

Springer Texts in Business and Economics

Geon Ho Choe

Quantitative Methods for Finance with Simulations I

An Introduction to Stochastic Analysis
and Option Pricing

 Springer

Springer Texts in Business and Economics

Springer Texts in Business and Economics (STBE) delivers high-quality instructional content for undergraduates and graduates in all areas of Business/Management Science and Economics. The series is comprised of self-contained books with a broad and comprehensive coverage that are suitable for class as well as for individual self-study. All texts are authored by established experts in their fields and offer a solid methodological background, often accompanied by problems and exercises.

Geon Ho Choe

Quantitative Methods for Finance with Simulations I

An Introduction to Stochastic Analysis
and Option Pricing

 Springer

Geon Ho Choe
Department of Mathematical Sciences
Korea Advanced Institute of Science
and Technology
Daejeon, Korea (Republic of)

ISSN 2192-4333 ISSN 2192-4341 (electronic)
Springer Texts in Business and Economics
ISBN 978-3-032-12326-8 ISBN 978-3-032-12327-5 (eBook)
<https://doi.org/10.1007/978-3-032-12327-5>

Mathematics Subject Classification: 91G30, 60H05, 91G20, 60J65

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

If disposing of this product, please recycle the paper.

In Memory of Loved Ones

Preface

This book is the first volume of a two volume set that provides an introduction to stochastic analysis and quantitative finance, including theoretical and computational methods, for advanced undergraduate and graduate students in mathematics and business, but not excluding practitioners in the finance industry. The two volume set is designed for readers who want to have a deeper understanding of the delicate theory of quantitative finance by doing computer simulations in addition to theoretical study. It is an expanded second edition of the previously published single volume edition titled *Stochastic Analysis for Finance with Simulations*. The first volume includes stochastic analysis, option pricing theory, optimal portfolio investment, bond pricing, with computer simulations for most theoretical ideas. The main topics of the second volume are numerical solution of the Black–Scholes–Merton partial differential equation and Monte Carlo methods for option pricing. Also included are ordinary and partial differential equations in general, stochastic differential equations, estimation of implied volatility, stochastic volatility models, and Fourier transform methods for option pricing. Basic measure theory is used as a tool to describe probabilistic phenomena. The level of familiarity with computer programming is kept to a minimum. To make the book accessible to a wider audience, some background mathematical facts are included in the early chapters of the book and also in the appendices.

Financial contracts are divided into two groups: The first group consists of primary assets such as shares of stock, bonds, commodities, and foreign currencies. The second group contains financial derivatives such as options and futures on the underlying assets belonging to the first group. A financial derivative is a contract that promises payment in cash or delivery of an asset contingent on the behavior of the underlying asset in the future. The goal of this book is to present mathematical methods for finding how much one should pay for a financial derivative. To understand the option pricing theory we need ideas from various disciplines ranging over pure and applied mathematics, not to mention finance itself. We try to bridge the gap between mathematics and finance by using diagrams, graphs and simulations in addition to theoretical exposition. The simulations in this book are not only used as a

computational method in quantitative finance, but can also facilitate an intuitive and deeper understanding of theoretical concepts.

Since the publications by Black, Scholes and Merton on option pricing in 1973, the theory of stochastic calculus, developed by Ito, has become the foundation for a new field called quantitative finance. In the first volume, stochastic calculus is presented starting from the theoretical first step. After introducing fundamental ideas in quantitative finance in Part I, we present mathematical prerequisites such as Lebesgue integration, probability, conditional expectation and stochastic processes in Part II. After that, basic properties of Brownian motion, Girsanov's theorem and the reflection principle are given in Part III. In Part IV we introduce the Ito integral and Ito's lemma, and then present the Feynman–Kac theorem. In Part V we present three methods for pricing options: the binomial tree method, the Black–Scholes–Merton partial differential equation, and the martingale method. In Part VI we analyze more examples of the martingale method, and study exotic options, American options and numeraires. In Part VII the variance minimization method for optimal portfolio investment is introduced for a discrete time model, and the Hamilton–Jacobi–Bellman equation is derived and used to solve optimization problems. In Part VIII some interest rate models are introduced and used in pricing bonds. In the appendices given in Part IX mathematical prerequisites are presented, including point set topology, linear algebra, ordinary differential equations, and partial differential equations, which are needed to understand the material in the main chapters. Basic MATLAB and Python commands are explained in Appendix A and Appendix G, respectively. The graphs and diagrams in this book were plotted by the author using tikz, MATLAB and Python, and the simulations were done using MATLAB and Python.

Comments from the readers are welcome. For corrections and updates please send emails to ghchoe@kaist.ac.kr.

Daejeon, Republic of Korea

Geon Ho Choe

Acknowledgements For the second edition the author wishes to thank Dong-Pyo Chi, Hyeong-In Choi, Dong Myung Chung, Dan Crisan, Will Geller, Ole Hald, Yuji Ito, Sun Young Jang, Intae Jeon, Duk Bin Jun, Byung-Chun Kim, Dohan Kim, Hong Jong Kim, Jaewan Kim, Kyung Soo Kim, Minhyong Kim, Soon Ki Kim, Sung Woo Kim, Tong Suk Kim, Yong Hoon Kwon, Dan Moore, Jong Hoon Oh, Adam Ostaszewski, Michael Tretyakov, Harry Zheng, and his colleagues in the Department of Mathematical Sciences and the Graduate School of Finance at KAIST.

He wishes to thank Youngho Ahn, Soeun Choi, Hyun Jin Jang, Myeong Geun Jeong, Mihyun Kang, Bong Jo Kim, Chihurn Kim, Dong Han Kim, Gunhee Kim, Minseok Kim, Sung Min Kim, Ki Hwan Koo, Soon Won Kwon, Dong Min Lee, Kyungsub Lee, Young Hoon Na, Jeong Yin Park, Jong Jun Park, Minseok Park, Byung Ki Seo, Min Hyeok Woo and Seongjun Yoon. He also thanks many students who took various courses on quantitative finance taught by him over the years.

The author thanks the editor Remi Lodh, the staff at Springer, and the anonymous referees whose many helpful suggestions contributed to the improvement of the quality of the book. Finally, he wishes to thank his wife for her love and patience.

Competing Interests The author has no competing interests to declare that are relevant to the content of this manuscript.

Contents

Part I Introduction to Financial Mathematics

1	Fundamental Concepts	3
1.1	Risk	3
1.2	Time Value of Money	4
1.3	No Arbitrage Principle	4
1.4	Arbitrage-Free Market	5
1.5	Risk-Neutral Pricing and Martingale Measures	10
1.6	The One Period Binomial Tree Model	12
1.7	Models in Finance	13
	Exercises	14
2	Financial Derivatives	17
2.1	Forward Contracts and Futures	18
2.2	Options	18
2.3	Put-Call Parity	21
2.4	Relations Among Option Pricing Methods	23
	Exercises	24

Part II Probability Theory

3	The Lebesgue Integral	27
3.1	Measures	27
3.2	Measurable Functions	30
3.3	The Lebesgue Integral	32
3.4	Inequalities	37
3.5	The Radon–Nikodym Theorem	39
	Exercises	41

4	Basic Probability Theory	43
4.1	Measure and Probability	43
4.2	Characteristic Functions	49
4.3	Independent Random Variables	53
4.4	Change of Variables	60
4.5	The Law of Large Numbers	63
4.6	The Central Limit Theorem	65
4.7	Statistical Ideas	66
4.8	Computer Experiments	68
	Exercises	73
5	Conditional Expectation	77
5.1	Conditional Expectation Given an Event	77
5.2	Conditional Expectation with Respect to a σ -Algebra	78
5.3	Conditional Expectation with Respect to a Random Variable	85
5.4	Conditional Variance	87
5.5	Computer Experiments	88
	Exercises	90
 Part III Stochastic Processes and Brownian Motion		
6	Stochastic Processes	95
6.1	Stochastic Processes	95
6.2	Predictable Processes	97
6.3	Martingales	98
6.4	Stopping Time	101
6.5	Computer Experiments	105
	Exercises	106
7	Brownian Motion	107
7.1	Brownian Motion as a Stochastic Process	107
7.2	Construction of Brownian Motion	115
7.3	Properties of Brownian Motion	119
7.4	Random Walk to the Horizon	122
7.5	Brownian Motion and Martingales	124
7.6	Computer Experiments	127
	Exercises	131
8	The Reflection Principle of Brownian Motion	133
8.1	The Reflection of Discrete Random Walks	133
8.1.1	Two One-Directional Random Walks	133
8.1.2	One-Dimensional Random Walks	135
8.2	The Reflection Property of Brownian Motion	136
8.3	The Maximum of Brownian Motion	138
8.4	The Maximum of Brownian Motion with Drift	140
8.5	Computer Experiments	145
	Exercises	146

Part IV Itô Calculus

9	The Itô Integral	149
9.1	Definition of the Itô Integral	149
9.2	Itô Isometry	151
9.3	Examples of the Itô Integral	155
9.3.1	A Correction Term for the Itô Integral	155
9.3.2	The Itô Integral of Deterministic Functions	158
9.4	The Martingale Property of the Itô Integral	158
9.5	Stochastic Integrals with Respect to a Martingale	162
9.6	The Martingale Representation Theorem	164
9.7	Computer Experiments	166
	Exercises	171
10	The Itô Formula	173
10.1	Motivation for the Itô Formula	173
10.2	The Itô Formula: Basic Form	178
10.3	The Itô Formula: General Form	185
10.4	Examples of the Itô Formula	188
10.4.1	Geometric Brownian Motion	188
10.4.2	Existence of Singularity	190
10.4.3	Brownian Bridge	191
10.4.4	Differentiability Condition	192
10.5	Computer Experiments	193
	Exercises	195
11	Girsanov's Theorem	197
11.1	Motivation	197
11.2	Equivalent Probability Measure	201
11.3	Brownian Motion with Drift	202
11.4	Girsanov's Theorem	203
11.5	The Martingale Property of the Radon–Nikodym Derivative	204
11.6	Computer Experiments	207
	Exercises	210
12	Stochastic Differential Equations	213
12.1	Strong Solutions	214
12.2	Weak Solutions	218
12.3	Brownian Bridges	220
12.4	Computer Experiments	227
	Exercises	231
13	The Feynman–Kac Theorem	233
13.1	The Feynman–Kac Theorem	233
13.2	Application to the Black–Scholes–Merton Equation	237

13.3	The Kolmogorov Equations	239
13.3.1	The Kolmogorov Backward Equation	239
13.3.2	The Kolmogorov Forward Equation	241
13.4	Computer Experiments	244

Part V Option Pricing Methods

14	The Binomial Tree Method for Option Pricing	249
14.1	Motivation for the Binomial Tree Method	249
14.2	The Single-Period Binomial Tree Method	250
14.2.1	Pricing by Hedging	250
14.2.2	Pricing by Replication	252
14.2.3	A Special Case: The Multiplicative Model	254
14.3	The Multiperiod Binomial Tree Method	254
14.3.1	Construction of the Multiperiod Binomial Tree	254
14.3.2	Delta-Hedging in the Multiperiod Binomial Tree Method	260
14.4	Convergence to the Black–Scholes–Merton Formula	262
14.5	Computer Experiments	268
	Exercises	279
15	The Black–Scholes–Merton Differential Equation	281
15.1	Derivation of the Black–Scholes–Merton Differential Equation	281
15.2	Price of a European Call Option	289
15.3	Stress Test for the Black–Scholes–Merton Formula	296
15.4	Greeks	298
15.5	Solution by the Laplace Transform	304
15.6	Stocks Paying Dividends	308
15.7	Computer Experiments	312
	Exercises	317
16	The Martingale Method	321
16.1	Option Pricing by the Martingale Method	321
16.1.1	Option Pricing by Hedging	322
16.1.2	Option Pricing by Replication	324
16.1.3	The Black–Scholes–Merton Equation Connection	326
16.2	The Probability Distribution of Asset Price	326
16.3	The Black–Scholes–Merton Formula	327
16.4	Derivation of the Black–Scholes–Merton Equation	328
16.5	Delta Hedging	329
16.6	Computer Experiments	331
	Exercises	333

Part VI Examples of Option Pricing

17 Pricing of Vanilla Options	337
17.1 Stocks with a Dividend	337
17.1.1 Continuous Dividends	337
17.1.2 Periodic Dividends	339
17.2 Bonds with Coupons	341
17.3 Binary Options	342
17.3.1 The Black–Scholes–Merton Formula for Binary Options	342
17.3.2 Delta of the Binary Option	352
17.3.3 The Binomial Tree Method for the Binary Option	357
17.4 Computer Experiments	358
Exercises	364
18 Pricing of Exotic Options	365
18.1 Asian Options	365
18.2 Barrier Options	368
18.3 Computer Experiments	377
19 American Options	381
19.1 American Call Options	381
19.2 American Put Options	382
19.3 Computer Experiments	387
Exercises	395

Part VII Portfolio Management

20 The Capital Asset Pricing Model	399
20.1 Return Rate and the Covariance Matrix	399
20.2 Portfolios of Two Assets and More	402
20.3 An Application of the Lagrange Multiplier Method	405
20.4 Minimum Variance Line	408
20.5 The Efficient Frontier	413
20.6 The Market Portfolio	415
20.7 The Beta Coefficient	419
20.8 Derivation of the Black–Scholes–Merton Equation	420
20.9 Computer Experiments	422
Exercises	426
21 Dynamic Programming	427
21.1 The Hamilton–Jacobi–Bellman Equation	427
21.2 Portfolio Management for Optimal Consumption	430
21.3 Computer Experiments	439

Part VIII Interest Rate Models

22	Bond Pricing	445
22.1	Periodic and Continuous Compounding	445
22.2	Zero Coupon Interest Rates	447
22.3	Term Structure of Interest Rates	448
22.4	Forward Rates	451
22.5	Yield to Maturity	452
22.6	Duration	453
22.7	Definitions of Various Interest Rates	456
22.8	The Fundamental Equation for Bond Pricing	458
22.9	Computer Experiments	466
	Exercises	474
23	Short-Rate Models	475
23.1	Affine Term Structure	475
23.2	The Vasicek Model	476
23.3	The Cox–Ingersoll–Ross Model	481
23.4	The Ho–Lee Model	483
23.5	The Hull–White Model	486
23.6	Computer Experiments	488
	Exercises	490
24	Numeraires	493
24.1	Change of Numeraire for a Binomial Tree Model	493
24.2	Change of Numeraire for Continuous Time	495
24.3	Numeraires for Pricing of Interest Rate Derivatives	501

Part IX Python Programming

25	Examples of Python Codes	507
25.1	Basic Commands	507
25.1.1	Arithmetic	507
25.1.2	Logic	508
25.1.3	Conditions for Loops	509
25.1.4	Random Numbers	510
25.1.5	Definition of a Function in Python	510
25.1.6	Set Operations	511
25.1.7	Sequences	512
25.1.8	Copies of a Sequence	513
25.1.9	Plotting of Graphs	514
25.1.10	Accumulation of Errors	517
25.2	Statistics	518
25.2.1	Moments of a Data Set	518
25.2.2	Correlation Coefficient and Scatter Plot	520

25.3	Linear Algebra	520
25.3.1	Arrays and Matrices	520
25.3.2	A System of Linear Equations	524
25.3.3	Eigenvalues and Eigenvectors	525
25.3.4	LU-Decomposition	526
25.3.5	Cholesky Decomposition	527
25.3.6	The Jacobi Method	528
25.3.7	The Gauss–Seidel Method	529
25.4	Probabilistic Experiments	529
25.4.1	The Monty Hall Problem	529
25.4.2	Random Walks	530
25.5	Brownian Motion	532
25.5.1	Brownian Motion	532
25.5.2	Geometric Brownian Motion	532
25.6	Probability Density of Geometric Brownian Motion	533
25.7	Option Pricing Methods	534
25.7.1	The Binomial Tree Model	534
25.7.2	The Black–Scholes–Merton Formula	536
25.7.3	Risk-Neutral Method for Option Pricing	537
25.8	Bond Pricing	537
25.8.1	Zero-Coupon Bond	538
25.8.2	Yield to Maturity	538
25.8.3	Pricing of Coupon Bonds	539
25.8.4	Duration	539
25.8.5	Convexity	540
25.8.6	The Vasicek Model	540
Appendix A: MATLAB Programming		543
Appendix B: Basic Analysis		555
Appendix C: Linear Algebra		563
Appendix D: Ordinary Differential Equations		569
Appendix E: Solution of the Heat Equation		577
Appendix F: Statistical Facts		585
Solutions for Selected Problems		589
Glossary		619
References		627
Index		631

Acronyms

ATM	At-the-money option
BSM	Black–Scholes–Merton
BTCS	Backward in time, central in space
CAPM	Capital Asset Pricing Model
cdf	Cumulative distribution function
CIR	Cox–Ingersoll–Ross model
EUR	Euro
FTCS	Forward in time, central in space
GARCH	Generalized ARCH model
GBP	British pound sterling
HJB	Hamilton–Jacobi–Bellman equation
ITM	In-the-money
LCG	Linear congruential generator
MA	Moving average model
MC	Monte Carlo method
ODE	Ordinary Differential Equation
OTC	Over-the-counter
OTM	Out-of-the-money
PDE	Partial Differential Equation
pdf	Probability Density Function
SDE	Stochastic Differential Equation
STRIPS	Separate trading of registered interest and principal securities
USD	US Dollar
YTM	Yield to maturity

Symbols

$\emptyset$	The empty set
∞	Infinity
$\mathbb{N}$	The set of natural numbers
$\mathbb{Z}$	The set of integers
$\mathbb{R}$	The set of real numbers
$\mathbb{C}$	The set of complex numbers
d	A metric
V	A vector space
$\ \cdot \ $	A norm
$\mathcal{F}$	A σ -algebra
$\mathcal{G}$	A sub- σ -algebra
$\{\mathcal{F}_t\}_{0 \leq t \leq T}$	A filtration
Ω	A probability space
A, B, E	Events, measurable subsets
X, Y	Random variables
$\mathbb{E}$	Expectation
$\text{Var}(X)$	Variance
$\text{Var}(\mathbf{X})$	Covariance matrix
Corr	Correlation
Π	A portfolio
Δ	Delta of an option
Γ	Gamma of an option
Θ	Theta of an option
B	Risk-free asset price
S	Risky asset price
r	Interest rate
$\{X_t\}, \{Y_t\}$	Stochastic processes
$\{W_t\}$	A Brownian motion
$\mathcal{H}_0^2$	The set of simple stochastic processes
$\mathcal{H}^2$	The closure of the set of simple stochastic processes
μ	Mean, drift coefficient

