq s_f = 0.408$, was used. These values are very close to the Gaussian probabilities of Eq. (3.35). The surprising fact is just that usually experimenters, given this type of results, *assume Gaussian confidence intervals already when errors are obtained from the linear combination of only two systematic errors*. This result also justifies the assumption of a triangular density for certain types of systematic errors, which come from experimental tests or simulations that combine a few systematic errors. In this case, in Eqs. (12.38) and (12.39), a value $\gamma = 1/24$ should be used, according to Eq. (12.41).

We now work out the problem of combining statistical and systematic errors. In the simplest case given by the linear superposition of two independent measures,

Table 12.2 Confidence levels of $1, 2, 3\sigma$ intervals for measurements where statistical Gaussian errors and systematic uniform errors are linearly combined. The intervals are parametrized as a function of the ratio Δ/σ and of the standard deviation $\sigma_m = \sqrt{\sigma^2 + \Delta^2/12}$, where σ is the statistical error and Δ is the total range of the systematic error

Δ/σ	$\pm\sigma_m$	$\pm 2\sigma_m$	$\pm 3\sigma_m$
1.0	68.3	95.4	99.7
3.0	67.1	95.8	99.8
5.0	64.9	96.7	100.0
10.0	60.9	98.6	100.0
100.0	57.8	100.0	100.0

one with Gaussian statistical error and the other with uniform systematic error within an interval Δ, the probability density of the result can be easily obtained from formula (5.34) derived in Exercise 5.1.

If we consider the systematic error within an interval ($a = -\Delta/2$, $b = +\Delta/2$), and introduce the standard variable $t = (z - \mu)/\sigma$, where σ is in this case the true statistical error, from Eqs. (3.40) and (3.44) the formula (5.34) becomes:

$$p(t) = \frac{\sigma}{\Delta}\left[\mathrm{E}\left(t + \frac{\Delta}{2\sigma}\right) - \mathrm{E}\left(t - \frac{\Delta}{2\sigma}\right)\right] . \tag{12.44}$$

This density is an even function with respect to the origin, since it is the convolution of two even functions, and has a standard deviation given by:

$$\sigma_m = \sqrt{\sigma^2 + \Delta^2/12}\,. \tag{12.45}$$

The corresponding probability levels can be evaluated by integrating Eq. (12.44). Table 12.2 shows the results obtained with our routine `Stasys(t,sigma, delta)`, where `t` is the quantile multiplying the error.

This table shows that, for $\sigma = \Delta$, i.e. for $\sigma > \Delta/\sqrt{12}$, the results coincide with the Gaussian levels (first row), while, for $\sigma \ll \Delta$, they tend to those of the uniform distribution (last row). In the intermediate cases, all in all, *the results do not differ much from the standard Gaussian levels*.

The results calculated with Eqs. (12.43) and (12.44) can be obtained with a few simulation lines, such as the following, in which $\sigma = 1$ and a systematic range $\Delta = 3\sigma$ are considered:

```
> vec <- rnorm(10000) + runif(10000,min=-1.5,max=1.5)
> error = sqrt(1 + 9/12)
> 1-(length(vec[vec<(-error)])+length(vec[vec>error]))/length(vec)
```

The obtained results coincide, within the statistical error of the simulation, with those of Table 12.2.

In general, the linear propagation method provides good results even in the case of measurement products or ratios. In this case Eq. (12.35) gives the *linear*

propagation of percentage errors, which is often used. Indeed, recalling Eq. (5.68), if we substitute the true standard deviations with statistical errors, we can write, with obvious notation:

$$Z = X_1\, X_2\ ,\ Z = \frac{X_1}{X_2}\ ,\ Z = \frac{X_2}{X_1} \quad \Longrightarrow \quad \frac{s_z^2}{z^2} = \frac{s_1^2}{x_1^2} + \frac{s_2^2}{x_2^2}\ . \tag{12.46}$$

This property can be also easily derived from Eq. (12.34) and can be immediately generalized to the case of n variables. Basically, *percentage or relative variances* are added together both in the product and in the division. The same property holds also for the maximum errors of Eq. (12.36):

$$Z = X_1\, X_2\ ,\ Z = \frac{X_1}{X_2}\ ,\ Z = \frac{X_2}{X_1} \quad \Longrightarrow \quad \frac{\Delta_z}{z} = \frac{\Delta_1}{x_1} + \frac{\Delta_2}{x_2}\ . \tag{12.47}$$

However, both the quadratic propagation of relative statistical errors (12.46) and the linear propagation of maximum systematic errors (12.47) should be used cautiously, because they are valid only if the measures x_1 and x_2 are independent. For example, it is easy to see that Eqs. (12.34) and (12.46) give different results when propagating the error for the ratio $x/(x+y)$. In this case only Eq. (12.34) is correct, because the presence of the variable x both in the numerator and in the denominator induces a correlation in their ratio, even if the measures x and y are independent.

Exercise 12.1
The measurement of the sides of a metal plate with a carpenter metre with millimetre marks provided the values:

$$b = 25.5 \pm 0.5 \quad \text{mm}\ , \qquad h = 34.5 \pm 0.5 \quad \text{mm}\ .$$

Find the value of the plate area A.

Answer A 1-mm systematic error (centred about the mean value of the interval which covers the measured value) has been attributed to the side measurements. According to Eq. (12.2), in this case one assumes that the true value is within the observed interval of width $2 \cdot 0.5 = 1$ mm with $CL = 100\%$ and probability given by the uniform distribution. Therefore, the standard deviation of the measures is given by:

$$s_b = s_h = \frac{2 \cdot 0.5}{\sqrt{12}} = 0.289 \quad \text{mm}\ .$$

(continued)

Exercise 12.1 (continued)
The relative deviations are:

$$\frac{s_b}{b} = \frac{0.289}{25.5} = 0.011 \simeq 1.1\% \,, \quad \frac{s_h}{h} = \frac{0.289}{34.5} = 0.008 \simeq 0.8\% \,.$$

From Eq. (12.46), the relative error on the area $A = bh$ results:

$$\frac{s(A)}{A} = \sqrt{(0.011)^2 + (0.008)^2} = 0.014 = 1.4\% \,.$$

Since obviously one has $A = bh = 25.5 \cdot 34.5 = 879.75\ \text{mm}^2$, the final result is:

$$A = 880 \pm 0.014 \cdot 880 = 880 \pm 12 \quad \text{mm}^2 \,.$$

Even in the simple cases that we have just developed in detail, the error propagation with analytical methods appears to be quite laborious. However, it can be replaced by the simpler direct computer simulation. The most common technique is the approximate algorithm described in Sect. 8.10 as the bootstrap method. Measured values and their errors are assumed to be the true distribution means and standard deviations; random variables are then computer sampled from these distributions and are combined as prescribed by the measurement, to obtain the simulated histogram of the final quantity (or sets of histograms, in the case of complex measurements).

The shape of the histogram gives an approximate evaluation of the probability density of the result, while the measurement errors are directly obtained as limits of histogram areas corresponding to the assigned confidence levels. In general, the central histogram values coincide or are very close, within the statistical error, to the measured values and therefore do not provide new information. In more complicated cases, which involve densities (usually multidimensional) not possessing symmetry properties such as Eq. (6.10), it is necessary to abandon the approximate bootstrap method and to use the rigorous Neyman's method, which implies the generation of simulated data for all possible values of the distribution and the assessment of the confidence regions as described in Sect. 8.10. These techniques are discussed in detail in [FC98, JLPe00].

The following example contains all the elements to understand the Monte Carlo techniques applied to the error propagation. Its simplicity is by no means limiting, since simulation methods have the great advantage of maintaining basically the same level of logical complication, regardless of the complexity of the analysed problem.

Exercise 12.2
An experimenter throws a stone into a well and with a manual stopwatch measures a 3-second falling time. By attributing to this measurement an uncertainty of ± 0.25 s, determine the depth of the well (neglecting the effects due to the sound speed).

Answer The central value of the well depth is calculated with the well-known law of falling bodies:

$$l = \frac{1}{2} g t^2 = 0.5 \cdot 9.81 \, \frac{m}{s^2} \cdot 9 \, s^2 = 44.14 \, m \, , \tag{12.48}$$

where $g = 9.81$ m/s^2 is the acceleration of gravity.

If we attribute a uniform distribution to the error inherent to the use of a manual stopwatch, the uncertainty of ± 0.25 s corresponds to a standard deviation value:

$$s_t = \frac{0.5}{\sqrt{12}} = 0.144 \simeq 0.14 \, s \, .$$

Then, the measurement error can be evaluated with Eq. (12.34) for the one-variable case:

$$s_l = \left(\frac{\mathrm{d}l}{\mathrm{d}t} \right) s_t = g \, t \, s_t = 9.81 \cdot 3 \cdot 0.144 = 4.24 \, m \, .$$

According to the analytical error propagation, the value of the well depth therefore lies in the interval:

$$l = (39.90, 48.38) = 44.1 \pm 4.2 \, m \, . \tag{12.49}$$

What is the confidence level of this interval? If we denote with τ the true value of the fall time, after measuring $t = 3 \, s$ with $\Delta = 0.5 \, s$, we know that $2.75 \leq \tau \leq 3.25 \, s$ with 100% probability. This value corresponds to a true length λ within the range:

$$\tau = \sqrt{2\lambda / g} \implies 37.09 \leq \lambda \leq 51.81 \, .$$

This interval contains the values of the λ parameter that, after the solution of the integrals (6.7), give the limits (λ_1, λ_2), of a specific confidence interval. If we search for the symmetric interval with $CL = 68.3\%$, Eq. (6.7) becomes:

$$\int_{t_0}^{\infty} p_t(t; \lambda_1) \, \mathrm{d}t = 0.158 \, , \quad \int_{-\infty}^{t_0} p_t(t; \lambda_2) \, \mathrm{d}t = 0.158 \, , \tag{12.50}$$

(continued)

Exercise 12.2 (continued)
where $t_0 = 3\,\mathrm{s}$ is the measured time. These integrals can be easily solved if one finds the cumulative of $p_t(t)$. If we consider values $T \sim U(\tau - 0.25, \tau + 0.25)$, we immediately obtain:

$$P\{T < t\} = 2(t - \tau + 0.25) = 2(t - \sqrt{2\lambda/g} + 0.25)\,,$$
$$\frac{g(t-0.25)^2}{2} \le \lambda \le \frac{g(t+0.25)^2}{2}\,.$$

Writing this equation with respect to λ, we can solve the integrals (12.50) with the simple condition:

$$\lambda_\alpha = \frac{g}{2}\left(t_0 - \frac{\alpha}{2} + 0.25\right)^2\,, \tag{12.51}$$

where the values $\alpha = 0.158$ and $\alpha = 0.841$ give the limits of the interval (λ_1, λ_2) with $CL = 68.3\%$. If we denote by l the true value of the length (this is the usual laboratory notation), we thus obtain the interval:

$$l = (39.28, 49.29) = 44.1^{+5.2}_{-4.8}\,, \quad CL = 68.3\%\,, \tag{12.52}$$

which is significantly larger than the approximate interval (12.49).

It is also useful to compare this result with that obtained from the simulation. From the bootstrap method, the fall times are randomly generated from the uniform density with mean equal to the measured value. The histogram with $N = 20\,000$ simulated measured lengths (evaluated with Eq. (12.48)) is shown in Fig. 12.7.

In this simple case, the simulation would not be necessary, because the derivation of the cumulative of length from the cumulative of time is straightforward:

$$\begin{aligned} F_l(l) = P\{L < l\} = P\left\{\frac{1}{2}gT^2 < l\right\} &= P\{T < \sqrt{2l/g}\} \\ &= 2(\sqrt{2l/g} - \tau + 0.25)\,, \\ \frac{1}{2}g(\tau - 0.25)^2 \le l \le \frac{1}{2}g(\tau + 0.25)^2\,, & \end{aligned} \tag{12.53}$$

and then to get, from its derivative, the p.d.f. of l:

$$p_l(l;\tau) = \sqrt{\frac{2}{g}}\frac{1}{\sqrt{l}}\,, \quad \frac{1}{2}g(\tau - 0.25)^2 \le l \le \frac{1}{2}g(\tau + 0.25)^2\,. \tag{12.54}$$

(continued)

Exercise 12.2 (continued)
The form of this function depends on τ through the limits of the definition range.

Let us now proceed to analyse the simulation results. Since the histogram contains 20,000 events, from Eq. (6.50) and from the data of the figure, the estimation interval of the mean is obtained:

$$\mu = 44.19 \pm \frac{4.22}{\sqrt{20000}} = 44.19 \pm 0.03 \,,$$

which is a result in agreement with the measured value of 44.14 metres. Given the density asymmetry of Fig. 12.7 and the nonlinearity of the physical law used, it is reasonable to expect a small deviation between the mean value and the measured one.

If we analyse the areas of the histogram centred around the measured value, it results that, as shown in Fig. 12.7, 68.3% of the depth values are within:

$$l = [39.24, 49.24] \simeq 44.1^{+5.1}_{-4.9}\text{m} \quad (CL = 0.68) \,, \tag{12.55}$$

a result coincident with the correct one (12.52). Figure 12.7 shows that the confidence levels are not Gaussian at all. We could say in this case that the measurement result is represented more precisely by Fig. 12.7 than by Eq. (12.55).

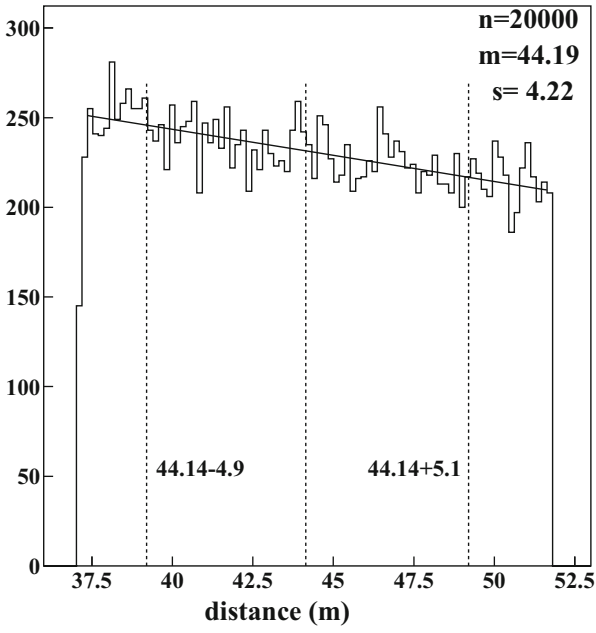


Fig. 12.7 Computer simulation of $N = 20\,000$ lengths $l = (1/2)gt^2$ where $g = 9.81\, m/s^2$ and t is a measured time of 3 s, distributed as the uniform density with $\Delta = 0.5$ s. The area between the two dashed lines, to the left and to the right of the measured value, contains 68% of the histogrammed events (lengths). The density shape (12.54) is well approximated with the best-fit straight line $f(l) = 343 - 2.5\, l$ shown as a solid line. By dividing this function by the histogram bin width $\Delta l = 0.17$, the density $Np(l) = 2018 - 14.70\, l$ is obtained, which allows the calculation of the areas under the curve (confidence levels)

The long discussion of this section can therefore be summarized in the following points:

- To evaluate statistical errors, it is necessary to estimate the sample variances and to apply Eq. (12.35). If errors are Gaussians, the results often follow the Gaussian density (since a linear combination of the effects is usually assumed), and the 3σ law (3.35) still holds.
- To evaluate systematic errors, Eq. (12.39) should be applied, and the resulting standard deviation often follows approximately Gaussian confidence levels. Equation (12.37) instead defines a maximum error, which must not be combined with other quantities representing estimated standard deviations.
- The combination of systematic and statistical errors must always be done in quadrature, using Eq. (12.35), where the variance of the systematic effects must be calculated with Eq. (12.39) (do not forget the $1/12$ or $1/24$ factors). The resulting standard deviation does not follow the 3σ law, because the corresponding density is not Gaussian (see Eq. (12.44)). However, Gaussian confidence levels are often assumed in practice. This assumption is generally all the more true the higher the number of combined errors is.
- The analytical procedure can lead to considerable inaccuracies in the case of large, correlated errors or of errors to be combined nonlinearly. In these cases simulation methods have to be used, since they usually allow us to solve any error propagation problem in a complete and satisfactory way. Thanks to simulation methods, *in recent years the results provided by experimental physics have remarkably improved their precision, accuracy and reliability*.

12.10 Measurement Types

The scheme of Fig. 12.1 suggests to classify measurements as represented in Fig. 12.8. This is our personal notation, you will not find it in other texts. The following examples clarify its meaning:

– $M(0,\ 0,\ \Delta)$ = measurement of a constant physical quantity with systematic errors
– $M(0,\ \sigma,\ 0)$ = measurement of a constant physical quantity with statistical errors
– $M(f,\ \sigma,\ \Delta)$ = measurement of a variable physical quantity in the presence of both statistical and systematic errors

Since each of the three symbols of the notation can assume two values, in total there are $(2 \times 2 \times 2 = 8)$ different types of measurements. However, the case $M(0,\ 0,\ 0)$, which refers to the error-free measurement of a fixed quantity, represents an ideal case without interest in this context. Therefore, *in practice seven different types of measurement must be considered*, that will be detailed in the next sections.

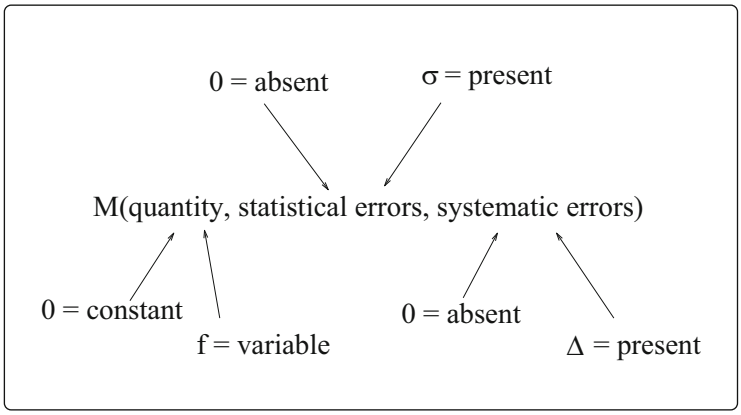


Fig. 12.8 Classification of the possible types of measurement

12.11 $M(0,\ 0,\ \Delta)$ Measurements

In this case, a constant quantity is measured, and the statistical errors are not present, because they are either totally absent or much smaller than the systematic error of width Δ.

This measure has no fluctuations, and all repeated measurements provide the same value x. The result is usually presented under the form (12.2):

$$x \pm \frac{\Delta(x)}{2}\,, \qquad (CL = 100\%)\,. \tag{12.56}$$

The error here has the meaning of maximum error, and the interval covers the true value with a 100% probability, that is, with certainty.

Often, based on the arguments of Sect. 12.4, systematic errors are associated with a uniform density. In this case the variance of the measure is given by (3.82):

$$s^2(x) = \frac{\Delta^2(x)}{12}\,, \tag{12.57}$$

whereas the standard deviation is usually written as:

$$s(x) = \frac{\Delta(x)}{2}\,\frac{1}{\sqrt{3}} \equiv \frac{U}{\sqrt{3}}\,, \tag{12.58}$$

where $U = \Delta/2$ is sometimes known as measure uncertainty.

The error deriving from the linear combination of several measures of this type must be propagated with Eqs. (12.38) and (12.39), and the confidence intervals tend to rapidly become Gaussian as the number of measures increases.

Exercise 12.3
The measure of an electrical resistance with a digital multimetre provided the value:

$$R = 235.4\ \Omega\ .$$

The "table of accuracy" of the instrument booklet gives an accuracy of:

$$\pm\ (0.1\%\ \text{rdg}\ +\ 1\ \text{dgt})$$

for resistance measurements. Find the result of the measurement.

Answer This is a measurement where only systematic errors are present as the statistical fluctuations due to time variations of the resistance value are below the multimetre sensitivity.

The instrument accuracy is 0.1% of the reading, (rdg) with the addition of one unit of the last right digit reported on the display (dgt), which in our case is 0.1 Ω. Therefore, we have:

$$\Delta(R) = \pm(235.4 \cdot 0.001 + 0.1) = \pm(0.2 + 0.1) = \pm 0.3\ \Omega$$

The result of the measurement is then:

$$R = 235.4 \pm 0.3\ \Omega\ ,$$

with 100% confidence level.

12.12 $M(0, \sigma, 0)$ Measurements

In this case, a constant physical quantity is still measured, but with a sensitivity interval of the apparatus much smaller than the statistical errors: $s \gg \Delta$.

Repeated measurements give different values, which generally, but not always, are distributed according to the Gaussian density. A sample of N measurements is thus obtained, from which the average m and the standard deviation s are calculated. Since, *in the absence of systematic errors*, it is assumed that the true value of the physical quantity coincides with the mean of the distribution of the measures (true mean), based on Eq. (6.50) the result of the measurement must be presented in the form:

$$x = m \pm \frac{s}{\sqrt{N}} \qquad (CL \simeq 68\% \text{ if } N > 10)\ , \tag{12.59}$$

which must be associated with a Student's confidence interval (for Gaussian data) or to a Gaussian one if $N > 100$ (for any data). This type of measurements then tends to have zero error.

The starting hypothesis, however, is to have a perfect instrument, that is, with $\Delta = 0$. In practice, the results are presented under the form (12.59) when the sample size is not sufficient to obtain a precision that can compete with the accuracy of the apparatus, i.e. when:

$$\frac{s}{\sqrt{N}} \gg \Delta \, .$$

We emphasize an important point: the interval (12.59) is different from the interval

$$m \pm s \, ,$$

which is an estimate of the interval $\mu \pm \sigma$, giving the probability to obtain a single measurement. Indeed, according to Definition 12.6, the standard deviation s is an estimate of the precision of a single measurement (measurement error), while the quantity $s/\sqrt{N}$ represents the precision of the global measurement. From Eq. (12.59) it results also that, if two measurements M_1 and M_2 have different errors, it is possible to obtain the same final precision from both of them if N_1 and N_2 obey to the relation:

$$\frac{N_1}{N_2} = \frac{s_1^2}{s_2^2} \, .$$

If, for example, $s_1 > s_2$, then $N_1 > N_2$, that is, the number of measures of the experiment with the larger error must be larger (see again Fig. 12.3).

If the x_i measures to be averaged come from different experiments, they will have different precisions s_i, and then the weighted mean formula (10.70) must be used with $m_i = x_i$, $n_i = 1$. In practice, a likelihood function must be considered here as the product of N Gaussian measures, which provided the results $x_i \pm s_i$. With the approximation $s_i \simeq \sigma_i$, we can write the probability of obtaining the observed result according to the likelihood (10.49):

$$L(\mu;\, \boldsymbol{x}) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}\, s_i} \exp\Big(-\frac{1}{2} \frac{(x_i - \mu)^2}{s_i^2} \Big) \right] . \tag{12.60}$$

The maximization of this likelihood is equivalent to the logarithmic negative likelihood (10.6) minimization that in this case coincides with the least squares minimization:

$$-\ln L(\mu) \equiv \mathcal{L}(\mu) = -\sum_{i=1}^{n} \ln \left(\frac{1}{\sqrt{2\pi}\, s_i} \right) + \frac{1}{2} \sum_{i=1}^{N} \frac{(x_i - \mu)^2}{s_i^2} \, . \tag{12.61}$$

To find the point of minimum $\hat{\mu}$, we set $d\mathcal{L}(\mu)/d\mu = d\chi^2/d\mu = 0$, and, with the same procedure of Sect. 10.8, we obtain the result:

$$x = \hat{\mu} \pm \mathrm{Var}[\hat{\mu}] = \frac{\sum_{i=1}^{N} x_i p_i}{\sum_{i=1}^{N} p_i} \pm \sqrt{\frac{1}{\sum_{i=1}^{N} p_i}}, \qquad p_i = \frac{1}{s_i^2}, \tag{12.62}$$

It should be remembered that this formula is strictly valid only for Gaussian measurements, where the interval (12.62) has a 68% coverage. However, even if the data were not Gaussian, for $N > 10$ the Central Limit Theorem holds, as in the case of Eq. (12.59).

Exercise 12.4
A series of measurements of the speed of light taken by different groups gave the following means:

$$\begin{aligned} c_1 &= 2.99 \pm 0.03 \quad 10^{10}\, cm/s \\ c_2 &= 2.996 \pm 0.001 \quad '' \\ c_3 &= 2.991 \pm 0.005 \quad '' \\ c_4 &= 2.97 \pm 0.02 \quad '' \\ c_5 &= 2.973 \pm 0.002 \quad '' \end{aligned}$$

From these data, determine the best estimate of the light speed.

Answer Since the fifth datum is incompatible with the first three, the most likely hypothesis is that the experimenter made a mistake and that only the first four data are correct.

The first four measures have weights:

$$\begin{aligned} p_1 &= 1\,111. \\ p_2 &= 10^6 \\ p_3 &= 40\,000. \\ p_4 &= 2\,500. \\ \sum_{i=1}^{4} p_i &= 1\,043\,611. \end{aligned}$$

(continued)

Exercise 12.4 (continued)
Note the higher values of the weights associated with the more precise data. Applying Eq. (12.62) we obtain:

$$c = (2.99574 \pm 0.00098)\ 10^{10}\ cm/s\ . \tag{12.63}$$

Here it is worth noting a general fact: the error of the weighted average is always lower than the smallest error in the original data. The result can also be presented in the rounded form:

$$c = (2.996 \pm 0.001)\ 10^{10}\ cm/s\ ,$$

which is identical to the result c_2 of the second measurement.

If we had not used the weighted average, we could have used (wrongly) the unweighted formula (12.59). Since the standard deviation of the first four data, based on the second of Eqs. (6.54), is:

$$s(c) = \sqrt{\frac{\sum_{i=1}^{4}(c_i - c)^2}{3}} = 0.01147\ 10^{10}\ cm/s\ ,$$

we would have obtained:

$$c = 2.98675 \pm \frac{s(c)}{\sqrt{4}} = (2.9867 \pm 0.0057)\ 10^{10}\ cm/s\ ,$$

which is a result quite different from the correct one (12.63).

12.13 $M(0,\ \sigma,\ \Delta)$ Measurements

Here a constant physical quantity is measured with an apparatus where both statistical and systematic errors are relevant.

When Δ is the sensitivity interval, repeated measurements provide different values, and the data sample is typically displayed in the histogram form. For obvious reasons, it makes no sense to choose the histogram bin width smaller than sensitivity interval Δ. Histograms can have many or few bins, depending on whether $s \gg \Delta$, $s \simeq \Delta$ or $s \ll \Delta$, where s is the standard deviation due to the random effects (see Fig. 12.9). This class of measurements also includes those of the type $M(0,\ \sigma,\ 0)$ when the data are displayed in a histogram with bin width Δ or measured with an instrument of equal sensitivity. This case corresponds to the systematic error

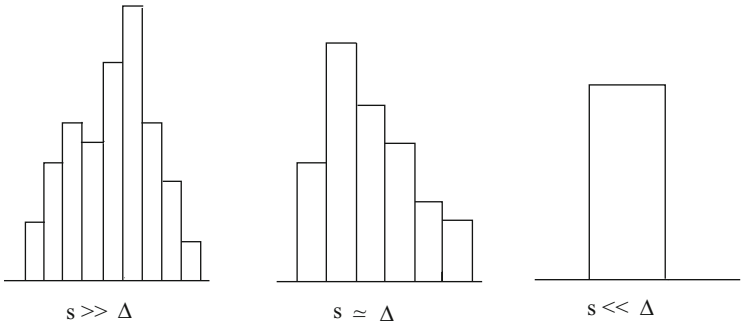


Fig. 12.9 Different types of histograms of bin width Δ when data have a statistical error s

of the first type discussed in Sect. 12.6, which requires only the application of Sheppard's correction (12.7) to the variance. However, it is useful to note that there is no complete agreement among researchers on the use or not of Sheppard's correction (indeed, we believe that many of them even ignore its existence . . .). In fact, the correction is valid only for Gaussian or bell-shaped distributions like that of Fig. 12.4, while for other different distributions, it generally tends to underestimate the variance.

Following the classification of Sect. 12.6, let us now consider the systematic effects of the second type, named offset or additive errors.

In the recent scientific literature, the measurements of the type $M(0\,,\sigma\,,\Delta)$ with offset systematic errors are often reported in the form:

$$x = m \pm \frac{s}{\sqrt{N}}\,(\text{stat}) \pm \frac{\Delta}{2}\,(\text{syst})\,, \tag{12.64}$$

where s is the standard deviation of the sample of N data and Δ is the uncertainty range of the systematic effect correction. The mean m is calculated from the observed mean m' corrected by the systematic effects, $m = m' - c$, where c is the best estimate of the offset value. The standard deviation s is estimated directly from the sample or after Sheppard's correction (12.7). Sometimes more sophisticated methods are used to find $s(\text{stat})$ and $\Delta(\text{sys})$, such as the simulation methods described in Sect. 12.9.

In Eq. (12.64) the confidence level remains undetermined, because the systematic errors are reported with $CL = 100\%$, whereas the statistical ones are usually given with $CL \simeq 68\%$ (if they are Gaussian). The way of combining the two errors is somewhat arbitrary but is generally the one suggested by Eq. (12.11), considering the term σ_x^2 as the variance of the mean of the observed data:

$$x = m \pm \sigma_x = m \pm \sqrt{\frac{s^2}{N} + \frac{\Delta^2}{12}}\,. \tag{12.65}$$

In this case the confidence intervals are approximately Gaussian and are given by Eq. (12.44) and in Table 12.2.

The third case considered in Sect. 12.6 is that related to multiplicative systematic effects, due to scaling or normalization factors. If c is the correction factor of the effect, all data must be divided by this factor. What remains is an uncertainty to which can be attributed a uniform distribution of amplitude $\pm\Delta/2$ and average $\mu_\Delta = 1$. After this correction, and using Eq. (12.19), the result of the measurement can be written as:

$$x = m \pm \sqrt{\frac{s^2}{N} + m^2\frac{\Delta^2}{12} + \frac{s^2}{N}\frac{\Delta^2}{12}}\ , \tag{12.66}$$

where s^2 is the variance of the data sample after the systematic effect correction. The last term under the square root is usually negligible.

As you can see, the procedure that combines statistical and systematic errors is neither univocal nor free from ambiguities. However, there are cases where the method does not present difficulties. A good example is the discovery of the nuclear particle Z^0, which gave to some of its finders the Nobel Prize in Physics in 1984. The two experiments (called UA1 and UA2), which simultaneously discovered the particle at the European CERN laboratories of Geneva, measured the following mass values, expressed in Giga electron-volts (GeV)[1] [Col83a, Col83b]:

$$M_Z = 95.2 \pm 2.5(\text{stat}) \pm 2.8(\text{syst})\ \text{GeV} \quad (\text{UA1})$$
$$M_Z = 91.9 \pm 1.3\ (\text{stat})\ \ \pm 1.4\,(\text{syst})\ \text{GeV} \quad (\text{UA2})\ ,$$

where, in both cases, the systematic error derives from the uncertainty in the absolute calibration in energy of the apparatus. On the other hand, the theory predicted the value

$$92.3 \pm 0.7\ \text{GeV}\ ,$$

where the error is due to the approximations used in the calculations.

In this case, the excellent agreement between theory and experiments is evident, regardless of the specific techniques of data analysis and error handling that could be possibly used.

12.14 $M(f, 0, 0)$ Measurements

Here we refer to the case of a random variable measured without any error.

[1] GeV is an energy unit used in particle physics and is equal to $1.6 \cdot 10^{-10}$ J.

The purpose of the measurement is *the determination of the distribution or statistical law that determines the considered physical phenomenon*. Sometimes it may be sufficient to determine mean and dispersion of this distribution, while in other cases, it is essential to know its precise functional form. For the latter case, think about the Maxwell and Boltzmann densities, which are the basis of statistical mechanics.

Basically, all the stochastic phenomena discussed in the previous chapters belong to this class of measurements and observations, and the methods to determine of means, variances and functional forms are precisely those that have been extensively described in this text. As a significant example, just remember the 10-coin experiment, which has been analysed in Exercise 10.7.

In physical sciences, this type of measurement includes all counting experiments, for which, as discussed in Sect. 3.7, the Poisson distribution plays a fundamental role.

In these cases, it often happens to first count N_{s+b} events from a source within a time period of length t_s; then, after removing the source, N_b background events are, in general, recorded for longer time t_b. The signal/background ratio can be estimated as the number n_σ of standard deviations of the signal over the background (i.e. as the standard Gaussian variable), normalizing the background counts to the time interval t_s. In this step, attention must be paid to the calculation of the background standard deviation, which is $\sigma[N_b] = \sqrt{N_b} t_s/t_b$. This means that the Poissonian error of the measured counts must be evaluated *before* the multiplication by the constant t_s/t_b, according to the error propagation law. In fact, an algebraically manipulated Poissonian variable no longer follows the original distribution.

Therefore, one obtains:

$$n_\sigma = \frac{N_{s+b} - N_b\, t_s/t_b}{\sqrt{N_{s+b} + N_b\, t_s^2/t_b^2}}\,. \tag{12.67}$$

In the search of new phenomena, physicists speak about *strong evidence* when $n_\sigma \simeq 3$, whereas a discovery is claimed when $n_\sigma > 5$. See Problem 12.3 for an application of this formula.

We will now discuss a typical nuclear physics counting experiment.

Exercise 12.5

In 1930 the physicist L.F. Curtiss performed an experiment to determine the statistical law describing the particle emission in the decay of a radioactive nucleus. Using a Geiger counter, he recorded the number of α particles emitted by a thin Polonium film. During the experiment, the number of particles counted in 3455 time intervals of equal length (a few minutes) was recorded. If we define:

(continued)

Exercise 12.5 (continued)

x = number of emitted particles (Geiger counts)

n_x = number of equal time intervals with x counts

The results of the experiment, reported in [Eva55], can be summarized as:

x	0	1	2	3	4	5	6	7
n_x	8	59	177	311	492	528	601	467
x	8	9	10	11	12	13	14	15
n_x	331	220	121	85	24	22	6	3

(12.68)

From this table it results, for instance, that there were 8 intervals without counts, 59 intervals with one count only, 177 intervals with two counts and so on, up to a limit of 3 intervals with 15 counts.

Perform the complete analysis of the experiment.

Answer The Polonium half-life is about 4 months, so the emission intensity of the source can be considered constant during the experimental data collection.

The Geiger counter is a gas tube under electric voltage, in which a discharge occurs when an ionizing α particle crosses the detector. The tube recharging time, needed to have a voltage value between electrodes sufficient to produce a new discharge, is of the order of one thousandth of a second. Since the experiment recorded less than 20 counts in a few minutes, the bias in the counts due to α particles entering the detector during the recharge (dead) time is absolutely negligible. Therefore, we are in the case of an experiment without errors in counting the number of emitted particles, that we previously labelled as $M(f, 0, 0)$.

Then, under the assumptions that the intensity of the Polonium source remains constant, and that all the nuclei of the sample are independent α particle emitters with the same constant probability over time (as in any nuclear model of radioactive decay), the experimental counts must be Poisson distributed. So let us verify this assumption.

We first note that, having grouped the data as in Eq. (12.68), the spectrum of the random variable under examination is given by the number x of recorded counts, while the event frequencies are given by the number of time intervals with a given number x of counts. The spectrum frequencies are therefore given by:

$$f_x = \frac{n_x}{N} , \quad N = \sum_x n_x = 3455 .$$

(continued)

Exercise 12.5 (continued)
Using Eqs. (2.53), (2.55), and (2.58), we can then calculate from the sample mean, variance and fourth-order moment (you can also open and use the R console on your computer):

$$m = \sum_{x=0}^{15} x f_x = 5.877$$

$$s^2 = \frac{N}{N-1} \sum_x (x-m)^2 f_x = 5.859$$

$$D_4 \simeq \frac{1}{N} \sum_x (x-m)^4 f_x = 107.07 \,.$$

The equality between mean and variance is evident, in agreement with the property of the Poisson distribution given by Eq. (3.16). In fact, the statistical estimate of the mean is given by Eq. (6.50):

$$\mu = m \pm \frac{s}{\sqrt{N}} = 5.877 \pm 0.041 \simeq 5.87 \pm 0.04 \,,$$

whereas that on the variance, from Eq. (6.63), is:

$$\sigma^2 = s^2 \pm \sqrt{\frac{D_4 - s^4}{N}} = 5.859 \pm 0.145 \simeq 5.86 \pm 0.14 \,.$$

For this calculation we used the general formula, since the mean value < 10 does allow us, strictly speaking, to apply Eq. (6.79), only valid for Gaussian variables (see Table 6.3). In this case, however, also Eqs. (6.68) (or (6.79)) give a statistically identical result:

$$s^2 \pm s^2 \sqrt{\frac{2}{N-1}} = 5.859 \pm 0.141 \simeq 5.86 \pm 0.14 \,.$$

Confidence intervals for mean and variance can be associated with Gaussian probability levels, because $N = 3455 \gg 100$. Then, we can proceed to the difference test of Eq. (7.6) (neglecting the covariance between M and S^2):

$$t = \frac{|m - s^2|}{\sqrt{s^2(m) + s^2(s^2)}} = \frac{|5.877 - 5.859|}{\sqrt{0.041^2 + 0.145^2}} = 0.12 \,,$$

(continued)

Exercise 12.5 (continued)
which gives a result compatible with $\mu = \sigma^2$ within 0.12 error (standard deviations) units.

From this preliminary analysis, we have obtained a first important verification in favour of the Poisson distribution. We can go on and perform the χ^2 minimization *with respect to* μ. The function to be minimized is given by Eq. (7.36):

$$\chi^2 = \sum \frac{(n_x - Np(x;\mu))^2}{n_x} ,$$

where:

$$p(x;\mu) = \frac{\mu^x}{x!} e^{-\mu}$$

is the Poisson density with an unknown mean μ to be determined.

With our routine `Nlinfit` (see Problem 12.16), we have obtained the following values for the mean and the minimal χ^2:

$$\mu = 5.866 \pm 0.041 , \qquad \chi^2_{min} = 18.64 .$$

To perform the χ^2 test, we first need to determine the number of degrees of freedom. The histogram has 16 channels and was obtained with a predetermined total number of $N = 3455$ events, since this is nothing more than the number of count tests performed by Curtiss. This decreases the degrees of freedom by one unit (if in doubt, re-read Sects. 6.14 and 7.5). Furthermore, the parameter μ has been estimated from the data, which decreases the number of degrees of freedom by another unit, which is therefore equal to $\nu = 16 - 2 = 14$. The reduced chi square χ^2_ν has the value:

$$\chi^2_{14} = \frac{18.64}{14} = 1.33 .$$

From Table E.3 we obtain a p-value of about 15%, in good agreement with the Poisson distribution.

The final data can be summarized in a table or, more briefly, in a graph, reporting, for each value of x, the value of n_x, its statistical error (the error

(continued)

Exercise 12.5 (continued)
bar) and Poisson's law prediction. For example, Eqs. (3.14) and (6.106) give for $x = 5$:

$$n_x \pm \sqrt{n_x} = 528 \pm 23 \,, \quad N\, p(5; 5.866) = 3\,455\, \frac{5.866^5}{5!} e^{-5.866} = 568 \,.$$

The fit result is shown in Fig. 12.10, where an excellent overall agreement between data and theory can be noticed.

The mean value found by the experiment, once normalized to the unit of time and divided by the number of radioactive nuclei present in the source, gives the Polonium α decay constant (with its statistical error). This constant, usually referred to as λ, is a fundamental intrinsic physical characteristic of the Polonium nucleus.

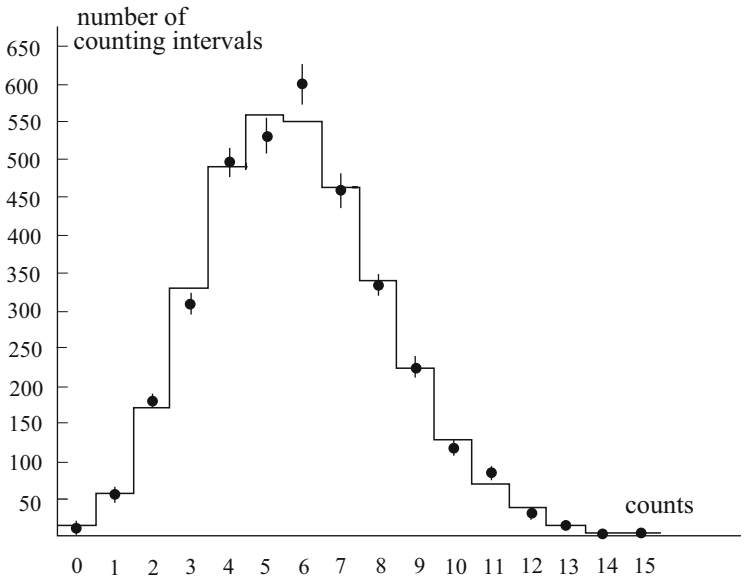


Fig. 12.10 Comparison between data (points with error bars) and estimated Poisson density (solid line histogram) in the case of the Polonium radioactive decay experiment

Exercise 12.6
Two counting experiments of the same phenomenon recorded 92 events in 100 s and 1025 events in 1000 s, respectively. Evaluate the weighted average of the two results.

Answer It is not possible to directly evaluate the weighted average of the raw data, because they refer to different counting times. However, since the counts come from the same source and the second measure is ten times larger, the data normalization can be performed as follows:

$$m_1 \in 92 \pm \sqrt{92} = 92 \pm 9.6$$
$$m_2 \in \frac{1}{10}\,1025 \pm \frac{1}{10}\sqrt{1025} = 102.5 \pm 3.2\,,$$

where Eq. (6.106) has been used and the second measure has been normalized (with error) to the first one. Note that the second measurement gives a more precise result, because, as explained in Sect. 6.14, for Poissonian events the relative statistical error goes as $\sqrt{n}/n$ and then decreases when the sample size increases.

Since the order of magnitude of the recorded counts is a hundred, these Poisson variables (having a mean $\gg 10$) can be considered Gaussian. We can therefore apply the weighted average formula.

The weights of m_1 and m_2 are given by:

$$p_1 = \frac{1}{9.6^2} = 0.011\,, \qquad p_2 = \frac{1}{3.2^2} = 0.098\,,$$

and the estimate of the true mean is:

$$\mu \in \frac{92 \cdot 0.011 + 102.5 \cdot 0.098}{0.011 + 0.098} \pm \sqrt{\frac{1}{0.011 + 0.098}}$$
$$= 101.4 \pm 3.0 \quad \text{counts in 100 seconds.}$$

To conclude, it is worth noting two very general facts: the final data is closer to the most precise measurement (the second, which has a weight ten times greater than the first), and the inclusion of the first measurement, even if much less precise, still reduces the error, even if only slightly.

Note also that, if one merges the counts of the two experiments, a total rate of $1117/1100 \pm \sqrt{(1117)}/1100 = 1.01 \pm 0.03$ counts/s is obtained, in agreement with the previous result.

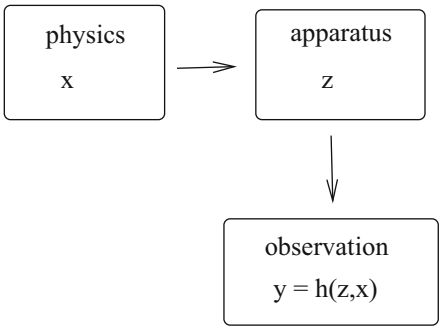


Fig. 12.11 The *folding* effect: the physical quantity x and the apparatus response z combine into a function h, to give a measured value y which is a function of these two variables

12.15 $M(f, \sigma, 0)$, $M(f, 0, \Delta)$ and $M(f, \sigma, \Delta)$ Measurements

The analysis of this class of experiments is very complicated, because the fluctuations of the values assumed by the phenomenon are coupled with the fluctuations and uncertainties due to the measurement apparatus. The goal of the analysis is to determine the function $f(x)$, dependent on one or more variables, which characterizes the physical statistical law describing the observed phenomenon. However, what is directly observed is a density g, dependent on the intrinsic fluctuations both of the observed quantity and of the measurement apparatus.[2] First of all, it is therefore mandatory to experimentally determine the response of the measurement device, called *instrument function* or *apparatus function*, which has the following meaning:

Definition 12.9 (Apparatus Function) The apparatus or instrument function $\delta(y, x)\,\mathrm{d}x\,\mathrm{d}y$ gives the probability that the value of the physical variable is within $[x, x + \mathrm{d}x]$ and that a value within $[y, y + \mathrm{d}y]$ is measured.

Basically, the apparatus function is the probability that the input value is x and the apparatus gives an output quantity y.

The density of the observed data $g(y)$ therefore depends on two variables (Z, X) (apparatus and physics), and the measured values of Y are linked to these variables by the relation (also schematized in Fig. 12.11):

$$y = h(z, x)\,, \quad z = h^{-1}(y, x)\,,$$

where the h function represents the link between z and x due to the measurement operations (a sum $x + z$, a product $x \cdot z$, or other functions).

If now $p_Z(z)$ and $p_X(x) \equiv f(x)$ are the p.d.f. associated with the measurement device and with physics, respectively, Z and X are usually independent, because it

[2] For brevity, we say we observe or measure a density g as a shorthand for observing or measuring a sample from g.

is reasonable to suppose that the behaviour of the apparatus is independent of that of the physical process. Then, Eq. (5.28) holds, and, for convenience, we rewrite it with the new notation now given to variables:

$$g(y) = \int p_Z\left(h^{-1}(y,x)\right) f(x) \frac{\partial h^{-1}}{\partial y} \, \mathrm{d}x \; .$$

The apparatus function can be now identified with the quantity:

$$p_Z\left(h^{-1}(y,x)\right) \frac{\partial h^{-1}}{\partial y} \equiv \delta(y,x) \; ,$$

which is just the density of the apparatus $p_Z(z)$ evaluated for $z = h^{-1}(y,x)$ and both connected to x and y through the measurement procedure. The derivative $\partial h^{-1}/\partial y$ is the Jacobian of the transformation.

Summarizing we can affirm that, based on the laws of compound and total probabilities, at the basis of Eq. (5.28), the three functions $f(x)$ (physics), $g(y)$ (observation) and $\delta(y,x)$ (apparatus) are linked by the relation:

$$g(y) = \int f(x)\, \delta(y,x) \, \mathrm{d}x \equiv f * \delta \; , \tag{12.69}$$

which can be interpreted as follows: the probability of observing a value y is given by the probability that the physical variable assumes a value x times the probability that the instrument, given an input value x, provides a value y. These probabilities must be added (integrated) over all the spectral values of x of the spectrum that can have y as observed value.

Equation (12.69) is called *folding integral*, and it is sometimes indicated with an asterisk symbol. This integral is very important both in physics and engineering. Since $g(y)$ is measured and $\delta(y,x)$ must be known, the experiment must determine the function $f(x)$ which is, so to speak, "trapped" or "wrapped" in the folding integral.

When the apparatus response is linear (as often happens):

$$y = h(z,x) = z + x \; , \quad z = h^{-1}(y,x) = y - x \; , \quad \frac{\partial h^{-1}}{\partial y} = 1 \; ,$$

the folding integral transforms into the convolution integral (5.27):

$$g(y) = \int f(x)\, \delta(y-x) \, \mathrm{d}x \; , \tag{12.70}$$

which is sketched in Fig. 12.12.

The techniques that extract $f(x)$ from the integrals (12.69) and (12.70) are called unfolding and deconvolution techniques. In principle, the technique of Fourier or

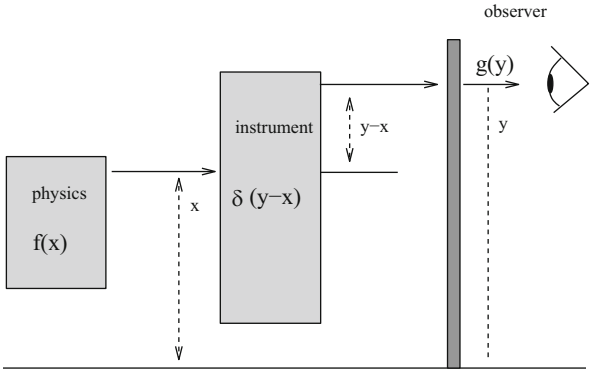


Fig. 12.12 Graphical representation of convolution

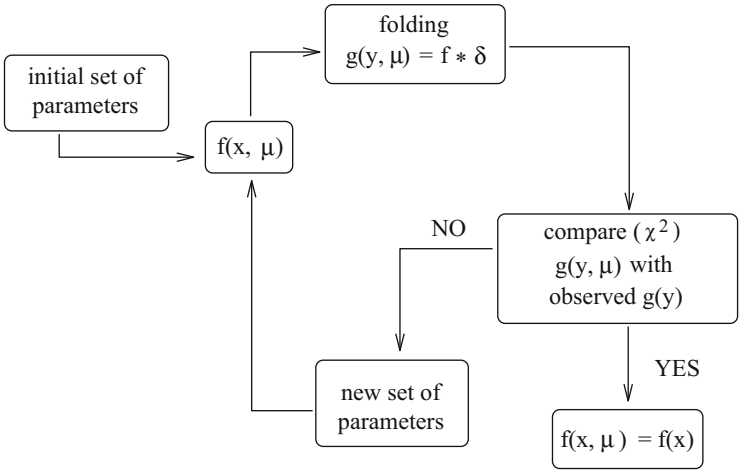


Fig. 12.13 Iterative method to solve a folding integral

Laplace transforms allows us to convert these integrals into easily solvable algebraic equations. However, the error propagation through these algorithms often creates considerable difficulties. Usually the iterative technique schematized in Fig. 12.13 is used. A function $f(x; \mu)$ dependent on one or more parameters μ is "injected" into the folding integral, which is solved with numerical or simulation methods; the obtained solution $g(y; \mu)$, also dependent on the same parameter set, is compared with the function $g(y)$ actually observed. If best-fit methods are used (as those used in the `Nlinfit` routine), the μ parameters at χ^2 minimum are evaluated with the negative gradient method, and the χ^2 test at the end of the procedure allows the determination of the agreement of the folding solutions with the observations. If agreement is not satisfactory, the previous procedure is repeated with a different function and a new set of initial parameters. In this way, we obtain the desired

function $f(x, \mu_0)$, with the optimal values of the parameters μ_0 and their associated error. This method, although rather computationally expensive, has the advantage of being quite flexible and general.

We do not intend to go into further detail here, since a very abundant specialized technical literature is available (see, e.g. [Blo84]). In our opinion, it is sufficient for you to be aware of the problem and to know that there are ways to solve it. If you follow the path of experimental scientific research career, then there is certainly a folding or convolution integral waiting for you. When you stumble on it, do not be discouraged, and search in the literature for the best method to solve that specific problem: it almost certainly exists. If the error calculation or the solution stability is important, remember to pay attention to the type of algorithm to use; this book perhaps will help you make the best choice.

We have shown that the complete and reliable determination of the density $f(x)$ is generally complicated due to the presence of the apparatus function $\delta(y, x)$, which takes into account the instrumental (Δ) and random (σ) effects present in the measurement process. In the case of convolution, i.e. when:

$$\text{observation} = \text{physics} + \text{apparatus} \implies y = x + (y - x) \implies y = x + d \,,$$

if not the complete density structure, at least the mean $\langle X \rangle$ and the variance $\text{Var}[X]$ (or the standard deviation) can be estimated from the sample quantities $m(x)$ and $s^2(x)$ with the known rules (5.67), (6.65), (6.71).

First of all, we note that the average deviation $m(d)$ and the dispersion caused by the measurement operations (the latter characterized by the standard deviation $s^2(d)$ and/or by the systematic error Δ) must be known with negligible error; otherwise the experiment is not feasible. Moreover, also $m(y)$ and $s(y)$ with their uncertainties are known from the observed data. Then, for the $M(f,\ \sigma,\ 0)$ measurements, we have:

$$\begin{aligned} s^2(x) &= s^2(y) - s^2(d) \,, \\ m(x) &= m(y) - m(d) \,, \\ \mu &= m(x) \pm \frac{s(x)}{\sqrt{N}} \,, \\ \sigma &\simeq s(x) \pm \frac{s(x)}{\sqrt{2N}} \,. \end{aligned} \tag{12.71}$$

In the case of the $M(f,\ 0,\ \Delta)$ measurement, we have instead:

$$\begin{aligned} s^2(x) &= s^2(y) - \frac{\Delta^2}{12} \,, \\ m(x) &= m(y) - m(d) \,, \end{aligned}$$

$$\mu = m(x) \pm \frac{s(x)}{\sqrt{N}} \pm \frac{\Delta}{2} \,, \tag{12.72}$$
$$\sigma \simeq s(x) \pm \frac{s(x)}{\sqrt{2N}} \,.$$

Finally, for $M(f,\ \sigma,\ \Delta)$, we have:

$$s^2(x) = s^2(y) - s^2(d) - \frac{\Delta^2}{12} \,,$$
$$m(x) = m(y) - m(d) \,,$$

while Eqs. (12.72) still hold for μ and σ.

12.16 A Case Study: Millikan's Experiments

We think that it is very instructive to apply some of the concepts introduced so far to the analysis of Millikan's famous experiments on the electron charge. In the past years, these experiments have attracted the attention of some historians and critics of science, opening up some controversy on the way scientists operate [Fra84, Fra97]. Let us see what it's about.