σ	Standard deviation, volatility
Σ, C	Covariance matrices
ρ	Correlation
A, B, C, M	Matrices
diag	Diagonal of a matrix
λ	An eigenvalue of a matrix
t	Index for time
K	Strike price of an option
T	Expiry date of an option, maturity date of a bond
$\mathbb{P}$	A (physical) probability
$\mathbb{Q}$	A risk-neutral probability
$N(x)$	Cdf of the standard normal distribution
$\int_{\Omega} X(\omega) \, d\mathbb{P}$	Lebesgue integral
$\int_0^T f(t, W_t) \, dt$	Itô integral
$\mathcal{L}$	Laplace transform
$\text{erf}(x)$	The error function
$\mathbb{Z}_p$	The field of integers modulo p for a prime number p
$\mathbb{Z}_n$	The error function
$f * g$	Convolution
$C(t_1, t_2; I_1, I_2)$	A cylinder subset of Brownian sample paths
d_1	$\{\log(S/K) + (r + \frac{1}{2}\sigma^2)T\}/(\sigma\sqrt{T})$
d_2	$\{\log(S/K) + (r - \frac{1}{2}\sigma^2)T\}/(\sigma\sqrt{T})$
U	A uniformly distributed random variable
Z	A standard normal variable

List of Figures

Fig. 1.1	The one period binomial tree for C and the risk-neutral probabilities	12
Fig. 1.2	The one period binomial tree for S and the risk-neutral probabilities	12
Fig. 1.3	The closed convex cone K in the proof of Farkas' lemma	14
Fig. 2.1	Payoffs of a European call (left) and a European put (right) . . .	19
Fig. 2.2	Payoffs of a European straddle (left) and a European strangle (right)	20
Fig. 2.3	Payoff of a European spread option	20
Fig. 2.4	Payoff of a butterfly	20
Fig. 2.5	A butterfly as a sum of two long calls and two contracts of a short call	21
Fig. 2.6	Payoff when we buy a European call option and sell a put option	22
Fig. 2.7	Logical relations among option pricing methods	23
Fig. 2.8	Sum of a short position on an asset and a call option	24
Fig. 3.1	The union and the intersection of two subsets A and B	28
Fig. 3.2	Inverse image of an interval (a, b) under X	31
Fig. 3.3	A simple function on a measurable space	32
Fig. 3.4	Partition of a domain for a Lebesgue integral	33
Fig. 3.5	Composition of a random variable X and a function ϕ	38
Fig. 4.1	An abstract model of a probability density function	45
Fig. 4.2	A model of a random variable X whose values are uniformly distributed in $[a, b]$	47
Fig. 4.3	Payoff of a European call option, $(S_T - K)^+$, as a composite function	48
Fig. 4.4	Probability density function of the standard normal distribution	48
Fig. 4.5	Cumulative distribution function of the standard normal distribution	50
Fig. 4.6	A disk $ z - 1 < 1$ for the power series expansion of $\log z$	52

Fig. 4.7	Independent random variables X and Y are viewed as projections	56
Fig. 4.8	Sum of two independent random variables	59
Fig. 4.9	A new probability density function obtained by change of variables	61
Fig. 4.10	The probability density function of $\exp X$ when $X \sim N(0, 1)$	62
Fig. 4.11	The composite random variable $F(X)$	63
Fig. 4.12	Scatter plots for independent standard normal variables Z_1 and Z_2 (<i>left</i>), and negatively correlated normal variables X_1 and X_2 (<i>right</i>)	68
Fig. 4.13	Uniform distribution of $F_X(X)$	69
Fig. 4.14	Equally spaced points along the diagonal (<i>left</i>) and a quantile-quantile plot for the standard normal distribution (<i>right</i>)	70
Fig. 4.15	The Central Limit Theorem: distribution of sample averages $(U_1 + \dots + U_n)/n$, $n = 2, 5, 20, 100$, for the uniform distribution $U(0, 1)$	72
Fig. 4.16	The composite random variable $F^{-1}(U)$	73
Fig. 4.17	A contour for $f(z) = \exp(-z^2/(2t))$ for $z \in \mathbb{C}$	74
Fig. 4.18	Support of the joint pdf (<i>left</i>) of uncorrelated normal variables X, Y in Exercise 4.18 and the graph of the joint pdf (<i>right</i>)	75
Fig. 4.19	The pdf and cdf of the product of two independent uniformly distributed random variables U and V	76
Fig. 5.1	Conditional expectation of X given an event A	78
Fig. 5.2	Conditional expectation when a probability space is decomposed into subsets of positive measure	79
Fig. 5.3	Comparison of $\mathbb{E}[X A]$, $\mathbb{E}[X B]$, $\mathbb{E}[X A \setminus B]$ where $B \subset A$	80
Fig. 5.4	Conditional expectation as an orthogonal projection	84
Fig. 5.5	Conditional expectation with respect to a sub- σ -algebra $\mathcal{G}$ on $\Omega = \{\omega_1, \omega_2, \omega_3\}$	85
Fig. 5.6	Conditional expectation of X with respect to Y	86
Fig. 5.7	Scatter plot of sample values of (Z^2, Z) marked by blue circles (<i>left</i>), and the plots of (Z, Z) marked by blue circles and $(Z, \mathbb{E}[Z Z^2] = 0)$ marked by red circles on symmetric sets $\{z, -z\}$ (<i>right</i>)	87
Fig. 5.8	A diagram for Bayes' theorem	91
Fig. 6.1	Martingale property of the symmetric random walk	100
Fig. 6.2	The first time T that the digit '0' appears in the binary expansion	103
Fig. 7.1	Sample paths of Brownian motion together with the parabola $t = W^2$ (<i>left</i>) and the probability density functions for W_t as t increases (<i>right</i>)	109

Fig. 7.2	Sample paths of Brownian motion for $0 \leq t \leq 10^{-8}$ and for $0 \leq t \leq 10^8$	109
Fig. 7.3	Exponential growth of the geometric Brownian motion: The sample average of S_t (<i>in red on the right</i>) is close to the exponential curve $y = S_0 e^{\mu t}$ (<i>dashed in black</i>) for $\mu = 0.2$ and $\sigma = 0.1$	113
Fig. 7.4	Exponential growth of the geometric Brownian motion: The sample average of S_t (<i>in red on the right</i>) is close to the exponential curve $y = S_0 e^{\mu t}$ (<i>dashed in black</i>) for $\mu = 0.2$ and $\sigma = 0.3$	113
Fig. 7.5	Construction of a sample path of Brownian motion	116
Fig. 7.6	A sample path of Brownian motion constructed by the bisection method	117
Fig. 7.7	A cylinder subset $C(t_1, \dots, t_n; I_1, \dots, I_n)$ for Brownian motion	118
Fig. 7.8	Brownian motion and a cylinder subset	119
Fig. 7.9	The Law of Large Numbers for Brownian motion	120
Fig. 7.10	The Law of Large Numbers for geometric Brownian motion for $\mu > 0$ and $\mu - \frac{1}{2}\sigma^2 < 0$ with exponentially increasing average	121
Fig. 7.11	The Law of Large Numbers for geometric Brownian motion and the martingale property for $\mu = 0$	121
Fig. 7.12	While an observer watches on a platform, a random walker staggers away	122
Fig. 7.13	The projected trajectory of a random walker converges to a point on the horizon	123
Fig. 7.14	Conditional expectation of W_t on a cylinder set $C(s; J)$	124
Fig. 8.1	A sample path ω of the two dimensional random walk arising from a sequence of payments at a ticket office (solid in black), and the reflected path ω' after ω hits a barrier $y = x + 1$ (dotted in red)	134
Fig. 8.2	A sample path of ω the random walk arising from a sequence of payments at a ticket office (<i>solid in black</i>), and the reflected path ω' after ω hits a barrier $y = -1$ (<i>dotted in red</i>)	135
Fig. 8.3	A Brownian path is stopped when it hits a barrier $m = 3$	136
Fig. 8.4	A Brownian path is reflected after it hits a barrier	137
Fig. 8.5	Range of (M_T, W_T)	138
Fig. 8.6	A scatter plot of (M_T, W_T) (<i>left</i>) and $f_{M_T, W_T}(m, w)$ (<i>right</i>) for $T = 2$	140
Fig. 8.7	Domain of integration for the double integral I_2	144
Fig. 9.1	A sample path of a simple process	151
Fig. 9.2	Approximation of a general stochastic process by a simple process	154

Fig. 9.3	A sample path of the Itô integral $\int_0^t W_s dW_s$ (solid) and a sample path of the exact solution $\frac{1}{2}W_t^2 - \frac{1}{2}t$ (dotted) in Sect. 9.3.1	156
Fig. 9.4	Sample paths of the stochastic process $\int_0^t W_s dW_s$, $0 \leq t \leq 3$, in Sect. 9.3.1	156
Fig. 9.5	Speed of L^2 -convergence of $I(f_n) = \sum W_i \delta W_i$ to $\frac{1}{2}W_T^2 - \frac{1}{2}T$ in Sect. 9.3.1	157
Fig. 9.6	For $T = 3$, normalized histograms of the probability distributions of random variables $X = \int_0^T t dW_t$ (left) and $Y = \int_0^T (T - t)dW_t$ (right) in Sect. 9.3.2 plotted together with the probability density function of the normal distribution $N(0, T^3/3)$ (in red)	159
Fig. 9.7	Scatter plot of (X, Y) where $X = \int_0^T t dW_t$ and $Y = \int_0^T (T - t) dW_t$ for $T = 3$ in Sect. 9.3.2	159
Fig. 9.8	The Itô integral $\int_0^t f_s dW_s$ has zero mean for $f(t, W_t) = \exp(W_t)$, $0 \leq t \leq 1$: five samples of increments dW_i of a Brownian motion, $0 \leq i \leq N - 1$; sample paths of a Brownian motion $W(t)$ observed at time $t = t_i$, $0 \leq i \leq N$; sample paths of a step process defined by $f(t_i, W(t_i))$, $t_i \leq t < t_{i+1}$, $0 \leq i \leq N - 1$; $f(t_i, W_i)dW_i$, $0 \leq i \leq N - 1$; sample paths of $\int_0^t f(u, W_u)dW_u$ observed at $t = t_i$, $0 \leq i \leq N$; and a numerical demonstration of $\mathbb{E}[\int_0^t f(s, W_s)dW_s] = 0$ at $t = t_i$, $0 \leq i \leq N$ (from top left to bottom right)	161
Fig. 10.1	Graph of a function and Itô's lemma	174
Fig. 10.2	Approximation errors of $f(t, x) = x^2$ by the second order Taylor expansion and the Itô formula	176
Fig. 10.3	Speed of L^2 -convergence in Itô formula for $f(t, x) = x^2$ (left) and $f(t, x) = \sin x$ (right)	178
Fig. 10.4	Sample paths of geometric Brownian motion together with the average given by the curve $S = S_0 e^{\mu t}$ (left) and the probability density functions for S_t (right)	189
Fig. 10.5	For $T = 3$, $\mathbb{E}[\int_0^t S_u dW_u] = 0$, $0 \leq t \leq T$	191
Fig. 11.1	Sample paths of a Brownian motion with drift $X_t = W_t + \theta t$ with drift coefficient $\theta = 0.5$ where the straight lines are $x = \theta t$ (solid in black) and $x = 0$ (dotted in blue)	198
Fig. 11.2	Probability density functions of $X_t = W_t + \theta t$ with respect to $\mathbb{P}$ (red) and $\mathbb{Q}$ (blue)	198
Fig. 11.3	The graph of $L(x) = \rho_T(x)$ (left), and the probability density functions of $X_T = W_T + \theta T$ with respect to $\mathbb{P}$ and $\mathbb{Q}$ (right), where $\theta = 0.5$ and $T = 5$	200

Fig. 11.4	The martingale property of $L_t = \exp(-\frac{1}{2}\theta^2 t - \theta W_t)$ is used in computing $\mathbb{E}^{\mathbb{P}}[W_s L_t]$, $0 \leq t \leq T$, where $\theta = 0.5$, $T = 5$ and $s = 3$	205
Fig. 11.5	The martingale property of $L_t = \exp(-\frac{1}{2}\theta^2 t - \theta W_t)$ is used in computing $\mathbb{E}^{\mathbb{P}}[W_s^2 L_t]$, $0 \leq t \leq T$, where $\theta = 0.5$, $T = 5$ and $s = 3$	206
Fig. 11.6	The martingale property of $L_t = \exp(-\frac{1}{2}\theta^2 t - \theta W_t)$ is used in computing $\mathbb{E}^{\mathbb{P}}[\exp(W_s) L_t]$, $0 \leq t \leq T$, where $\theta = 0.5$, $T = 5$ and $s = 3$	207
Fig. 12.1	Sample paths of the Ornstein–Uhlenbeck process	217
Fig. 12.2	Sample paths of the Brownian bridge $X_t = X_t^{0 \rightarrow 0}$ from $a = 0$ to $b = 0$ for $T = 3$	221
Fig. 12.3	A numerical solution of $dX_t = 2W_t dW_t + dt$ and the exact solution $X_t = W_t^2 + X_0$	228
Fig. 12.4	Sample paths of $Y_t = (T - t) \int_0^t \frac{1}{T-s} dW_s$ for $0 \leq t \leq 3$	230
Fig. 12.5	Brownian bridge: sample paths of $dY_t = \frac{1}{T-t} (b - Y_t) dt + dW_t$, $Y_0 = a$ and $Y_T = b$, for $0 \leq t \leq T = 3$	231
Fig. 13.1	The Feynman–Kac Theorem: Expectation of the final value $h(X_T)$ is the present value $F(t, x)$	234
Fig. 13.2	The Feynman–Kac theorem for a Brownian motion with drift given $X_t = x$	237
Fig. 13.3	Derivation of the Kolmogorov backward equation	240
Fig. 14.1	A one period binomial tree and real probabilities	250
Fig. 14.2	Prices of risk-free and risky assets in the one period binomial tree model	251
Fig. 14.3	A single-period binomial tree and the risk-neutral probability	252
Fig. 14.4	A multiperiod binomial tree	256
Fig. 14.5	Payoff of a European call option in a binomial tree	257
Fig. 14.6	The first step in pricing a European call option by a binomial tree	257
Fig. 14.7	Pricing of a European call option by a binomial tree	258
Fig. 14.8	The binomial tree for the asset price with $S_0 = 110$ and $K = 110$	259
Fig. 14.9	The binomial tree for the European call option price with $S_0 = 110$ and $K = 110$	259
Fig. 14.10	The binomial tree for the Delta of the European call option with $S_0 = 110$ and $K = 110$	260
Fig. 14.11	Delta-hedging in the multiperiod binomial tree method for a European call option using a self-financing portfolio $\Pi = -V + B + \Delta S$ with $T = 3$, $S_0 = 10$, $K = 15$ and $N = 15$ (<i>from top left to bottom right</i>)	261

Fig. 14.12	Convergence of the price obtained by the binomial tree method to the Black–Scholes–Merton price	265
Fig. 14.13	The binomial trees for the asset price (<i>left, red dots</i>) and the European call option price (<i>right, blue dots</i>) together with the graph of the Black–Scholes–Merton formula	267
Fig. 14.14	The binomial tree for the asset price with $S_0 = 110$	269
Fig. 14.15	The binomial tree for the European put option price	270
Fig. 14.16	The binomial tree for the Delta of the European put option	270
Fig. 14.17	Asset price S in the binomial tree model	280
Fig. 14.18	Option price V in the binomial tree model	280
Fig. 15.1	Adjustment of Δ_t in discrete time hedging	283
Fig. 15.2	Components of $\delta\Pi$ in Delta hedging where $\Pi = -V + D + \Delta S$	284
Fig. 15.3	In-the-money European call option: Delta hedging using a self-financing portfolio $\Pi = -V + D + \Delta S$ (<i>from top left to bottom right</i>)	286
Fig. 15.4	Out-of-the-money European call option: Delta hedging using a self-financing portfolio $\Pi = -V + D + \Delta S$ (<i>from top left to bottom right</i>)	287
Fig. 15.5	At-the-money European call option: Delta hedging using a self-financing portfolio $\Pi = -V + D + \Delta S$ (<i>from top left to bottom right</i>)	288
Fig. 15.6	The boundary condition of the Black–Scholes–Merton equation for a European call option	289
Fig. 15.7	Price of a European call option	295
Fig. 15.8	Price of a European put option	295
Fig. 15.9	Prices of European call and put options as functions of asset price with different times to expiration $\tau = T - t$	295
Fig. 15.10	Prices of European call and put options as functions of time to expiry $\tau = T - t$ with different strike prices	296
Fig. 15.11	Delta of European call and put options	303
Fig. 15.12	Gamma of European call and put options	303
Fig. 15.13	Rho of European call and put options	304
Fig. 15.14	Vega of European call and put options	304
Fig. 15.15	Theta of European call and put options	305
Fig. 15.16	Boundary condition for the Laplace transformation of a European call price	308
Fig. 15.17	Readjustment of the bank deposit in discrete time hedging	313
Fig. 15.18	The boundary condition of the Black–Scholes–Merton equation for a European put option	318
Fig. 16.1	Time discretization for discounted bond price $\tilde{D}_t$	323
Fig. 16.2	Convergence to the Black–Scholes–Merton price	332
Fig. 16.3	Convergence to the Black–Scholes–Merton formula for Delta	332

Fig. 17.1	Payoffs of cash-or-nothing call and put (clo)	343
Fig. 17.2	Payoffs of asset-or-nothing call and put	343
Fig. 17.3	Price of a cash-or-nothing call option	344
Fig. 17.4	Boundary conditions of a cash-or-nothing call option	344
Fig. 17.5	Difference of two European call options	347
Fig. 17.6	Approximation of a cash-or-nothing call option	347
Fig. 17.7	Difference of two binary call options	348
Fig. 17.8	Delta of a cash-or-nothing call option as a function of t and S with $K = 5$ and $T = 1$	353
Fig. 17.9	Delta of a cash-or-nothing call option as a function of S with $K = 5$ and $T = 1$ as $t \rightarrow T$	353
Fig. 17.10	Delta-hedging of the at-the-money digital call option using a self-financing portfolio $\Pi = -V + D + \Delta S$: The option price fluctuates between 0 and 1, and Delta can become large as $t \rightarrow T$ (<i>from top left to bottom right</i>)	355
Fig. 17.11	Difference of two binary options	356
Fig. 17.12	$z = d_2(x)$ as a function of strike price x where $S = 10$, $r = 0.05$, $\sigma = 0.4$ and $T = 0.25$	357
Fig. 17.13	Prices obtained by the binomial tree method and the Black–Scholes–Merton formula for binary call options where $K < S_0$ (<i>left</i>) and $K > S_0$ (<i>right</i>)	358
Fig. 18.1	An asset price path for Asian option with six monitoring times	366
Fig. 18.2	A knock-out barrier option with five sample asset paths	369
Fig. 18.3	A double barrier option	369
Fig. 18.4	Boundary and final conditions for an up-and-out call barrier option	371
Fig. 18.5	Price of a down-and-out put option	372
Fig. 19.1	Optimal exercise boundary for an American put option	384
Fig. 19.2	The price P of an American put option when the asset price S_t takes the boundary price S_t^* at time t	384
Fig. 19.3	The price V of the perpetual American put option when the asset price S takes the optimal boundary price L_0	385
Fig. 19.4	The graph of the function $V(S, L)$ given in Theorem 19.3 where $K = 100$, $r = 0.05$, $\sigma = 0.3$	387
Fig. 19.5	The graph of the function $V(S, L)$, $0 \leq L \leq K$, given in Theorem 19.3, for various values of S where $K = 100$, $r = 0.05$, $\sigma = 0.3$ and $L_0 = 52.63$	388
Fig. 19.6	The graph of the function $V(S, L)$, $0 \leq S \leq 200$, given in Theorem 19.3, for various values of L where $K = 100$, $r = 0.05$, $\sigma = 0.3$ and $L_0 = 52.63$	388
Fig. 19.7	The binomial tree for the asset price with $S_0 = 110$ and exercise price $K = 140$	389
Fig. 19.8	The binomial tree for the American put option price with $S_0 = 110$ and exercise price $K = 140$	389