Robert Millikan was an American physicist who became famous for his measurements on the electron charge. For these researches he obtained the Nobel Prize in 1923 and numerous other awards, including 20 honorary degrees. In 1910, when he was a professor at the University of Chicago, Millikan published the first results of his experiments, obtained by studying the fall of oil droplets in an electric field. The experimental set-up is shown in Fig. 12.14. Exploiting the Venturi effect, an air flow A captures tiny droplets emitted from the oil ampoule O; these droplets fall by gravity into a hole made in a capacitor C, where there is a static electric field produced by the voltage generator P. The apparatus is kept depressurized and controlled by a pressure gauge. The droplets become negatively electrified by contact with the air flow, and their motion is therefore influenced by the presence of the electric field of the capacitor, whose polarity slows down the motion of the drop, since the lower plate has a negative potential.

It is known that the fall of a body in a viscous fluid (air) becomes rapidly a uniform motion, with constant velocity, when the viscous friction force, which according to Stokes' law depends linearly on the speed, becomes equal to the weight. This is the same motion of raindrops, or of a skydiver, both in free fall and with the parachute open. Using therefore Stokes' law and the equation of the electric field of a capacitor, Millikan wrote the equations of motion of the drops with

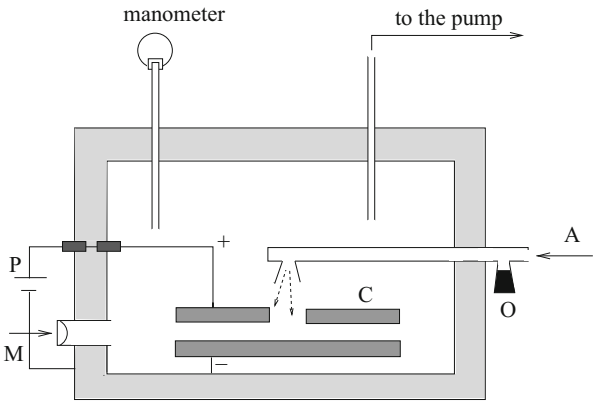


Fig. 12.14 Millikan apparatus for the measurement of the electron charge. A = air flux, C = capacitor, M = microscope, P = high-voltage generator, O = oil ampoule

and without a field as:

$$6\pi\eta v_0 r = \frac{4}{3}\pi r^3 \rho g \quad \text{(without electric field)}\,, \tag{12.73}$$

$$6\pi\eta v r = \frac{4}{3}\pi r^3 \rho g - \frac{V}{h}q \quad \text{(with electric field)}\,, \tag{12.74}$$

where η is the air viscosity coefficient; v and v_0 the drop falling velocity with and without the field, respectively; r is the droplet radius; ρ the oil density; g the gravity acceleration; V the applied voltage; h the distance between the capacitor plates; and q the droplet charge, which is the quantity to be measured.

By measuring the fall velocities v and v_0 with and without field, from Eqs. (12.73) and (12.74), it is possible to obtain the unknowns r and q. The oil drop velocity was evaluated by measuring the time of fall of each drop with a microscope, looking through a lens with notches. The drop velocity could be adjusted at will by changing the voltage V or the air pressure. It was also possible to keep the droplets suspended in the air ($v = 0$) by applying a suitable negative potential to the lower plate of the capacitor.

In this way Millikan noticed that the charge q was always an integer multiple of a base value q_0. He wrote all of his observations in a log that is still available today. The measurement error was assumed to be purely statistical, deriving from the uncertainty in the manual determination of the falling times with the microscope and a clock. Following our notation, we are therefore in the presence of a $M(0,\ \sigma,\ 0)$ measurement.

Millikan averaged a first group of 23 measurements, calculated the error of the average as prescribed by Eq. (12.59) and published in 1913 the value of the elementary electron charge:

$$q_0 \equiv e = 4.778 \pm 0.002\ 10^{-10} \quad \text{esu} \quad . \tag{12.75}$$

A controversy concerning the stability of the measurement and the small error assigned to the value of e prompted Millikan to publish, in 1923, a new study of the fall of 58 drops, in which he emphasized that the data were not belonging "to a selected group of falls, but represented all falls occurring in 60 consecutive days". The value of this second measurement was:

$$e = 4.780 \pm 0.002\ 10^{-10} \quad \text{esu} \quad . \tag{12.76}$$

After a re-examination of Millikan's notes, it appears however that in the first measurement (that of the 23 drops), he excluded from publication 7 measures judged as "bad", while in the second measurement (that of the 58 drops), he even excluded 82 values. Based on these findings, Millikan is often cited as a negative example of scientific dishonesty. However, as he has been resting in peace for many years now, it is not known whether the drops were discarded on good grounds (i.e. measures not properly carried out by him or his assistants) or only with the intention to minimize the error. However, a re-analysis of Millikan's data showed that the discarded drops do not modify the published result in a statistically significant way [Fra84]. At this point you might have wondered what the true value of the electron charge is. Obviously, we do not know the true value, but the weighted average of all the measurements performed to date, which turns out to be:

$$e = 4.803\,206\,8 \pm 0.000\,001\,5\ 10^{-10} \quad \text{esu} \ . \tag{12.77}$$

This value is known with a relative error of $3 \cdot 10^{-7}$ (0.3 parts per million) and can be considered as the true reference value for the present discussion. The comparison between this value and that of Millikan (12.75) can be made with the standard variable:

$$\frac{|\text{Millikan value} - \text{true value}|}{\text{Millikan error}} = \frac{|4.778 - 4.803|}{0.002} = 12.5 \ .$$

The Millikan value is 12.5 error units away from the true one, in absolute violation of the 3σ law. From a formal and methodological point of view, Millikan performed a wrong measurement, because the true value is not covered by the interval (12.75) according to the 3σ law.

However, it should be noted, in favour of the scientist, that Millikan demonstrated for the first time that the electric charge is quantized and determined its value with a relative error of:

$$\frac{|4.778 - 4.803|}{4.803} \simeq 0.005 = 0.5\% \ ,$$

which seems to us a very respectable result, also considering that period of time, the originality of the measurement and the type of apparatus used.

However, the anomalous difference (outside the statistical error) among Millikan results (both original and correct ones) and the true value remains to be clarified.

The reason lies in the physical model used for the measurement, represented by Eqs. (12.73) and (12.74): an approximate η coefficient of air viscosity was used, and, in determining the final value, Millikan did not take into account the systematic error resulting from this approximation.

We conclude this discussion with the words of another Nobel Prize for physics, Richard Feynman [Fey18], who quoted the Millikan measurements as an example of the so-called bandwagon effect, which we will discuss in the next section:

> We have learned a lot from experience about how to handle some of the ways we fool ourselves. One example: Millikan measured the charge on an electron by an experiment with falling oil drops, and got an answer which we now know not to be quite right. It's a little bit off, because he had the incorrect value for the viscosity of air. It's interesting to look at the history of measurements of the charge of the electron, after Millikan. If you plot them as a function of time, you find that one is a little bigger than Millikan's, and the next one's a little bit bigger than that, and the next one's a little bit bigger than that, until finally they settle down to a number which is higher. Why didn't they discover that the new number was higher right away? It's a thing that scientists are ashamed of this history because it's apparent that people did things like this: when they got a number that was too high above Millikan's, they thought something must be wrong and they would look for and find a reason why something might be wrong. When they got a number closer to Millikan's value they didn't look so hard. And so they eliminated the numbers that were too far off, and did other things like that. We've learned those tricks nowadays, and now we don't have that kind of a disease.

12.17 Some Remarks on the Scientific Method

After the description of the technical contents of the measurement operations, we want to conclude the chapter (and the book) with some general observations on the scientific method.

As we mentioned at the beginning, the scientific method is based on the observation of nature, through the measurement of physical observables with the procedures and techniques described in the previous sections.

This approach is based on a principle without which all modern sciences could not exist: the *postulate of objectivity*. In other words, *nature exists by itself*, with its own laws, and is not a projection of the human mind. It is therefore not allowed to attribute to nature the aims and purposes of the subjective and cultural world of the experimenter. Objectivity must be considered as a postulate, since it cannot be demonstrated on the basis of more obvious or elementary assumptions.

However, when this principle is not valid, as happens in some medical experiments, *scientists are able to recognize this fact*. We refer to the so-called placebo effect, which consists in the spontaneous healing of patients who are made to believe that they are being treated with an effective drug, while in reality they have taken a substance (such as water or sugar), called placebo, with no therapeutic effect. As experimentally verified, particular psychological conditions induced in confident patients can lead, for some pathologies, to a sizeable percentage of healings [Bro13].

So how do you check the effectiveness of a drug? The method is precisely that of the contingency tables explained in Exercise 7.8: a placebo is administered to sample A of patients, and the drug to sample B; the χ^2 test is then applied to verify if drug gives statistically different effects from placebo. This methodology is known as "double-blind experiment", because neither the doctor nor the patient know if they are part of the group that is testing the drug or the placebo. Homeopathy and pranotherapy are among the most popular practices that have never passed the double-blind test. It is therefore correct to state that these "therapies" heal; however, it should always be remembered that in these cases therapeutic effects significantly different from the placebo were never measured.

Apart from these situations related to biomedical experimentation, scientific results are absolutely independent of any conscious or unconscious experimenter wish. Of course, it is not excluded that this situation may change one day or another, but up to now, this has never happened in any measurement correctly carried out in the fields of chemistry, physics and biology.

In non-technical language, understandable to any thoughtful people, we could say that the scientific method can be summarized in two equivalent and completely obvious principles, even if unfortunately little used: theories collapse in front of facts (and not vice versa), or similarly reality must be analysed as it is and not as one would like it to be.

Technically, this methodology corresponds to an attitude that in science, following Popper [Pop59], is known as *falsification procedure*: if you have a theory, you must try to prove that it is false; if even a single experimental fact disagrees with it, this theory must be revised. On the other hand, if all the experimental results agree with it, the theory must be (temporarily) accepted. Therefore, there are no true theories, but only incorrect theories (or valid only for a limited range of phenomena), as they have been falsified by one or more experiments, and valid theories that are not yet falsified. The latter, like the theory of relativity and quantum mechanics, are part of the current scientific heritage. We invite you to re-read the epigraphs of Fermi and Einstein at the beginning of Sect. 7.

At this point it should be quite clear that the statistical comparison between an experimental result and a model, through the calculation of the significance level and the evaluation of the probability of making a mistake by rejecting (falsifying) a true hypothesis (type I error), is *a particular but fundamental technical aspect of the general falsification procedure of modern science*. As an example, you can review Exercises 3.17, 7.2, and 12.5.

Another important aspect is that the theories and models, to have scientific validity, *must be falsifiable*, that is, it must be possible in principle to establish that they are false. If I state that when the patient heals the treatment works, while if the patient does not heal it is because he/she is not tuned with cosmic energy, it is clear that the treatment I propose is not falsifiable, since it cannot be disproved by any experiment.

The distinctive feature of the falsification process, and of all persons having a scientific mentality and culture, is to focus attention on rejections and failures rather than on the successes of a theory or hypothesis. This point is fundamental

and by no means obvious. Suppose you make a horoscope and predict that persons with a particular zodiac sign will get a flat tire during a certain week. Among (let's say) 5000 people reading this prediction, it will happen by chance (say) to 10 people. These ten people, if not scientific-minded, will be impressed by your prediction, will become your followers and will spread their positive experience to others. By continuing to make horoscopes, your predictions will always be crowned with success (in a statistical sense); in this way you will gradually accumulate a considerable number of fans and could perhaps become a famous (and rich) astrologer.

The falsification process is not to be understood in a rigid and schematic sense: if an experiment falsifies a theory that has withstood hundreds or thousands of previous experiments, it is reasonably more likely that that single experiment is wrong and the theory valid, rather than the other way around. Frequently, amateur scientists claim sensational discoveries, such as that of perpetual motion or the non-conservation of the electric charge; if these experiments were true, the whole edifice of modern physics would collapse. A lot of carefulness is therefore needed in judging and evaluating new sensational scientific results.

Science can go ahead according to the scheme described so far only if scientists are in good faith and have a reciprocal control and there are mechanisms verifying scientific information. The rules adopted by the professional scientist community require that scientific results, to be considered as such, must be published in "reliable" journals, which are edited by some of the world's leading experts of the subject (all those working in a specific field know who are these experts).

Theories can thus be checked and experiments repeated: then a transitory phase is triggered at the international level, which we could define as validation, at the end of which the theoretical or experimental scientific results can be considered acquired. Statistical data analysis and the comparison among different experiments almost always play a leading role at this point. Scientists therefore operate in a sort of free zone, self-managing scientific information in absolute autonomy. This, in our opinion, is one of the most significant positive features of our times. In order to safeguard that autonomy that science has been able to acquire, it is then very important for every researcher to be aware how important it is to publish a scientific result in a journal or communicate it in a congress and operate with care and correctness.

After the description of the system physiology, we now want to discuss some possible pathologies. Concerning the execution of the experiments, they are essentially of three types:

- Use fraud.
- Treat data incorrectly (data cooking, trimming . . .).
- Fall into the expectation bias (named also bandwagon effect).

We will now briefly review these three pathologies.

The use of fraud in science, that is, the existence of unscrupulous scientists, fortunately proved to be a secondary aspect. The university selection mechanisms and the cross-checking between scientists, through conferences and scientific

journals, have so far made it possible to promptly reveal frauds and tricks. They have mainly occurred in the biomedical sector, where there is a more frequent interplay between science and economic interests. The case of W.T. Summerlin, who in the mid-1970s painted black patches on the fur of white mice to simulate grafts, is often cited as an extreme example [Hix76].

The case of E. Rupp is instead famous in physics. As he later admitted himself, he literally invented, at the beginning of the 1930s, some results on the polarization of double electron scattering. It should be noted that the results of Rupp undoubtedly falsified Dirac's relativistic theory of the electron, which already had several experimental confirmations. However, the physicist community, particularly active and lively, quickly discovered the deception, forcing Rupp to retract his results [Fra84, vD07].

The second pathology, the incorrect data treatment, consists in cleaning operations (*trimming*) and manipulation (*cooking*) of the results and is much more subtle than the previous one, as it is much more difficult to discover. It is not due so much to the bad faith of the experimenter, as to the ignorance of data analysis and processing techniques. The most frequent mistakes concern the failure to correct systematic effects and *the incorrect application of statistics and of techniques that calculate and propagate measurement errors*. Millikan's experiments, discussed in the previous section, fall into this category. In fact, we do not know if Millikan dishonestly discarded some measures; however, it is certain that he did not take into account important systematic effects.

Another aspect of this pathology is the overestimation of measurement errors, to confirm previous measures or theories that are considered as reliable. In this way, possible significant discrepancies between theory and experiment are kept hidden, totally nullifying the principle of falsification. This too demonstrates, once again, how important it is for an experimenter *to be able to correctly evaluate the measurement uncertainties*.

Multiple repetitions of experiments and cross-checks among scientists are the weapons used successfully by the scientific community to identify incorrectly performed experiments or technically wrong theories. But sometimes they are not enough. As an example, let us examine the modern evaluations of one of the most important physical quantities, the Newtonian constant of gravitation G, taken from [RS17] and shown in Fig. 12.15. It is evident that the "Big G" measures are incompatible. Are discrepancies due to miscalculated errors or to a dependence of G on the experimental materials? This is a still open point, linked to the possible presence of an unknown gravitational component of very weak intensity (the so-called "fifth force"), which would deprive G of the characteristic of universality.

And now we come to the most subtle and dangerous effect of all, the expectation bias effect, often referred to as *bandwagon effect* [Jen06], which has already been clearly described in the quote by Feynman reported at the end of the previous section. As you have surely understood, this is the psychological influence on the experimenter of the pre-existing scientific situation. In fact, researchers obtaining results in strong disagreement to accepted theories or previous measurements usually scramble to search for possible experimental biases until when they get

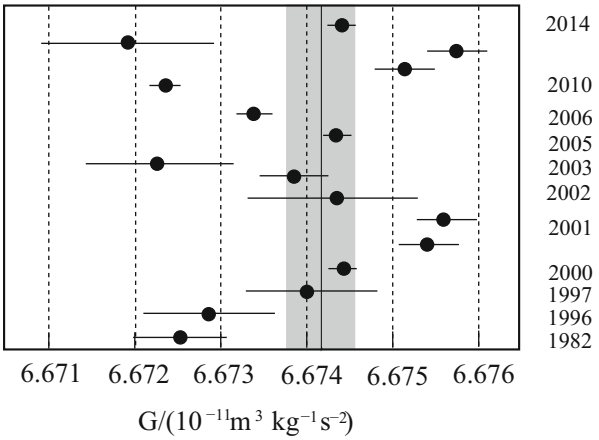


Fig. 12.15 Some recent measurements of the universal gravitation constant G and their measurement uncertainties. The numbers on the right denote the years when the results were published. The vertical black line and the grey band give the recommended value (2014) and its 1σ uncertainty, respectively, (from [RS17])

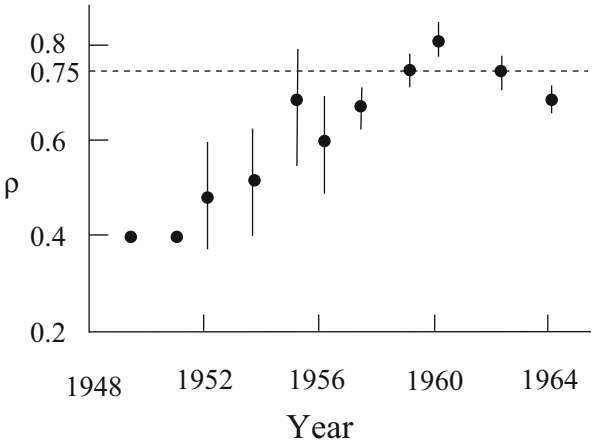


Fig. 12.16 Michel's ρ parameter measurements as a function of the year in which they were performed. Within errors, any measurement is in agreement with the previous one (adapted from [LW65])

a result close to expectations. At this point they are satisfied and become less careful. This, however, can generate situations where the experimental results are all in agreement with each other *and all wrong*. A good example is Fig. 12.16, taken from [LW65], which shows the successive measures of Michel's ρ parameter, characterizing the energy distribution of the electrons emitted in some nuclear decays. Each experiment is in agreement with the previous one, but the first ones are incompatible with the value of 0.75, which is the expected theoretical value.

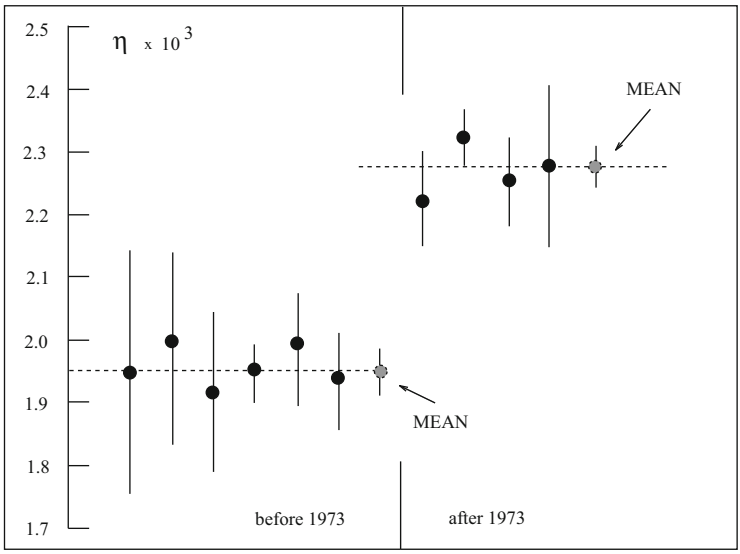


Fig. 12.17 Measures of the parameter η_{+-} ordered following their publication date (from [Jen06])

Moreover, measurements are approaching the theoretical expected value in an increasing monotonous way. Very likely, these results are affected by the bandwagon effect.

Another example, reported in [Fra84], is the measurement of the parameter η_{+-}, which is the ratio between the decays of some unstable nuclear particles, the K mesons. This parameter has a crucial importance in elementary particle physics, and, despite of the experimental difficulties, it has been measured several times over the years. The obtained results are shown in Fig. 12.17: their averages before and after 1973 are incompatible. Also here a bandwagon effect within the two groups of measures seems evident. The currently accepted value (2020, see [ZeaPDG20]) is the weighted average of the second group of measurements:

$$\eta_{+-} = (2.232 \pm 0.011)\, 10^{-3}\,.$$

Unlike previous effects, the expectation bias does not imply bad faith or ignorance of the experimenter. On the contrary, it is a psychological effect which can affect also well trained, very experienced and excellent researchers. As pointed out in Feynman's quote, probably the best remedy against this effect is to be aware that it exists.

We think we have demonstrated that scientists, contrary to popular belief, are not infallible, nor do they consider themselves to be so. However, although we are aware that we are about to make a very demanding statement, we would like to end this book by saying that the *scientific method*, with its technical and fundamental

principles, is flawless in the long run and is able to be immune from possible mistakes of individual scientists.

12.18 Problems

12.1 If the mass m and the acceleration a are independently measured with a relative statistical error of 4% and 5%, respectively, which is the relative statistical error on the force $F = ma$?

12.2 A radioactive source having a mean life $\tau = 5$ days, and with an initial activity of $I_0 = 1000$ decays/s known with negligible uncertainty, has now a residual activity of $I = 10$ decays/s, with an uncertainty of $s_I = \pm 1$ decay/s. Given the decay law $I/I_0 = \exp(-t/\tau)$, determine the elapsed time with the corresponding error.

12.3 A radiation monitor records $I = 157$ counts in 1 s in the presence of the source and $F = 620$ background counts in 10 s without the source. Determine the number of source counts, with its relative error, after background subtraction and the signal/background ratio.

12.4 The expected background of a counting experiment (accurately measured in the calibration phase) is ten events/s. In one test (background plus signal), 25 counts are recorded in a second. Using the Gaussian approximation, find the upper limit of the signal counts only with $CL = 95\%$.

12.5 A scientific result is reported as 5.05 ± 0.04, specifying that it is the average of four measurements and that a $CL = 95\%$ is associated with this interval. Find the accuracy of the single measurement.

12.6 The volume of a cylinder $V = (\pi R^2)L$ is obtained by measuring the base radius R and the height L. To improve the accuracy of the indirect measurement of V, is it more important to reduce the percentage error on R or on L?

12.7 An angle has been measured as $\theta = (30 \pm 2)^\circ$. Calculate $\sin\theta$ together with its error.

12.8 A polarized particle beam is scattered onto a target, and the polarization percentage is measured as $P = (N_+ - N_-)/(N_+ + N_-)$, where N_+ is the number of particles deviated "up", N_- is the number of those deviated "down" and $N = N_+ + N_-$ is the total number of scattered particles. It is well known to physicists that the measured polarization should be written as $P \pm s(P) = P \pm \sqrt{(1 - P^2)/N}$. Check the error formula, both for fixed and variable N. Use the results of Sect. 6.14.

12.9 A voltage is measured with a class 1 instrument as $V = 10.00 \pm 0.05$ V, where the error is the sensitivity interval. Determine the value of the function $f(V) = \exp(0.1\ \text{V})$ with its error. Discuss the confidence levels.

12.10 The output voltage E_2 of a divider with two resistors R_1 and R_2 and input voltage E_1 is given by $E_2 = E_1\, R_1/(R_1 + R_2)$. If $R_1 = R_2 = 1000 \pm 50\ \Omega$ and E_1 is measured, with 1% accuracy, as $E = 10.00 \pm 0.05$ V, calculate E_2. Check the confidence levels with a simulation. The measure at 1% of the output voltage is $E_2 = 4.91 \pm 0.02$ V. Verify if this value is in agreement with the predictions. Note: all uncertainties are maximum sensitivity errors ($CL = 100\%$).

12.11 The constant speed of a slide moving horizontally on a cushion of air is measured as $v = s/t$. The space covered is $s = 2$ m, with an uncertainty estimated at ± 1 mm due to the sensitivity of the used metric tape. The time t, measured by repeating the test 20 times, shows random variations, and its histogram has parameters $m_t = 5.35$ and $s_t = 0.05$ s. Determine the slide speed and comment the obtained result.

12.12 Write the matrix V of Eq. (12.24) for a common systematic error proportional to the measured value x_i ($x_i^S = \epsilon \cdot x_i$).

12.13 Verify Eq. (12.7) using simulation.

12.14 Consider a measurement of the variable $Z = XY$ where $X, Y \sim U(0, 1)$. Find the coverage probabilities of the intervals $\mu \pm K\sigma$ with $K = 1, 2, 3$.

12.15 A rectangular flat counter records 1750 events in 1 s. The counter dimensions are $a = 30 \pm 0.5$ cm and $b = 50 \pm 0.5$ cm, where the error is systematic. Find the counting frequency and its error in ($1/\text{m}^2\text{s}$) units.

12.16 Use the `Nlinfit` routine to estimate the value μ of the Poissonian density of Fig. 12.10.

12.17 The energy of photons from an atomic transition is transformed into electrical pulses whose peak voltage is measured with a multi-channel analyser. On the display monitor, the spectral line shape is a Gaussian with $\sigma \simeq 20$ channels. The calibration of the analyser with a fixed voltage value delivered by a pulse generator gives a Gaussian curve of $\sigma_0 \simeq 5$ channels. This value represents the dispersion introduced by the measurement. Determine the actual line width σ_r.

12.18 Perform the linear best fit to the data of Eq. (12.27) with the covariance method (see Eq. (12.21)) assuming as σ_{sys} a multiplicative error of 20%. Compare the result with that of Table 12.1.