Fig. 19.9	The early exercise domain for an American put option marked by ‘\$’ (<i>in blue</i>) and the no-exercise domain marked by ‘.’ (<i>in red</i>)	392
Fig. 20.1	A risk-free asset (<i>left</i>) and a risky asset (<i>right</i>)	400
Fig. 20.2	A line representing portfolios for $\rho_{12} = 1$ (<i>left</i>) and a line for $\rho_{12} = -1$ (<i>right</i>)	403
Fig. 20.3	A curve representing portfolios of two risky assets for $0 < \rho_{12} < 1$ (<i>left</i>) and a curve for $-1 < \rho_{12} < 0$ (<i>right</i>) . . .	404
Fig. 20.4	Portfolios consisting of a risk-free asset and a risky asset	404
Fig. 20.5	The set of attainable portfolios consisting of three risky assets	405
Fig. 20.6	The image of the triangle $w_1 + w_2 + w_3 = 1$, $w_1, w_2, w_3 \geq 0$	406
Fig. 20.7	Two optimization problems: Find the maximum of expected return when σ_0 is given, or find the minimum of variance when μ_0 is given	409
Fig. 20.8	The minimum variance line for risky assets and the point (σ_0, μ_0) representing the portfolio with the minimum variance among all portfolios (<i>left</i>), and asymptotes and a tangent line (<i>right</i>)	412
Fig. 20.9	The first asset dominates the second asset	413
Fig. 20.10	The efficient frontier	414
Fig. 20.11	A portfolio with a risk-free asset	416
Fig. 20.12	The capital market line and the market portfolio	417
Fig. 20.13	The relation between the beta coefficient and expected return	420
Fig. 21.1	Partition of time interval for total utility	429
Fig. 21.2	Sample paths of optimal portfolio X_t with power utility for $0 \leq t \leq T$	435
Fig. 21.3	The proportion of spending $h(t)^{1/(\gamma-1)}$ for $0 \leq t \leq 30$	435
Fig. 21.4	Optimal consumption and portfolio for $0 \leq t \leq 30$	436
Fig. 21.5	Optimal portfolio X_t with log utility for $0 \leq t < \infty$	437
Fig. 22.1	Monotone convergence of $(1 + 1/n)^n$ to the limit $e = 2.71828 \dots$ as $n \rightarrow \infty$	446
Fig. 22.2	Cash flow of a coupon bond	448
Fig. 22.3	Cash flow of a coupon bond	449
Fig. 22.4	Computation of the yield to maturity when the present value is given	453
Fig. 22.5	Delta hedging for interest rate derivatives V_1 and V_2 presented in Remark 22.5 using a self-financing portfolio $\Pi(r, t) = V_1(r, t) - \Delta V_2(r, t) + C$, $0 \leq t \leq 3$ under an interest rate model from Sect. 23.2 (<i>from top left to bottom right</i>)	462
Fig. 22.6	Monotone convergence of $(1 - 1/n)^n$ to the limit $1/e = 0.367879 \dots$ as $n \rightarrow \infty$	467

Fig. 22.7	Probability distributions of $\int_0^T r_t dt$ and $\exp(-\int_0^T r_t dt)$ in the Vasicek model	473
Fig. 22.8	Probability distributions of $\int_0^T r_t dt$ and $\exp(-\int_0^T r_t dt)$ in the CIR model	474
Fig. 23.1	Sample paths of the interest rate in the Vasicek model	478
Fig. 23.2	$B(\tau)/\tau$ as a monotonically decreasing function of τ in the Vasicek model	480
Fig. 23.3	Term structure of interest rates in the Vasicek model	480
Fig. 23.4	Bond price by the Vasicek model	481
Fig. 23.5	Sample paths of the interest rate in the CIR model	482
Fig. 23.6	Probability distribution of the interest rate r_t , $t = 1, 2, 5, 50$, in the CIR model: $r_0 = 0.01$ (left) and $r_0 = 0.3$ (right)	483
Fig. 23.7	Term structure of interest rates in the CIR model	484
Fig. 23.8	Bond price by the CIR model	484
Fig. 23.9	Volatility decreases as time to maturity τ increases in the Hull–White model	487
Fig. 23.10	Limit of the pdf of the interest rate r_t in the CIR model with $2\alpha\bar{r} = \sigma^2$	490
Fig. 24.1	The bank deposit and a risky asset price	494
Fig. 24.2	The discounted bank deposit and a discounted risky asset price when the risky asset is a numeraire	494
Fig. 25.1	Graphs of trigonometric functions	515
Fig. 25.2	Plot of $Z = XY$ over a mesh grid	515
Fig. 25.3	Histogram for the standard normal distribution	519
Fig. 25.4	Scatter plot of two correlated sets of data	521
Fig. 25.5	Sample paths of random walks	531
Fig. 25.6	Interest rates by the Vasicek model	541
Fig. 25.7	Interest rates by the Vasicek model	541
Fig. A.1	Polynomial fitting of data	550
Fig. C.1	The line of best fit and the least squares method	567
Fig. F.1	Probability density functions of $\chi^2(n)$ for $1 \leq n \leq 6$	586
Fig. F.2	Probability density functions of noncentral $\chi^2(n)$ for $n = 2, 5$ and $\lambda = 1, 2, 3$	587
Fig. F.3	The product of two independent random variables X and Y	597
Fig. F.4	A partition of the plane where the subsets $A_1, \dots, A_8$ have the same probability $\frac{1}{8}$ with respect to the two-dimensional normal density due to the rotational symmetry	604
Fig. F.5	A single-period binomial tree and the risk-neutral probability	611
Fig. F.6	The sublinearity of the payoff function $f(x) = (x - K)^+$ of a European call	612

List of Tables

Table 2.1	Comparison of forward and futures	18
Table 2.2	Comparison of two portfolios	23
Table 4.1	Comparison of terminology and notation	44
Table 7.1	Limiting behaviors of geometric Brownian motion depending on the signs of μ and $\mu - \frac{1}{2}\sigma^2$	122
Table 10.1	Multiplication rule for Itô calculus	174
Table 10.2	Multiplication rule for general Itô calculus	188
Table 14.1	A spreadsheet for a multiperiod binomial tree	256
Table 14.2	Option pricing by risk-neutral method	258
Table 14.3	The spreadsheet corresponding to the binomial tree for the underlying asset	268
Table 14.4	The spreadsheet corresponding to the binomial tree for the European call option price	268
Table 15.1	The Greek alphabet	299
Table 18.1	Definitions of average for Asian options	366
Table 19.1	Viewpoint of the holder of an American call option	382
Table 22.1	Market data for bonds	447
Table 22.2	Continuously compounded zero rates determined by Table 22.1	449
Table 22.3	Continuously compounded zero rates determined by Table 22.1	449
Table 22.4	Present interest rates	450
Table 22.5	New interest rates	450
Table 22.6	Calculation of forward rates	451
Table 22.7	Calculation of duration: a bond with coupons	455
Table 22.8	Calculation of duration	455
Table 24.1	Equivalent measures, Brownian motions and martingales	500
Table C.1	Comparison of Lebesgue integral on a finite space and linear algebra	566
Table F.1	Bayes' rule and probabilities	599

List of Simulations

Simulation 4.1	Uniform Distribution	68
Simulation 4.2	Equally Spaced Points in the Unit Interval	69
Simulation 4.3	Quantile-Quantile Plot	70
Simulation 4.4	The Central Limit Theorem	71
Simulation 4.5	Correlated Normal Variables	71
Simulation 5.1	Monty Hall Problem	88
Simulation 5.2	Conditional Expectation	90
Simulation 6.1	Estimation of Exponential	105
Simulation 6.2	Estimation of Exponential	105
Simulation 7.1	Sample Paths of Brownian Motion	127
Simulation 7.2	Average of Geometric Brownian Motion	128
Simulation 7.3	Brownian Motion with Boundary Condition	129
Simulation 7.4	Law of Large Numbers for Brownian Motion	130
Simulation 7.5	Random Walk to the Horizon	130
Simulation 8.1	The First Hitting Time	145
Simulation 8.2	The Maximum of Brownian Motion	145
Simulation 9.1	Itô Integral	166
Simulation 9.2	Itô Integral	166
Simulation 9.3	Convergence	167
Simulation 9.4	Normal Distribution of Itô Integral	167
Simulation 9.5	Expectation of Itô Integral is Zero	169
Simulation 10.1	Itô Formula	193
Simulation 10.2	Geometric Brownian Motion	194
Simulation 11.1	Brownian Motion with Drift	207
Simulation 11.2	Moments of the Brownian Motion with Drift	208
Simulation 11.3	The Martingale Property of the Radon–Nikodym Derivative	209
Simulation 12.1	Numerical Solution of an SDE	227
Simulation 12.2	Ornstein–Uhlenbeck Process	228
Simulation 12.3	Brownian Bridge	229
Simulation 12.4	Brownian Bridge Given by an Integral	229

Simulation 12.5	Brownian Bridge Given by an SDE	230
Simulation 13.1	Feynman–Kac Theorem	244
Simulation 14.1	The Binomial Tree Method for a European Call Price	268
Simulation 14.2	The Binomial Trees for the Asset, the European Put Price and the Delta	269
Simulation 14.3	Delta-hedging in the Binomial Tree Method	272
Simulation 14.4	Convergence of the Binomial Tree Method	275
Simulation 14.5	The Binomial Tree Method in Agreement with the Black–Scholes–Merton Partial Differential Equation	276
Simulation 14.6	American Call Option: the Binomial Tree Method	278
Simulation 15.1	Delta Hedging	312
Simulation 15.2	Option Price Surface	315
Simulation 15.3	The Greeks by the Monte Carlo Method	316
Simulation 16.1	Martingale Method	331
Simulation 17.1	Cash-or-Nothing Call	358
Simulation 17.2	Graph of the Delta of the At-the-Money Option	359
Simulation 17.3	Brownian Bridge and At-the-Money Option	360
Simulation 17.4	Convergence of the Price by the Binomial Tree Method to the Black–Scholes–Merton Formula	363
Simulation 18.1	Geometric Average Asian Option	377
Simulation 18.2	Stopped Asset Price Process	378
Simulation 18.3	Down-and-Out Put	379
Simulation 19.1	The Binomial Trees for the Asset and the American Put Price	387
Simulation 19.2	American Put Option: the Binomial Tree Method	391
Simulation 19.3	American Put Option: the Early Exercise Domain	392
Simulation 19.4	Perpetual American Put Option	394
Simulation 20.1	Portfolios of Two Assets	422
Simulation 20.2	Portfolios of Three Assets	423
Simulation 20.3	Capital Market Line	424
Simulation 21.1	Power Utility	439
Simulation 21.2	Log Utility	440
Simulation 22.1	Monotone Convergence to $e = 2.71828$	466
Simulation 22.2	Yield to Maturity	467
Simulation 22.3	The Fundamental Equation for Pricing	468
Simulation 22.4	Discounted Feynman–Kac Theorem	471
Simulation 22.5	Distribution of the Discount Factor	472
Simulation 23.1	Vasicek Model	488
Simulation 23.2	Cox–Ingersoll–Ross Model	488
Simulation 23.3	CIR Model: Probability Distribution	489
Simulation F.1	χ^2 -Distribution	587
Simulation F.2	Noncentral χ^2 -Distribution	588

Part I
Introduction to Financial Mathematics

Chapter 1

Fundamental Concepts



A stock exchange is an organization of brokers and financial companies which has the purpose of providing the facilities for trade of company stocks and other financial instruments. Trade on an exchange is by members only. In Europe, stock exchanges are often called bourses. The trading of stocks on stock exchanges, physical or electronic, is called the stock market.

A *portfolio* is a collection of financial assets such as government bonds, stocks, commodities and financial derivatives. Mathematically speaking, an asset is a time-indexed sequence of random variables, and a portfolio is a linear combination of assets. For example, consider a bank deposit with value 1 at time $t = 0$ and B_t at time t , and a stock with price S_t at time t per share. Then a portfolio consisting of a units of bond and b shares of stock has value $aB_t + bS_t$ at time t if there is no trading before or at t . The coefficients a and b can be negative, which means borrowing bonds or stocks from someone else. A *contingent claim* is a claim that can be made depending on whether one or more specified outcomes occur.

In this chapter we introduce three fundamental concepts in financial mathematics: risk, time value of money, and the no arbitrage principle. For more information consult Capinski and Zastawniak (2003), Hoek and Elliott (2006), Shreve (2004a).

1.1 Risk

Risk is the possibility of exposure to uncertain losses in the future. It is different from danger. Investment in stock is said to be risky, not dangerous. Investors prefer less risky investment if all other conditions are identical. The basic principle of investment is that the payoff should be large if the risk level is high. To reduce risk in investment, it is better to diversify: Don't put all your eggs in one basket, as the proverb says. For that purpose, we usually construct a portfolio consisting of assets which are not strongly correlated or even negatively correlated.

There are several different types of risk. First, credit risk is the risk of not receiving back a loan from the borrower. Second, market risk is the risk caused by unexpected asset price movement. Finally, operational risk means the risks from malfunctioning of a computer system, or fraudulent activity of employees, etc.

Risk management refers to management activities to protect asset values from the risks in investment. Regulating agencies are concerned with risk management of the whole financial system as well as individual financial companies. If financial derivatives are used as tools for excessive speculation, financial crisis can occur.

When we completely eliminate the risk in an investment, we call it *hedging*. Consult Bernstein (1996) for the viewpoint that probability is a tool for understanding risks in investment. For general risks caused by natural catastrophes consult Woo (1999).

1.2 Time Value of Money

A dollar today is usually worth more than a dollar in the future due to the opportunity to lend it and receive interest or to invest in a business and make profit. When there is inflation, a dollar in the future has less purchasing power than a dollar today. Furthermore, there is credit risk: receiving money in the future rather than now has uncertainty in recovering the money. Time value of money is expressed in terms of interest rate. For example, \$100 today, invested for one year at a 3% interest rate, will be \$103 one year from now. Interest rates are usually positive, however, zero or negative interest rates are possible, especially in a time of deflation or economic recession. As an economic policy, the central bank in a country may lower its interest rate to stimulate the national economy. Throughout the book the risk-free interest rate is assumed to be positive unless stated otherwise.

1.3 No Arbitrage Principle

An opportunity to make a profit without risking any future loss is called an *arbitrage*, which exists as a result of market inefficiencies. An *arbitrage trade* means a transaction generating risk-free profit by buying and selling related assets with no net investment of cash. For example, if the price of a product is \$10 in New York, but in Los Angeles, the same product is selling for \$15, and if someone buys the merchandise in New York and sells it in Los Angeles, he/she can profit from the difference without any risk. When expenses such as transportation cost and taxes are considered there might be no overall profit to the arbitrager, and no arbitrage exists in this case. If the market functions perfectly, there would be no arbitrage opportunities. All the participants in the financial market look for any possibility of arbitrage and make riskless profit if there exists one. Arbitrage chances only last for a very

short period of time. Once an arbitrage opportunity is exploited, there would be no remaining arbitrage opportunity for others.

The concept of arbitrage may be defined in a formal and precise way in the following simple model of the financial market in which we consider only two time points $t = 0$ and $t = T > 0$. The values of a portfolio V at $t = 0$ and $t = T$ are denoted by V_0 and V_T , respectively. At $t = 0$, which represents the present time, all asset values and portfolio values are known and there is no uncertainty. The asset prices other than the risk-free bank deposit at time T are unknown and random, and hence the value V_T of a portfolio containing a risky asset is a random variable.

Definition 1.1 (*Arbitrage*) We say that there exists an *arbitrage* if there is a portfolio V satisfying one of the following conditions:

- (i) $V_0 = 0$, and $V_T \geq 0$ with probability one and $V_T > 0$ with positive probability,
or
- (ii) $V_0 < 0$, and $V_T \geq 0$ with probability one.

See also Definition 1.2. In more general cases, we assume that there exists a generally accepted concept of arbitrage even when it is not stated explicitly. The *no arbitrage principle* means that one cannot create positive value out of nothing, i.e., there is no free lunch.