12.19 Which of the two conjectures, (a) "some elephants fly" and (b) "all elephants can fly", has scientific validity?

Appendix A
Table of Symbols

Symbol	Meaning
$\mathbb{R}$, $\mathbb{R}^1$	Real numbers
$\mathbb{R}^n$	Real n-ple
$X, Y, Z, \ldots$	Random variables
$x, y, z, \ldots$	Random variate: observed value (occurrence or outcome) of a random variable in an experiment
$\underline{X}$	n-dimensional random sample of X: $\underline{X} = (X_1, X_2, \ldots X_n)$
$\boldsymbol{X}$	Vector of m random variables: $\boldsymbol{X} = (X_1, X_2, \ldots X_m)$
$\underline{\boldsymbol{X}}$	n-dimensional random sample of $\boldsymbol{X}$: $\underline{\boldsymbol{X}} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \boldsymbol{X}_n)$
$\underline{\boldsymbol{x}}$, $\underline{x}$	Occurrence of a random sample
$\langle X \rangle$, $E[X]$	Mean or expected value operator of X
μ_x	Value of the operator $\langle X \rangle$
$\mathrm{Var}[X]$	Variance operator of X
σ_x^2	Value of the operator $\mathrm{Var}[X]$
$\sigma[X]$	Standard deviation of X: $\sqrt{\mathrm{Var}[X]} \equiv \sigma[X]$
σ_x	Value of the operator $\sigma[X]$
$\mathrm{Cov}[X, Y]$	Covariance operator of X and Y
σ_{xy}	Value of the operator $\mathrm{Cov}[X, Y]$
M	Sample mean
m, $\langle x \rangle$	Observed value of M
S^2	Sample variance
s^2	Value of S^2 after an experiment
S	Sample standard deviation
s	Value of the standard deviation after an experiment or statistical error (called also root mean square)

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2 139, https://doi.org/10.1007/978-3-031-09429-3

Symbol	Meaning
$s_{xy},\ s(x,y)$	Value of Cov$[X,Y]$ after an experiment
$\rho(x,y)$	True value of the correlation coefficient
$r(x,y)$	Sample value of the correlation coefficient
$\simeq$	Approximatively equal
$\sim$	Distributed as
$\#(A)$	Number of elements satisfying the condition A
$\Longrightarrow$	Implies
$p_X(x)$	Probability density function of X
p.d.f.	Probability density function
$F_X(x)$	Cumulative (distribution) function of X
$f(x)$	A function of x
$f(\cdot)$	Functional form of a variable
$N(\mu,\sigma^2)$	Normal or Gaussian distribution of parameters μ e σ
$g(x;\mu,\sigma)$	Probability density function of $X\sim N(\mu,\sigma^2)$
$\Phi(x)$	Cumulative function of $X\sim N(0,1)$
t_α	Quantile of the Student's or Gaussian distributions
$U(a,b)$	Uniform distribution in $[a,b]$
$\chi^2(\nu)$	Chi-square distribution with ν degrees of freedom
$Q(\nu)$	Variable distributed as $\chi^2(\nu)$
χ^2	Values assumed by $Q(\nu)$ after a trial
χ^2_α	Quantile of the χ^2 density
$\chi^2_R(\nu)$	Reduced chi-square distribution with ν degrees of freedom
$Q_R(\nu)$	Variable distributed as $\chi^2_R(\nu)$; $Q_R(\nu)=Q(\nu)/\nu$
χ^2_R	Values assumed by $Q_R(\nu)$ after a trial
$\chi^2_{R\alpha}$	Quantile of the χ^2_R density
$L(\theta;\ \underline{x})$	Likelihood function
$\hat{\theta}$	Maximum likelihood (ML) or least squares (LS) point estimate of a parameter θ

Appendix B
R Software

The R software was created in 1994 in the Statistics Department of the University of Auckland, New Zealand, as an evolution of the S language, a statistical software developed in the Bell laboratories by a research group led by John Chambers. Since 1997, it has become an international standard and is distributed as the Comprehensive R Archive Network (CRAN). To date, it is the most popular and most used statistical software; it is free and open source, i.e. public.

The codes follow standards similar to structured C++ programming, and the routines are extremely easy to use. R now contains thousands of routines and allows you to perform the same calculation in many different ways. We assume that you have installed R on your computer and found one of the many good manuals available on the web, so that you are able to use the R routines and to create simple R scripts.

In this Appendix we collect some methods that we have frequently used throughout the book. They are certainly not the only ones and are perhaps not the best to be used. Nevertheless, they well satisfy one of the purposes of the book, that is, to verify and describe, using simulation and data science techniques, the basic concepts gradually introduced in this text.

The best thing one can do to solve a problem with *R* is to query a web search engine: usually the answer is easily found, because that problem has already been addressed and solved with R by someone else. For example, to sort a vector `v`, the function `sort(v)` can be used; if, on the other hand, one is only interested in having a vector of indices in ascending order, a simple web query will lead to the `order(v)` routine, which gives the sequence of the indices corresponding to the ordered elements of the `v` vector. By typing `?order` in the R console, one will have access to the R html manual which will explain the details of the routine. The R's prevailing type of learning is inductive and bottom-up, as is often the case in computer science.

Let us now briefly summarize some ways of using R presented in the book.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2 139, https://doi.org/10.1007/978-3-031-09429-3

R's prompt is >, and comments are identified by the character #. The base object of R is the vector, which is sized by what is on the right, which often uses the `c()` function. For example, with the command:

```
> x <- c(1,3,2,5,7)
```

a numeric vector `x` of five entries is defined. The `<-` symbol is often used for assignment, which is equivalent to =. With the command `?c()`, the numerous possibilities of this function are listed. The vector `x` thus created can be mathematically manipulated without difficulty.

For example, the command `y <- x*x` or `y <- x^2` creates a vector `y` containing the squares of the entries of x, whereas `y <- sqrt(x)` creates a vector of square roots and so on. The symbols `"..."` and `'...'` are equivalent. There are many useful functions that manipulate vectors. Some of them are reported in Table B.1.

Matrices can be created in different ways; the simplest one is perhaps to use the function `matrix()`. For example, the command;

```
> x <- matrix(c(1,2,3,4,5,6),nrow=2,byrow=TRUE)
```

generates the matrix:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} . \tag{B.1}$$

The command `?matrix()` lists all the methods available to generate a matrix. In the text we have sometimes used the functions `cbind()` and `rbind()`, which join matrices by columns and rows, respectively. As an example, the matrix (B.1) can be generated with the command:

```
> x <- rbind(c(1,2,3),c(4,5,6)) .
```

In statistics, it is often necessary to sum across the rows or columns of a matrix or table. In this case it is better to use the function `apply()`, as in our routine `CorrelEstH`. For example, the command `sr <- apply(x,1,sum)` creates

Table B.1 Some R functions acting on a vector `x`

Function	Result
`range(x)`	Two-entry vector `c(min(x),max(x))`
`which.min(x)`	Integer, index of the minimum
`which.max(x)`	Integer, index of the maximum
`sort(x)`	Vector sorted in ascending order
`sort(x,dec=T)`	Vector ordered in descending order
`order(x)`	Vector of indices of `x` giving the increasing order
`length(x)`	Integer, length of `x`
`sum(x)`	Sum of the entries of `x`
`prod(x)`	Product of the entries of `x`
`mean(x)`	Mean of the entries of `x`
`var(x)`	Variance of the entries of `x`

Table B.2 R functions for statistical distributions

Prefixes	Function
`d`	Density value
`p`	Cumulative function value
`q`	Quantile value
`r`	Random variate
Suffixes	*Distribution*
`unif`	Uniform
`binom`	Binomial
`norm`	Gaussian
`pois(lambda)`	Poissonian with mean `lambda`
`chisq`	χ^2
`gamma(shape=1)`	Negative exponential
`gamma(shape=n)`	Erlang distribution of order `n`
`rayleigh(scale)`	Rayleigh distribution with $\sigma =$ `scale`

a vector `sr = c(6,15)` with the sum across the columns of the matrix (B.1), whereas the command `> sc <- apply(x,2,sum)` creates a vector `sc = c(5,7,9)` with the sum across the rows.

The R software has a lot of graphics options; in the book we have frequently used the routine `hist(x)` which creates the histogram from the raw data contained in `x` and the function `plot()`. These routines have many options that can be listed by typing `?plot` or `?hist`. The simplest command to have a line plot of a function with abscissae and ordinates respectively contained into the vectors `x` and `y` is `plot(x,y,type="l")`. The R software obviously contains functions handling all the most frequently used p.d.f. The name semantics is explained in Table B.2: the prefixes `d, p, q, r` are positioned at the beginning of the names, followed by the suffix which indicates the distribution type. For example, if `x` is a vector containing the support values, `dnorm(x)` gives a vector containing the corresponding ordinates of the standard Gaussian; `x <- rpois(100,lambda=2)` generates a vector `x` of 100 Poissonian random variates with mean $\mu = 2$; `qnorm(0.95)` gives as output 1.6448, which is the Gaussian quantile corresponding to a 95% cumulative probability. Most of these routines have default values that can be changed by following the instruction manual.

For the standard Gaussian, `..norm()` has the default values $\mu = 0, \sigma = 1$, which, for example, can be modified by writing `..norm(..,mean=3,sd=2)`.

While studying distributions, we often used the function, `density(x)`, a powerful routine which produces density functions starting from a raw data vector `x`,

by applying smoothing techniques.[1] Without going into the complex mathematical details of the routine, which can be found in the R manual, it is sufficient for the user to know that smoothing can be controlled with the parameter `adj`, which is 1 by default. A value `adj = 0.01` practically reproduces the original data distribution, which is gradually smoothed by increasing the parameter value. The following instructions visualize the uniform distribution, essentially a smooth histogram connected by continuous lines, of 200 data randomly generated from the $U(0, 1)$ p.d.f.:

```
x <- runif(200)
plot(density(x,adj=0.01),type='l')
grid()
```

With the command `grid()`, a grid is superimposed to the plot. It is instructive to vary the `adj` parameter to verify how it affects the smoothing.

In the case of joint two-dimensional distributions, if the raw data are contained in a two-column array `x` of pairs, the routine `bkde2D` can be used as follows:

```
    # 1000 gaussian pairs with
    # meanx=5, meany=10, varx=vary=3 and varxy=-2
require(mvtnorm)
x <- rmvnorm(1000,c(5,10),sigma=rbind(c(3,-2),c(-2,3)))
    # plot with persp(x1,x2,fhat)
require(KernSmooth)
hsm <- bkde2D(x,adj=0.01)
persp(hsm$x1,hsm$x2,hsm$fhat)
```

where again `adj` is the smoothing parameter and `persp(x,y,freq)` draws the bidimensional plot of the curve.

Often in R the list of objects is used: for example, an object `alis` can be defined as `alis<-list(a=1,b=2,c=3)`, and then the list members can be recalled using the symbol `$`; with the command `alis$c`, the answer is `3`. Often the routines give an output list as `return`: it is therefore necessary to list the names of these variables (usually declared as `Value` in the user manual) and access them out with `$`. For example, if we make the histogram of a data vector `x`, the command `hist(x)$counts` will return a vector with the number of events in each bin.

[1] The smoothing procedure replaces the original point of a function with the mean of the neighboring points. The function, therefore, appears "smoother" as the number of points used for the average increases.

Appendix C
Moment-Generating Functions

In this appendix we briefly mention one of the most important methods for the study of distributions, the method of generating functions, which allows to solve many problems in a concise and elegant way. The method associates a new variable to the random variable X, defined by the function e^{tX}, where t is a real variable that has no particular statistical meaning. The average value $\left\langle e^{tX} \right\rangle$ is a function $G_X(t)$ which takes the name of moment-generating function (Mgf):

$$G_X(t) = \left\langle e^{tX} \right\rangle = \begin{cases} \sum_i p_i e^{t x_i} & \text{discrete variables} \\ \\ \int e^{tx} p(x)\, \mathrm{d}x & \text{continuous variables} \end{cases} \tag{C.1}$$

For $t = 0$, the sum and the integral are obviously always convergent. For $t \neq 0$, Eq. (C.1) may not converge. Here we will deal with some cases, related to the most important distributions, in which there is convergence for values $|t| < M$, where M is an arbitrary positive number. For continuous variables the generating function, apart from a sign in the exponent appearing in the second of Eq. (C.1), it is the Laplace transform of the density of X. Expanding G_X into series and exploiting the linearity properties of the mean, one obtains:

$$G_X(t) = 1 + t \langle X \rangle + \frac{t^2}{2} \left\langle X^2 \right\rangle + \frac{t^3}{6} \left\langle X^3 \right\rangle + \dots . \tag{C.2}$$

From this result the important relation:

$$\left\langle X^n \right\rangle = \left. \frac{\mathrm{d}^n G_X}{\mathrm{d}t^n} \right|_0 = G_X^{(n)}(0) , \tag{C.3}$$

follows, which connects the n-th derivative of G at the origin to the n-th order moment of X.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3

Table C.1
Moment-generating function (Mgf) of some probability distributions

p.d.f.	Generating function $G_X(t)$
Binomial $b(x; n, p)$	$[pe^t + (1-p)]^n$
Poissonian $p(x; \mu)$	$\exp[\mu(e^t - 1)]$
Exponential $e(x; \lambda)$	$\lambda/(\lambda - t)\ ,\ \ t < \lambda$
Gaussian $g(x; \mu, \sigma)$	$\exp\left(\mu t + \sigma^2 t^2/2\right)$
Chi-square $\chi^2(\nu)$	$(1-2t)^{-\nu/2}$

By solving the integrals or sums of Eq. (C.1), one can easily obtain the generating functions of the most important probability distributions. Some of these are shown in Table C.1.

The importance of the moment-generating function is due to the following property: if $Y = X_1 + X_2$ and the random variables X_1 and X_2 are independent, then from Eq. (4.9) it follows that the Mgf of Y:

$$G_Y(t) = \left\langle e^{t(X_1+X_2)} \right\rangle = \left\langle e^{tX_1} \right\rangle \left\langle e^{tX_2} \right\rangle = G_{X_1}(t) G_{X_2}(t)\ , \tag{C.4}$$

is the product of the Mgf of X_1 and X_2. Obviously this property also holds for a sum of k densities, so that the Mgf $G_{T_k}(y)$ of the Erlang density (3.57), which is the sum of k exponential times, results in:

$$G_{T_k}(t) = \prod_{i=1}^{k} G_{T_i}(t) = \left(\frac{\lambda}{\lambda - t}\right)^k\ ,\ \lambda > t\ , \tag{C.5}$$

where the exponential Mgf of Table C.1 has been used. This property can also be checked directly by solving the integral of Eq. (C.1) for the Erlang distribution (3.57). If we denote the time variable by x instead of t to avoid notation conflicts with the Mgf variable, we can write:

$$G_{T_k}(t) = \frac{\lambda}{(k-1)!} \int_0^\infty (\lambda x)^{k-1} \exp[-(\lambda - t)x]\, \mathrm{d}x = \left(\frac{\lambda}{\lambda - t}\right)^k\ , \tag{C.6}$$

where the standard result has been used:

$$\int_0^\infty x^k \exp(-ax)\, \mathrm{d}x = a^{-(k+1)}(k+1)!\ ,\ a > 0\ ,\ k \text{ integer}\ .$$

The Central Limit Theorem 3.1 for a sum of n variables X_i all having the same density of parameters μ and σ can be easily proved with the use of Mgf. Indeed, if:

$$Y = \frac{\sum_i X_i/n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \frac{\sum_i (X_i - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_i Z_i\ ,\ \ Z_i = \frac{X_i - \mu}{\sigma}\ , \tag{C.7}$$

from Eq. (C.4) one obtains:

$$G_Y(t) = \left\langle e^{tY} \right\rangle = \prod_i \left\langle \exp\left(\frac{t}{\sqrt{n}} Z_i\right) \right\rangle$$
$$= \left[1 + \frac{t}{\sqrt{n}} \langle Z \rangle + \left(\frac{t}{\sqrt{n}}\right)^2 \frac{\langle Z^2 \rangle}{2!} + \left(\frac{t}{\sqrt{n}}\right)^3 \frac{\langle Z^3 \rangle}{3!} + \dots \right]^n . \quad \text{(C.8)}$$

Since $\langle Z \rangle = 0$ and $\langle Z^2 \rangle = 1$, passing to the limit the result:

$$\lim_{n\to\infty} G_Y(t) = \lim_{n\to\infty} \left(1 + \frac{1}{n}\frac{t^2}{2}\right)^n = e^{t^2/2} \quad \text{(C.9)}$$

is obtained showing that, for large n, the Mgf of Y tends to that of a standard Gaussian density (see Table C.1).

Also the additivity theorem (3.4) for χ^2 variables can be easily proved with Mgf. If:

$$Y = Q(\nu_1) + Q(\nu_2) \quad \text{(C.10)}$$

and $Q(\nu_1) \sim \chi^2(\nu_1)$, $Q(\nu_2) \sim \chi^2(\nu_2)$ are independent, then $G_{\nu_1}(t) G_{\nu_2}(t) = G_Y(t)$ so that

$$G_Y(t) = (1-2t)^{-(\nu_1+\nu_2)/2} \implies Y \sim \chi^2(\nu_3) \quad \nu_3 = \nu_1 + \nu_2 \quad \text{(C.11)}$$

from Table C.1. If instead $Q(\nu_1) \sim \chi^2(\nu_1)$ and $Q(\nu_2) \sim \chi^2(\nu_2)$, it is easy to show, by inverting Eq. (C.4), that, when Y and $Q(\nu_2)$ are independent, the following property holds:

$$Q(\nu_1) = Y + Q(\nu_2) \implies Y \sim \chi^2(\nu_1 - \nu_2) . \quad \text{(C.12)}$$

Appendix D
Solutions of Problems

Chapter 1

1.1. If C is the change of door, A "is the car behind the first chosen door" and $\bar{C}$ and $\bar{A}$ the complementary events, from the partition theorem, one has: $P(C) = P(C|A)P(A) + P(C|\bar{A})P(\bar{A}) = 0 \cdot 1/3 + 1 \cdot 2/3 = 2/3$, $P(\bar{C}) = P(\bar{C}|A)P(A) + P(\bar{C}|\bar{A})P(\bar{A}) = 1 \cdot 1/3 + 0 \cdot 2/3 = 1/3$. It is better to change the door.

1.2. From Eq. (1.31) it results that the number of possible games is $52!/(13!)^4 \simeq 5.36 \cdot 10^{28}$. Since the number of games played is $\simeq 5.475 \cdot 10^{14}$, $P \simeq 1.02 \cdot 10^{-14}$.

1.3. $P = p_1[1 - (1 - p_2)(1 - p_3)] = 0.776$.

1.4. (a) Elements 1 and 2 are in parallel, and they are in series with the parallel combination of elements 3 and 4. (b) $P = [1 - (1 - p)^2]^2 = 0.922$.

1.5. $P = 1 - (5/6)^3 = 0.421$.

1.6. (a) $P = 1 - 7/10 = 0.30$; (b) $P = 1 - 120/720 = 0.83$.

1.7. We outline the not simple solution: one can imagine the arrival of one of the two friends as a point x or y located at random within a 60 min long interval. X and Y do not meet if the $\{X \notin [y, y + 12], Y \notin [x, x + 10]\}$ event occurs. So $P = 1 - (48/60)(50/60) = 1 - 2/3 = 1/3$.

1.8. From the total probabilities formula: $P(T) = 0.14$. From Bayes theorem: $P(B|T) = 0.678$. Also the graphic method of Fig. 1.6 can be used.

1.9. $P\{X \leq Y\} = 1/2$. Indeed, it is reasonable to assume the probability as the ratio of the area above the diagonal to the total area of a unit square.

1.10. With obvious notation:

$$P(A)P(D|A) + P(B)P(D|B) + P(C)P(D|C) = 0.165 \rightarrow 16.5\%$$

1.11. $P(C \geq R|C \geq 150, R \leq 155)P(C \geq 150, R \leq 155) = (1/2)(5/15)(5/20) = 0.0417$.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3

1.12. One obtains $P(H_1) \simeq P(H_2) \simeq 0$, $P(H_3) = 0.03$, $P(H_4) = 0.22$, $P(H_5) = 0.47$, $P(H_6) = 0.28$. Compare the results with those of Table 1.2.

1.13. From Eq. (1.41), if V_n is the event where the friend wins n consecutive games ($V_1 \equiv V$), defining the hypotheses B = cheat and O = honest, assuming $P(B) = P(O) = 0.5$, and $P(V|O) = 0.5$, $P(V|B) = 1$, the following recursive formula is obtained: $P(B|V_n) = P(B|V_{n-1})/[P(B|V_{n-1}) + 0.5(1 - P(B|V_{n-1}))]$. One obtains $P(B|V_5) = 0.97$, $P(B|V_{10}) = 0.999$, $P(B|V_{15}) = 0.99997$. See how the results change as $P(B)$ and $P(O)$ change.

1.14. (1) True, (2) false, (3) true, (4) false.

1.15. $P(21|20)P(20) = 0.2 \cdot 0.4 = 0.08$, $P(21|21)P(21) = 0.6 \cdot 0.4 = 0.24$, $P(21|22)P(22) = 0.2 \cdot 0.1 = 0.02$, from which $P(20|21) = 0.24$, $P(21|21) = 0.70$, $P(22|21) = 0.06$ from the Bayes formula.

1.16. We need to find the probability $P(H|V)$ that a winner is honest, knowing that the probability that an honest man will be a winner is $P(V|H) = 10^{-6}$. The two probabilities are connected by the Bayes theorem: $P(H|V) = P(V|H)P(H)/[P(V|H)P(H)+P(V|D)P(D)]$. The probability of winning dishonestly is $P(V|D) = 1$; however, if we assume that there are no dishonest players cheating, $P(D) = 0$, and it turns out that, as it should be, $P(H|V) = 1$.

1.17. We must find the probability $P(C|DNA)$ that a positive person is guilty, in a test where the probability for an innocent is $P(DNA|I) = 10^{-4}$. Assuming that in the tested population the probability that a person is guilty holds true is $P(C) = 0.5 \cdot 10^{-4}$ and that is innocent is instead $P(I) = 1 - 10^{-4} \simeq 1$, from Bayes theorem one obtains $P(C|DNA) = P(DNA|C)P(C)/[P(DNA|C)P(C)+P(DNA|I)P(I)] = 1 \cdot 0.5 \cdot 10^{-4}/[1 \cdot 0.5 \cdot 10^{-4} + 10^{-4} \cdot 1] = 0.33$.

1.18. The probability to extract 2 identical cards (apart from the suits) from a deck of 52 cards is given by the hypergeometric law (1.33) with $a = 4$, $b = 52$, $k = 2, n = 2, n - k = 0$:

$$\frac{4!}{2!\,2!}\frac{2!50!}{52!} = \frac{4!}{2!\,52 \cdot 51} = \frac{12}{52 \cdot 51}\,.$$

The probability to get a particular face throwing a dice is $1/6$. The probability of the event is then the product of the probabilities:

$$\frac{12}{52 \cdot 51}\frac{1}{6} = \frac{2}{52 \cdot 51} = \frac{1}{1326} \simeq 7.54 \cdot 10^{-4}\,.$$

Chapter 2

2.1. $b(2; 3, 1/6) = 5/72$.

2.2. The number of favourable cases is $10!/(3!7!) = 120$ and that of the possible ones is $2^{10} = 1024$. The probability is $720/1024 = 0.117$, according to the binomial density.

2.3. From the binomial density: $1-b(0;\,10,\,0.1)-b(1;\,10,\,0.1)-b(2;\,10,\,0.1)=0.0702$, that is about 7%.

2.4. $\langle X+Y\rangle=\langle X\rangle+\langle Y\rangle=100\cdot 18/37+0\cdot 19/37+360\cdot 1/37+0\cdot 36/37=58.4$. If you imagine a very large set (tending to ∞) of N players, each with a starting capital of 60 euros, after a bet the average capital of this set will have dropped to 58.4 euros, with an average profit in favour of the dealer of 1.6 euro per player.

2.5. $\langle X\rangle=6$, $\sigma[X]=2.83$, $\langle Y\rangle=13$, $\sigma[Y]=5.66$.

2.6. The solution is given by the hypergeometric law (1.33). The R routine `dhyper` can be used with the call `dyper(k,7,3,3)` with `k=0,1,2,3`. The results of the following table are obtained:

r	0	1	2	3
$p(r)$	6/720	126/720	378/720	210/720

This is an example of a hypergeometric distribution. Note that the density is normalized. $\langle R\rangle=0.9$, $\sigma[R]=0.7$.

2.7. $p(x)=2x$, $F(x)=x^2$, $\langle X\rangle=2/3$, $\sigma[X]=1/(3\sqrt{2})$.

2.8. $F(x_{0.25})=x_{0.25}^2=0.25$, from which $x_{0.25}=0.5$. The probability to observe values ≤ 0.5 is 25%.

2.9. The journey takes 6 h, 4 on the way and 2 h on the way back. If the velocity V is considered as a two-valued variable, one has $\langle V\rangle=25\cdot 4/6+50\cdot 2/6=200/6=33.3$ km/h. Note that the result $\langle V\rangle=(50+25)/2=37.5$ km/h is wrong.

2.10. The trials are 1000 (instead of 100), and the spectrum values are two: $x=0,\,1$ (heads or tails, instead of the 11 values $x=0,\,1,\,\ldots\,10$). From Table 2.2, by summing the products of the first column by the second one $(0\cdot 0+1\cdot 0+2\cdot 5+3\cdot 13+\ldots)$, we obtain 521 heads and, obviously, 479 tails, to be compared with an expected value of $1000\cdot b(1,0.5,0)=1000\cdot b(1,0.5,1)=1000\cdot 0.5=500$.

2.11. 100 sorted binomial variates are generated with the calls:

```
x <−sort(rbinom(100, size = 20, prob = 0.3)),
y <−sort(rbinom(100, size = 20, prob = 0.3))
```

and the plot is generated with the call `qqplot(x,y)`, in which the irregular structure is due to the presence of repeated discrete values. The comparison with the theoretical distribution requires first the construction of a vector `x1` with cumulative probabilities associated with the discrete values `x`. This can be achieved with the call:

```
for(k in 1:length(x))x1[k]=length(which(x<=x[k]))
/length(x).
```

The expected quantiles are obtained with `qth<-qbinom(x1,size=20, prob=0.3)`.
With the command `qth<-qth[1:(length(qth)-1)]`, the last value (= 1 by construction) is excluded. Finally, the plot is obtained with `qqplot(qth,x)`. It is useful, using the plot, to examine the effect of varying `prob` in the calculation of `qth`.

Chapter 3

3.1. $\sum_x P\{Y \geq X | X = x\} P\{X = x\} \to \int_0^1 (1 - x)\, \mathrm{d}x = 1/2$.

3.2. The density of X is binomial (Gaussian), so that $\langle X \rangle = 500 \cdot 0.5 = 250$, $\sigma[X] = \sqrt{500 \cdot 0.25} = 11.18$. The 3-sigma law is valid.

3.3. After n steps, the path X has a spectrum $X = n, n - 2, \ldots, -n$ of discrete values which differ from each other by two units. The variable $Y = (X+n)/2$ is binomial and assumes integer values within the interval $[0, n]$. Therefore, one has $\langle Y \rangle = (\langle X \rangle + n)/2 = np = n/2$ and $\mathrm{Var}[Y] = \mathrm{Var}[X]/4 = np(1 - p) = 125$. Finally, it results: $\langle X \rangle = 0$ and $\sigma[X] = 2\sqrt{125} = 22.36$. The 3-sigma law is valid.

3.4. (a) 0.01576; (b) 0.016.

3.5. $\mu = 200 \cdot 0.2 = 40$, $\sigma = \sqrt{200 \cdot 0.2 \cdot 0.8} = 5.66$. Using the Gaussian approximation and Table E.1, one has: $P\{-1.77 < t < 1.77\} = 0.923$.

3.6.

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left[-\frac{1}{2\sigma^2}(\ln y - \mu)^2\right], \quad y > 0\,.$$

It is enough to note that $\ln Y$ is Gaussian and to remember that $\mathrm{d} \ln y = \mathrm{d}y/y$. This p.d.f. is known as log-normal distribution.