1.4 Arbitrage-Free Market

Consider a one-period model of a financial market which is observed at times $t = 0$ and $t = T$. Assume that there are N assets with prices $S_t^1, \dots, S_t^N$ at $t = 0, T$. All the prices of N assets are already known at $t = 0$, and $S_T^1, \dots, S_T^N$ are random. Uncertainty in the financial market at T is modeled using a finite number of possible future states. Suppose that there are M states at time T , which is represented by a sample space $\Omega = \{\omega_1, \dots, \omega_M\}$. The probability that the future state is ω_j is equal to $p_j > 0$, $1 \leq j \leq M$, $p_1 + \dots + p_M = 1$, and S_T^i takes one of the values $S_T^i(\omega_1), \dots, S_T^i(\omega_M)$ depending on the future state. Let

$$\mathbf{S}_t = \begin{bmatrix} S_t^1 \\ \vdots \\ S_t^N \end{bmatrix}$$

and define the *matrix of securities* by

$$D = \begin{bmatrix} S_T^1(\omega_1) & \cdots & S_T^1(\omega_M) \\ \vdots & & \vdots \\ S_T^N(\omega_1) & \cdots & S_T^N(\omega_M) \end{bmatrix}.$$

An interpretation of D is given by the diagram

$$\begin{array}{ccc} & j\text{th state} & \\ & \downarrow & \\ i\text{th asset} \rightarrow & \left[\begin{array}{c} D_{ij} \end{array} \right] & \end{array}$$

i.e., the entry D_{ij} represents the value (including dividend payments) of the i th asset in the j th state at time T . Let $\mathbf{v} \cdot \mathbf{w}$ denote the scalar product of $\mathbf{v}$, $\mathbf{w}$. A *portfolio* V_t^π of securities is defined by $\pi \cdot \mathbf{S}_t$, where

$$\pi = \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_N \end{bmatrix} \in \mathbb{R}^N$$

and π_i is the number of units of the i th asset held in the portfolio. It is assumed to be constant over the time interval $[0, T]$. If $\pi_i > 0$, then the investor has a *long position* in the i th asset and will receive cash flow $\pi_i D_{ij}$ at time T depending on the outcome. If $\pi_i < 0$ for some i , then we have a *short position* in the i th asset, i.e., we borrow π_i units of S^i at $t = 0$ and have the obligation to return it at time T . We assume that any investor can take short and long positions in arbitrary amounts of securities. That is, even though in the real financial market the values π_i are integers, we assume that π_i are real numbers in our mathematical model. With abuse of language, π is also called a portfolio since there is a one-to-one correspondence between the collections of V_t^π and π . The total value of the portfolio defined by π is equal to

$$V_t^\pi = \sum_{i=1}^N \pi_i S_t^i = \pi \cdot \mathbf{S}_t$$

for $t = 0, T$. In particular, for $t = T$, the cash flow of the portfolio in the j th state is

$$V_T^\pi(\omega_j) = \sum_{i=1}^N \pi_i S_T^i(\omega_j) = \sum_{i=1}^N \pi_i D_{ij}.$$

Now the definition of arbitrage given in Definition 1.1 can be restated as follows:

Definition 1.2 (*Arbitrage Portfolio*) Let $\mathbf{d}_1, \dots, \mathbf{d}_M$ denote the columns of D . An *arbitrage portfolio* is a portfolio π satisfying one of the following conditions:

- (i) $\pi \cdot \mathbf{S}_0 = 0$ (zero cost), and $\pi \cdot \mathbf{d}_j \geq 0$ for all j (no loss) and $\pi \cdot \mathbf{d}_k > 0$ for some k (possibility of profit),
or
- (ii) $\pi \cdot \mathbf{S}_0 < 0$, and $\pi \cdot \mathbf{d}_j \geq 0$ for all j .

If there does not exist any arbitrage portfolio, we say that the market is *arbitrage-free*.

The second case can be reduced to the first one if an investor can buy risk-free bonds and lock in a profit in the following way: Buy risk-free bonds using the borrowed money $-\pi \cdot \mathbf{S}_0 > 0$ and receive $e^{rT}(-\pi \cdot \mathbf{S}_0)$ at time T , making a risk-free profit of

$$e^{rT}(-\pi \cdot \mathbf{S}_0) - (-\pi \cdot \mathbf{S}_0) > 0 ,$$

where $r > 0$ is the risk-free interest rate. An efficient market does not allow arbitrage opportunities to last for long.

Definition 1.3 (*Arrow–Debreu Security*) If a security θ_i^i , $1 \leq i \leq M$, satisfies the condition that $\theta_T^i(\omega_i) = 1$ and $\theta_T^i(\omega_j) = 0$ for $j \neq i$, then it is called an *Arrow–Debreu security*.

A contingent claim at time T can be expressed as a linear combination of the Arrow–Debreu securities if they exist. A portfolio $\pi = (1, \dots, 1)$ consisting of one unit of each Arrow–Debreu security is a risk-free asset that pays 1 in any future state. It may be regarded as a bond. The difference of prices of π at times 0 and T is regarded as interest payment.

Definition 1.4 (*State Price Vector*) A column vector $\psi = (\psi_1, \dots, \psi_M)^t$ is called a *state price vector* if $\psi_i > 0$ for every i , $1 \leq i \leq M$, and

$$\mathbf{S}_0 = D \psi . \quad (1.1)$$

Equivalently, there exist positive constants $\psi_1, \dots, \psi_M$ such that

$$S_0^i = \psi_1 S_T^i(\omega_1) + \dots + \psi_M S_T^i(\omega_M) \quad (1.2)$$

for every $1 \leq i \leq N$, or

$$\mathbf{S}_0 = \psi_1 \mathbf{d}_1 + \dots + \psi_M \mathbf{d}_M . \quad (1.3)$$

Remark 1.1 Note that (1.2) implies that the asset price today is equal to a weighted average of its future values, with weights given by the state price vector ψ . The weights ψ_i , $1 \leq i \leq M$, are called the state prices, and ψ_i is the value of one unit of the Arrow–Debreu security θ^i at time 0, i.e., $\psi_i = \theta_0^i$, $1 \leq i \leq M$.

Throughout the rest of the section we use the notation $(x_1, \dots, x_n) \geq 0$ to mean that $x_i \geq 0$ for every i , $1 \leq i \leq n$. Similarly, $(x_1, \dots, x_n) > 0$ means that $x_i > 0$ for every i , $1 \leq i \leq n$. The following fact implies that asset prices and cash-flows satisfy certain relations in a no-arbitrage financial world.

Theorem 1.1 *A market is arbitrage-free if and only if there exists a state price vector.*

Proof ($\Leftarrow$) Suppose that there exists an arbitrage portfolio π . Then either it satisfies (i) or (ii) in Definition 1.2. In the first case the condition that $\pi \cdot \mathbf{d}_j \geq 0$ for all j is equivalent to the condition that $\pi^t D \geq 0$ with at least one of its entries being positive. Since $\psi > 0$, we have

$$0 = \pi \cdot \mathbf{S}_0 = \pi \cdot (D \psi) = \pi^t (D \psi) = (\pi^t D) \psi > 0 ,$$

and thus we have a contradiction. Now we consider the second case. The condition that $\pi \cdot \mathbf{d}_j \geq 0$ for all j is equivalent to the condition that $\pi^t D \geq 0$. Since $\psi > 0$, we have

$$0 > \pi \cdot \mathbf{S}_0 = \pi \cdot (D \psi) = \pi^t (D \psi) = (\pi^t D) \psi \geq 0 ,$$

and thus we have a contradiction as well.

($\Rightarrow$) Let

$$\mathbb{R}_+^{M+1} = \{\mathbf{x} = (x_0, x_1, \dots, x_n) \in \mathbb{R}^{M+1} : \mathbf{x} \geq 0\} ,$$

and let

$$W = \{(-\pi \cdot \mathbf{S}_0, \pi \cdot \mathbf{d}_1, \dots, \pi \cdot \mathbf{d}_M) \in \mathbb{R}^{M+1} : \pi \in \mathbb{R}^N\} .$$

Let $\tilde{D}$ be an $N \times (M+1)$ matrix with its columns given by $-\mathbf{S}_0, \mathbf{d}_1, \dots, \mathbf{d}_M$. Then $W = \{\pi^t \tilde{D} : \pi \in \mathbb{R}^N\}$. Since an arbitrage portfolio does not exist by the assumption, the subspace W and the cone $\mathbb{R}_+^{M+1}$ intersect only at the origin $(0, \dots, 0)$. For, if there exists (a portfolio) $\pi \in \mathbb{R}^N$ such that $\pi^t \tilde{D} \geq 0$ and $\pi^t \tilde{D} \neq \mathbf{0}$, then either (i) $\pi \cdot \mathbf{S}_0 \leq 0, \pi \cdot \mathbf{d}_j \geq 0$ for every j , and there exists $1 \leq j \leq M$ such that $\pi \cdot \mathbf{d}_j > 0$, or (ii) $\pi \cdot \mathbf{S}_0 < 0, \pi \cdot \mathbf{d}_j = 0$ for every j , both of which are impossible.

Hence there exists $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_M) \in \mathbb{R}^{M+1}$ and a hyperplane

$$H = \{\mathbf{x} \in \mathbb{R}^{M+1} : \lambda \cdot \mathbf{x} = 0\}$$

that separates $\mathbb{R}_+^{M+1} \setminus \{\mathbf{0}\}$ and W in such a way that $W \subset H$ and $\lambda \cdot \mathbf{x} > 0$ for every $\mathbf{x} \in \mathbb{R}_+^{M+1} \setminus \{\mathbf{0}\}$, which holds if and only if $\lambda_j > 0$ for every j .

Since W is contained in H , for every $\pi \in \mathbb{R}^N$ we have

$$-\lambda_0 \pi \cdot \mathbf{S}_0 + \sum_{j=1}^M \lambda_j \pi \cdot \mathbf{d}_j = 0 ,$$

which implies that

$$-\lambda_0 \mathbf{S}_0 + \sum_{j=1}^M \lambda_j \mathbf{d}_j = \mathbf{0} ,$$

or

$$\mathbf{S}_0 = \sum_{j=1}^M \frac{\lambda_j}{\lambda_0} \mathbf{d}_j = \sum_{j=1}^M \psi_j \mathbf{d}_j ,$$

where $\psi_j = \lambda_j / \lambda_0 > 0$, $1 \leq j \leq M$. Thus (1.3) holds. $\square$

For more details of the proof, consult Avellaneda and Laurence (1999), Etheridge (2002). For an equivalent version, see Corollary 1.1.

Definition 1.5 (*Complete Market*) If a contingent claim at time T is expressed as a linear combination of the assets traded in the market, we say that the claim is *replicated*, or *hedged*. If any contingent claim at T can be replicated, then we say that the market is *complete*.

Theorem 1.2 *The market is complete if and only if the rank of the matrix D of securities is equal to M .*

Proof ($\Rightarrow$) Let θ_i^j , $1 \leq i \leq M$, be the Arrow–Debreu securities. Since the market is complete, there exist a_{ij} , $1 \leq i \leq N$, $1 \leq j \leq M$, such that

$$\begin{cases} \theta_T^1 = a_{11}S_T^1 + \cdots + a_{N1}S_T^N \\ \vdots \\ \theta_T^M = a_{1M}S_T^1 + \cdots + a_{NM}S_T^N. \end{cases}$$

Let $A = [a_{ij}]_{1 \leq i \leq N, 1 \leq j \leq M}$. Thus, if the future is in state ω_j , then

$$\begin{bmatrix} \theta_T^1(\omega_j) \\ \vdots \\ \theta_T^M(\omega_j) \end{bmatrix} = A^t \begin{bmatrix} S_T^1(\omega_j) \\ \vdots \\ S_T^N(\omega_j) \end{bmatrix} ,$$

and hence

$$\begin{bmatrix} \theta_T^1(\omega_1) & \cdots & \theta_T^1(\omega_M) \\ \vdots & & \vdots \\ \theta_T^M(\omega_1) & \cdots & \theta_T^M(\omega_M) \end{bmatrix} = A^t \begin{bmatrix} S_T^1(\omega_1) & \cdots & S_T^1(\omega_M) \\ \vdots & & \vdots \\ S_T^N(\omega_1) & \cdots & S_T^N(\omega_M) \end{bmatrix} .$$

Since $\theta_T^i(\omega_j) = 0$ or 1 depending on whether $i = j$ or not, we have

$$I = A^t D , \tag{1.4}$$

where I denotes the $M \times M$ identity matrix. Hence

$$M = \text{rank}(I) \leq \min\{\text{rank}(A^t), \text{rank}(D)\} \leq \min\{N, M\} ,$$

and thus $M \leq N$. Since

$$\text{rank}(D) \geq \text{rank}(I) = M$$

from (1.4), and since D is an $N \times M$ matrix, we conclude that $\text{rank}(D) = M$.

($\Leftarrow$) To prove that the market is complete, it suffices to find a matrix $A = [a_{ij}]$ satisfying (1.4). Taking the transposes of (1.4), we have $I = D^t A$. $\square$

Note that if the market is complete then $N \geq M$. For more information, consult Avellaneda and Laurence (1999), Baz and Chacko (2004), Bjork (2009), Cvitanic and Zapatero (2004), Etheridge (2002), Gianin and Sgarra (2013), Tretyakov (2013). For continuous time market models the conditions for no-arbitrage and completeness are essentially the same, even though the necessary mathematical techniques are sophisticated. Interested readers are referred to Bingham and Kiesel (2004).

1.5 Risk-Neutral Pricing and Martingale Measures

Theorem 1.3 (Discount Rate) *Let $\psi = (\psi_1, \dots, \psi_M)$ be a state price vector. Assume that there exists a risk-free security, i.e., a risk-free bond, S^0 such that*

$$S_0^0 = 1 \quad \text{and} \quad S_T^0 = 1 + r$$

regardless of future state, where $r > 0$ is a constant that is regarded as the risk-free interest rate. Then

$$\sum_{j=1}^M \psi_j = \frac{1}{1+r}.$$

Proof Since

$$S_0^0 = \psi_1 S_T^0(\omega_1) + \dots + \psi_M S_T^0(\omega_M),$$

we have

$$1 = \psi_1(1+r) + \dots + \psi_M(1+r),$$

and the proof is complete. $\square$

Definition 1.6 (Risk-Neutral Probability or Martingale Measure) *Let*

$$p_j^* = \frac{\psi_j}{\sum_{k=1}^M \psi_k} = (1+r)\psi_j > 0, \quad 1 \leq j \leq M.$$

Then the positive constants $p_1^*, \dots, p_M^*$ add up to 1, and define a probability on the future states $\omega_1, \dots, \omega_M$, which is called a *risk-neutral probability* or a *martingale measure*. It is not a real probability, but a formal probability representing the coefficients of a convex linear combination of given values.

We can rewrite (1.2) as

$$S_0^i = \sum_{k=1}^M p_k^* \frac{S_T^i(\omega_k)}{1+r} = \frac{1}{1+r} \mathbb{E}^*[S_T^i], \quad (1.5)$$

where $\mathbb{E}^*$ denotes the expectation under the martingale measure $(p_1^*, \dots, p_M^*)$. S_0^i is the discounted asset price. The idea given in (1.5) is called the *risk-neutral pricing* of financial claims, and is the central theme for the rest of the book.

Remark 1.2 The state prices ψ_i , $1 \leq i \leq M$, are the prices at time $t = 0$ of the Arrow–Debreu securities θ^j , $1 \leq j \leq M$. To see why, note that

$$\theta_0^j = \mathbb{E}^* \left[\frac{1}{1+r} \theta_T^j \right] = \frac{1}{1+r} \sum_{k=1}^M p_k^* \theta_T^j(\omega_k) = \frac{1}{1+r} p_j^* = \psi_j.$$

This is why ψ is called the state price vector.

Corollary 1.1 *The following statements are equivalent:*

- (i) *There exists a state price vector.*
- (ii) *There exists a martingale measure.*
- (iii) *The market is arbitrage-free.*

Proof Use Theorem 1.1. □

Theorem 1.4 (Uniqueness of Martingale Measure) *Suppose that the market is arbitrage-free. Then the market is complete if and only if there exists a unique martingale measure.*

Proof ($\Rightarrow$) Recall that $M \leq N$ and $\text{rank}(D) = M$. Since D is an $N \times M$ matrix, and since $\text{rank}(D) + \text{nullity}(D) = M$, we have $\text{nullity}(D) = 0$. Suppose that there exist two distinct state price vectors ψ and ϕ . Since they are solutions of (1.1), we have $D(\psi - \phi) = \mathbf{0}$, and hence $D\mathbf{x} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$, which is a contradiction.

($\Leftarrow$) Let $\psi > 0$ be the unique price. If the market were not complete, then $\text{rank}(D) < M$, and hence there exists $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^M$ such that $D\mathbf{v} = \mathbf{0}$. Hence $D(\psi + \alpha\mathbf{v}) = D\psi + \alpha D\mathbf{v} = S_0$ and $\psi + \alpha\mathbf{v} > 0$ for $|\alpha| < \delta$ and $\delta > 0$ sufficiently small. If we choose $\alpha \neq 0$, then $\psi + \alpha\mathbf{v}$ is another state price vector, which contradicts the assumption. □

1.6 The One Period Binomial Tree Model

Consider the case when $N = 2$ and $D = 2$. There are two future states: up state ω_u and down state ω_d . Let C be any contingent claim with cash flows C^u in state ω_u and C^d in state ω_d at time T . (See Fig. 1.1.) If there are no arbitrage opportunities, then there exists a martingale measure (p_u^*, p_d^*) such that the price C_0 at time 0 is given by

$$C_0 = \frac{1}{1+r} \mathbb{E}^*[C_T] = \frac{1}{1+r} (p_u^* C^u + p_d^* C^d) . \quad (1.6)$$

To see why, note that C can be replicated by a linear combination of two given assets, each of which has price given by (1.5). (See Exercise 1.3(i), which shows that the market is complete.)

We assume that the market consists of a risky asset S and a risk-free asset B such as a bond. Furthermore, S_T has values uS_0 and dS_0 , where $0 < d < u$ depending on whether the future state is up or down. (See Fig. 1.2.)

Then, by (1.5) or (1.6) we have

$$S_0 = \frac{1}{1+r} \mathbb{E}^*[S_T] = \frac{1}{1+r} (p_u^* uS_0 + p_d^* dS_0) \quad (1.7)$$

and

$$B_0 = \frac{1}{1+r} \mathbb{E}^*[C_T] = \frac{1}{1+r} (p_u^* (1+r)B_0 + p_d^* (1+r)B_0) = B_0 . \quad (1.8)$$

We obtain no new information from (1.8). From (1.7), we have

$$p_u^* u + p_d^* d = 1 + r . \quad (1.9)$$

Fig. 1.1 The one period binomial tree for C and the risk-neutral probabilities

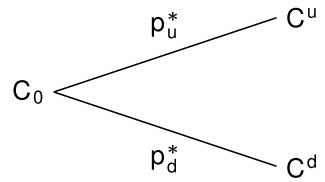
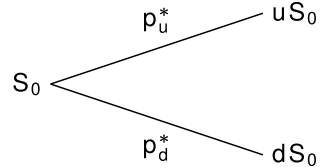


Fig. 1.2 The one period binomial tree for S and the risk-neutral probabilities



Recall that

$$p_u^* + p_d^* = 1 . \quad (1.10)$$

From (1.9) and (1.10), we finally obtain

$$p_u^* = \frac{1 + r - d}{u - d} \quad \text{and} \quad p_d^* = \frac{u - (1 + r)}{u - d} .$$

Thus $p_u^* > 0$ and $p_d^* > 0$ if and only if $d < 1 + r < u$. Note that $p_u^* > 0$ and $p_d^* > 0$ is the condition for the existence of a martingale measure. Also note that the inequality $d < 1 + r < u$ is the no arbitrage condition since if $d < u \leq 1 + r$, for example, then every investor would buy the risk-free asset B making a risk-free profit that is at least the maximum possible profit from investment in the risky asset S . For more information, consult Bjork (2009), Pliska (1997). The one period binomial tree model is extended to the multiperiod binomial tree model in Chap. 14. For continuous time models, consult Bingham and Kiesel (2004), Klebaner (2005).

1.7 Models in Finance

Why do we need new financial models in addition to existing ones? A model reflects only some aspects of reality. Information obtained from a model is a shadow on the cave wall. The real financial market represented by a model can change over time, and models should be updated constantly. Technological advances allow better models, and increased computing power enables us to develop more computationally intensive models. New models are needed constantly, and are invented for new problems and techniques.