3.7. FWHM=$2.355\,\sigma$.

3.8. (a) $0.0695 \simeq 7\%$; (b) $0.7165 \simeq 70\%$, independently of the previous 8 months.

3.9. $\chi^2/\nu = 7/10 = 0.7$, corresponding to a probability (significance level) $\simeq 28\%$, given by Table E.3.

3.10. $(8500 - 8100)/\sqrt{8500} = 4.34$. The decrease is significant.

3.11. $1 - \exp(-100 \cdot 0.001) = 0.0952 \simeq 9.5\%$.

3.12. From Poisson density: $P\{X = 0\} = \exp(-10/2) = 0.0067$. The probability to be wrong is then $\simeq 0.7\%$.

3.13. If $\{X_1 + X_2 = n\}$ is the sum of counts recorded in disjoint time intervals, from Newton's binomial formula, it is possible to prove that $P\{X_1 + X_2 = n\} = \sum_k [P\{X_1 = k\} P\{X_2 = n - k\}]$ when the probabilities are calculated with the Poisson density (3.47).

3.14. $[-0.675, 0.675]$, by linear interpolation from Table E.1.

3.15. From the interpolation of Table E.1, one obtains $P\{t > 0.806\} = 0.21$ and $P\{t > 1.55\} = 0.06$. From the values $(4.41 - \mu)/\sigma = 0.806$ and $(6.66 - \mu)/\sigma = 1.55$, one obtains $\mu = 2$ and $\sigma = 3$. A numerical check can be done with the R function `pnorm`.

3.16. Using the hour as time unit, for $t = 1$ hour, $\lambda t = 100/120 = 0.8333$, and hence, from the Poisson p.d.f: $p(4) = (\lambda t)^4 \exp(-\lambda t)/4! = 0.00873$. The waiting time is then given by 1/0.00873=114.5 hours corresponding to $4.77 \simeq 5$ days.

3.17. The cumulative function is $F(x) = \int_1^x (2x - 2)\,\mathrm{d}x = x^2 - 2x + 1$. From Theorem 3.5, $X^2 - 2X + 1 = R$ where $R \sim U(0, 1)$. The second degree equation, for $1 \leq x \leq 2$, gives the solution $X = 1 + \sqrt{R}$. X is generated according to the assigned density with the command $\texttt{X} = \texttt{1} + \sqrt{\texttt{Random}}$, where $0 \leq \texttt{Random} \leq 1$ has a uniform distribution.

3.18. (a) 4.6% from Table E.1; (b) since $\mathrm{Var}[\sum_{i=1}^{10} x_i] = 10\,\sigma^2[x] = 250$, one has $\sigma[\sum_{i=1}^{10} x_i] = 15.8$ and $t = |1050 - 1000|/15.8 = 3.16$ corresponding to a probability $8\,10^{-4}$; (c) $< 25\%$, from Chebyshev's inequality.

3.19. $< X >= 1.5$, $< X^2 >= 2.9$. $< 30\,X >= 45$, $\sigma[30\,X] = 24$; $(80-45)/24 = 1.46$, corresponding, from Table E.1, to a probability of 7.2%.

3.20. The number of defects is $np \pm \sqrt{np(1-p)} = 10.0 \pm 3.1$, so that $t = |16 - 10|/3.1 = 1.93$. From Table E.1, it results that to this value gives a right-tail probability of 2.7%, which is the requested percentage.

Chapter 4

4.1. $p(x, y)\,\mathrm{d}x\,\mathrm{d}y = P\{x \leq X \leq x + \mathrm{d}x, y \leq Y \leq y + \mathrm{d}y\} = \mathrm{d}x\,\mathrm{d}y/(ab)$ for $0 \leq x \leq a, 0 \leq y \leq b$; $p(x, y) = 0$ otherwise.

4.2. (a) $(X_1 - \mu_1)^2/\sigma_1^2 = 2.71$; (b) $\sum_{i=1}^{2}(X_i - \mu_i)^2/\sigma_i^2 = 4.61$; (c) $\sum_{i=1}^{3}(X_i - \mu_i)^2/\sigma_i^2 = 6.25$. The values are obtained from Table E.4.

4.3. From Theorem 4.1 and Eqs. (4.14) and (4.16), it results that the inequality is valid.

4.4. The tangents as in Fig. 4.5 are traced, and the intersection points (x_2, y_2) and (x_1, y_1) are found. From Eq. (4.55) one then has $(\sigma_x/\sigma_y)\,(y_2 - y_1)/(x_2 - x_1) = \rho$.

4.5. $\mathrm{Cov}[Z, U] = \langle ZU \rangle - \langle Z \rangle \langle U \rangle = \langle (aX + b)(cY + d) \rangle - (a \langle X \rangle + b)(c \langle Y \rangle + d) = ac(\langle XY \rangle - \langle X \rangle \langle Y \rangle)$. Since $\sigma[Z] = a\sigma[X]$ and $\sigma[U] = c\sigma[Y]$, one has $\rho[Z, U] = (\langle XY \rangle - \langle X \rangle \langle Y \rangle)/(\sigma[X]\sigma[Y]) = \rho[X, Y]$.

4.6. It is enough to use Eq. (4.79) with different limits: $P\{X \geq 180,\ Y \geq 170\} = [0.5 - \mathrm{E}(0.625)] \cdot [0.5 - \mathrm{E}(0.833)] = 0.266 \cdot 0.202 = 0.0537 \simeq 5.4\%$. The values of the function $\mathrm{E}(\ldots)$ can be obtained from Table E.1.

4.7. It would be necessary to integrate Eq. (4.40) over the region $X \in [180, +\infty)$, $Y \in [170, +\infty)$. Notice also that the concentration ellipse does not give the correct solution in this case. See Problem 8.11 to evaluate a solution using Monte Carlo methods.

4.8. The density of p_{XY} is given by the table at the top of the next page, while its marginal densities are in the table below. From these tables it results that the variable X, corresponding to the first die, is uniformly distributed, while Y, which is correlated to X, has a different density. From Eqs. (4.17) and (4.18) it is possible to calculate means and standard deviations: $\mu_x = 3.5$, $\mu_y = 3.25$, $\sigma_x = 1.708$, $\sigma_y = 1.726$. Finally, the covariance can be calculated with

(x, y)	p_{XY}	(x, y)	p_{XY}	(x, y)	p_{XY}
(1, 1)	4/36	(3, 1)	1/36	(5, 1)	1/36
(1, 3)	1/36	(3, 3)	4/36	(5, 3)	1/36
(1, 5)	1/36	(3, 5)	1/36	(5, 5)	4/36
(2, 1)	1/36	(4, 1)	1/36	(6, 1)	1/36
(2, 2)	3/36	(4, 3)	1/36	(6, 3)	1/36
(2, 3)	1/36	(4, 4)	3/36	(6, 5)	1/36
(2, 5)	1/36	(4, 5)	1/36	(6, 6)	3/36

x, y	1	2	3	4	5	6
p_X	1/6	1/6	1/6	1/6	1/6	1/6
p_Y	1/4	1/12	1/4	1/12	1/4	1/12

Eq. (4.21) and from the tabulated values of p_{XY}. The value $\sigma_{xy} = 1.458$ is thus obtained.

Chapter 5

5.1. Since, for a uniform variable X, one can write $P\{Y = -2\ln X > y\} = P\{X < e^{-y/2}\} = e^{-y/2}$, the cumulative function of Y is $F(y) = 1 - e^{-y/2}$. The derivative of the cumulative provides the requested density: $p(y) = \mathrm{d}F/\mathrm{d}x = e^{-y/2}/2$, $0 \leq y < \infty$. Equation (3.67) shows that a χ^2 distribution with 2 degrees of freedom has been obtained.

5.2. For $x \geq 0$ there is one root $x = \sqrt{z}$. Then, one obtains $p_Z(z) = 2(1 - \sqrt{z})/(2\sqrt{z}) = 1/\sqrt{z} - 1$, $0 < z \leq 1$.

5.3. Defining the auxiliary variable $W = Y$, one has $X = ZW$ and $Y = W$; the derivatives $\partial f_1^{-1}/\partial Z = W$, $\partial f_1^{-1}/\partial W = z$, $\partial f_2^{-1}/\partial Z = 0$, $\partial f_2^{-1}/\partial W = 1$ allow the calculation of the Jacobian $|J| = |W|$. The result is then $p_Z = \int |w|\, p_{XY}(zw, w)\, \mathrm{d}w$.

5.4. From Eq. (5.32) one has $p_Z(z) = \int_{-\infty}^{+\infty} p_Y(z-x)\, p_X(x)\, \mathrm{d}x$, with $x \geq 0$, $(z - x) \geq 0$, so that $p_Z(z) = \int_0^z p_Y(z - x)\, p_X(x)\, \mathrm{d}x = z \exp[-z]$.

5.5. The procedure of the previous problem must be extended by induction, and the Erlang or gamma distribution $\Gamma(n, \lambda)$ is found, which is $e_k(t) = \lambda(\lambda t)^{n-1} e^{-\lambda t}/(n-1)!$.

5.6. The conditions $0 \leq Z \leq 1$, $W \geq 0$ holds, and the inverse functions are $X = ZW$ and $Y = W(1 - Z)$. The Jacobian is $|J| = |W| = W$, and the requested density is $p_{ZW}(z, w)\, \mathrm{d}z\, \mathrm{d}w = w \exp[-w]\, \mathrm{d}z\, \mathrm{d}w$. From Theorem 4.1, Z and W are independent, and Z is a uniform random variable.

5.7. From Eq. (5.27), taking into account that $z = f(x, y) = xy$, $f^{-1}(z, y) = z/y$, $\partial f^{-1}/\partial z = 1/y$ and that in the ratio $x = z/y$ the condition $y \geq z$ must hold to assure the limits $0 \leq x \leq 1$, one obtains $p_Z(z) = \int_z^1 (1/y)\, \mathrm{d}y = -\ln z$, $for 0 \leq z \leq 1$. From the integral $\int z^n \ln z\, \mathrm{d}z = \ln z\, z^{n+1}/(n + 1) - z^{n+1}/(n + 1)^2$, one easily obtains $\langle Z \rangle = 1/4 = 0.25$ and $\mathrm{Var}[Z] = \langle Z^2 \rangle - \langle Z \rangle^2 = (1/9) - (1/16) \simeq 0.049$.

5.8. If T_1 and T_2 are the failure times, the system stops working at a time $T = T_1 + T_2$. The p.d.f. is then given by the convolution of two exponential densities. One then has $\lambda^2 \int \exp[-\lambda(t-u)] \exp(-\lambda u)\, du = \lambda^2 t \exp(-\lambda t)$, which is the gamma density (3.57). The mean life time of the system is $\langle T \rangle = 2/\lambda$. Compare this result with that of the Exercise 4.1.

5.9. Case (a): $\langle Z_1 \rangle = 0$, $\langle Z_2 \rangle = 0$, $\mathrm{Var}[Z_1] = 13$ and, from Eq. (5.69), $\mathrm{Var}[Z_2] = 1$. Notice that in the linear approximation one should have, from Eq. (5.66), $\mathrm{Var}[Z_2] = 0$. Using the method of Eq. (5.84), one easily obtains $\mathrm{Cov}[Z_1, Z_2] = 0$. The covariance and the correlation coefficient are null, even if a relation between the variable exists. Case (b): $\langle Z_1 \rangle = 5$, $\langle Z_2 \rangle = 1$, $\mathrm{Var}[Z_1] = 13$, $\mathrm{Var}[Z_2] = 3$, $\mathrm{Cov}[Z_1, Z_2] = 5$, $\rho[Z_1, Z_2] = 0.8$.

5.10. Since X and Y are Gaussians, from Exercise 5.3 it results the same also for $Y - X$. From Eq. (5.66) it results $\sigma[Y - X] = \sqrt{0.020^2 + 0.020^2} = 0.0283$. Using the standard variable, we obtain the values $t_1 = (0.050 - 0.100)/0.0283 = -1.768$, $t_2 = (0.150 - 0.100)/0.0283 = 1.768$, corresponding to an interval of probability $P\{0.050 \le (Y - X) \le 0.150\} = 2 \cdot 0.4614 = 0.923$, by interpolating from Table E.1. The requested percentage is then $1 - 0.923 = 0.077 = 7.7\%$.

5.11. From Exercise 5.4, the density of the number of vehicles is Poissonian with mean $\mu + \lambda$. From the binomial density, one then obtains $P(k) = n!/[k!(n-k!)][\mu/(\mu+\lambda)]^k\, [\lambda/(\mu+\lambda)]^{(n-k)}$.

5.12. `X<- rnorm(1000), y<- rnorm(1000), Z1<- X+3*Y, Z2<- 5*X+Y, cov(Z1,Z2)/sqrt(var(Z1)*var(Z2))`.

Chapter 6

6.1. (a) $13 \le R \le 120$, (b) $26 \le R \le 133$, (c) $17 \le R \le 116$.

6.2. The solution is the value of μ satisfying the equation $\mu - 1.28\sqrt{\mu} - 20 = 0$ with $\mu > 20$, so that $\mu = 26.6$.

6.3. The expected value of the root mean square (standard deviation) coincides with the true value : $\langle S \rangle = 12$ kg. Instead, one has $\sigma[S] \simeq \sigma/\sqrt{2N} = 0.6$ kg.

6.4. From the equation $Np = 215 \pm \sqrt{215(1-0.2)} = 215 \pm 13$, it follows $N = 215/0.2 \pm 1.65 \cdot 13/0.2 = 1075 \pm 108$, $CL = 90\%$.

6.5. The result follows from Theorem 6.1, since the sample variance S^2 is a function of the deviation, and from Theorem 4.3. It can also be directly shown that $\mathrm{Cov}[M, (X_i - M)] = \langle X_i^2 \rangle / N - \langle M^2 \rangle = \sigma^2/N - \sigma^2/N = 0$, since in a random sample the extractions are independent: $\langle X_i X_j \rangle = 0$ if $i \ne j$.

6.6. (a) 9.8 ± 0.2; (b) $(x - 245)/5 \simeq 1.645$, from which $x = 253.2$, corresponding to an upper limit of 10.13 for the mean. One can also write $(\mu - 9.8)/0.2 \simeq 1.645$, from which $\mu < 10.13$, $CL = 95\%$.

6.7. $\rho \in [-0.013, 0.305]$. From the approximate formula (6.122), one obtains $\rho \in [-0.011, 0.311]$.

6.8. From error propagation and from the formula $\nu = (x - b)/\epsilon$, with $b = 5.30 \pm 0.23$ counts/s, $\epsilon = 0.90 \pm 0.10$, $x = 16.7 \pm 1.3$ counts/s, one obtains $\nu = 12.7 \pm 2.0$ counts/s.

6.9. The exponential time distribution has $\mu = \sigma = \tau$. The mean sum of the times is $N\tau$, and the standard deviation is $\sigma = \sqrt{N}\tau$. From the inequality, $(1000 - 100N)/(\sqrt{N}100) \leq 1.64$ follows the second degree equation, $N^2 - 22.7N + 100 = 0$, whose solution is $N = 11.6 \simeq 12$.

6.10. The upper limit with $CL = 95\%$ is given by the equation $25 + 1.65\sqrt{\mu} = \mu$, from which $\mu = 34.7 \simeq 35$. The limit of the signal is then $35 - 10 = 25$ counts/s. Note: it is important to subtract the background *after* the calculations.

6.11. The quantile of the reduced χ^2 distribution is $\chi^2(24)_{R(0.95)} = 0.58$. We must then set $s^2/\sigma^2 \leq 0.58$, from which $\sigma^2 \geq s^2/0.58 = 31.03$, $s = \sqrt{(}18) = 4, 24$, $\sigma \geq 5.56$.

6.12. One has: 30/20=1.5 defects/km. The error is obtained from Eq. (6.44) with $t = 1$, or with `PoissApp(30)`, because, if the confidence level is not given, by default $CL = 0.68$. One then has: $\mu = [24.54, 36.54] = 30^{+6.54}_{-5.46}$. The number of defects per km is obtained dividing by 20 result and error: $\mu = 1.5^{+0.33}_{-0.27}$. In an approximate way, from Eq. (6.45), one has $\mu = 30/20 \pm \sqrt{30}/20 = 1.5 \pm 0.27$. Important remark: to find the error as $\sqrt{30/20} = 1.22$ is wrong. Indeed, the rule $\sigma \approx \sqrt{x}$ holds for *original* Poissonian variables only. A Poisson variable, if manipulated in any way, for example, by dividing or multiplying by some value, is no longer Poisson distributed.

6.13. One obtains `poisson.test(35,conf=0.95,alt="greater")=25.87`, `PoissApp(35,conf=0.95,alt="low")=25.78`.

6.14. Analytically, one has $\text{Cov}[x, y] = 0$, because x and y are independent. Instead $\text{Cov}[x, y1] = \langle (x - \langle x \rangle)(y1 - \langle y1 \rangle) \rangle = \langle x(3x + y - \langle 3x + y \rangle) \rangle = \langle 3x^2 + xy - 5x \rangle = 3\langle x^2 \rangle = 3\langle x \rangle^2 + 3\sigma_x^2 = 3$. Recall that, since x and y are independent, $\langle xy \rangle = \langle x \rangle \langle y \rangle = 0 \cdot 5 = 0$. For the correlation coefficient, one has $\text{Cov[x, y1]}/(\text{s}_\text{x}\text{s}_\text{y1}) = 3/\sqrt{9\sigma_x^2 + 1} = 3/\sqrt{10} = 0.942$.
From the routine `CorrelEst(x,y1)`, one has $\text{Cov}[x, y1] = 2.91 \pm 0.57$ and $r = 0.941^{+0.019}_{-0.014}$.

6.15. Analytically, one has $\langle x \rangle = \langle y \rangle = 1/2$, $\sigma_x = \sigma_y = 1/\sqrt{12} = 0.289$, $\langle y1 \rangle = 3/2 = 1.5$ and $\sigma_{y1} = \sqrt{4\sigma_x^2 + \sigma_y^2} = \sqrt{5/12} = 0.645$.
Hence, $\text{Cov}[x, y1] = \langle (x - 1/2)(y1 - 3/2) \rangle, = \langle (x - 1/2)(2x + y - 3/2) \rangle = \langle 2x^2 - x + xy - y/2 - 3x/2 + 3/4 \rangle = 1/6 = 0.166$, since $\langle xy \rangle = \langle x \rangle \langle y \rangle = 1/4$, because x and y are independent. For the correlation coefficient, one has $\text{Cov[x, y1]}/(s_x s_{y1}) = 0.166/(0.289 \cdot 0.645) = 0.890$.
From the routine `CorrelEst(x,y1)`, one has $\text{Cov}[x, y1] = 0.166 \pm 0.011$ and $r = 0.90 \pm 0.02$.

6.16. Applying logarithms, the variance additivity formula (5.66) and the Wald formula for the frequency variance for big samples (6.33), one obtains $\text{Var}[\ln(OR)]] = (1 - f_1)/(f_1 N_1) + (1 - f2)/(f_2 N_2)$, where $N_1 = 11037$ and $N_2 = 11034$. The $CL = 90\%$ interval, under the Gaussian linear approximation, must be multiplied by the quantile $t = 1.65$. We then obtain the logarithmic interval $\ln(OR) \pm 1.65\sqrt{\text{Var}[\ln(OR)]} = -0.597 \pm 0.200 = [-0.797, -0.397]$.

Returning to the original variables, one has
$OR = [\exp(-0.797), \exp(-0.397)] = [0.451, 0.672] = 0.55^{+0.12}_{-0.10}$.

Chapter 7

7.1. To use the routine `TdiffMean`, one needs the two sample standard deviations that are $0.05\sqrt{10} = 0.158$. With the call
`TdiffMean(c(5.36,5.21),c(0.158,0.158),c(10,10), alt='two')`, a p-value of 0.048 is obtained. The same result is obtained also under the hypothesis of equal variances, by setting`var=TRUE` in the routine call. The 5% homogeneity test is (narrowly) not passed.

7.2. Assuming that differences are due to chance (null hypothesis), Eqs. (7.42) and (7.43) provide the value $\chi^2_R(1) = 0.993$, corresponding to $SL \simeq 34\%$ (interpolated from Table E.3).
The R routine `chisq.test(rbind(c(40,28),c(30,30),cont=F)` gives a p-value of 31.7%. It is not possible to discard the hypothesis; therefore, it cannot be stated that the drug is effective.

7.3. One uses Eq. (7.32), where $Np_i = 100\, m^{x_i} \exp[-m]/x_i!$ ed $m = 4.58$ is the value of the sample mean from the data. It is necessary to group the first two and the last three channels, to have $Np_i > 5$. One finds $\chi^2 = 2.925$ with 6 degrees of freedom, corresponding to $\chi^2_R(6) = 0.49$. From Table E.3, one finds $P\{Q_R \geq 0.49\} \simeq 0.82$, giving a p-value of $2(1-0.82) \simeq 0.36$. It results that, on average, 4.6 buses arrive in 5 min and that the data are compatible with the Poisson distribution, because we would have a high probability to be wrong if we discard this hypothesis when true.

7.4. $\chi^2(1) = (356-375)^2/375+(144-125)^2/125 = 3.85$. $P\{Q > 3.85\} < 0.05$ from Table E.3. The model is rejected.

7.5. (a) The non-parametric method of contingency tables gives, for 5 degrees of freedom, $\chi^2_R = \chi^2(5)/5 = 0.84/5 \simeq 0.2$, $P\{Q_R \geq \chi^2_R\} \simeq 0.96$, from Table E.3. The compatibility test is passed, but the data should be viewed with suspicion, because the χ^2 value is too small. (b) Applying Eq. (7.32) with $Np_i = 100$ for all the 12 bins, one obtains $\chi^2 = 2.52$. Each die contributes with 5 degrees of freedom, so that $\chi^2_R(10) = 2.52/10 = 0.252$, $P\{Q_R \geq \chi^2_R\} \simeq 0.99$. If we say that the dice or rolls are rigged, we have a chance of being wrong $\leq 1\%$. In fact, with 100 events expected per channel, the 1σ statistical fluctuations are $\simeq 100 \pm \sqrt{100} = 100 \pm 10$, and all the 12 observed bins all fluctuate only within 1σ. It is therefore reasonable to discard the hypothesis, because the "data fluctuates too little".

7.6. The difference test gives $t = (60 - 33)/\sqrt{60 + 33} = 2.8$, from which, under the Gaussian approximation, $P\{T > 2.8\} = 2.6\, 10^{-3}$. This is the probability to be wrong in rejecting the limit ineffectiveness hypothesis if it were true.

7.7. The χ^2 is $(29 - 19)^2/19 + (18 - 19)^2/19 + \ldots = 11.7$ with 5 degrees of freedom, corresponding to a $< 5\%$ level. The test fails.

7.8. From the difference test, one obtains $(10500-2400-6700)/\sqrt{10500 + 91000} = 4.4$. The two results are incompatible. Note that, after multiplication by 10, the variances of the first two measures are *not* $10 \cdot 240$ and $10 \cdot 670$, because

the scaled variables are no longer Poisson. The variance of the sum has been instead calculated as $[10 \cdot \sqrt{240}]^2 + [10 \cdot \sqrt{670}]^2 = 91000$.

7.9. The χ^2 test gives the result

$$\chi^2 = (22-21.2)^2/21.2+(12-11.6)^2/11.6+(7-6.37)^2/6.37+(6-3.49)^2/3.49 = 1.91$$

with 3 degrees of freedom. Since `1-pchisq(1.91,3)` gives a p-value$=$ 0.59, there is agreement between data and model.

7.10. Integrating the exponential law over the experimental time intervals, the following values are obtained: 393, 239, 233, 135. The χ^2 test gives:

$$\frac{(368-393)^2}{393} + \frac{(266-239)^2}{239} + \ldots = 7.65\,.$$

We have 4 degrees of freedom, corresponding to a significance level $SL \simeq$ 13%. The model is accepted.

7.11. Since $\chi^2 = 2.5$ with 3 degrees of freedom, $SL \simeq 40\%$, and the measures are compatible.

7.12. The standard variable is $(58-55)/(10/\sqrt{10}) \simeq 1$. In order not to pass the test, a value of at least 1.65 is required, so the obtained value does not exceed the limits at the required level.

7.13. The χ^2 value is $= (13-16)^2/16 + (25-34)^2/34 + (44-34)^2/34 + (16-16)^2/16 = 5.88$, corresponding to $SL \simeq 20\%$. The hypothesis is accepted.

7.14. The χ^2 value is:

$$(18-15)^2/15 + (14-15)^2/15 + .. = 6.7\,,$$

corresponding to a p-value, given by `1-pchisq(6.7,6)`, of about 35%. The results agree each other.

7.15. The 15 values are loaded into a vector `pv`. Then the routine is called, for example, as `MultiTest(pv,alpha=0.05,method='bonferroni')`. The problem is solved by changing the `alpha` values and invoking also the `'fdr'` method. The data are already sorted in ascending order. The number of discarded hypotheses due to significantly different parameters after the drug intake is listed in the table:

Method	α	*False hypotheses*
'bonferroni'	0.05	3
	0.01	2
'fdr'	0.05	4
	0.01	3

It can be deduced that the first three parameters are certainly meaningful, most likely also parameter 4, which is spotted by 'fdr', with a maximum probability of false discovery of 5 %, given by the value of α.

7.16. The data are present inside the vector `breaks`. It is necessary to calculate means and standard deviations of the three samples with the commands `m1<-mean(breaks[1:9])` and `s1<-sd(breaks[1:9])` for sample one, and so on for the others, with their values contained in `[10:18]` and `[19:27]`. From the values `m1,m2,m3,s1,s2,s3` thus obtained and given the dimensions `n1=n2=n3=9`, we can perform the tests of Sect. 7.3 using the routine `TdiffMean`.
With the call `TdiffMean(m=c(m1,m3),s=c(s1,s3),n=c(9,9), var=TRUE)`, a p-value $\simeq 2\%$ is obtained for the difference between samples with high and low tension (`m1, m3`) and with medium and high tension (`m2,m3`). These results change only slightly with `var=FALSE`. A similar result is obtained ($p \simeq 1.6\%$) also with the *pooled* variance of Table 7.6 $s = \sqrt{(70.03/9)} = 2.79$ and using the approximate formula (7.6) with $s = s1 = s2$ and $x1 = m1$ and $x2 = m3$. Notice that Tukey's test gives p-values between 4 and 6% (Table 7.8). The value is higher because Tukey's test carries out three cross-tests and takes into account the test multiplicity. Consistent results are obtained if Bonferroni correction is applied to the values obtained with the t-test.

7.17. From Table 7.6 one sees that the *pooled* variance is 70.03. Tukey's quantiles (7.66) are then:

`q1 = 9.444/sqrt(70.03/9)` = 3.385,
`q2 = 10.0/sqrt(70.03/9)` = 3.585,
`q3 = 9.444/sqrt(70.03/9)` = 0.199.

With the command `1-ptukey(q1,nmeans=3,df=24)`, the first p-value of the table is obtained; replacing `q1` with the other values `q2, q3`, the calculation is completed. (Notice that the parameter `nmeans` of `ptukey` in the online R handbook is erroneously defined as the group dimension, whereas it is the number of groups.)

7.18. The days are loaded into a vector `days` of 25 components and the balm treatments into a vector `balm` with the command `balm<-rep(c('A','B', 'P'), c(8,8,9))`. The table is created with `test<-data.frame (day,balm)`, and the ANOVA test is performed with `fit<-aov(days balm,data=test)` and `summary(fit)`. A p-value= 0.0096 is obtained, indicating the effectiveness of at least one treatment. Tukey's test `TukeyHSD(fit)` indicates the effectiveness of A and an inconclusive value for B.

Chapter 8[1]

8.1. No.

8.2. The probability value is 2/3.

8.3. It is necessary to generate two uniform variables $X = 60 * \texttt{rndm}$, $Y = 60 * \texttt{rndm}$ and to count the number of times $-10 \le X - Y \le 12$. The result must be statistically compatible with 1/3.

8.4. The exact result, which can be obtained with non-trivial geometric considerations, is $1 - 3\sqrt{3}/(4\pi) = 0.586503$. Using the method of Problem 8.9, we obtained a fraction of pairs 3 614 289/6 164 195, corresponding to a probability of 0.58634 ± 0.00020.

8.5. The highest generation efficiency ($\simeq 78\%$) is obtained for the minimum k value satisfying the relation $ke^x \ge p(x) \quad \forall x \ge 0$; this condition is verified for $k \ge \sqrt{2e/\pi}$. To obtain a variable $Z \sim N(0, 1)$, it is necessary to generate a new number ξ and set $z = -x$ if $0 \le \xi \le 0.5$; $z = x$ if $0.5 < \xi \le 1$. Alternatively, also the method described in [Mor84] can be employed.

8.6. After a loop of $n = 5$ attempts is executed, the frequency $f = x/n$ is calculated, with x equal to the number events satisfying the additional condition $0 \le \texttt{rndm} \le 0.25$. The counters t_1, t_2, t_3 are incremented by one unit when p is inside the interval (6.31), calculated with the obtained value of f and with $t = 1, 2, 3$, respectively. This cycle must be repeated for a large number N of times, and the final values of t_1/N, t_2/N and t_3/N provide the requested levels. With $N = 10\,000$, we obtained $t_1 = 6570$, $t_2 = 9852$, $t_3 = 9994$. Calculate the resulting frequencies with their error and compare them with the probability levels of the 3σ law.

8.7. With uniform variables, $\mu = 0.5$, and the uncertainty of the sample mean is $\sigma = 1/\sqrt{12N}$. If $N = 12$, the standard variable value $t = \sum_{i=1}^{12} \xi_i - 6$ is obtained. This algorithm is implemented in the routine `MCgauss1`, present in our website.

8.8. $Y = \rho X + \sqrt{1-\rho^2}\, Y_R$, where X and Y_R are standard variables from the `gauss2(0,1)` routine.

8.9. The variables $-R \le X \le R$ and $-R \le Y \le R$ are uniformly and randomly generated and are accepted as coordinates when $\sqrt{X^2 + Y^2} \le R$.

8.10. The histogram comes from a population of density $p_Z(z) = \frac{1}{\pi(1+z^2)}$, known as Cauchy's distribution, which has no mean and infinite variance.

8.11. We need to generate two Gaussian variables of given mean, standard deviation and correlation: $X = 8 \cdot g_1 + 175$, $Y = (\rho \cdot g_1 + \sqrt{1-\rho^2} \cdot g_2) \cdot 6 + 165$, where g_1 and g_2 are standard normal variates from the `gauss` or `gauss2` routines. Then, the percentage of the generated pairs having $X \ge 180$ and $Y \ge 170$ is evaluated. With 10 millions of pairs, we obtained $(10.71 \pm 0.01)\%$.

[1] The values obtained by the reader with simulation codes must be statistically compatible with the solutions reported here. Compatibility must be verified with the statistical tests described in particular in Sects. 6.12, 7.1, and 7.2.

8.12. The results must be compatible with the solution of Problem (5.9).

8.13. Two uniform variates $x = \xi_1$ and $\phi = 2\pi\xi_2$ are generated. The number n of successes $x + \cos(\phi) < 0$ or $x + \cos(\phi) > 1$ over a total of N attempts is calculated. The estimate of π is $2N/n \pm (2N/n^2)\sqrt{n(1 - n/N)}$. See also our `Buffon` routine.

8.14. A possible solution is our routine `MCasimm`. Remember that the uncertainty on the standard deviation of N_e events is $\simeq \sigma/\sqrt{2N_e}$ (see Sect. 6).

8.15. After excluding the extreme values $p < 0.01$ and $p > 0.99$, the condition is reached for $n \gtrapprox 200$.

8.16. The condition is reached for $\mu \gtrapprox 70$.

8.17. $P\{r^* > 0\} \simeq 6\%$ (r = sample correlation coefficient). The model must be rejected even if this value is only slightly above the limit.

8.18. The routine `BootOdds` solves the problem by generating 1000 values f_1 and f_2 using `rbinom(1,size=N1,prob=f1)`, `rbinom(1,size=N2, prob=f2)`, from which a sample `oddsboot` of values f_1/f_2 is determined. With the command
`quantile(oddsboot,c(0.05,0.95),names=FALSE)`,
the values [0.444, 0.668] are obtained. Compare this result with solution D.

8.19. Tukey's test considers the means of m Gaussian samples of size n and uses the quantiles of the statistic $q = (\bar{Y}_{\max} - \bar{Y}_{\min})/(\sqrt{MS_E/n})$, where $MS_E = \sum_{ij}(y_{ij} - \bar{y}_{i\cdot})^2/(mn - m)$ is the pooled variance of the sample set. The solution is our routine `MCTukey`, which evaluates the quantiles in the following way:

```
for(j in 1:Nsim){
   for(k in  1:groups){
     vg <- rnorm(nmeans)  # sample of nmeans data
     meang[k] = mean(vg)
     devg[k] = var(vg)*(nmeans-1) # deviance
   }
   pool = sum(devg)/dof  # pooled sample variance
   delta[j]=(max(meang)-min(meang))/sqrt(pool/nmeans) #T. stat
  }
  qtl= 1.-alpha
  qtuk = quantile(delta,probs=qtl,names=FALSE) #Tukey quantile
```

Chapter 9

9.1. A possible solution is our code `MCdelta`, which generates $n = 100$ Gaussian variates and calculates the difference $\Delta = x_{\max} - x_{\min}$ between the maximum and minimum values. The procedure is repeated a very large number N of times to obtain, at the end of the loop, the histogram of the variable Δ. An asymmetric histogram is obtained, coming from a population of unknown analytic form. With $N = 50{,}000$, we obtained, for $n = 100$, a sample with mean and standard deviation $m = 2.508 \pm 0.001$, $s = 0.302 \pm 0.001$. From the graphic study of the histogram, we then determined the quantile value $\Delta_{0.99} = 3.32$. The quality control must discard the batch when $\Delta > 3.32$. Note that the method is insensitive to the shift of the mean $\langle X \rangle$.

9.2. The 5 times are randomly generated from the exponential law, and $m_1 = \min(t_1, t_3, t_5)$, $m_2 = \min(t_2, t_3, t_4)$, $m_3 = \min(t_1, t_4)$ and $m_4 = \min(t_2, t_5)$ are determined. The machine stops after a time $t = \max(m_1, m_2, m_3, m_4)$. By repeating the cycle 10,000 times, we obtained an asymmetric histogram with parameters $m = 2.52 \pm 0.02$ and $s = 1.70 \pm 0.01$ days. About 73% of operating times from the histogram are included in the $m \pm s$ interval.

9.3. It is necessary to modify the function `Funint` inside the routine `MCinteg` and to set `Mode3=0` to exclude importance sampling. The upper limit x is the variable `i2` which has to be modified to the appropriate value. The solution is the routine `Mcinteg1` of our website. If stratified sampling is applied, our routine `MCintopt` can be used after the definition: `funint<- function(x) exp(-x*x/2)/(2*pi)`, and with the call `MCintopt(funint,lower=0,upper=2)`, when, for example, the integral between 0 and 2 is evaluated. It is also instructive to perform a very accurate standard numerical integration with the R routine `integrate` and compare its results with the MC ones.

9.4. The exact value of I can be computed analytically and is $I = 2\log 2 - 1 = 0.38629\ldots$. For $N = 1000$, the crude method gives $\sigma \simeq 6.25 \cdot 10^{-3}$; the hit or miss method $\sigma \simeq 1.1 \cdot 10^{-2}$; the importance sampling (choosing $g(x) = x$) $\sigma \simeq 1 \cdot 10^{-3}$; and the stratified sampling (with $k = 20$ layers) $\sigma \simeq 0.3 \cdot 10^{-3}$.

9.5. The result must be statistically compatible with the value $I = 8$. With the crude method and $N = 1000$, we obtained $I = 8.01 \pm 0.13$. After enclosing the integrand function into the region defined by the conditions $-1 \le x_1, x_2, x_3 \le 1$, $0 \le y \le 3$, we have obtained $I = 7.89 \pm 0.36$ with the hit of miss method (again with $N = 1000$) .

9.6. This equation represents an ellipse centred at the origin and with semi-major and semi-minor axes x and y of length $\sqrt{2}$ and 1, respectively. The rectangle surrounding the ellipse has an area $A = 4\sqrt{2}$. The resolution code randomly extracts a point within the rectangle and accepts it if $x^2/2 + y^2 \le 1$. The ratio between the accepted points N_S and the generated ones N gives the ellipse area with an error given by Eqs. (9.42), and (9.43). With the routine `MCellipse` and 100,000 points, we obtained an area of 4.433 ± 0.007.

9.7. The tests can be performed with our routine `MCmetrop`. Generating 10,000 variates and using the last 5000, we obtained $m = 0.029 \pm 0.018$, $s = 0.875 \pm 0.010$ with $a = 2$, and $m = 0.017 \pm 0.024$, $s = 1.002 \pm 0.016$ with $a = 3$. One notes that with $a = 3$ the error increases, but the biased estimate of σ is corrected. The results do not substantially change with $a > 3$, but longer sequences are needed to stabilize the result. The value $a = 3.2$ seems optimal. With 5000 standard variates from the routine `runif`, we obtained $m = 0.023 \pm 0.014$, $s = 0.994 \pm 0.010$, which is a more accurate result than that evaluated using the Metropolis algorithm with $a = 3$.

9.8. Referring to the figure, one has to write a code generating the emission point with the formulae $x = -2 + 4\xi$, $y = -3 + 6\xi$ and the flight direction as $\cos\theta = 1 - 2\xi$, $\phi = 2\pi\,\xi$. One then has $a = h\,\mathrm{tg}\theta$, $r =$

$\sqrt{x^2+y^2}$, $R = \sqrt{r^2+a^2-2ra\cos\phi}$. If R is less than the radius of the window R_d, the particle is counted by the detector. If n is the number of counted particles over a total of N emitted particles, the efficiency is given by $\epsilon = n/N \pm \sqrt{n(1-n/N)}/N$. Using the given input data, our routine `MCdetector` gives as result $\epsilon = 0.059 \pm 0.002$.

9.9. A possible solution is given by our routine `MCvmises`, which uniformly samples within $-\pi \le x \le \pi$, with the value of c given as an input parameter.

9.10. The solution depends on the subjective evaluation of the evolution graph. The reader should be able to verify that, for n too high, the algorithm is not reliable in predicting the expected number of atoms with spin = 1.

Chapter 10

10.1. The probability to extract a black marble is 1/3 or 2/3. From the binomial density, the following table is obtained:

	$x=0$	$x=1$	$x=2$	$x=3$	$x=4$
$b(x;4,p=1/3)$	16/81	32/81	24/81	8/81	1/81
$b(x;4,p=2/3)$	1/81	8/81	24/81	32/81	16/81

The ML estimate of p is then $\hat{p} = 1/3$ if $0 \le x \le 1$, $\hat{p} = 2/3$ if $3 \le x \le 4$. If $x = 2$, the likelihood function (binomial p.d.f.) has no maximum, so that the ML estimator is undefined in this case.

10.2.

	$x=0$	$x=1$	$x=2$	$x=3$
$\hat{p}$	0.1	0.3	0.7	0.9

10.3. From Theorem 10.2, $\hat{x}_\alpha$ is such that $\int_{-\infty}^{\hat{x}_\alpha} p(x;\hat{\mu},\hat{\sigma})\,\mathrm{d}x = \alpha$.

10.4. $\mathrm{d}\mathcal{L}/\mathrm{d}\lambda = n/\hat{\lambda} - \sum_i t_i = 0$, from which $1/\hat{\lambda} = \sum_i t_i/n = m$.

10.5. (a) From Eq. (1.33), we have $P(x;N) = A(x)[(N-n)!]^2/[N!(N-2n+x)!]$, where $A(x)$ contains factors independent of N. Since $P(x;N) \equiv L(N)$, the study of the function for discrete values of N shows that $L(N+1) \ge L(N)$ until $N \le n^2/x - 1$. The maximum of $L(N)$ occurs for $\hat{N} = 1 + \mathrm{int}(n^2/x - 1)$, where int is the integer part of the argument. From the data, $\hat{N} = 609$ is obtained. (b) Using Stirling approximation and Eq. (3.21), one obtains $\mathrm{d}\mathcal{L}/\mathrm{d}N = -\mathrm{d}\ln L(N)/\mathrm{d}N = \ln[N(N-2n+x)/(N-n)^2]$, from which $\hat{N} = n^2/x = 608$. Since N is a discrete variable, the Cramér-Rao bound cannot be used. Moreover, Eq. (10.29) is difficult to apply, even using the Stirling approximation and assuming N as a continuous variable. The process can then be simulated by generating a series of x values with $N = 600$ and analysing the histogram of the estimated $\hat{N}$. Repeating the experiment for 5000 times, we obtain an asymmetric histogram (with a long right tail) of parameters $m = 618 \pm 1$, $s = 80.0 \pm 0.8$. The maximum value is at $\hat{N} = 600$. A fraction of 72% of the values is inside

the interval 608 ± 80. Hence, a reasonable estimate is $N \in 608 \pm 80$, $CL = 72\%$. Since x follows the binomial density, under the Gaussian approximation ($x > 10$, $n-x > 10$) Eq. (5.57) can be used, giving the result $\mathrm{Var}[\hat{N}] = \mathrm{Var}[n^2/x] \simeq (n^4/x^3)[(1 - x/n) + (2/x)(1 - x/n)^2]$ (note that the second term is negligible). An interval $N \in 608 \pm 88$ is thus obtained, in good agreement with simulation.

10.6. The likelihood function of two trials is $L = p^{x_1+x_2}(1-p)^{2-(x_1+x_2)} = L(p; S)$. S is a sufficient statistic; P is not. In fact, p can only be estimated from the sum of the successes, and not from the product.

10.7. Neglecting constant factors, $\ln L = -(n/2)[\ln\sigma^2 + s^2/\sigma^2]$, where $\hat{\sigma^2} \equiv s^2 = w/n$ (the mean is known). Notice that $\langle s^2 \rangle = \sigma^2$ and that w is a sufficient statistic. By computing the first and second derivatives of $\ln L$, we find that the expected information is $nI(\theta) = n/(2\sigma^4)$. The same result is obtained by applying the suggested methods (a) and (b), from $[|\ln L|/\sqrt{nI}] \leq t_\alpha|$ and $[|s^2 - \sigma^2|/\sqrt{(nI)^{-1}}] \leq t_\alpha|$, where t_α is the selected quantile. The result is also identical to that of Eq. (6.70).

10.8. If $w = \sum_i x_i^2$ and $n = 6$, the logarithm of the likelihood obtained from the Gaussian with zero mean is $\ln L(\sigma^2; w) = -(n/2)\ln\sigma^2 - w/(2\sigma^2)$. The maximum value is $\hat{\sigma}^2 = w/n = 39.0/6 = 6.5$, which is the ML estimate of σ^2. To apply Eq. (10.45), it is necessary to redefine L, so that $\Delta[\ln L(\hat{\sigma}^2; x)] = 0$. One then obtains $\Delta[\ln L(\sigma^2; w)] = +(n/2)[\ln\sigma^2 + w/(n\sigma^2) - (n/2)[\ln(w/n) + 1]$. The value $CL = 95.4\%$ is obtained when $2\Delta[\ln L(\sigma^2; x)] = 4$. The numerical study of the function gives the interval $\sigma^2 \in [2.5, 27.1] = 6.5^{+20.6}_{-4}$.

The χ^2_R value (see Eq. (6.76)) is calculated with 6 degrees of freedom. From Table (E.3) the following values are obtained by interpolation: $\chi^2_{R0.023}(6) = 2.46$ and $\chi^2_{R0.977}(6) = 0.20$. The confidence interval with $CL = 95.4\%$ is $\sigma^2 \in [2.6, 32.6]$. This estimate is better than the previous one, which holds only asymptotically.

The artificial data of this exercise were generated from a Gaussian with $\sigma^2 = 9$.

10.9. The two methods give the same result numerically: $\sigma^2 \in [8.0, 14.0]$. Also these artificial data were generated from a Gaussian with $\sigma^2 = 9$.

10.10. The interval estimate of the two means is $\mu_1 \in 2.08 \pm 0.016\,\mathrm{cm}$, $\mu_2 \in 2.05 \pm 0.011\,\mathrm{cm}$. The two means are compatible, because the difference test (7.11) provides a standard value $t = 1.58$. The ML estimate is the weighted mean: $\hat{\mu} = 2.0596$. The interval estimate is given by Eq. (10.70): $\mu \in 2.0596 \pm 0.0091$.

10.11. If the random variable X is the number of negative samples over a total a priori fixed number of N examined samples, the likelihood function is binomial: $L(\mu; x) = N!/[x!(N-x)!]\,[\exp(-\mu)]^x[1-\exp(-\mu)]^{N-x}$, where $\exp(-\mu)$ is the Poissonian probability of observing zero events when the mean is μ. The ML estimate of μ gives $\exp(-\hat{\mu}) = x/N$, from which

$\hat{\mu} = \ln(N/x) = \ln(50/5) = 2.3$. Compare this result with Table 6.2 and Eqs. (6.41).

10.12. It is necessary to group the last three bins to have $n(t) \geq 5$: the result is *ten* events (times) between 14 and 20 s. The expected probability values within each bin $\Delta t = (t_1, t_2)$ are given by $p_i(\theta) \equiv p_i(\lambda) = \int_{\Delta t} e(t)\,\mathrm{d}t = \exp(-\lambda t_1) - \exp(-\lambda t_2)$. By using Eq. (10.58), and a non-linear fit program, one obtains $\lambda \in 0.337 \pm 0.011$, $\chi^2_R = 9.55/7 = 1.36$. The χ^2 must be divided by 7 degrees of freedom, since a parameter estimated from the data (λ) was used. The observed SL (p-value) is $\simeq 24\%$, by interpolation from Table E.3. From Eq. (10.60) one then obtains $\lambda \in 0.339 \pm 0.011$, $\chi^2_R = 10.32/7 = 1.47$, $SL \simeq 19\%$. The data were artificially generated from an exponential density with $\lambda = 1/3$.

10.13. Test level: $\alpha = P\{X = 1;\, H_0\} = \varepsilon$. Power of the test: $1 - \beta = 1 - P\{X = 0;\, H_1\} = P\{X = 1,\, H_1\} = 1 - \varepsilon$.

10.14. Test level: $\alpha = P\{X_1 = 1,\, X_2 = 1;\, H_0\} + (1/2)P\{X_1 + X_2 = 1;\, H_0\} = \varepsilon$. Power of the test: $1-\beta = 1-P\{X_1 = 0,\, X_2 = 0;\, H_1\}-(1/2)P\{X_1+X_2 = 1;\, H_1\} = 1 - \varepsilon$. The result is the same as in the single trial. However, if possible, it is advisable to carry out both tests, because in extreme cases the observed significance level and its power assume more favourable values. For example: $\alpha = P\{X_1 = 1,\, X_2 = 1;\, H_0\} = \varepsilon^2$, $1 - \beta = 1 - P\{X_1 = 0,\, X_1 = 0;\, H_1\} = 1 - \varepsilon^2$.

10.15. Since $p = (0.050 - 0.029)/(0.057 - 0.029) = 0.750$, for $S = 16$, a number $0 \leq \texttt{random} \leq 1$ is generated, and the hypothesis is accepted if $0.75 \leq \texttt{random} \leq 1$. In practice, one lot out of four is accepted.

10.16. The requested test levels are $\alpha = 0.01$ and $\beta = 0.05$. (a) If $\lambda_0 = 1/100$, $\lambda_1 = 1/110$ and $t_n = (\sum_1^n t_i)/n$ is the mean of the observed times, from the equations $(t_n - 100)/(100/\sqrt{n}) = 2.326$ and $(t_n - 110)/(110/\sqrt{n}) = -1.645$, the values $n = 1710$ and $t_n = 105.6$ are obtained. It is necessary to allocate a sample of 1710 suspensions, to calculate the average of the downtimes and to accept the supplier's declaration if the average exceeds 105.6 h.

(b) From the ratio: $L_0/L_1 = [\lambda_0 \exp(-\lambda_0 t_n)]/[\lambda_1 \exp(-\lambda_1 t_n)]$, and using logarithms, one can write the indecision interval (10.110) as $-4.55/n - \ln(\lambda_0/\lambda_1) < (\lambda_1 - \lambda_0)t_n < 2.98/n - \ln(\lambda_0/\lambda_1)$. The hypothesis H_1 (H_0) is chosen when the condition is unsatisfied to the left (right) tail. Simulation shows that, if the suspensions are really better, the mean of a sample with $n \in 934 \pm 6$ suspensions and a sample mean of $t_n \in 112.7 \pm 0.1$ hours are enough to confirm the manufacturer's claim. If the quality of the new suspensions is the same as the old ones, on average, a sample of $n = 670$ suspensions and an average of $t_n \in 97.1 \pm 0.1$ hours is enough to take the right decision. Warning: these are average values, and it can happen, with a single test, to exceed the value $n = 1710$ previously obtained with the sample of fixed size and the normal approximation. The region of indecision is included, in the plane (n, t_n), between two hyperbolas. We suggest you to carefully examine the simulated histograms!

It would seem convenient to start with the sequential test and use method (a) when the number of pieces exceeds 1710. However, this procedure is incorrect, because deciding which type of test to use *after* having obtained some (full or partial) results alters the a priori levels of probability. The type of test (a) or (b) must therefore be decided before performing the test.

10.17. The variable $Q_{\lambda_1} = 2\lambda_1 n T_n$, where T_n is the mean value of n exponential times, follows the $\chi^2(2n)$ distribution. One then has $1 - \beta(\lambda_1) = 1 - P\{Q \le q_{\lambda_1}\}$, where Q has $\chi^2(2n)$ distribution. Since n is large, the normal approximation can be used, and the Gaussian test on the mean is valid: $1 - \beta(\lambda_1) = 1 - \Phi[(q_{\lambda_1} - 2n)/(2\sqrt{n})] = 1 - \Phi[(t_n - 1/\lambda_1)/(1/(\lambda_1\sqrt{n}))]$.

10.18. Look at Fig. 10.6, and assume that the hypotheses $H_0 : p_0 = 0.5$ and $H_1 : p_1 = 0.3$ are associated to Gaussian densities. Since, from Table E.1, it results $P\{t \le -(0.5 - \alpha) = -0.4\} = -1.28$, $P\{t \ge (0.5 - \beta) = 0.4\} = +1.28$, and we have a one-tailed test ($t_{\alpha/2} \to t_\alpha$), it is enough to use Eq. (10.83) with $|t_\alpha| = |t_\beta| = 1.28$, $p_0 = 0.5$ and $p_1 = 0.3$. One thus obtains $n = 38$. The critical value is $x = 38(0.3 + 1.28\sqrt{0.3 \cdot 0.7/38}) \simeq 15$. A sample of 38 elements is needed; if $x \le 15$, H_1 is accepted; otherwise, H_0 is kept.

10.19. The ratio (10.99) is $R = L_0/L_1 = (0.5^x \cdot 0.5^{n-x})/(0.3^x \cdot 0.7^{n-x})$. From Eqs. (10.104) and (10.106), since $(1 - \alpha)/\beta = 9$ and $\alpha/(1 - \beta) = 1/9$, one obtains, passing to logarithms, the following result: $p = 0.5$ is accepted if $x \ge 0.397n + 2.593$, and $p = 0.3$ is chosen if $x \le 0.397n - 2.593$. The plane band between these two lines is the uncertainty region.
Repeating the test 10,000 times with the simulation of a fair coin ($p = 0.5$), we obtained, for the n values, an exponential histogram with parameters $m \in 23.3 \pm 0.2$ and $\sigma \in 17.3 \pm 0.1$. The wrong hypothesis has been chosen in 786 cases over 10,000. The average number of attempts (23) is smaller than the result of Problem 10.18 (38). However, in 1484 cases the simulated sequential test required a number of attempts > 38.

10.20. One easily finds $\langle X \rangle = (\theta + 1)/(\theta + 2)$, $\mathrm{Var}[X] = (\theta + 1)/[(\theta + 2)^2(\theta + 3)]$ and the cumulative function $F_X(x) = x^{\theta+1}$. The Neyman-Pearson ratio evaluates the best critical region as $\{-\ln r_\alpha + n \ln[(1 + \theta_0)/(1 + \theta_1)]\}/(\theta_1 - \theta_0) \le \sum_i \ln x_i \le 0$. If $Z = \ln X$, $F_z(z) = P\{Z \le z\} = P\{X \le e^z\} = e^{(1+\theta)z}$. The p.d.f. of Z is then $\mathrm{d}F_z/\mathrm{d}z = p_Z(z) = (1 + \theta)\exp[(1 + \theta)z]$, and hence $\langle \sum_i \ln X_i \rangle = -n/(1 + \theta)$, $\mathrm{Var}[\sum_i \ln X_i] = n/(1 + \theta)^2$. Since $n = 100$, the normal approximation can be used; for $\alpha = 0.05$, $t_{1-\alpha/2} = 1.64$, and, if $\theta = \theta_0 = 1$, the null hypothesis rejection occurs for $[\sum_i \ln x_i - (-n/2)]/(\sqrt{n}/2) \ge 1.64$, that is for: $\sum_i x_i \ge -41.8$. This test on $\sum_i \ln x_i$ has the maximum power for $\alpha = 0.05$. It also results that $r_\alpha = 3.49$, but this value is not necessary for the test, if the normal approximation is used. The power curve is $1 - \beta = 1 - \Phi[(-41.8 + n/(1 + \theta_1))/(\sqrt{n}/(1 + \theta_1))]$, $\theta_1 > \theta_0$. For $n = 100$ and $\theta_1 = 2$, one has $1 - \beta = 0.996$. It is useful to verify that the test on the sum $\sum x_i$ gives, under the normal approximation, only a slightly smaller power, $1 - \beta = 0.989$.