The more the better. A model is an approximation, and parameters in the model are not precisely known. Some models are practically oriented, while others give insights. A model should be stable under small perturbations. A slightly modified model of a classical model (e.g. the Black–Scholes–Merton differential equation) should produce a result that is close to a conventional solution. Using a modification of an existing classical model, we can test the existing model. Different models provide different viewpoints. A practitioner chooses a useful and convenient model.

Exercises

1.1 (Farkas' Lemma) We are given $M + 1$ vectors $\mathbf{d}_0, \mathbf{d}_1, \dots, \mathbf{d}_M$ in $\mathbb{R}^N$. Then exactly one of following two statements is true:

- (i) There exist nonnegative constants $\lambda_1, \dots, \lambda_M$ such that

$$\mathbf{d}_0 = \lambda_1 \mathbf{d}_1 + \dots + \lambda_M \mathbf{d}_M .$$

- (ii) There exists a $\boldsymbol{\pi} \in \mathbb{R}^N$ such that

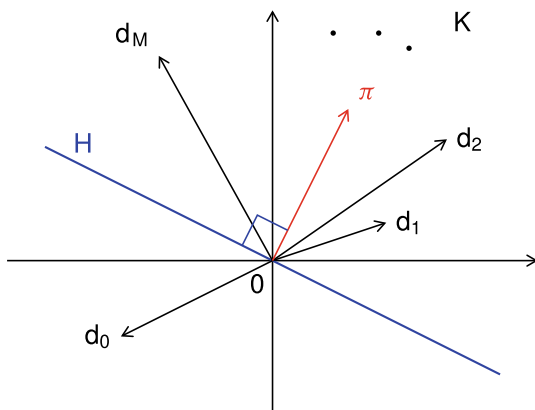
$$\boldsymbol{\pi} \cdot \mathbf{d}_0 < 0 \quad \text{and} \quad \boldsymbol{\pi} \cdot \mathbf{d}_j \geq 0$$

for every $1 \leq j \leq M$.

(Hint: Recall that if $\sum_{i=1}^n \lambda_i = 1$, $\lambda_i \geq 0$, then $\sum_{i=1}^n \lambda_i \mathbf{x}_i$, $\mathbf{x}_i \in \mathbb{R}^N$, is called a convex linear combination of $\mathbf{x}_i$. Let K be the closed cone given by all convex linear combinations of $\mathbf{d}_1, \dots, \mathbf{d}_M$ with nonnegative coefficients. Then either $\mathbf{d}_0 \in K$ or $\mathbf{d}_0 \notin K$. If $\mathbf{d}_0 \in K$, then (i) has a solution. If $\mathbf{d}_0 \notin K$, then there exists a hyperplane H in $\mathbb{R}^N$ that separates $\mathbf{d}_0$ and K . Choose a normal vector $\boldsymbol{\pi} \neq \mathbf{0}$ to H and pointing in the direction where K is located. See Fig. 1.3.)

1.2 Suppose that the market is arbitrage-free as given in Definition 1.2. Show that $\mathbf{S}_0 = \lambda_1 \mathbf{d}_1 + \dots + \lambda_M \mathbf{d}_M$ for some $\lambda_1, \dots, \lambda_M \geq 0$.

Fig. 1.3 The closed convex cone K in the proof of Farkas' lemma



1.3 Consider the binomial tree model in Sect. 1.6.

- (i) Find a condition under which there exists a unique solution (a, b) such that

$$\begin{cases} C^u = aS_T^1(\omega_u) + bS_T^2(\omega_u) \\ C^d = aS_T^1(\omega_d) + bS_T^2(\omega_d) \end{cases} .$$

- (ii) Let B denote the risk-free bond. For $d < u$, show that there exists a unique solution (a, b) such that

$$\begin{cases} C^u = aS_T(\omega_u) + bB_T(\omega_u) \\ C^d = aS_T(\omega_d) + bB_T(\omega_d) \end{cases} .$$

Chapter 2

Financial Derivatives



Financial assets are divided into two categories: The first group consists of primary assets including shares of a stock company, bonds issued by companies or governments, foreign currencies, and commodities such as crude oil, metal and agricultural products. The second group consists of financial contracts that promise some future payment of cash or future delivery of the primary assets contingent on an event in the future date. The event specified in the financial contract is usually defined in terms of the behavior of an asset belonging to the first category, and such an asset is called an *underlying asset* or simply an *underlying*. The financial assets belonging to the second category are called derivatives since their values are derived from the values of the underlying asset belonging to the first category. *Securities* are tradable financial instruments such as stocks and bonds.

How much should one pay for a financial derivative? The payoff of a financial derivative depends upon the underlying asset, and that is why it is called a derivative. The buyer of a financial contract is said to hold the *long* position, and the seller the *short* position.

Short selling means selling securities one does not own. A trader borrows the securities from someone else and sells them in the market. At some stage one must buy back the securities so they can be returned to the lender. The short-seller must pay dividends and other benefits the owner of the securities would normally receive. For more information see Kumar (2015). In this chapter we introduce financial derivatives and their fundamental properties such as the put-call parity. Also presented is a flowchart of ideas in option pricing. For a general introduction to finance, consult Hull (2011), Tretyakov (2013).

2.1 Forward Contracts and Futures

A *forward* contract is an agreement to buy or sell a fixed quantity of a specific asset on a specified future date, called the maturity date, for the forward price specified at the time the contract is initiated. There is no initial payment at the time the contract is signed, and there is no daily settlement during the lifetime of the contract except at the end of the life of the contract. At the maturity date, one party buys the asset for the agreed price from the other party.

A *futures* contract is the same as a forward contract except that futures are traded on exchanges and the exchange specifies certain standard features of the contract and a particular form of settlement. No money changes hands initially. However, futures contracts are settled daily, and the difference between the price of the previous day and the present day is calculated and added/subtracted to/from the account by the end of the trading day. Futures are available on a wide range of underlyings, and they are traded on an exchange. We need to specify what can be delivered, where it can be delivered, and when it can be delivered. A *margin* is cash or marketable securities deposited as collateral by an investor with his or her broker. The balance in the margin account is adjusted to reflect daily settlement. Margins minimize the risk of a default on a contract either by a buyer when the price falls or a seller when the price rises. Closing out a futures position involves entering into an offsetting trade and most futures contracts are closed out before maturity. If a futures contract is not closed out before maturity, it is usually settled by delivering the underlying asset. Forward and futures prices are equal if the interest rate is constant. For the proof, see Hull (2011). For the comparison of forward and futures contracts see Table 2.1.

2.2 Options

An option is a financial contract that allows the option holder (or a buyer) a right to ask the issuer (or writer or seller) to do what is stated in the contract. The issuer has the obligation to do the work. For example, there are options on the right to sell or buy a stock at a fixed price, called an exercise price or a strike price, at a fixed date, called the expiry date or maturity date. A relatively recent type of option is on credit

Table 2.1 Comparison of forward and futures

Forward	Futures
Private contract between two parties	Traded on an exchange
Not standardized	Standardized
Settled at end of contract	Settled daily
Delivery or final settlement	Usually closed out prior to maturity
Credit risk exists	Practically no credit risk

risk for a bond issued by a company. If the company is in good shape and the default risk level is low, then the issuer of the option on credit risk can make a substantial profit. On the other hand, if default risk level is high, then the issuer of the option is exposed to a high level of loss.

Options have financial values derived from underlying assets such as stocks, bonds, foreign exchange and commodities, and hence they are called *derivatives*. European options can be exercised only at expiry date, while American options can be exercised before or at expiry date. The value of an option if it can be exercised immediately is called its *intrinsic value*. Standardized options are traded at a regulated exchange, while over-the-counter (OTC) options are customized, and not traded on exchanges.

A call option gives the holder the right to buy an underlying asset at a strike price (or exercise price), and a put option gives the holder the right to sell. In Fig. 2.1 payoff functions at the expiry date T of a European call option and a European put option are plotted as functions of asset price S_T . A call option is insurance for an asset we plan to buy in the future, and a put option is insurance for an asset we already own.

European call options are bought when at expiry date the underlying asset price is expected to rise above the exercise price, and European put options are bought when the underlying asset price is expected to fall below the exercise price. If we expect that the asset price will rise or fall above or below the exercise price and buy a call option and a put option of the same expiry date T and exercise price K , then that is equivalent to buying one option, called a straddle, with a payoff given on the left in Fig. 2.2. Therefore, its price is given by the sum of the prices of a European call option and a European put option. Since the set of all payoffs of European options of the same expiry date is a vector space, by taking linear combinations of several payoffs we can construct a new option as seen in the previous example.

If we expect that the asset price will move sufficiently far away from the strike price at expiry date, it is preferable to buy an option with a payoff, called a strangle given on the right in Fig. 2.2. In this case, the payoff is zero if the asset price at expiry date is between K_1 and K_2 . Hence its price is the sum of prices of a put option with strike price K_1 and a call option with strike price K_2 .

If we expect that at expiry date the asset price will rise above the exercise price but not too much, then it is preferable to buy an option, called a *spread*, which is a

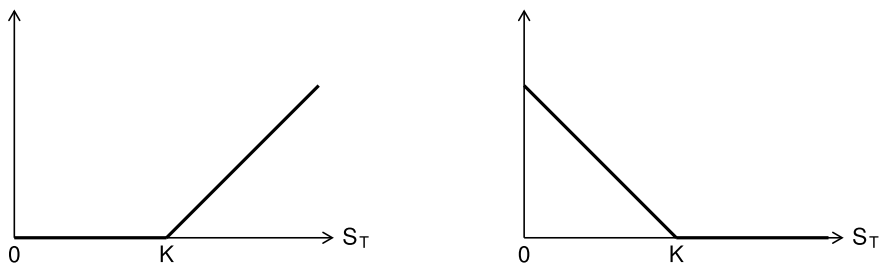


Fig. 2.1 Payoffs of a European call (left) and a European put (right)

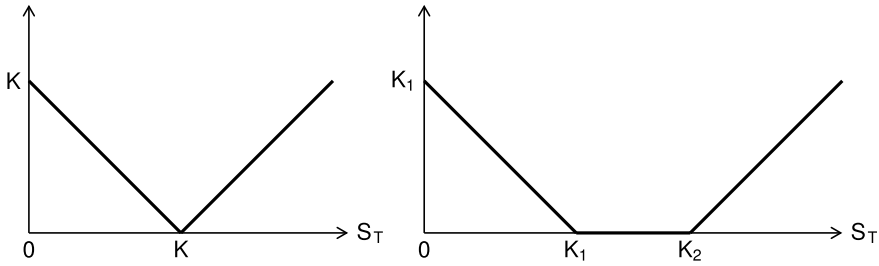


Fig. 2.2 Payoffs of a European straddle (left) and a European strangle (right)

Fig. 2.3 Payoff of a European spread option

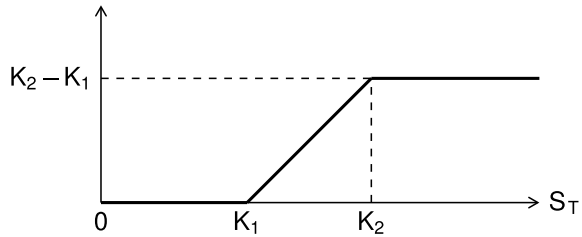
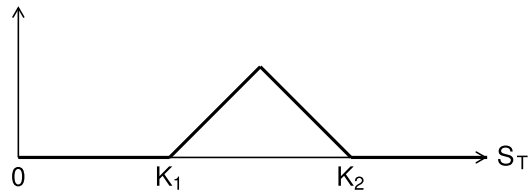


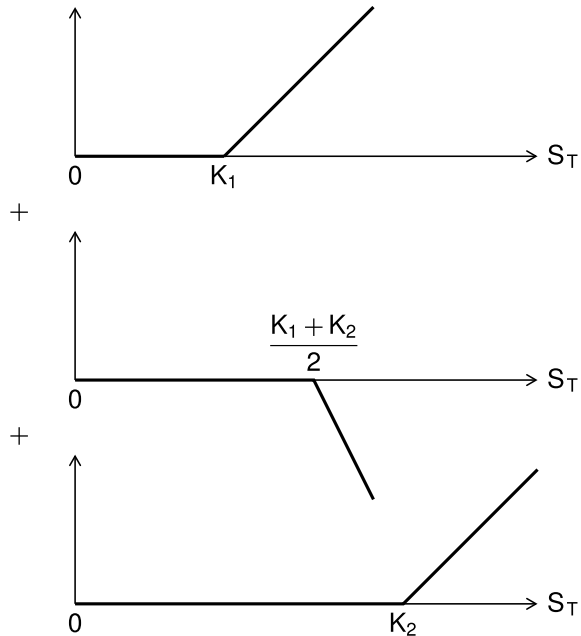
Fig. 2.4 Payoff of a butterfly



combination of two or more options (both calls or both puts) on the same underlying asset with different strikes. Buying a spread option is equivalent to buying a call option of strike K_1 and selling a call option of strike K_2 . See Fig. 2.3. Therefore, the price of a spread is the difference between the prices of a European call option with exercise price K_1 and a European call option with exercise price K_2 . From the viewpoint of the seller of the option, the maximum of possible loss is bounded.

A European option called a *butterfly* has a payoff given in Fig. 2.4. To find its price, we first express the given payoff as a linear combination of payoffs of other options with their prices known. For example, we take three European call options with strike prices $K_1 < \frac{1}{2}(K_1 + K_2) < K_2$. We buy two call options, one with strike K_1 and one with strike K_2 , and sell two contracts of a call option with strike price $\frac{1}{2}(K_1 + K_2)$. In other words, long two calls with strike prices K_1 and K_2 , respectively, and short two contracts of a call with strike price $\frac{1}{2}(K_1 + K_2)$. See Fig. 2.5 where a butterfly is decomposed into a sum of long and short positions.

Fig. 2.5 A butterfly as a sum of two long calls and two contracts of a short call



Simple options are called *vanilla*¹ options, and complicated options are called *exotic* options. There are many exotic options, and new varieties are being invented.

2.3 Put-Call Parity

The *put-call parity* is an equality which describes a relationship between a European put option and a European call option of the same expiry date and strike price on the underlying asset S . Thus, if we know the price of a call, we can find the price of a put with the same expiry date and exercise price, and vice versa.

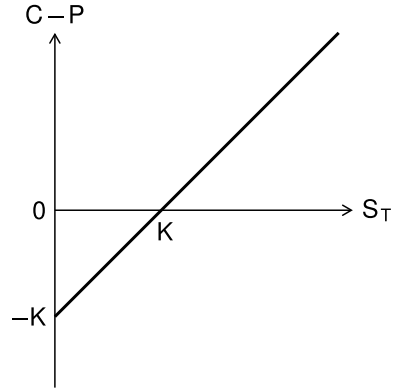
Suppose that we buy a European call option and sell a European put option with the same expiry date T and exercise price K on the underlying asset S with the price process $\{S_t\}_{0 \leq t \leq T}$. Then the overall payoff at T equals

$$\max\{S_T - K, 0\} - \max\{K - S_T, 0\} = S_T - K.$$

See the graph in Fig. 2.6 and also Fig. 2.1.

¹ Vanilla is an orchid with runners climbing up a tree. Its pods contain fragrant material. Although natural vanilla is expensive, synthesized vanilla fragrance is widely used, and thus the word ‘vanilla’ is a synonym for ‘common’.

Fig. 2.6 Payoff when we buy a European call option and sell a put option



Theorem 2.1 (Put-Call Parity) *Let $r \geq 0$ be a risk-free interest rate. Let C_t and P_t respectively denote the prices of European call and put options at time $0 < t < T$ on the underlying asset S_t with expiry date T and exercise price K . Then*

$$C_0 - P_0 = S_0 - Ke^{-rT}.$$

In general, we have

$$C_t - P_t = S_t - Ke^{-r(T-t)},$$

where $T - t$ is called the time to expiry.

Proof We will construct portfolios Π_t^1 and Π_t^2 with equal payoff at maturity. The portfolio Π^1 consists of a call and risk-free deposit Ke^{-rT} at $t = 0$. Its payoff at T is equal to

$$\Pi_T^1 = \max\{S_T - K, 0\} + K = \max\{S_T, K\}.$$

The portfolio Π^2 consists of a put and a share of the stock. Its payoff is equal to

$$\Pi_T^2 = \max\{K - S_T, 0\} + S_T = \max\{K, S_T\}.$$

See Table 2.2. Define a new portfolio $\Pi_t^3 = \Pi_t^1 - \Pi_t^2$. Since $\Pi_T^3 = \Pi_T^1 - \Pi_T^2 = 0$ with probability 1, the no arbitrage principle implies that $\Pi_0^3 = 0$, and hence $\Pi_0^1 - \Pi_0^2 = 0$. (See Definition 1.1.) Hence $C_0 + Ke^{-rT} = P_0 + S_0$. $\square$

Remark 2.1 The put-call parity makes sense even in the following extreme cases:

- (i) If $K = 0$, then holding a call option is equivalent to owning a share of the stock, and hence $C_0 = S_0$. Since it is impossible for the asset price to fall below 0, the put option has no value. Hence $P_0 = 0$.
- (ii) If $T = 0$, i.e., the option is exercised immediately, then $C_0 = \max\{S_0 - K, 0\}$ and $P_0 = \max\{K - S_0, 0\}$.

Table 2.2 Comparison of two portfolios

Time	0	T
Π^1	Call option and bank deposit Ke^{-rT}	$\max\{S_T - K, 0\} + K$
Π^2	Put option and a share of stock S_0	$\max\{K - S_T, 0\} + S_T$

2.4 Relations Among Option Pricing Methods

In the rest of this book we study several different methods for option pricing: First, the binomial tree method for discrete time models where the underlying asset can move only upward and downward, second, the partial differential equation approach introduced by F. Black, M. Scholes and R. Merton, and third, the martingale method that is the most sophisticated mathematically and is the theoretical basis for the Monte Carlo method. The Girsanov theorem is at the center of the method. The logical relations among them are presented in Fig. 2.7. The N -period binomial tree method has a solution which converges to the Black–Scholes–Merton formula as

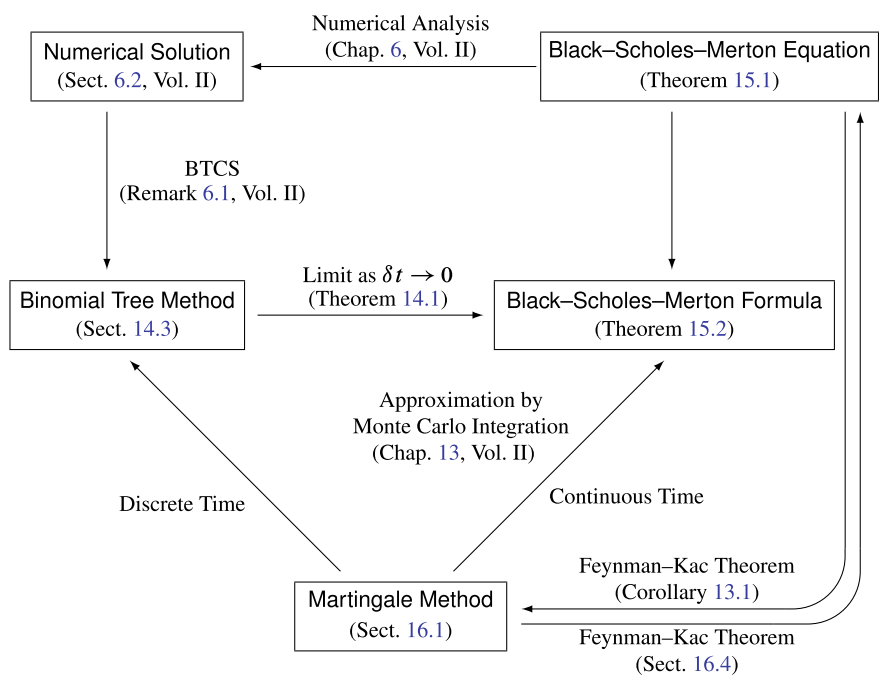


Fig. 2.7 Logical relations among option pricing methods

$N \rightarrow \infty$ for some special choice of parameters as mentioned in Sect. 14.4. The partial differential equation approach is combined with numerical methods when it is hard to find closed form solutions. The Feynman–Kac theorem provides the bridge between the partial differential equations approach and the martingale method.

Exercises

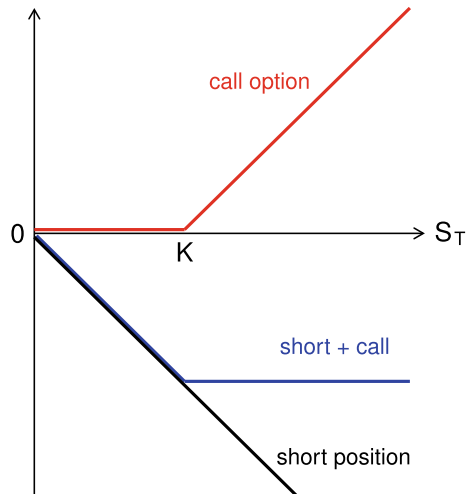
2.1 Write down the formula for a payoff of a portfolio consisting of a short position on an asset and a call option on the same asset. (Hint: Consult the graph for a payoff in Fig. 2.8.)

2.2 Show that a call option is insurance for a short position while a put option is insurance for a long position.

2.3 Express a strangle as a linear combination of other options.

2.4 Using Fig. 2.5 as a hint, express the price of the given butterfly option as a sum of prices of calls and a put.

Fig. 2.8 Sum of a short position on an asset and a call option



Part II

Probability Theory

Chapter 3

The Lebesgue Integral



Given an abstract set Ω , how do we measure the size of one of its subsets A ? When Ω has finite or countably infinite elements, it is natural to count the number of elements in A . However, if A is an uncountable set, we need a rigorous and systematic method. In many interesting cases there is no logical way to measure the sizes of *all* the subsets of Ω , however, we define the concept of size for sufficiently many subsets. Subsets whose sizes can be determined are called measurable subsets, and the collection of these measurable subsets is called a σ -algebra, where the set operations such as union and intersection resemble operations on numbers such as addition and multiplication. After measures are introduced we define the Lebesgue integral of measurable functions. Also presented at the end of the chapter is the Radon–Nikodym theorem, which is needed in understanding Girsanov’s theorem in Chap. 11. For more information, consult Adams and Guillemin (1996), Royden (1989).

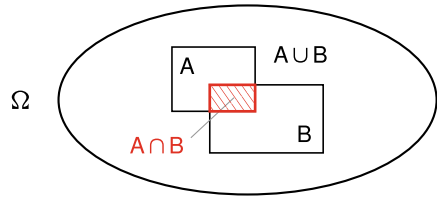
3.1 Measures

When a set is endowed with two operations which resemble the addition and the multiplication of numbers, we call the given set an algebra or a field. If the set under consideration is given by a collection of subsets of a given set, then the operations are usually given by set union and set intersection. (See Fig. 3.1.)

Definition 3.1 (*Algebra*) A collection $\mathcal{F}_0$ of subsets of a set Ω is called an *algebra* on Ω if it satisfies the following conditions:

- (i) $\emptyset, \Omega \in \mathcal{F}_0$.
- (ii) If $A \in \mathcal{F}_0$, then $X \setminus A \in \mathcal{F}_0$.
- (iii) If $A, B \in \mathcal{F}_0$, then $A \cup B \in \mathcal{F}_0$.

Fig. 3.1 The union and the intersection of two subsets A and B



Definition 3.2 (σ -Algebra) A collection $\mathcal{F}$ of subsets of a set Ω is called a σ -algebra on Ω if it satisfies the following conditions:

- (i) $\emptyset, \Omega \in \mathcal{F}$.
- (ii) If $A \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$.
- (iii) If $A_1, A_2, A_3, \dots \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

A subset belonging to $\mathcal{F}$ is called an $\mathcal{F}$ -measurable subset.

Remark 3.1 (i) A σ -algebra is an algebra.

- (ii) If $\mathcal{P} = \{B_1, \dots, B_k\}$ is a finite partition of Ω , then there exists an associated σ -algebra $\mathcal{F}$ whose measurable subsets are unions of some subsets in $\mathcal{P}$. More precisely, if $A \in \mathcal{F}$ then $A = C_1 \cup \dots \cup C_k$ where $C_i = B_i$ or $C_i = \emptyset$ for $1 \leq i \leq k$.
- (iii) A σ -algebra is not necessarily closed under uncountable unions. The prefix σ indicates ‘countable’ just as Σ represents countable summation.
- (iv) If $\mathcal{F}_\gamma$ is a σ -algebra for every $\gamma \in \Gamma$, then $\bigcap_{\gamma \in \Gamma} \mathcal{F}_\gamma$ is also a σ -algebra.
- (v) For a collection $\mathcal{S}$ of subsets of a set Ω there exists the smallest σ -algebra among all the σ -algebras containing $\mathcal{S}$, which is denoted by $\sigma(\mathcal{S})$ and called the σ -algebra generated by $\mathcal{S}$.
- (vi) For any function $X : \Omega \rightarrow \mathbb{R}^1$, let $\sigma(X)$ denote the collection of the subsets of the form $X^{-1}(B)$ where $B \subset \mathbb{R}^1$ is a Borel subset. It can be shown that $\sigma(X)$ is a σ -algebra, called the σ -algebra generated by X . Since the intervals generate the Borel measurable subsets of $\mathbb{R}$, $\sigma(X)$ is the smallest σ -algebra containing the inverse images of all the intervals.
- (vii) When a σ -algebra $\mathcal{F}$ is given on Ω , we call it a measurable space and write $(\Omega, \mathcal{F})$. More than one σ -algebra may be defined on the same set Ω .

Definition 3.3 (Measure) A measure is a mapping which assigns a nonnegative number or $+\infty$ to each measurable subset. More precisely, a measure μ on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \rightarrow [0, +\infty) \cup \{+\infty\}$ satisfying the following conditions:

- (i) $\mu(\emptyset) = 0$.
- (ii) If $A_1, A_2, A_3, \dots \in \mathcal{F}$ are pairwise disjoint, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$

When $\mu(\Omega) = 1$, then μ is called a probability measure.

When a measure μ is defined on $(\Omega, \mathcal{F})$, we call Ω a measure space, and write $(\Omega, \mathcal{F}, \mu)$ or (Ω, μ) . If μ is a probability measure, then we use the symbol $\mathbb{P}$ instead of μ to denote the measure, and the triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability measure space, or a probability space.

Example 3.1 (*Lebesgue measure*) In $\mathbb{R}^n$ a set of the form $[a_1, b_1] \times \cdots \times [a_n, b_n]$ is called an n -dimensional rectangle. Consider the collection $\mathcal{R}$ of n -dimensional rectangles. Using the concept of volume, define $\mu_{\mathcal{R}} : \mathcal{R} \rightarrow [0, +\infty) \cup \{+\infty\}$ by

$$\mu_{\mathcal{R}}([a_1, b_1] \times \cdots \times [a_n, b_n]) = (b_1 - a_1) \times \cdots \times (b_n - a_n).$$

Let $\mathcal{B}_0$ be the σ -algebra generated by $\mathcal{R}$. Then we can extend $\mu_{\mathcal{R}}$ to a measure μ_0 defined on $\mathcal{B}_0$. (We call $\mathcal{B}_0$ the Borel σ -algebra, and a measurable set belonging to $\mathcal{B}_0$ a Borel measurable set.) Next, extend the σ -algebra $\mathcal{B}_0$ by including all the subsets $N \subset A$ such that $\mu_0(A) = 0$, and form a new σ -algebra $\mathcal{B}$. Also extend μ_0 to $\mathcal{B}$ as a measure on $\mathcal{B}$, which is denoted by μ . The extended measure μ is called Lebesgue measure. If $A \in \mathcal{B}$ and $\mu(A) = 0$ and if $B \subset A$, then $B \in \mathcal{B}$ and $\mu(B) = 0$. Such a measure is said to be complete.

Fact 3.1 (Monotonicity of Measure) Consider a sequence of measurable subsets $A_1, A_2, A_3, \dots$ in a probability measure space $(\Omega, \mathcal{F}, \mathbb{P})$.

- (i) If $A_1 \supset A_2 \supset A_3 \supset \cdots$, then $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$.
- (ii) If $A_1 \subset A_2 \subset A_3 \subset \cdots$, then $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right)$.

Fact 3.2 (Kolmogorov Extension Theorem) For a finite set S , consider an infinite product $\Omega = \prod_{i=1}^{\infty} S$. An element of Ω is a sequence $\omega = (\omega_1, \omega_2, \omega_3, \dots)$ where $\omega_i \in S$. If μ is a function defined on the collection of all cylinder subsets satisfying the consistency conditions such as

$$\mu([a_1, \dots, a_{n-1}]) = \sum_{a \in S} \mu([a_1, \dots, a_{n-1}, a])$$

then it can be extended to a measure on the σ -algebra generated by the cylinder subsets.

Example 3.2 (*Binary expansion*) Consider the product $\Omega = \prod_{i=1}^{\infty} \{0, 1\}$. An element of Ω is a binary sequence $\omega = (\omega_1, \omega_2, \omega_3, \dots)$ where $\omega_i = 0$ or 1 . By the binary expansion a real number $x \in [0, 1]$ can be written as $x = \sum_{i=1}^{\infty} a_i 2^{-i}$, $a_i \in \{0, 1\}$, which is identified with the sequence $(a_1, a_2, a_3, \dots) \in \Omega$. Define a cylinder subset of length n by $[a_1, \dots, a_n] = \{\omega \in \Omega : \omega_1 = a_1, \dots, \omega_n = a_n\}$. The collection of all cylinder sets is denoted by $\mathcal{R}$. For $0 \leq p \leq 1$ define a set function $\mu_p : \mathcal{R} \rightarrow [0, \infty)$ by

$$\mu_p([a_1, \dots, a_n]) = p^k (1 - p)^{n-k}$$

where k is the number of times that the symbol '0' appears in the string $a_1, \dots, a_n$. Then the Kolmogorov Extension Theorem implies that μ_p extends to the σ -algebra generated by $\mathcal{R}$, which is still denoted by μ_p .

Definition 3.4 (*Conditional Measure*) Given a measurable subset A of positive measure on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, define a new measure $\mathbb{P}_A$ by

$$\mathbb{P}_A(B) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} .$$

Then $\mathbb{P}_A$ is called a *conditional measure*, usually denoted by $\mathbb{P}(\cdot | A)$, i.e., $\mathbb{P}(B|A) = \mathbb{P}_A(B)$. See Fig. 3.1. Or, we may define a σ -algebra on A by $\mathcal{F}_A = \{B \in \mathcal{F} : B \subset A\}$, and define a probability measure $\mathbb{P}_A$ on $(A, \mathcal{F}_A)$ by $\mathbb{P}_A(B) = \mathbb{P}(B)/\mathbb{P}(A)$.

Definition 3.5 Let $(\Omega, \mathbb{P})$ be a probability space. Suppose that a statement H depends on $\omega \in \Omega$, i.e., H is true or false depending ω . If there exists an A such that $\mathbb{P}(\Omega \setminus A) = 0$ and if H holds for every $\omega \in A$, then we say that H holds almost everywhere or for almost every ω with respect to $\mathbb{P}$. In probability theory we say that H holds almost surely, or with probability 1.

3.2 Measurable Functions

Definition 3.6 (*Measurable Function*) Let $(\Omega, \mathcal{F})$ be a measurable space. A function $X : (\Omega, \mathcal{F}) \rightarrow \mathbb{R}$ is said to be *measurable* (or $X : \Omega \rightarrow \mathbb{R}$ is *measurable* with respect to $\mathcal{F}$, or $X : \Omega \rightarrow \mathbb{R}$ is $\mathcal{F}$ -measurable) if

$$X^{-1}((a, b)) \in \mathcal{F} \quad \text{for every } a < b .$$

(See Fig. 3.2.) Some other equivalent conditions are

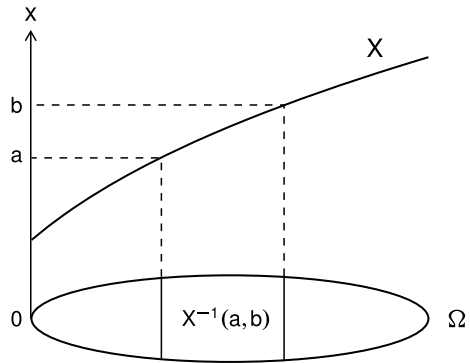
$$\begin{aligned} X^{-1}((-\infty, a]) &\in \mathcal{F} \quad \text{for every } a , \\ X^{-1}((-\infty, a)) &\in \mathcal{F} \quad \text{for every } a , \\ X^{-1}((a, +\infty)) &\in \mathcal{F} \quad \text{for every } a , \\ X^{-1}([a, +\infty)) &\in \mathcal{F} \quad \text{for every } a . \end{aligned}$$

In probability theory a measurable function is called a *random variable*. (The definition can be extended to complex-valued functions.)

Example 3.3 Let $\Omega = \{a, b, c\}$ and consider a σ -algebra $\mathcal{F} = \{\{a, b\}, \{c\}, \emptyset, \Omega\}$. Then a function $X : \Omega \rightarrow \mathbb{R}^1$ is $\mathcal{F}$ -measurable if and only if $X(a) = X(b)$.

Remark 3.2 Let $(\Omega, \mathcal{A})$ be a measurable space and let $\mathcal{B}$ be a sub- σ -algebra of $\mathcal{A}$. If X is measurable with respect to $\mathcal{B}$, then it is also measurable with respect to $\mathcal{A}$.

Fig. 3.2 Inverse image of an interval (a, b) under X



Definition 3.7 Let $X : \Omega \rightarrow \mathbb{R}^1$ be a measurable function. Let $\sigma(X)$ denote the collection of all the subsets of the form $X^{-1}(B)$ where $B \subset \mathbb{R}^1$ is a Borel subset. We can show that $\sigma(X)$ is a sub- σ -algebra of $\mathcal{F}$, which is called the *sub- σ -algebra generated by X* . Since the intervals generate the Borel measurable subsets of $\mathbb{R}$, $\sigma(X)$ is the smallest sub- σ -algebra containing the inverse images of all the intervals. For measurable functions X_γ , $\gamma \in \Gamma$, let $\sigma(\{X_\gamma : \gamma \in \Gamma\})$ denote the smallest sub- σ -algebra containing all the sub- σ -algebras $\sigma(X_\gamma)$.

Definition 3.8 (Indicator Function) Let $A \subset \Omega$. A function $\mathbb{1}_A : \Omega \rightarrow \mathbb{R}^1$ defined by $\mathbb{1}_A(\omega) = 1$ for $\omega \in A$ and $\mathbb{1}_A(\omega) = 0$ for $\omega \notin A$ is called an *indicator function*. If A is a measurable subset, then $\mathbb{1}_A$ is a measurable function.

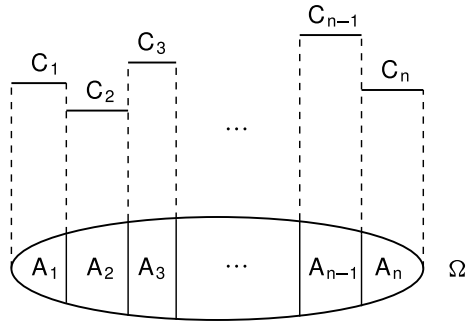
Remark 3.3 If σ -algebras $\mathcal{F}$ and $\mathcal{G}$ satisfy $\mathcal{G} \subset \mathcal{F}$, i.e., there are additional measurable subsets in $\mathcal{F}$, then a $\mathcal{G}$ -measurable function is also $\mathcal{F}$ -measurable. A function is more likely to be measurable with respect to a larger σ -algebra.

Definition 3.9 (Simple Function) A measurable function $s : \Omega \rightarrow \mathbb{R}$ is called a *simple function* if it takes finitely many values, i.e., it is of the form

$$s(\omega) = \sum_{k=1}^n c_k \mathbb{1}_{A_k}(\omega)$$

where A_k is a measurable set and c_k is constant for $1 \leq k \leq n$. We may assume that A_k are pairwise disjoint. See Fig. 3.3. Note that a simple function is a linear combination of indicator functions.

Fig. 3.3 A simple function on a measurable space



3.3 The Lebesgue Integral

H. Lebesgue in the early 1900s invented a new integration technique, which extended the classical Riemann integral in the sense that the domain of integrand for the new method can be an arbitrary set, not necessarily a subset of a Euclidean space. In contrast to the partitioning of the x -axis for the Riemann integral, the Lebesgue integral partitions the y -axis. In the Lebesgue integral the concept corresponding to length, area, or volume in Euclidean spaces is measure. One of the advantages of the Lebesgue integral is that the operations of taking integral and taking limit can be done interchangeably.

When the Lebesgue integral was first invented, it was regarded as an unnecessarily abstract mathematical concept compared to the Riemann integral, but in 1933 A.N. Kolmogorov used it in his book (Kolmogorov, 1933) to build an axiomatic foundation for probability theory, and the Lebesgue integral has become the language of probability theory. However, there exists a linguistic gap between real analysis, which is the area built upon the Lebesgue integral, and probability theory, and both disciplines still have their own terminology. For example, a measurable function in real analysis corresponds to a random variable in probability theory. Consult Table 4.1.