Chapter 11

11.1. From Eq. (5.65), under the hypothesis $\mathrm{Cov}[X, Z] = 0$, one obtains $\sigma_y^2 = b^2\,\mathrm{Var}[X] + \mathrm{Var}[Z] + 2b\,\mathrm{Cov}[X, Z] = b^2\,\mathrm{Var}[X] + \mathrm{Var}[Z]$. Defining $\Delta X = X - \mu_x$ and $\Delta Y = Y - \mu_y$, the covariance between these variables can be computed from Eq. (5.83), as in Exercise 4.3: $\mathrm{Cov}[X, Y] = \langle \Delta X\, \Delta Y \rangle = b\left\langle \Delta^2 X \right\rangle = b\,\sigma_x^2$. We then obtain the result $\rho = \pm\sigma[f(X)]/\sigma[Y] = b\sigma_x/\sigma_y = b\sigma_x^2/\sigma_x\,\sigma_y = \sigma_{xy}/(\sigma_x\,\sigma_y)$.

11.2. From the relation: $s_x = 0.10x/\sqrt{12}$ (the same for y), and using the effective variance formula for the s_E calculation, the following table is obtained:

x	10	20	30	40	50	60	70
s_x	0.3	0.6	0.9	1.2	1.4	1.7	2.0
y	21.4	38.8	52.2	88.1	99.5	120.4	158.3
s_y	0.6	1.1	1.5	2.6	2.9	3.5	4.6

.

The minimization of Eq. (11.16), where $f(x; a, b) = a + bx$ and the effective variance in the denominator is $s_y^2 + b^2 s_x^2$, gives the values $a \in -0.13 \pm 1.15$, $b \in 2.04 \pm 0.05$. With two iterations of the linear fit, one obtains $a \in 0.23 \pm 1.12$, $b \in 2.02 \pm 0.05$. The data were generated with a simulation starting from $Y = 1 + 2X$. We also get $\chi^2(\nu)/\nu = 28.0/5 = 5.6$. However, the χ^2 test is not meaningful, because data are not Gaussian.

11.3. A non-linear fit with the straight line, $f(x; a, b) = a + bx$, and effective variance $s_y^2 + b^2 s_x^2$ give the values $a \in -0.64 \pm 1.14$, $b = 2.10 \pm 0.05$. With a two-step linear fit, one obtains $a \in -0.62 \pm 1.12$, $b \in 2.10 \pm 0.05$. The same result is obtained also with the routine `FitLineBoth`. The data have been simulated starting from $Y = 1 + 2X$. The χ^2 test is meaningful, because data are Gaussian. One obtains $\chi^2(\nu)/\nu = 2.05/5 = 0.41$, corresponding, from Table E.3, to a left-tailed $SL \simeq 20\%$. In this case, the straight line is a good model.

11.4. The straight lines passing through (x_0, y_0) must satisfy the constraint $y_0 = a_i + b_i x_0$. If a_i and b_i are known, it is convenient to represent them in the so-called dual plane, where a line is represented by a point of coordinates (b, a). In this plane, the straight line fit of the points (b_i, a_i) performed with the function $a = (-x_0)b + y_0$ evaluates the vertex coordinates (x_0, y_0) and their error. In physics, this method is used to determine the emission vertex of electrically charged nuclear particles from their reconstructed trajectories inside specific detectors. In case of curved trajectories (as the ones inside magnetic fields), one can proceed by successive steps, using straight lines tangent to the particle trajectories.

11.5. The data were obtained from a simulation assuming $Y = 5 + 0.8X^2 + Y_R$, where $\sigma[Y_R] = 0.5$.

11.6. Both X and Y have $\sigma = 0.5$, and a correlation $f(X) = 5X + 0.2X^2$ has been simulated between them.

11.7. The error $\sigma[Y]$ is estimated as $s(y_i) = 0.10 \cdot y_i$. In this way the following values are obtained:

x	2	4	6	8	10
y	7.9	11.9	17.0	25.5	23.8
$s(y)$	0.8	1.2	1.7	2.5	2.4

These data are loaded into vectors `x,y,sy` and passed to the routine `FitPolin(x=x,y=y,dy=sy,fitfun=y x,ww='ABS')`, which determines the estimates $\hat{a}$ and $\hat{b}$ of the functional relation $y = \hat{a} + \hat{b}\,x$. The code provides the result $\hat{a} \pm s(\hat{a}) = 3.31 \pm 1.08$, $\hat{b} \pm s(\hat{b}) = 2.26 \pm 0.24$, $r(\hat{a}, \hat{b}) = -0.845$, $\chi^2 = 3.66$. The LS estimates are compatible with the true values $a = 5$ and $b = 2$, used to generate the artificial simulated data. The reduced chi-square $\chi^2_R(3) = 1.22$ corresponds, from Table E.3, to one-tailed significance level of about 30%. Their covariance, obtained from the covariance matrix provided by the routine, is $s(\hat{a}, \hat{b}) = -0.22$. It is very instructive to verify the fit result with the simulation method described in Sect. 8.10. Using the LS estimates $\hat{a}$ and $\hat{b}$ and the experimental data x_i, we generate artificial data y'_i with errors $s(y'_i)$ according to the algorithm $y'_i = (1+0.10\cdot\xi_i)\,(\hat{a}+\hat{b}\,x_i)$, $s(y'_i) = 0.10\cdot y'_i$, where ξ is a standard Gaussian variate. Note that the straight line is calculated with the estimated values and that the error is always calculated in an approximate way as a percentage of the observed data, as in the case of a real experiment. Repeating the fit 20,000 times, the histograms of $\hat{a}$, $\hat{b}$, $s(\hat{a})$, $s(\hat{b})$ and of χ^2 are obtained. *The histogram widths* allow you to directly check that the errors of $s(\hat{a})$ and $s(\hat{b})$ from the fit coincide within 3 decimal digits with the standard deviations of the simulated histograms of $\hat{a}$ and $\hat{b}$. The densities of the estimates of a and b are practically perfect Gaussian, as might be expected. The 20 000 simulated χ^2 values are also perfectly distributed as the χ^2 density with three degrees of freedom.

11.8. At first, it is necessary to evaluate the errors $s(y_i) = (\Delta/\sqrt{12})y_i = 2(0.10/3.46)y_i = 0.058y_i$ and, analogously, $s(x_i) = 0.058x_i$. Also in this case, the estimate is approximate, because it is obtained as a percentage of the observed values, instead of the true ones. Loading the values x_i, y_i, $s(x_i)$ and $s(y_i)$ to vectors `x,y,sx,sy` and defining the functions `fun<-function(par,x){par[1]+par[2]*x}` and `dfun<-function(par,x){par[2]}`, where `par<-c(0.5,0.5)`, the routine `FitLineBoth(x,y,sx,sy,par=par,fun=fun,dfun=dfun)` is called, minimizing Eq. (11.16). We obtain $\hat{a} \pm s(\hat{a}) = 5.09 \pm 0.55$, $\hat{b} \pm s(\hat{b}) = 2.176 \pm 0.13$, $s(\hat{a}, \hat{b}) = -0.0616$, $r(\hat{a}, \hat{b}) = -0.842$, $\chi^2 = 3.22$. The reduced χ^2 value is $\chi^2/3 = 1.07$, indicating a good agreement between model and data. The estimates are compatible with the true values $a = 5$ and $b = 2$, which are those used to generate the simulated data. As in the previous exercise, we now repeat the fit for 20,000 times.

Using the experimental data x_i and the LS parameter estimates, we can generate artificial data x_i', y_i' with errors $s(x_i')$ and $s(y_i')$ according to the algorithm $y_i' = [1 + 0.10\,(1 - 2\,\xi_i)]\,(\hat{a} + \hat{b}x_i)$, $x_i' = x_i + 0.10\,(1 - 2\,\xi_i)\,x_i$, $s(x_i') = 0.058\,x_i'$, $s(y_i') = 0.058\,y_i'$. At each iteration, the variance s_E^2 is computed, and the fit is performed. From this simulation we obtain that *the standard deviations of the histograms of a and b c coincide with the values estimated by the LS method*. This agreement increases our confidence in the least squares algorithms, even with non-Gaussian data. Simulation thus demonstrates an important fact: *the deviations from the Gaussian model are small*. Ultimately, a negligible error is introduced by applying the intervals of the 3σ law to the regression result.

Chapter 12

12.1. $\sigma[F]/F \simeq \sqrt{0.04^2 + 0.05^2} = 0.064$, that is, 6.4%.

12.2. Passing to logarithms, one has $t = -\tau \ln(I/I_0) = 23$ days. Propagating the error on I, one obtains $s_t = \pm\tau s_I/I = \pm 0.5$ days.

12.3. Since the process is Poissonian, the experimental count number is an estimate of variance. From the error propagation law, the background value is $F = 620/10 = 62$ counts/s, with $\sigma[F] = \sqrt{620}/10 = 2.5$ counts/s. Then $(I - F) \pm \sigma[I - F] \simeq (157 - 62) \pm \sqrt{157 + 620/100} = 95 \pm 13$, counts/s, $CL \simeq 68\%$. The signal to background ratio is "7 sigma": $n_\sigma = 95/13 \simeq 7.3$.

12.4. The upper limit with $CL = 95\%$ is given by the equation $25 + 1.65\sqrt{\mu} = \mu$, from which $\mu = 34.7 \simeq 35$. The upper limit for the signal is then $35 - 10 = 25$ counts/s. Note: it is important to subtract background *after* the limit calculation.

12.5. Student's quantiles with 3 degrees of freedom must be associated to these four values. From Table E.2 (and for a probability of 0.975), it results that the statistical error of the mean is $0.04/3.18 = 0.012$. The precision is then $0.012\sqrt{4} \simeq 0.02$.

12.6. The percentage error of R propagates on V twice as much as the percentage error of L. Therefore, it is better to improve the measurement of R.

12.7. $\mathrm{Var}[\mathrm{sen}\,\theta] \simeq (\cos^2\theta)\,s_\theta^2$, from which $\mathrm{sen}\,\theta = 0.50 \pm 0.03$. Remember to transform the angle values to radians!

12.8. It is immediate to derive that $(1 - P^2)/N = (4N_+N_-)/N^3$. If N is a random Poissonian variable, $N_\pm$ are Poissonian independent random variables, $s^2(N_\pm) = N_\pm$, and, since $(\partial P/\partial N_\pm) = \pm 2\,N_\mp/N^2$, it follows that the error propagation procedure gives $s^2(P) = (4N_+N_-)/N^3$. If N is fixed before the measurement, $N_\pm$ are values of correlated binomial variables. Since $N_- = N - N_+$, one has $P = 2(N_+/N) - 1$, $\partial P/\partial N_+ = 2/N$, $s^2(N_+) = N_+(1 - N_+/N)$ and hence $s^2(P) = 4(N_+/N^2)(1 - N_+/N) = 4N_+N_-/N^3$. In conclusion, the uncertainty $s(P) = \sqrt{(1 - P^2)/N}$ holds for any experimental condition.

12.9. $f(V) = 2.718 \pm 0.008$, where the error is the standard deviation. The distribution of $f(V)$ is nearly uniform, as it is easy to verify both analytically

and with a simulation. Then, it is possible to write $f(V) = 2.718 \pm 0.008\sqrt{12}/2 = 2.718 \pm 0.013$, $CL = 100\%$.

12.10. Since $s^2(E_2) = (0.05)^2(\partial E_2/\partial E_1)^2 + (100/\sqrt{12})^2(\partial E_2/\partial R_1)^2 + (100/\sqrt{12})^2\ (\partial E_2/\partial R_2)^2 = (0.103)^2$, one has $E_2 = 5.00 \pm 0.10$ V. The simulation randomly generates uniform variates E_1, R_1, R_2 (e.g. $R_1 = 1000 + (2\,\texttt{rndm} - 1)\,50$) and provides the histogram of E_2. A distribution close to the triangular p.d.f. is obtained, with a standard deviation $s = 0.103$, coincident with the one calculated analytically. The interval $m \pm s$ includes 65% of the values. The $CL = 68\%$ is achieved for $s = 0.110$. The direct measurement of E_2 is in agreement with this calculation.

12.11. $v = 2.00/5.35 = 0.3738$ m/s. $s_t = 0.05/\sqrt{20} = 0.011$ s; $s_l = 0.002/\sqrt{12} = 5.8\,10^{-4}$ m. Since $s_v^2 = v^2[s_l^2/l^2 + s_t^2/t^2] = 6.2\,10^{-7}$, one obtains $v = 0.3738 \pm 0.0008$ m/s. The error is dominated by the time uncertainty. Since v is the ratio between a uniform and a Gaussian variable, it is better to verify the CL with a simulation. A Gaussian histogram is thus obtained, with standard deviation coinciding with the calculated one. A $CL \simeq 68\%$ can then be associated with the result.

12.12. In this case, $\text{Var}[Y_i] = \sigma_i^2 + \epsilon^2 x_i^2$ and $\text{Cov}[Y_i, Y_j] = \epsilon^2 x_i x_j$ for $i \neq j$. One then obtains:

$$V = \begin{pmatrix} \sigma_1^2 + \epsilon^2 x_1^2 & \epsilon^2 x_1 x_2 & \dots & \epsilon^2 x_1 x_n \\ \epsilon^2 x_1 x_2 & \sigma_2^2 + \epsilon^2 x_2^2 & \dots & \epsilon^2 x_2 x_n \\ \dots & \dots & \dots & \dots \\ \epsilon^2 x_1 x_n & \epsilon^2 x_2 x_n & \dots & \sigma_n^2 + \epsilon^2 x_n^2 \end{pmatrix} .$$

12.13. We can generate 10,000 standard Gaussian variates and create an histogram to calculate their mean and variance. The interactive code lines are

```
x<-rnorm(10000); fx<-hist(x)$counts
xbin<-hist(x)$mids
meanx=sum(fx*xbin)/sum(fx)
varx=sum((xbin-meanx)^2*fx)/sum(fx).
```

It is better to verify the CL with a simulation. A Gaussian histogram is thus obtained, with standard deviation coinciding with the calculated one. A $CL \simeq 68\%$ can then be associated with the result.

12.14. Using the solution of the Problem 5.7, it results that the variable $Z = XY$ has logarithmic density $p_Z(z) = -\ln z$, with mean and standard deviation $\mu \pm \sigma = 0.25 \pm \sqrt{(0.49)}$. The probabilities can be found by integrating the density in the intervals $(\mu \pm K\sigma)$, where K is a real number. Integrating between the limits $\alpha = \max(0, \mu - K\sigma)$ and $\beta = \min(1, \mu + K\sigma)$, one obtains $\int_\alpha^\beta -\ln z = [z - z\,\ln z]_\alpha^\beta = \beta - \beta\ln(\beta) - \alpha + \alpha\ln(\alpha)$; hence, $P(Z < |\mu - K\sigma|) = 0.689,\ 0.946,\ 1$ for $K = 1, 2, 3$. Even though the distribution is logarithmic, the levels are close to those of the 3σ law.

12.15. The frequency is obtained by the formula $f = c/(a\,b)$, where $c = 1750$, $a = 0.3$ m and $b = 0.5$ m. The corresponding errors are $\sigma_c = \sqrt{1750} = 44$,

$\sigma_a = \sigma_b = 2 \cdot 0.005/\sqrt{12} = 0.00289$ m. From the error propagation, one has:

$$\sigma_f^2 = [1/(ab)]^2\sigma_c^2 + [c/(a^2b)]^2\sigma_a^2 + [c/(ab^2)]^2\sigma_b^2 = (306)^2 \,,$$

from which: $f = 1750 \pm 306$ counts/(m^{2}s).

12.16. If we assume to have the values of x and n_x of Eq. (12.68) loaded into the vectors `x` and `nx`, together with a vector of errors `s <- sqrt(nx)`, the R instructions are

```
db<-data.frame(x,nx,s);
class(fitexp <- nx ~ sum(nx)*(mu^x/factorial(x))
*exp(-mu));
sv<-list(mu=5.5);
Nlinfit(nlfitfcn=fitexp,database=db,startvalues=sv,
weight='ABS').
```

12.17. $1\,\sigma_r \simeq \sqrt{20^2 - 5^2} \simeq 19$ bins, from convolution properties.

12.18. The fit result is $\hat{a} = 1.86 \pm 0.58$; $\hat{b} = 0.38 \pm 0.10$; it is distorted with respect to the result of Table 12.1.

12.19. It is statement (b), because it is falsifiable.

Appendix E
Tables

E.1 Integral of the Gaussian Density

The standard Gaussian density, described in Sect. 3.5, is given by:

$$g(t; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) . \tag{E.1}$$

The values of the integral probability:

$$\mathrm{E}(t) = \int_0^t g(t; 0, 1)\, \mathrm{d}t \tag{E.2}$$

to obtain, in a random sampling, a value inside the interval $[0, t]$ are reported in Table E.1 for $t \in [0.0, 3.99]$ in steps of 0.01. The first two digits of t are read in the first bold column on the left, the third digit in the bold row at the top. The values of the integral (E.2) are read at the intersection of the rows and columns. By exploiting the symmetry of $\mathrm{E}(t)$ around the value $t = 0$ (see Eq. (3.40)) and using this table, it is possible to calculate the cumulative integral within any interval.

The integral (E.2) can also be obtained, with the R instruction `pnorm(t)-0.5` for any positive `t`; hence, the same can be done for all the values in the table. In particular, $t_{1-\alpha/2}$ is found by looking for $1-\alpha/2-0.5$ in the table and its matching t value. For example, $1 - \alpha = CL = 0.99\%$ corresponds to $1 - \alpha/2 = 0.995$, so $1 - \alpha/2 - 0.5 = 0.495 \simeq 0.4949$ and the matching t is 2.57.

We recall that the functions of R which compute the fundamental densities tabulated in this Appendix are listed in Table B.2.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3

E.2 Quantiles of the Student's Density

The quantile values t_P of the variable T, corresponding to different values of the integral probability not to exceed a given value t_P:

$$P = P\{T \leq t_P\} = \int_{-\infty}^{t_P} s_\nu(t)\,\mathrm{d}t\ , \tag{E.3}$$

where $s_\nu(t)$ is the Student's density (5.42), are reported in Table E.2 for different ν values. This distribution has been obtained in Exercise 5.5.

The last row of the table, where $\nu = \infty$, gives exactly the same quantile values of the Gaussian density.

Quantile values corresponding to a cumulative probability `p` for `df` degrees of freedom can also be obtained with the R instruction `qt(p, df)`.

E.3 Integrals of the Reduced χ^2 Density

The reduced chi-square p.d.f. $p_\nu(\chi^2)$ of the variable $Q(\nu)$ with ν degrees of freedom is given by Eq. (3.67).

The values of the variable $Q_R(\nu) = Q(\nu)/\nu$, corresponding to different values of the integral probability:

$$P = P\{Q_R(\nu) > \chi^2_{R\nu}\} = \int_{\chi^2_\nu}^{\infty} p_\nu(\chi^2_\nu)\,\mathrm{d}\chi^2_\nu \tag{E.4}$$

to exceed a given value of reduced χ^2, are reported in Table E.3 for ν values between 0 and 100. When $\nu > 100$, $p_\nu(\chi^2)$ tends to a Gaussian with mean and standard deviation $\mu = 1, \sigma = \sqrt{2/\nu}$, respectively.

The table values corresponding to a probability `p` for `df` degrees of freedom can be also be obtained, with the R instruction `qchisq(1-p,df)/df`.

E.4 Quantile Values of the Non-Reduced χ^2 Density

Table E.4 gives the quantile values of the non-reduced χ^2 density:

$$P = P\{0 \leq Q(\nu) \leq \chi^2\} = \int_0^{\chi^2} p_\nu(\chi^2)\,\mathrm{d}\chi^2\ .$$

These values can be used for the χ^2 test instead of Table E.3. The use of Table E.3 of reduced χ^2 for given significance levels is equivalent to the use of Table E.4 of

the quantile values of non-reduced χ^2. The choice of one or the other depends on the preferences of the reader or on the type of problem.

When $\nu > 100$, the $\chi^2(\nu)$ density tends to be Gaussian distributed with mean and standard deviation given by $\mu = \nu$ and $\sigma = \sqrt{2\nu}$, respectively.

The quantile values of Table E.4 are to be preferred for the calculation of the confidence levels corresponding to the value of $\Delta\chi^2$ in the equation:

$$\chi^2 = \chi^2_{min} + \Delta\chi^2 \,, \tag{E.5}$$

where χ^2_{min} is calculated with the true (or best fit) parameter values of the model. In this case the degrees of freedom ν represent the number of χ^2 free parameters.

For example, if χ^2 has ten free parameters, the concentration ellipse that has $\Delta\chi^2 \simeq 16$ as boundary contains 90% of the values, that is, it includes the 90% of hyperspace, corresponding to $CL = 0.90$.

The quantile values corresponding to a probability `p` for `df` degrees of freedom can be also obtained with the R instruction `qchisq(p,df)`.

E.5 Quantiles of the *F* Density

The p.d.f. $p_{\mu,\nu}$ of the Snedecor's $\boldsymbol{F}$ variable has been obtained in Exercise 5.6, Eq. (5.46).

The probability to obtain a value $\{\boldsymbol{F} \leq F_\alpha\}$ in a random sample with ν_1 and ν_2 degrees of freedom is given by the integral:

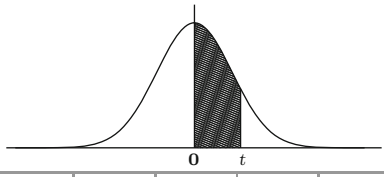
$$P\{\boldsymbol{F} \leq F_\alpha\} = \int_0^{F_\alpha} p_{\nu_1,\nu_2}(F)\, \mathrm{d}F \,. \tag{E.6}$$

The values of F_α, corresponding to right-tailed significance levels of 5% ($\alpha = 0.95$) and 1% ($\alpha = 0.99$), are shown in Tables E.5 and E.6, respectively, for different values of ν_1 and ν_2. The left-tailed $F_{1-\alpha}$ values can be obtained from Eq. (5.49):

$$F_\alpha(\nu_1, \nu_2) = \frac{1}{F_{1-\alpha}(\nu_2, \nu_1)} \,. \tag{E.7}$$

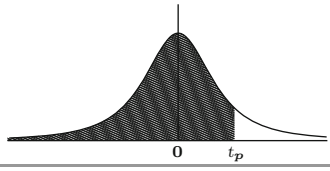
The quantile values corresponding to a probability `p` for `df1` and `df2` degrees of freedom can be also obtained with the R instruction `qf(p,df1,df2)`.

Table E.1 Integral of the standard Gaussian $g(t; 0, 1)$ as a function of the standard variable t



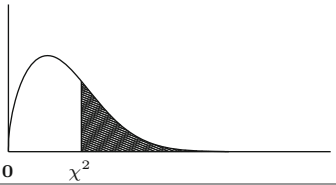
t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.9	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000	0.5000

Table E.2 Quantile values t_P of the Student's t variable for ν degrees of freedom



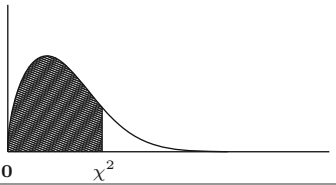
	P										
ν	0.60	0.70	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995	0.9995
1	0.325	0.727	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.65	632.0
2	0.289	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.60
3	0.277	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.92
4	0.271	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.267	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.265	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	5.959
7	0.263	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.262	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	5.041
9	0.261	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	4.781
10	0.260	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.260	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.259	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	4.318
13	0.259	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	4.221
14	0.258	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	4.140
15	0.258	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	4.073
16	0.258	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	4.015
17	0.257	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.965
18	0.257	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.922
19	0.257	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.883
20	0.257	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.849
21	0.257	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.819
22	0.256	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.792
23	0.256	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.768
24	0.256	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.745
25	0.256	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.725
26	0.256	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.707
27	0.256	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.690
28	0.256	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.674
29	0.256	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.659
30	0.256	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.255	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
50	0.255	0.528	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	3.496
60	0.254	0.527	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	3.460
70	0.254	0.527	0.678	0.847	1.044	1.294	1.667	1.994	2.381	2.648	3.435
80	0.254	0.526	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	3.416
90	0.254	0.526	0.677	0.846	1.042	1.291	1.662	1.987	2.368	2.632	3.402
100	0.254	0.526	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	3.390
∞	0.253	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Table E.3 Values of the χ^2/ν variable having a probability P to be exceeded in a sampling