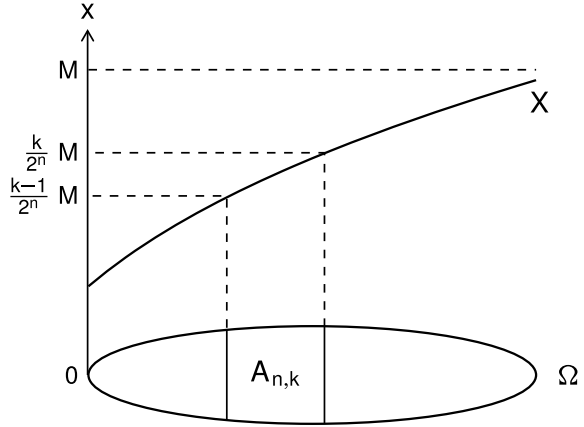
Now we define the Lebesgue integral of a measurable function X defined on Ω with respect to a measure μ . First, we consider a simple function s of the form

$$s(\omega) = \sum_{k=1}^n c_k \mathbb{1}_{A_k}(\omega)$$

where $\mu(A_k) < \infty$ for $1 \leq k \leq n$. Its Lebesgue integral $I(s)$ is defined by

$$I(s) = \sum_{k=1}^n c_k \mu(A_k) .$$

Fig. 3.4 Partition of a domain for a Lebesgue integral



Next, we consider a nonnegative bounded measurable function X whose support $\{\omega \in \Omega : X(\omega) \neq 0\}$ has finite measure. Since X is bounded, there exists a constant $M > 0$ such that $0 \leq X(\omega) \leq M$ for $\omega \in \Omega$. Put

$$A_{n,k} = \left\{ \omega \in \Omega : \frac{k-1}{2^n} M < X(\omega) \leq \frac{k}{2^n} M \right\}, \quad 1 \leq k \leq 2^n.$$

Since X is a measurable function, $A_{n,k}$ is a measurable set. (See Fig. 3.4.) Note that the sequence of simple functions

$$s_n(\omega) = \sum_{k=1}^{2^n} \frac{k-1}{2^n} M \mathbb{1}_{A_{n,k}}(\omega)$$

is monotonically increasing and uniformly converges to X at every ω as $n \rightarrow \infty$. Now the limit of the monotonically increasing sequence

$$I(s_n) = \sum_{k=1}^{2^n} \frac{k-1}{2^n} M \times \mu(A_{n,k})$$

is called the Lebesgue integral of X on (Ω, μ) , and is denoted by

$$\int_{\Omega} X \, d\mu.$$

For a nonnegative measurable function X that is not necessarily bounded, we define its Lebesgue integral by

$$\int_{\Omega} X d\mu = \sup_Y \int_{\Omega} Y d\mu$$

where the supremum is taken over all bounded measurable functions such that $0 \leq Y \leq X$. If the supremum is finite, then we say that X is integrable.

Finally, we define the Lebesgue integral of a real-valued measurable function X . Note that X can be written as

$$X = X^+ - X^-$$

where

$$X^+ = \max(X, 0) \geq 0 \quad \text{and} \quad X^- = \max(-X, 0) \geq 0.$$

Thus we define the integral of X by

$$\int_{\Omega} X d\mu = \int_{\Omega} X^+ d\mu - \int_{\Omega} X^- d\mu$$

if both integrals on the right are finite.

In defining a Riemann integral of a continuous function the domain is a subset of a Euclidean space such as an interval or a rectangle. On the other hand, in Lebesgue integration theory the domain of an integrand is not necessarily a subset of a Euclidean space and it is not possible to partition the domain in a natural way. Instead, the vertical real axis representing the function values is partitioned into sufficiently short intervals, and their inverse images are used to partition the domain of the integrand. Since the integrand is measurable, the inverse images are measurable subsets, and their sizes are measured by a given measure. Even when the domain is a bounded interval, Lebesgue theory is more powerful than Riemann integral theory. (See Exercise 3.5 for an example.) There are many interesting and useful facts that hold only for Lebesgue integrals, and some of them are presented in this section.

In probability theory the Lebesgue integral of X is called the expectation, and denoted by $\mathbb{E}[X]$ or $\mathbb{E}^{\mathbb{P}}[X]$ where $\mathbb{P}$ is a probability measure. For more details see Sect. 4.1.

Let $\mathbb{1}_A$ denote the indicator function of a measurable subset A . Then $\mathbb{1}_A$ is a measurable function, and we have

$$\int_{\Omega} \mathbb{1}_A d\mu = \mu(A).$$

We define the integral of X on A by

$$\int_A X d\mu = \int_{\Omega} \mathbb{1}_A X d\mu.$$

If two measurable sets A_1 and A_2 satisfy $\mu(A_1 \triangle A_2) = 0$ where $\triangle$ denotes symmetric difference, then from the viewpoint of measure theory A_1 and A_2 may be

regarded as the same sets, and we say that $A_1 = A_2$ modulo measure zero sets. When the integral of a measurable function X is finite, then X is said to be integrable. If the values of two measurable functions are different only on a subset of measure 0, then two functions have the same integral.

Example 3.4 For a finite or countably infinite set $\Omega = \{\omega_1, \dots, \omega_n, \dots\}$ we define its σ -algebra as the collection of all subsets of Ω . Then every function is measurable on Ω . If a measure μ on Ω is given by $\mu(\{\omega_i\}) = p_i$ for every i , then

$$\int_{\Omega} X \, d\mu = \sum_i p_i X(\omega_i) .$$

Such a measure is said to be *discrete*. It is a probability measure if and only if $\sum_i p_i = 1$.

The smallest σ -algebra generated by open subsets in a topological space such as a Euclidean space is called the Borel σ -algebra, and a measure defined there is called a Borel measure. If we extend the Borel σ -algebra by including all the subsets of Borel measure zero subsets, then it is called a completion. The extension of Borel measure in the case of a Euclidean space is called Lebesgue measure, which is regarded as an extension of the concept of volume in Euclidean space to subsets of irregular shape. On a finite closed interval the Riemann integral of a continuous function is equal to the Lebesgue integral with respect to Lebesgue measure. On a general measure space, if every subset of every measure zero set is measurable, the measure and the corresponding measure space are said to be complete. Given a possibly incomplete measure space, there is an extension that is complete. The smallest such extension is called the completion of the measure space. In what follows we assume that the given measure space is complete if needed.

Remark 3.4 Consider a measure μ on the real line $\mathbb{R}$. If there exists a measurable function $\alpha(x) \geq 0$ such that

$$\mu(A) = \int_A \alpha(x) \, dx$$

for every measurable set A where dx denotes Lebesgue measure, then μ is said to be *absolutely continuous*, and we write $d\mu = \alpha \, dx$. In this case, for any measurable function $f : \mathbb{R} \rightarrow \mathbb{R}$ we have

$$\int_{\mathbb{R}} f(x) \, d\mu = \int_{\mathbb{R}} f(x) \alpha(x) \, dx .$$

Here are some important properties of the Lebesgue integral.

Fact 3.3 (Monotone Convergence Theorem) Let $0 \leq X_1 \leq X_2 \leq \dots$ be a monotonically increasing sequence of measurable functions on a measure space (Ω, μ) . If X_n converges to X almost surely, i.e., $X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)$ for almost every $\omega \in \Omega$, then

$$\int_{\Omega} X \, d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} X_n \, d\mu .$$

Fact 3.4 (Fatou's Lemma) Let $X_n \geq 0$, $n \geq 1$, be a sequence of measurable functions on a measure space (Ω, μ) . Then

$$\int_{\Omega} \liminf_{n \rightarrow \infty} X_n \, d\mu \leq \liminf_{n \rightarrow \infty} \int_{\Omega} X_n \, d\mu .$$

Fact 3.5 (Lebesgue Dominated Convergence Theorem) Let $X_n, n \geq 1$, be a sequence of measurable functions on a measure space (Ω, μ) . Suppose that the limit $\lim_{n \rightarrow \infty} X_n(\omega)$ exists at almost every $\omega \in \Omega$, and that there exists an integrable function $Y \geq 0$ such that $|X_n(\omega)| \leq Y(\omega)$ for every n . Then

$$\int_{\Omega} \lim_{n \rightarrow \infty} X_n \, d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} X_n \, d\mu .$$

Definition 3.10 (Norm) Let X denote a measurable function on a measure space $(\Omega, \mathcal{F}, \mu)$. For $1 \leq p < \infty$ define the L^p -norm $\|\cdot\|_p$ and the Lebesgue space L^p by

$$\|X\|_p = \left(\int_{\Omega} |X|^p \, d\mu \right)^{1/p} ,$$

and $L^p(\Omega, \mu) = \{X : \|X\|_p < \infty\}$. For $p = \infty$ define

$$\|X\|_{\infty} = \inf \{M : |X(\omega)| \leq M \text{ for almost every } \omega\} ,$$

and $L^{\infty}(\Omega, \mu) = \{X : \|X\|_{\infty} < \infty\}$. Then $L^p(\Omega, \mu)$ is a vector space, and $\|\cdot\|_p$ is a norm. For $1 \leq p \leq \infty$ the normed space L^p is complete. If $\mu(\Omega) < \infty$ and $1 \leq r < p$, then $L^p(\Omega) \subset L^r(\Omega)$, and hence $L^p(\Omega) \subset L^1(\Omega)$.

Definition 3.11 (Convergence)

- (i) A sequence of functions $X_n \in L^p$ is said to *converge in L^p* to some $X \in L^p$ if $\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0$.
- (ii) A sequence of measurable functions Y_n *converges in measure* to Y if for every $\varepsilon > 0$ we have $\lim_{n \rightarrow \infty} \mu(\{\omega : |Y_n(\omega) - Y(\omega)| > \varepsilon\}) = 0$. If (Ω, μ) is a probability space, convergence in measure coincides with convergence in probability defined by the condition $\lim_{n \rightarrow \infty} \mathbb{P}(|Y_n - Y| > \varepsilon) = 0$ for every $\varepsilon > 0$.

Example 3.5 Take $\Omega = (0, 1]$. Let $A_n = (0, \frac{1}{n}]$, $n \geq 1$. Define a sequence of indicator functions $X_n = n \times \mathbb{1}_{A_n}$, $n \geq 1$. Then X_n converges to 0 as $n \rightarrow \infty$ at every point, but $\|X_n - 0\|_1 = 1$ and $\|X_n - 0\|_2 = \sqrt{n} \rightarrow \infty$.

Now we compare various modes of convergence.

Fact 3.6 For $1 \leq p < \infty$ the following statements hold:

- (i) If X_n converges to X in L^p , then X_n converges to X in probability.
- (ii) If X_n converges to X in probability, and if $|X_n| \leq Y$ for every n for some $Y \in L^p$, then X_n converges to X in L^p .
- (iii) If $X_n, X \in L^p$ and if X_n converges to X almost everywhere, then the L^p -convergence of X_n to X is equivalent to $\|X_n\|_p \rightarrow \|X\|_p$.

Fact 3.7 Assume that $\mu(\Omega) < \infty$.

- (i) For $1 \leq p \leq \infty$, if X_n converges to X in L^p , then X_n converges to X in measure.
- (ii) If $X_n(\omega)$ converges to $X(\omega)$ at almost every $\omega \in \Omega$, then X_n converges to X in probability.

Fact 3.8 Let $1 \leq p \leq \infty$. If X_n converges to X in L^p , then there exists an increasing sequence of natural numbers $\{n_k\}_{k=1}^\infty$ such that $X_{n_k}(\omega)$ converges to $X(\omega)$ for almost every ω as $k \rightarrow \infty$.

3.4 Inequalities

Definition 3.12 (*Convex Function*) A function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is said to be *convex* if

$$\phi\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i \phi(x_i)$$

for $\lambda_1, \dots, \lambda_n \geq 0$ such that $\sum_{i=1}^n \lambda_i = 1$ and for $x_1, \dots, x_n \in \mathbb{R}$. If the inequality is in the opposite direction, then it is called a concave function.

If ϕ is convex, then $-\phi$ is concave. Convex and concave functions are continuous.

Fact 3.9 (Jensen's Inequality) Let $X : \Omega \rightarrow \mathbb{R}$ be a measurable function on a probability measure space $(\Omega, \mathbb{P})$. If $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is convex and $\phi \circ X$ is integrable, then

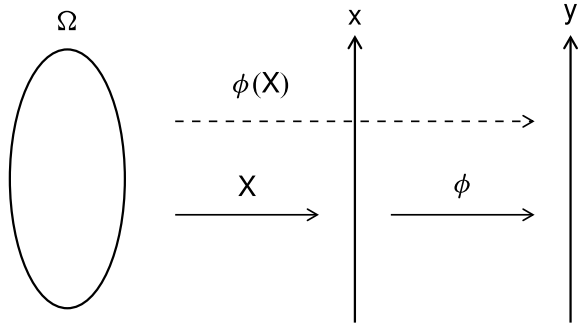
$$\phi\left(\int_{\Omega} X \, d\mathbb{P}\right) \leq \int_{\Omega} \phi \circ X \, d\mathbb{P}.$$

(See Fig. 3.5.) For a concave function the inequality is in the opposite direction.

Example 3.6 Since $\phi(x) = |x|$ is a convex function, we have

$$\left| \int_{\Omega} X \, d\mathbb{P} \right| \leq \int_{\Omega} |X| \, d\mathbb{P}.$$

Fig. 3.5 Composition of a random variable X and a function ϕ



Example 3.7 Since $\phi(x) = \frac{1}{x}$ is a convex function on $(0, \infty)$, we have

$$\frac{1}{\int_{\Omega} X \, d\mathbb{P}} \leq \int_{\Omega} \frac{1}{X} \, d\mathbb{P}$$

for $X > 0$.

Example 3.8 Since $\phi(x) = \log x$ is a concave function on $(0, \infty)$, we have

$$\int_{\Omega} \log X \, d\mathbb{P} \leq \log \left(\int_{\Omega} X \, d\mathbb{P} \right)$$

for $X > 0$. Now we consider Example 3.4, and choose $X(\omega_i) = \frac{1}{p_i}$ for $1 \leq i \leq n$. Since $\int_{\Omega} X \, d\mathbb{P} = \sum_i p_i \frac{1}{p_i} = n$, Jensen's inequality implies that

$$-\sum_{i=1}^n p_i \log p_i \leq \log n.$$

(The left-hand side is called the entropy of the probability distribution.)

Example 3.9 Take a convex function $\phi(x) = \max\{x - K, 0\}$ for a constant K , then we have

$$\max \left\{ \int_{\Omega} X \, d\mathbb{P} - K, 0 \right\} \leq \int_{\Omega} \max\{X - K, 0\} \, d\mathbb{P}$$

for a random variable X on a probability space $(\Omega, \mathbb{P})$. In option pricing theory, X is the asset price at expiry date and $\phi(X)$ is the payoff of a European call option with strike price K .

Fact 3.10 (Hölder's Inequality) Let X, Y be measurable functions on a measure space (Ω, μ) such that $X \in L^p$, $Y \in L^q$, $XY \in L^1$. If p, q satisfy $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\|XY\|_1 \leq \|X\|_p \|Y\|_q.$$

Equality holds if and only if there exist two constants $C_1 \geq 0$ and $C_2 \geq 0$, not both zero, such that for almost every $\omega \in \Omega$ $C_1|X(\omega)|^p = C_2|Y(\omega)|^q$.

Hölder's inequality for $p = 2$ is called the Cauchy–Schwarz inequality.

Fact 3.11 (Minkowski's Inequality) For $p \geq 1$, let X and Y be measurable functions on a measure space (Ω, μ) such that $X, Y \in L^p$. Then

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p .$$

For $1 < p < \infty$ a necessary and sufficient condition for equality is that there exist constants $C_1 \geq 0$ and $C_2 \geq 0$, not both 0, such that for almost every ω we have $C_1 X(\omega) = C_2 Y(\omega)$. For $p = 1$ a necessary and sufficient condition for equality is that there exists a measurable function $Z \geq 0$ such that $X(\omega)Z(\omega) = Y(\omega)$ for almost every ω satisfying $X(\omega)Y(\omega) \neq 0$.

Minkowski's inequality implies that an L^p space is a normed space.

Fact 3.12 (Chebyshev's Inequality) Let (Ω, μ) be a measure space. If $p \geq 1$ and $X \in L^p(\Omega, \mu)$, then for every $a > 0$ we have

$$\mu(\{\omega \in \Omega : |X(\omega)| > a\}) \leq \frac{1}{a^p} \int_{\Omega} |X|^p d\mu .$$

Proof

$$\mu(\{\omega : |X(\omega)| \geq a\}) = \frac{1}{a^p} \int_{\{|X| \geq a\}} a^p d\mu \leq \frac{1}{a^p} \int_{\{|X| \geq a\}} |X|^p d\mu \leq \frac{1}{a^p} \int_{\Omega} |X|^p d\mu .$$

3.5 The Radon–Nikodym Theorem

Definition 3.13 A measure μ on Ω is *discrete* if there is a countable set $B \subset \Omega$ such that $\mathbb{P}(B^c) = 0$. A measure μ on Ω is *continuous* if $\mu(\{\omega\}) = 0$ for any single element $\omega \in \Omega$. If μ is continuous, then $\mu(A) = 0$ for any countable set A .

Let $X \geq 0$ be an integrable function on a measure space $(\Omega, \mathcal{F}, \mu)$. For arbitrary $A \in \mathcal{F}$ define

$$v(A) = \int_A X d\mu . \tag{3.1}$$

Then v is also a measure on $(\Omega, \mathcal{F})$. If $\int_{\Omega} X d\mu = 1$, then v is a probability measure. If $\mu(A) = 0$, then $v(A) = 0$.

Definition 3.14 (*Absolute Continuity*)

- (i) Given measures $\mathbb{P}$ and $\mathbb{Q}$ on a measurable space $(\Omega, \mathcal{F})$, we say that $\mathbb{Q}$ is *absolutely continuous* with respect to $\mathbb{P}$ if $\mathbb{P}(A) = 0$ implies $\mathbb{Q}(A) = 0$, and write $\mathbb{Q} \ll \mathbb{P}$. Equivalently, for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\mathbb{P}(A) < \delta$ implies $\mathbb{Q}(A) < \varepsilon$. If $\mathbb{Q}$ by $\mathbb{Q}(A) = \mathbb{E}[\mathbb{1}_A X]$ as in (3.1), then $\mathbb{Q} \ll \mathbb{P}$. Theorem 3.1 states that the converse is also true.
- (ii) If $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$, then we say that $\mathbb{P}$ and $\mathbb{Q}$ are *equivalent*, and write $\mathbb{Q} \approx \mathbb{P}$.

Theorem 3.1 (Radon–Nikodym) *Let $\mathbb{P}$ and $\mathbb{Q}$ be finite measures on a measurable space $(\Omega, \mathcal{F})$. If $\mathbb{Q} \ll \mathbb{P}$, then there exists a nonnegative measurable function $X : \Omega \rightarrow \mathbb{R}^1$ such that for arbitrary $A \in \mathcal{F}$ we have*

$$\mathbb{Q}(A) = \int_A X d\mathbb{P}.$$

We write $X = \frac{d\mathbb{Q}}{d\mathbb{P}}$ and call it the *Radon–Nikodym derivative* of $\mathbb{Q}$ with respect to $\mathbb{P}$.