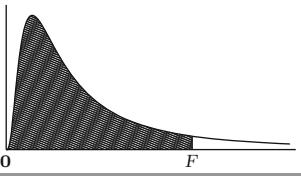
	P												
ν	0.005	0.01	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975	0.99	0.995
1	7.88	6.63	5.02	3.84	2.71	1.32	0.45	0.10	0.02	0.004	0.001	0.000	0.000
2	5.30	4.61	3.69	3.00	2.30	1.39	0.69	0.29	0.11	0.05	0.03	0.01	0.005
3	4.28	3.78	3.12	2.61	2.08	1.37	0.79	0.40	0.19	0.12	0.07	0.04	0.02
4	3.71	3.32	2.79	2.37	1.95	1.35	0.84	0.48	0.27	0.18	0.12	0.07	0.05
5	3.35	3.02	2.57	2.21	1.85	1.33	0.87	0.53	0.32	0.23	0.17	0.11	0.08
6	3.09	2.80	2.41	2.10	1.77	1.31	0.89	0.58	0.37	0.27	0.21	0.15	0.11
7	2.90	2.64	2.29	2.01	1.72	1.29	0.91	0.61	0.40	0.31	0.24	0.18	0.14
8	2.74	2.51	2.19	1.94	1.67	1.28	0.92	0.63	0.44	0.34	0.27	0.21	0.17
9	2.62	2.41	2.11	1.88	1.63	1.27	0.93	0.66	0.46	0.37	0.30	0.23	0.19
10	2.52	2.32	2.05	1.83	1.60	1.25	0.93	0.67	0.49	0.39	0.32	0.26	0.22
11	2.43	2.25	1.99	1.79	1.57	1.25	0.94	0.69	0.51	0.42	0.35	0.28	0.24
12	2.36	2.18	1.94	1.75	1.55	1.24	0.95	0.70	0.53	0.44	0.37	0.30	0.26
13	2.29	2.13	1.90	1.72	1.52	1.23	0.95	0.72	0.54	0.45	0.39	0.32	0.27
14	2.24	2.08	1.87	1.69	1.50	1.22	0.95	0.73	0.56	0.47	0.40	0.33	0.29
15	2.19	2.04	1.83	1.67	1.49	1.22	0.96	0.74	0.57	0.48	0.42	0.35	0.31
16	2.14	2.00	1.80	1.64	1.47	1.21	0.96	0.74	0.58	0.50	0.43	0.36	0.32
17	2.10	1.97	1.78	1.62	1.46	1.21	0.96	0.75	0.59	0.51	0.44	0.38	0.34
18	2.06	1.93	1.75	1.60	1.44	1.20	0.96	0.76	0.60	0.52	0.46	0.39	0.35
19	2.03	1.90	1.73	1.59	1.43	1.20	0.97	0.77	0.61	0.53	0.47	0.40	0.36
20	2.00	1.88	1.71	1.57	1.42	1.19	0.97	0.77	0.62	0.54	0.48	0.41	0.37
21	1.97	1.85	1.69	1.56	1.41	1.19	0.97	0.78	0.63	0.55	0.49	0.42	0.38
22	1.95	1.83	1.67	1.54	1.40	1.18	0.97	0.78	0.64	0.56	0.50	0.43	0.39
23	1.92	1.81	1.66	1.53	1.39	1.18	0.97	0.79	0.65	0.57	0.51	0.44	0.40
24	1.90	1.79	1.64	1.52	1.38	1.18	0.97	0.79	0.65	0.58	0.52	0.45	0.41
25	1.88	1.77	1.63	1.51	1.38	1.17	0.97	0.80	0.66	0.58	0.52	0.46	0.42
26	1.86	1.76	1.61	1.50	1.37	1.17	0.97	0.80	0.67	0.59	0.53	0.47	0.43
27	1.84	1.74	1.60	1.49	1.36	1.17	0.98	0.81	0.67	0.60	0.54	0.48	0.44
28	1.82	1.72	1.59	1.48	1.35	1.17	0.98	0.81	0.68	0.60	0.55	0.48	0.45
29	1.80	1.71	1.58	1.47	1.35	1.16	0.98	0.81	0.68	0.61	0.55	0.49	0.45
30	1.79	1.70	1.57	1.46	1.34	1.16	0.98	0.82	0.69	0.62	0.56	0.50	0.46
40	1.67	1.59	1.48	1.39	1.30	1.14	0.98	0.84	0.73	0.66	0.61	0.55	0.52
50	1.59	1.52	1.43	1.35	1.26	1.13	0.99	0.86	0.75	0.70	0.65	0.59	0.56
70	1.49	1.43	1.36	1.29	1.22	1.11	0.99	0.88	0.79	0.74	0.70	0.65	0.62
80	1.45	1.40	1.33	1.27	1.21	1.10	0.99	0.89	0.80	0.75	0.71	0.67	0.64
90	1.43	1.38	1.31	1.26	1.20	1.10	0.99	0.90	0.81	0.77	0.73	0.69	0.66
100	1.40	1.36	1.30	1.24	1.18	1.09	0.99	0.90	0.82	0.78	0.74	0.70	0.67

Table E.4 Quantile values χ^2_P of the non-reduced χ^2 variable for ν degrees of freedom



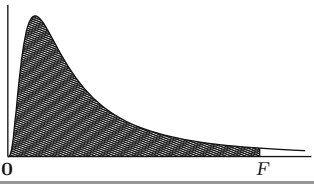
	P											
ν	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.975	0.99	0.995
1	0.06	0.15	0.27	0.45	0.71	1.07	1.64	2.71	3.84	5.02	6.63	7.88
2	0.45	0.71	1.02	1.39	1.83	2.41	3.22	4.61	5.99	7.38	9.21	10.60
3	1.00	1.42	1.87	2.37	2.95	3.67	4.64	6.25	7.82	9.35	11.34	12.84
4	1.65	2.19	2.75	3.36	4.04	4.88	5.99	7.78	9.49	11.14	13.28	14.86
5	2.34	3.00	3.66	4.35	5.13	6.06	7.29	9.24	11.07	12.83	15.09	16.75
6	3.07	3.83	4.57	5.35	6.21	7.23	8.56	10.64	12.59	14.45	16.81	18.55
7	3.82	4.67	5.49	6.35	7.28	8.38	9.80	12.02	14.07	16.01	18.47	20.28
8	4.59	5.53	6.42	7.34	8.35	9.52	11.03	13.36	15.51	17.53	20.09	21.95
9	5.38	6.39	7.36	8.34	9.41	10.66	12.24	14.68	16.92	19.02	21.67	23.59
10	6.18	7.27	8.30	9.34	10.47	11.78	13.44	15.99	18.31	20.48	23.21	25.19
11	6.99	8.15	9.24	10.34	11.53	12.90	14.63	17.27	19.68	21.92	24.72	26.76
12	7.81	9.03	10.18	11.34	12.58	14.01	15.81	18.55	21.03	23.34	26.22	28.30
13	8.63	9.93	11.13	12.34	13.64	15.12	16.98	19.81	22.36	24.74	27.69	29.82
14	9.47	10.82	12.08	13.34	14.69	16.22	18.15	21.06	23.68	26.12	29.14	31.32
15	10.31	11.72	13.03	14.34	15.73	17.32	19.31	22.31	25.00	27.49	30.58	32.80
16	11.15	12.62	13.98	15.34	16.78	18.42	20.47	23.54	26.30	28.85	32.00	34.27
17	12.00	13.53	14.94	16.34	17.82	19.51	21.61	24.77	27.59	30.19	33.41	35.72
18	12.86	14.44	15.89	17.34	18.87	20.60	22.76	25.99	28.87	31.53	34.81	37.16
19	13.72	15.35	16.85	18.34	19.91	21.69	23.90	27.20	30.14	32.85	36.19	38.58
20	14.58	16.27	17.81	19.34	20.95	22.77	25.04	28.41	31.41	34.17	37.57	40.00
21	15.44	17.18	18.77	20.34	21.99	23.86	26.17	29.62	32.67	35.48	38.93	41.40
22	16.31	18.10	19.73	21.34	23.03	24.94	27.30	30.81	33.92	36.78	40.29	42.80
23	17.19	19.02	20.69	22.34	24.07	26.02	28.43	32.01	35.17	38.08	41.64	44.18
24	18.06	19.94	21.65	23.34	25.11	27.10	29.55	33.20	36.42	39.36	42.98	45.56
25	18.94	20.87	22.62	24.34	26.14	28.17	30.68	34.38	37.65	40.65	44.31	46.93
26	19.82	21.79	23.58	25.34	27.18	29.25	31.79	35.56	38.89	41.92	45.64	48.29
27	20.70	22.72	24.54	26.34	28.21	30.32	32.91	36.74	40.11	43.19	46.96	49.64
28	21.59	23.65	25.51	27.34	29.25	31.39	34.03	37.92	41.34	44.46	48.28	50.99
29	22.48	24.58	26.48	28.34	30.28	32.46	35.14	39.09	42.56	45.72	49.59	52.34
30	23.36	25.51	27.44	29.34	31.32	33.53	36.25	40.26	43.77	46.98	50.89	53.67
40	32.34	34.87	37.13	39.34	41.62	44.16	47.27	51.81	55.76	59.34	63.69	66.77
50	41.45	44.31	46.86	49.33	51.89	54.72	58.16	63.17	67.50	71.42	76.15	79.49
60	50.64	53.81	56.62	59.33	62.13	65.23	68.97	74.40	79.08	83.30	88.38	91.95
70	59.90	63.35	66.40	69.33	72.36	75.69	79.71	85.53	90.53	95.02	100.4	104.2
80	69.21	72.92	76.19	79.33	82.57	86.12	90.41	96.58	101.88	106.6	112.3	116.3
90	78.56	82.51	85.99	89.33	92.76	96.52	101.1	107.6	113.2	118.1	124.1	128.3
100	87.95	92.13	95.81	99.33	103.0	106.9	111.7	118.5	124.3	129.6	135.8	140.2

Table E.5 95% quantile values of the Snedecor's $F(\nu_1, \nu_2)$ variable



	ν_1													
ν_2	1	2	3	4	6	8	10	15	20	30	40	60	120	∞
1	161	200	216	225	234	239	242	246	248	250	251	252	253	254
2	18.6	19.0	19.2	19.2	19.3	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	8.94	8.85	8.79	8.70	8.66	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.16	6.04	5.96	5.86	5.80	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	4.95	4.82	4.74	4.62	4.56	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.28	4.15	4.06	3.94	3.87	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.87	3.73	3.64	3.51	3.45	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.58	3.44	3.35	3.22	3.15	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.37	3.23	3.14	3.01	2.94	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.22	3.07	2.98	2.85	2.77	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.09	2.95	2.85	2.72	2.65	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.00	2.85	2.75	2.62	2.54	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	2.92	2.77	2.67	2.53	2.46	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.85	2.70	2.60	2.46	2.39	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.79	2.64	2.54	2.40	2.33	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.74	2.59	2.49	2.35	2.28	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.70	2.55	2.45	2.31	2.23	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.66	2.51	2.41	2.27	2.19	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.63	2.48	2.38	2.23	2.16	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.60	2.45	2.35	2.20	2.12	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.57	2.42	2.32	2.18	2.10	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.55	2.40	2.30	2.15	2.07	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.53	2.38	2.27	2.13	2.05	1.96	1.91	1.87	1.81	1.76
24	4.26	3.40	3.01	2.78	2.51	2.36	2.25	2.11	2.03	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.49	2.34	2.24	2.09	2.01	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.47	2.32	2.22	2.07	1.99	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.46	2.31	2.20	2.06	1.97	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.45	2.29	2.19	2.04	1.96	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.43	2.28	2.18	2.03	1.94	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.42	2.27	2.16	2.02	1.93	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.34	2.18	2.08	1.92	1.84	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.25	2.10	1.99	1.84	1.75	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.18	2.02	1.91	1.75	1.66	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.10	1.94	1.83	1.67	1.57	1.46	1.39	1.32	1.22	1.00

Table E.6 99% quantile values of the Snedecor's $F(\nu_1, \nu_2)$ variable



	ν_1													
ν_2	1	2	3	4	6	8	10	15	20	30	40	60	120	∞
1	4052	5000	5403	5625	5859	5981	6056	6157	6209	6261	6287	6313	6339	6366
2	98.5	99.0	99.3	99.2	99.3	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5
3	34.1	30.8	29.5	28.7	27.9	27.5	27.2	26.9	26.7	26.5	26.4	26.3	26.2	26.1
4	21.2	18.0	16.7	16.0	15.2	14.8	14.6	14.2	14.0	13.8	13.8	13.7	13.6	13.5
5	16.3	13.3	12.1	11.4	10.7	10.3	10.1	9.72	9.55	9.38	9.29	9.20	9.11	9.02
6	13.7	10.9	9.78	9.15	8.47	8.10	7.87	7.56	7.40	7.23	7.14	7.06	6.97	6.88
7	12.2	9.55	8.45	7.85	7.19	6.84	6.62	6.31	6.16	5.99	5.91	5.82	5.74	5.65
8	11.3	8.65	7.59	7.01	6.37	6.03	5.81	5.52	5.36	5.20	5.12	5.03	4.95	4.86
9	10.6	8.02	6.99	6.42	5.80	5.47	5.26	4.96	4.81	4.65	4.57	4.48	4.40	4.31
10	10.0	7.56	6.55	5.99	5.39	5.06	4.85	4.56	4.41	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.07	4.75	4.54	4.25	4.10	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.96	5.41	4.82	4.50	4.30	4.01	3.86	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.62	4.30	4.10	3.82	3.66	3.51	3.43	3.34	3.25	3.17
14	8.86	6.52	5.56	5.04	4.46	4.14	3.94	3.66	3.51	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.41	4.89	4.32	4.00	3.81	3.52	3.37	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.20	3.89	3.69	3.41	3.26	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.10	3.79	3.59	3.31	3.16	3.00	2.92	2.84	2.75	2.65
18	8.29	6.01	5.09	4.58	4.02	3.71	3.51	3.23	3.08	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	3.94	3.63	3.43	3.15	3.00	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	3.87	3.56	3.37	3.09	2.94	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	3.81	3.51	3.31	3.03	2.88	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.76	3.45	3.26	2.98	2.83	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.71	3.41	3.21	2.93	2.78	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.67	3.36	3.17	2.89	2.74	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.63	3.32	3.13	2.85	2.70	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.59	3.29	3.09	2.82	2.66	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.56	3.26	3.06	2.78	2.63	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.53	3.23	3.03	2.75	2.60	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.05	3.50	3.20	3.00	2.73	2.57	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.47	3.17	2.98	2.70	2.55	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.29	2.99	2.80	2.52	2.37	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.12	2.82	2.63	2.35	2.20	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	2.96	2.66	2.47	2.19	2.03	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	2.80	2.51	2.32	2.04	1.88	1.70	1.59	1.47	1.32	1.00

Bibliography

[A+03] S. Agostinelli et al. Geant4-a simulation toolkit. *Nuclear Instruments and Methods A*, 506:250–303, 2003.

[AAN+07] V.S. Anishchenko, V. Astakov, A. Neiman, T. Vadivasova, and L. Schimansky-Geier. *Non linear Dynamics of Chaotic and Stochastic Systems*. Springer, 2007.

[Azz96] A. Azzalini. *Statistical Inference Based on the Likelihood*. Chapman & Hall/CRC, Boca Raton, Florida (USA), 1996.

[Bar89] R. Barlow. *Statistics*. J. Wiley and Sons, New York, 1989.

[BCD01] L.D. Brown, T.T. Cai, and A. DasGupta. Interval Estimation for a Binomial Proportion. *Statistical Science*, 16:101–133, 2001.

[BFS87] P. Bratley, B.L. Fox, and L.E Schrage. *A Guide to Simulation*. Academic Press, New York, second edition, 1987.

[BH95] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society B*, 57:289–300, 1995.

[Blo84] V. Blobel. Unfolding Methods in High-Energy Physics Experiments. Technical Report DESY 84-118, DESY, Amburgo, 1984.

[BR92] P.R. Bevington and D.K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York, 1992.

[Bro13] W.A. Brown. *The Placeo Effect in Clinical Practice*. Oxford University Press, Oxford, 2013.

[BS91] M. Berblinger and C. Schlier. Monte Carlo integration with quasi-random numbers: some experience. *Computer Physics Communications*, 66:157–166, 1991.

[Buc84] S.T. Buckland. Monte Carlo Confidence Intervals. *Biometrics*, 40:811–817, 1984.

[Bun86] B.D. Bunday. *Basic Queueing Theory*. Edward Arnold, London, 1986.

[Car85] J. Carlson. A double blind test for astrology. *Nature*, 318:419–425, 1985.

[CB90] G. Casella and R.L. Berger. *Statistical Inference*. Wadsworth & Brook-Cole, Pacific Grove, 1990.

[Cha75] G.J. Chaitin. Randomness and Matemathical Proof. *Scientific American*, 232:47–53, May 1975.

[Coc77] W.G. Cochran. *Sampling Techniques*. J.Wiley and Sons, New York, third edition, 1977.

[Col83a] UA1 Collaboration. Experimental Observations of Lepton Pairs of Invariant Mass around 95 GeV/c at the CERN SPS Collider. *Physics Letters B*, 126:398–410, 1983.

[Col83b] UA2 Collaboration. Evidence of $Z^0 \to e^+e^-$ at the CERN Collider. *Physics Letters B*, 129:130–140, 1983.

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2
139, https://doi.org/10.1007/978-3-031-09429-3

[Com91] A. Compagner. Definitions of randomness. *American Journal of Physics*, 59:700–705, 1991.

[Cou95] R.D. Cousins. Why isn't every physicist a bayesian? *American Journal of Physics*, 63:398–410, 1995.

[Cra51] H. Cramer. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, 1951.

[CS61] D.R. Cox and W.C. Smith. *Queues*. Chapman and Hall, London, 1961.

[D'A94] G. D'Agostini. On the use of the covariance matrix to fit the correlated data. *Nuclear Instruments and Methods in Physics Research A*, 346:306–311, 1994.

[D'A99] G. D'Agostini. Bayesian Reasoning in High-Energy Physics: Principles and Applications. Technical Report CERN 99-03, CERN, Ginevra, 1999.

[Dav08] A.C. Davison. *Statistical Models*. Cambridge University Press, New York, 2008.

[DE83] P. Diaconis and B. Efron. Computer-Intensive Methods in Statistics. *Scientific American*, 248(5):116–130, 1983.

[DF33] B. De Finetti. Sul concetto di probabilità. *Rivista Italiana di Statistica, Economia e Finanza*, V-4:723–747, 1933. English translation of the paper published in: *Statistica*, LXXVI-3, 2016.

[DH99] A.C. Davison and D.V. Hinkley. *Bootstrap Methods and their Applications*. Cambridge University Press, Cambridge, 1999.

[Efr79] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7:1–26, 1979.

[Efr82] B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Filadelfia, 1982.

[ET93] B. Efron and R.J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, London, 1993.

[Eva55] R.D. Evans. *The Atomic Nucleus*. Mc Graw-Hill, New York, 1955.

[F+18] L. Falcioni et al. Report of final results regarding brain and heart tumors in Sprague-Dawley rats exposed from prenatal life until natural death to mobile phone radiofrequency field representative of a 1.8 GHz GSM base station environmental emission. *Environmental Research*, 165:496–503, 2018.

[FC98] G.J. Feldman and R.D. Cousins. A unified approach to the classic statistical analysis of small signals. *Physical Review D*, 57:5873–5889, 1998.

[Fel47] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley and Sons, New York, second edition, 1947.

[Fey18] R.P. Feynman. *Surely You're Joking, Mr. Feynman!: Adventures of a Curious Character*. Norton & Co, London, 2018.

[Fis41] R.A. Fisher. *Statistical Methods for Research Workers*. Eighth, London, 1941.

[Fis96] G.S. Fishman. *Monte Carlo Concepts, Algorithms, and Applications*. Springer-Verlag, New York, 1996.

[FJ10] M.J. Flegal and G.L. Jones. Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38:1034–1070, 2010.

[FLJW92] A.M. Ferrenberg, D.P. Landau, and Y. Joanna Wong. Monte Carlo Simulation: hidden errors from "good" random numbers generators. *Physical Review Letters*, 69:3382–3384, 1992.

[Fra84] A. Franklin. Forging, cooking, trimming, and riding on the bandwagon. *American Journal of Physics*, 52:786–793, 1984.

[Fra97] A. Franklin. Millikan's Oil-Drop Experiments. *The Chemical Educator*, 2:1–14, 1997.

[fSI93] International Organization for Standardization (ISO). Guide to the expression of uncertainty in measurement. Technical report, ISO, Ginevra, 1993.

[Gne76] B.V. Gnedenko. *The Theory of Probability*. Mir, Moscow, 1976.

[Gre06] P. Gregory. *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press, Cambridge, 2006.

[GS92] G.R. Grimmet and D.R. Stirzaker. *Probability and Random Processes*. Clarendon Press, Oxford, 1992.

[Has70] W.K. Hastings. Monte Carlo samplings methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[HH64] J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods*. Chapman and Hall, London, 1964.

[Hix76] J.R. Hixson. *The Patchwork Mouse*. Anchor Press, New York, 1976.

[HL07] J. Heinrich and L. Lyons. Systematic errors. *Annual Review of Nuclear and Particle Science*, 57:145–169, 2007.

[Jam80] F. James. Monte Carlo theory and practice. *Reports on Progress in Physics*, 43:1145–1189, 1980.

[Jam90] F. James. A review of pseudorandom number generators. *Computer Physics Communications*, 60:329–344, 1990.

[Jam92] F. James. Minuit Reference Manual. Technical Report CERN D506, CERN, Geneva, 1992.

[Jam08] F. James. *Statistical Methods in Experimental Physics*. World Scientific, London, 2008.

[Jen06] M. Jeng. A selected history of expectation bias in physics. *American Journal of Physics*, 74:578–583, 2006.

[JLPe00] F. James, L. Lyons, and Y. Perrin (editors). Proceedings of "Workshop on confidence limits". Technical Report CERN 2000-005, CERN, Geneva, 2000.

[Kah56] H. Kahn. Use of different Monte Carlo sampling techniques. In H.A. Meyer, editor, *Symposium on Monte Carlo Methods*, pages 146–190. J.Wiley and Sons, New York, 1956.

[KH96] K. Kacperski and A. Holyst. Phase Transition and Hysteresis in a Cellular Automata-Based Model of Opinion Formation. *Journal of Statistical Physics*, 84:168–189, 1996.

[Knu81] D.E. Knuth. *The Art of Computer Programming*, volume 2: Seminumerical Algorithms, chapter 2. Addison-Wesley, Reading, second edition, 1981.

[Kol33] A. Kolmogorov. Sulla determinazione empirica di una legge di distributione. *Giornale dell' Istituto Italiano degli Attuari*, 4:83–91, 1933. English translation of the paper published in: A. N. Shiryayev (ed.), *Selected Works of A. N. Kolmogorov*, volume II, 139–146, Springer, 1992.

[KR88] B.W. Kernighan and D.M. Ritchie. *The C Programming Language*. Pearson Education, USA, 1988.

[KS73] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*, volume II. Griffin, London, 1973.

[KW86] M.H. Kalos and P. Whitlock. *Monte Carlo Methods*, volume One: Basics. J.Wiley and Sons, New York, 1986.

[Lam66] J.R. Lamarsh. *Nuclear Reactor Theory*. Addison-Wesley, Reading (Mass.), 1966.

[Lev60] H. Levene. *Robust Tests for Equality of Variances. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. Stanford University Press, 1960.

[LSH+90] N. Lenssen, G. Schmidt, J. Hansen, M. Menne, A. Persin, R. Ruedy, and D. Zyss. Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres*, 124:6307–6326, 1990.

[LW65] T.D. Lee and C.S. Wu. Weak interactions. *Annual Review of Nuclear Science*, 15:381–476, 1965.

[LW18] L. Lyons and N. Wardle. Statistical issues in searches for new phenomena in High Energy Physics. *Journal of Physics G; Nuclear and Particle Physics*, 45(3):033001, 2018.

[Lyb84] M. Lybanon. Comment on "Least squares when both variables have uncertainties". *American Journal of Physics*, 52:276–278, 1984.

[Lyo13] L. Lyons. Discovering the Significance of 5 sigma. Eprint: https://doi.org/10.48550/arXiv.1310.1284, 2013.

[MGB73] A.M. Mood, F.A. Graybill, and D.C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, New York, 1973.

[MNZ90] G. Marsaglia, B. Narashimhan, and A. Zaman. A random number generator for PC's. *Computer Physics Communications*, 60:345–349, 1990.

[Mon03] D. C. Montgomery. *Design and analysis of experiments*. John Wiley & sons, New York, 2003.

[Mor84] B.J.T. Morgan. *Elements of Simulations*. Chapman and Hall, London, 1984.

[MRR+53] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equations of State Calculations by Fast Computing Machines. *the Journal of Chemical Physics*, 21:1087–1092, 1953.

[Ney37] J. Neyman. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Phil. Trans. Royal Society London*, 236:333–380, 1937.

[Ons44] L. Onsager. A Two Dimensional Model with an Order-Disorder Transition. *Physical Review*, 65:117–149, 1944.

[Ore82] J. Orear. Least squares when both variables have uncertainties. *American Journal of Physics*, 50:912–916, 1982.

[PFTW92] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Wetterling. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, second edition, 1992.

[Pop59] J. R. Popper. *The Logic of Scientific Discovery*. Hutchinson & Co, London, 1959.

[PS20] P. Pedroni and S. Sconfietti. A new Monte Carlo-based fitting method. *Journal of Physics G; Nuclear and Particle Physics*, 47(5):05401, 2020.

[PUP02] A. Papoulis and S. Unnikrishna Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, Europe, 2002.

[RC99] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Verlag, New York, 1999.

[Rip86] B. Ripley. *Stochastic Simulation*. J.Wiley and Sons, New York, 1986.

[Ros96] S.B. Ross. *Simulation*. Academic Press, London, second edition, 1996.

[Rot10] A. Rotondi. On frequency and efficiency measurements in counting experiments. *Nuclear Instruments and Methods in Physics Research A*, 614:106–119, 2010.

[RPP] A. Rotondi, P. Pedroni, and A. Pievatolo. Text web site: https://tinyurl.com/ProbStatSimul.

[RS17] C. Rothleitner and S. Schlamminger. Invited Review Article: Measurements of the Newtonian constant of gravitation, G. *Review of Scientific Instruments*, 88:111101.25–111101.28, 2017.

[Rub81] R.Y. Rubinstein. *Simulation and Monte Carlo Method*. J. Wiley and Sons, New York, 1981.

[Rue96] D. Ruelle. *Chance and Caos*. Princeton Science Library, Princeton, 1996.

[RvN63] R.D. Richmeyer and J. von Neumann. Statistical methods in neutron diffusion. In A.H. Taub, editor, *John von Neumann collected works*, volume V, pages 751–767. Pergamon Press, Oxford, 1963.

[Spi61] M.R. Spiegel. *Statistics*. McGraw-Hill, New York, second edition, 1961.

[Ste97] I. Stewart. *Does God Play Dice? The New Mathematics of Chaos*. Penguin Books, London, 1997.

[SW89] G.A.F. Seber and C.J. Wild. *Nonlinear Regression*. Wiley Interscience, New York, 1989.

[TC93] R. Toral and A. Chakrabarti. Generation of gaussian distributed random numbers by using a numerical inversion method. *Computer Physics Communications*, 74:327–334, 1993.

[Tea22] GISTEMP Team-2022. *GISS Surface Temperature Analysis (GISTEMP), version 4.* NASA Goddard Institute for Space Studies; https://data.giss.nasa.gov/gistemp/, 2022. [Online; dataset accessed 2022-02-07].

[Tuk49] J. Tukey. Comparing Individual Means in the Analysis of Variance. *Biometrics*, 5:99–114, 1949.

[vD07] J. van Dongen. Emil Rupp, Albert Einstein, and the canal ray experiments on wave-particle duality: Scientific fraud and theoretical bias. *Historical Studies in the Physical and Biological Sciences*, 37, Supplement:73–120, 2007.

[Voi03] J. Voit. *The Statistical Mechanics of Financial Markets*. Springer-Verlag, Berlin, second edition, 2003.

[W+18] C.J. Werner et al. MCNP Version 6.2 Release Notes. Technical Report LA-UR-18-20808, Los Alamos Laboratories, Los Alamos, 2018.

[Wel47] B.L. Welch. The generalization of "Student's" problems when several different population variances are involved. *Biometrika*, 34:28–35, 1947.

[Wik22] Wikipedia. *Sally Clark*—Wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Sally_Clark, 2022. [Online; accessed 2022-07-02].

[ZeaPDG20] P.A. Zyla et al. (Particle Data Group). Review of Particle Physics. *Progress in Theoretical and Experimental Physics*, 2020:083C01, 2020.

Index

A. Rotondi et al., *Probability, Statistics and Simulation*, La Matematica per il 3+2 139, https://doi.org/10.1007/978-3-031-09429-3

S