Remark 3.5 (i) If $\mathbb{P}$ and $\mathbb{Q}$ are probability measures on Ω , then for any measurable function $Y : \Omega \rightarrow \mathbb{R}^1$ we have

$$\int_{\Omega} Y d\mathbb{Q} = \int_{\Omega} Y \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P}.$$

- (ii) We say that two measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent when $\mathbb{P}(A) = 0$ if and only if $\mathbb{Q}(A) = 0$. In this case we have $\frac{d\mathbb{P}}{d\mathbb{Q}} = \left(\frac{d\mathbb{Q}}{d\mathbb{P}}\right)^{-1}$.
- (iii) The Radon–Nikodym Theorem is an essential concept in the martingale method for option pricing. Instead of the real world probability $\mathbb{P}$ for the stock price movement we find an equivalent probability $\mathbb{Q}$ and compute the expectation.
- (iv) A measure ν on $\mathbb{R}$ is absolutely continuous (with respect to Lebesgue measure) if there exists an integrable function $f(x) \geq 0$ such that $d\nu = f dx$ where dx is Lebesgue measure. Note that f is the Radon–Nikodym derivative of ν with respect to Lebesgue measure. An absolutely continuous measure is continuous. A measure ν on $\mathbb{R}$ is *singular continuous* (with respect to Lebesgue measure) if ν is continuous and there exists a set A of Lebesgue measure zero such that $\nu(A^c) = 0$. A measure ν can be decomposed into a sum $\nu = \nu_{ac} + \nu_{sc} + \nu_d$, the components respectively representing the absolutely continuous part, singular continuous part and discrete part of ν . If ν is absolutely continuous, then $\nu = \nu_{ac}$, i.e., $\nu_{sc} = 0$ and $\nu_d = 0$.

Exercises

3.1 Let $X : \Omega \rightarrow \mathbb{R}^1$ be a function defined on a measurable space $(\Omega, \mathcal{F})$.

- (i) Show that $\sigma(X + \alpha) = \sigma(X)$ for any constant α .
- (ii) Show that $\sigma(\beta X) = \sigma(X)$ for any constant $\beta \neq 0$.
- (iii) Let $\phi : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ be a continuous function. Show that $\sigma(\phi(X))$ is a sub- σ -algebra of $\sigma(X)$. For example, $\sigma(X^2)$ is a sub- σ -algebra of $\sigma(X)$.

3.2 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability measure space and let $A_n, n \geq 1$, be a sequence of measurable subsets such that $\mathbb{P}(A_n) = 1$ for $n \geq 1$. Show that $\mathbb{P}(\bigcap_{n=1}^{\infty} A_n) = 1$.

3.3 Let $A \subset [0, 1]$ be a set consisting of points $\sum_{n=1}^{\infty} \frac{a_n}{2^n}$ where $a_n = 0, 1, n \geq 1$, and $a_n = 0$ except for finitely many n . Find the Lebesgue measure of A .

3.4 Let $(\Omega, \mathbb{P})$ be a probability measure space. Suppose that $X \geq 0$ and $Y \geq 0$ are measurable functions on Ω such that $XY \geq 1$. Show that $\int_{\Omega} X d\mathbb{P} \int_{\Omega} Y d\mathbb{P} \geq 1$. (For an example in finance, see Example 4.4.)

3.5 Define a function $f : [0, 1] \rightarrow \mathbb{R}$ by $f(x) = 0$ if x is rational, and $f(x) = x$ if x is irrational. Show that f is not Riemann integrable. Find the Lebesgue integral $\int_{[0,1]} f dm$ where m denotes Lebesgue measure.

3.6 Let (Ω, μ) be a measure space. Show that a nonnegative measurable function X is integrable if $\sum_{n=0}^{\infty} \mu(\{\omega \in \Omega : X(\omega) \geq n\}) < +\infty$. (See also Exercise 4.1.)

3.7 Evaluate

$$\lim_{n \rightarrow \infty} \int_0^{\infty} \left(\frac{\sin x}{x} \right)^n dx .$$

Chapter 4

Basic Probability Theory



When Henri Lebesgue (1875–1941) invented Lebesgue integral, it was regarded as an abstract concept without applications. It was Andrey Nikolaevich Kolmogorov (1903–1987) who first showed in Kolmogorov (1933) how to formulate the axiomatic probability theory based on Lebesgue integration. For example, probability corresponds to a measure, and the expectation of a random variable in probability theory is the Lebesgue integral of a measurable function. In this chapter we introduce basic probabilistic and statistical facts, including independence of random variables, change of variables, the Law of Large Numbers and the Central Limit Theorem. For more information consult (Adams and Guillemin 1996; Billingsley 2012; Durrett 2019; Pfeiffer 1990; Resnick 1998).

4.1 Measure and Probability

There is a linguistic gap between terminologies in Lebesgue integral theory and probability theory. To compare the definitions, a measure in Lebesgue integral theory is called probability, a measure space is a sample space, a measurable subset is an event, a σ -algebra means a σ -field, a measurable function corresponds to a random variable, and a Lebesgue integral is called an expectation. The characteristic function χ_E of a subset E is called an indicator function, and denoted by $\mathbb{1}_E$. In set theory, the domain of a function is usually denoted by X , and a variable denoted by x , however, in probability theory a probability measure space is denoted by Ω and the values of a random variable X denoted by x . If we may repeat, a random variable is a function.

Once we get used to two different sets of notations, we find them convenient in practice. While in real analysis such expressions as ‘almost everywhere’ or ‘for almost every point’ are used, in probability theory more intuitive expressions such as ‘almost surely’ or ‘with probability 1’ are used. Consult Table 4.1.

Table 4.1 Comparison of terminology and notation

Probability theory	Lebesgue integral
Sample space Ω	Measure space X
$\omega \in \Omega$	$x \in X$
σ -field $\mathcal{F}$	σ -algebra $\mathcal{A}$
Event A	Measurable set A
Random variable X	Measurable function f
$\{X \leq x\}$	$f^{-1}((-\infty, x])$
Probability $\mathbb{P}$	Measure μ
Indicator function $\mathbb{1}_A$	Characteristic function χ_A
Expectation $\mathbb{E}[X]$	Lebesgue integral $\int_X f d\mu$
Almost surely, with probability 1	For almost every point
Conditional probability $\mathbb{P}(\cdot A)$	Conditional measure on a subset μ_A

More confusing is the fact that in probability theory a Fourier transform is called a characteristic function. In this case, the Fourier transform of a random variable $X : \Omega \rightarrow \mathbb{R}$ is defined by

$$\mathbb{E}[e^{itX}] = \int_{\Omega} e^{itX(\omega)} d\mathbb{P}(\omega), \quad -\infty < t < \infty,$$

where i denotes a complex number satisfying $i^2 = -1$. (We may define the Fourier transform by using $e^{-itX(\omega)}$.)

For an event A of positive probability, we define the conditional probability $\mathbb{P}(B|A)$ that the event B occurs on the condition that the event A occurs by

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

The conditional probability measure $\mathbb{P}_A$ is defined by $\mathbb{P}_A(B) = \mathbb{P}(B|A)$.

Definition 4.1 Given a probability measure space $(\Omega, \mathcal{F}, \mathbb{P})$, a measurable function $X : (\Omega, \mathcal{F}) \rightarrow \mathbb{R}$ is called a *random variable*. We define a probability measure μ_X on $\mathbb{R}$ by

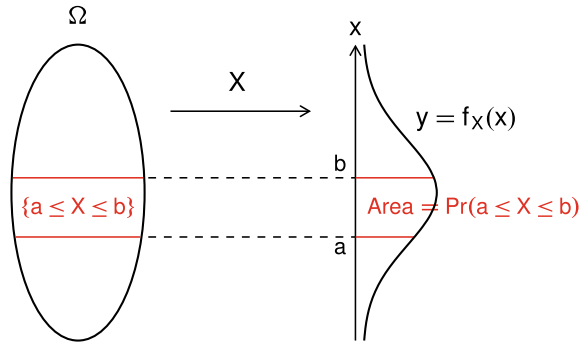
$$\mu_X(A) = \mathbb{P}(X^{-1}(A)),$$

where A is a Borel measurable subset of $\mathbb{R}$. If there exists an $f_X \geq 0$ on $\mathbb{R}$ such that

$$\int_A f_X(x) dx = \mu_X(A)$$

for every A , we call f_X the *probability density function*, or pdf for short, of X .

Fig. 4.1 An abstract model of a probability density function



Note that $\mu_X(A)$ is the probability that a value of X belongs to A , i.e., $\mu_X(A) = \Pr(X \in A)$. The probability measure μ_X has all the information on the distribution of the values of X , and we focus our attention on μ_X instead of Ω .

Example 4.1 (*Coin Tossing*) To solve a given problem we may choose any convenient probability space. For example, when we consider the problem of tossing a coin once, if we let H denote heads, T tails, then we may take $\Omega = \Omega_1 = \{H, T\}$. However, in the problem of tossing a coin twice or tossing two coins once, we take $\Omega = \Omega_2 = \{HH, HT, TH, TT\}$. Even in the first problem of tossing a coin once, it is possible to use Ω_2 , but it is simpler and more convenient to use Ω_1 . For example, suppose that a random variable X takes the value $X = 1$ if a coin turns up heads and takes the value $X = -1$ if a coin turn up tails. In the case of employing Ω_2 we define $X(HH) = X(HT) = 1$, $X(TH) = X(TT) = -1$, which is more complicated than the case of Ω_1 with $X(H) = 1$, $X(T) = -1$.

In Fig. 4.1 the probability of a subset of Ω on the left is equal to the integral of the probability density function on an interval $[a, b]$ on the right. The mapping X is a probability preserving transformation. In the diagram it is intuitive to regard the random variable X as a function that measures the level of the domain Ω .

For a measurable function $h : \mathbb{R} \rightarrow \mathbb{R}$ we have

$$\int_{\Omega} h(X(\omega)) \, d\mathbb{P}(\omega) = \int_{-\infty}^{\infty} h(x) \, d\mu_X(x) .$$

If we take $h(x) = |x|^p$, then

$$\mathbb{E}[|X|^p] = \int_{\omega} |X(\omega)|^p \, d\mathbb{P}(\omega) = \int_{-\infty}^{\infty} |x|^p \, d\mu_X(x) .$$

The *expectation* or *mean* of X is defined by

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \, d\mu_X(x) ,$$

and its *variance* is defined by

$$\sigma^2(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 d\mu_X(x).$$

The square root of the variance is called the *standard deviation*, and denoted by σ . If two random variables X and Y have the same distribution, in other words, if μ_X and μ_Y are equal on $\mathbb{R}$, then they have the same probabilistic properties. For example, their expectations are equal.

Example 4.2 (*Coin Tossing*) To model coin tossing we take $\Omega = \{H, T\}$ and define a σ -algebra $\mathcal{F} = \{\emptyset, \Omega, \{H\}, \{T\}\}$, and a probability measure $\mathbb{P}$ by $\mathbb{P}(\{H\}) = p$, $\mathbb{P}(\{T\}) = 1 - p$, where the probability of obtaining heads and tails are equal to p and $1 - p$, respectively. Define a random variable $X : \Omega \rightarrow \mathbb{R}$ by $X(H) = 0$, $X(T) = 1$. Then

$$\mathbb{E}[X] = \mathbb{E}[X^2] = p \times 0 + (1 - p) \times 1 = 1 - p$$

and

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1 - p - (1 - p)^2 = p - p^2.$$

Example 4.3 (*Uniform Distribution*) If the values of a random variable $X : \Omega \rightarrow \mathbb{R}$ are uniformly distributed in the interval $[a, b]$, then the probability density function of X is given by

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases}$$

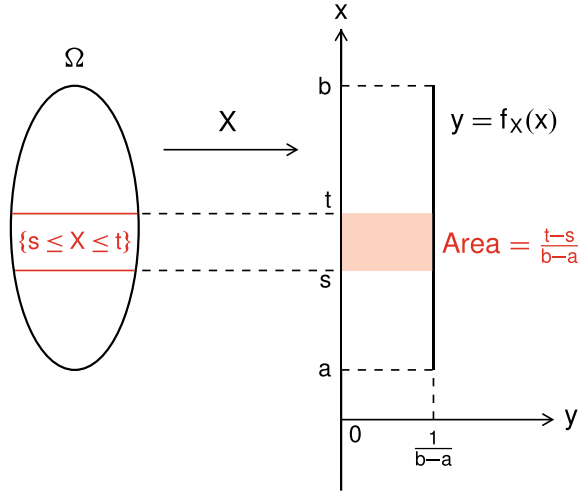
See Fig. 4.2. For $a = 0$ and $b = 1$, we have $\mathbb{E}[X] = \int_0^1 x dx = \frac{1}{2}$ and $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \int_0^1 x^2 dx - (\frac{1}{2})^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$.

Example 4.4 (*Siegel's Paradox*) We consider a seemingly contradictory fact on the exchange rate between two currencies. Let X represent the future exchange rate between US dollar and euro, i.e., if a dollar will be equal to X euro in the future then Jensen's inequality in Theorem 3.9 implies

$$\frac{1}{\mathbb{E}[X]} \leq \mathbb{E}\left[\frac{1}{X}\right],$$

where $1/X$ represents the exchange rate for converting one euro into dollars. Note that $1/\mathbb{E}[X] \neq \mathbb{E}[1/X]$ in general. For a concrete example, consider the case when the present exchange rate 1.00 (EUR/USD) changes to either $X = 2.00$ (EUR/USD) with probability $\frac{1}{2}$ or $X = 0.50$ (EUR/USD) with probability $\frac{1}{2}$ in a year. The opposite

Fig. 4.2 A model of a random variable X whose values are uniformly distributed in $[a, b]$



exchange rate $1/X$ is currently 1.00 (USD/EUR) and in a year it will change to 0.50 (USD/EUR) or 2.00 (USD/EUR) with probability $\frac{1}{2}$. Hence

$$\mathbb{E}[X] = \frac{1}{2} \times 2.00 + \frac{1}{2} \times 0.50 = 1.25 \text{ (EUR/USD) ,}$$

$$\mathbb{E}\left[\frac{1}{X}\right] = \frac{1}{2} \times 0.50 + \frac{1}{2} \times 2.00 = 1.25 \text{ (USD/EUR) .}$$

The exchange rate will rise 1.25 times in a year whether we exchange euro with USD or do the opposite. In other words, we have the same expected return in any case, which is called Siegel's paradox. It is not a paradox, but seems to be counterintuitive at first glance. Of course, if the exchange rate at future date is determined, one side gains profit while the other side suffers loss. For more information, see Kritzman (2000).

Example 4.5 (*Payoff of an Option*) Figure 4.3 illustrates the payoff $(S_T - K)^+$ of a European call option as a composite function of a random variable $S_T : \Omega \rightarrow [0, \infty)$ representing the stock price at time T and a function

$$(x - K)^+ = \max\{x - K, 0\} : [0, \infty) \rightarrow [0, \infty) .$$

The new composite function is again a random variable.

Example 4.6 If X is a random variable with normal probability distribution with mean 0 and variance σ^2 , then

$$\mathbb{E}[X^{2k+1}] = \int_{-\infty}^{\infty} x^{2k+1} \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)} dx = 0$$

Fig. 4.3 Payoff of a European call option, $(S_T - K)^+$, as a composite function

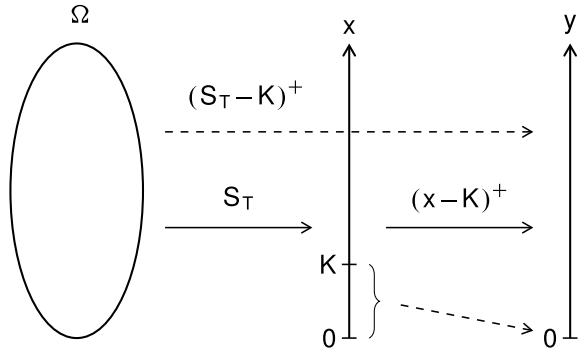
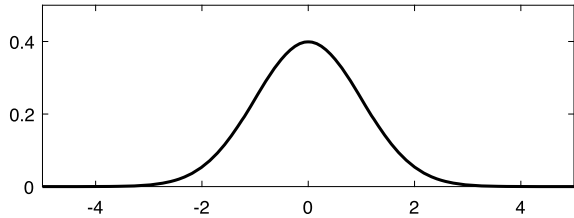


Fig. 4.4 Probability density function of the standard normal distribution



for every $k \geq 0$ since the integrand is an odd function. (See also Example 4.9.)

Example 4.7 For $\sigma = 1$ in Example 4.6 we have the standard normal distribution with its probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

(See Fig. 4.4.) Let $a_0 = 1$ and $a_n = \mathbb{E}[Z^{2n}]$, $n \geq 1$. Then

$$\begin{aligned} a_{n+1} &= \mathbb{E}[Z^{2(n+1)}] \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2(n+1)} e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} z e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \left\{ \left[-z^{2n+1} e^{-z^2/2} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (2n+1) z^{2n} (-e^{-z^2/2}) dz \right\} \\ &= (2n+1) a_n. \end{aligned}$$

Hence $a_n = 1 \times 3 \times \cdots \times (2n-1)$, $n \geq 1$. For example, $\mathbb{E}[Z^2] = 1$, $\mathbb{E}[Z^4] = 3$ and $\mathbb{E}[Z^6] = 15$. (See also Example 4.9. For a numerical experiment see Appendix A.8.)

Remark 4.1 Let $t > 0$ and $f(x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right)$. Then $\int_{-\infty}^{\infty} x^4 f(x) dx = 3t^2$.

Definition 4.2 (*The Error Function*) For $-\infty < x < \infty$ define the *error function* by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du .$$

Note that $\lim_{x \rightarrow \infty} \operatorname{erf}(x) = 1$. See also (E.2) in Appendix E.

4.2 Characteristic Functions

Definition 4.3 (*Cumulative distribution function*) Given a probability measure space $(\Omega, \mathbb{P})$ and a random variable $X : \Omega \rightarrow \mathbb{R}^1$, let $\Pr(X \in A)$ denote the probability that the value of X belongs to A for an interval $A \subset \mathbb{R}^1$. For example, $\Pr(X \leq x) = \Pr(X \in (-\infty, x])$. The *cumulative distribution function* (cdf for short) of X is defined by

$$F(x) = \Pr(X \leq x) = \mathbb{P}(\{X \leq x\}) , \quad -\infty < x < \infty .$$

If a cumulative distribution function satisfies

$$F(x) = \int_{-\infty}^x f(t) dt , \quad -\infty < x < \infty ,$$

for some integrable function $f \geq 0$, then $f(x)$ is called a *probability density function* (pdf for short) for X .

If a probability density function $f(x)$ is continuous at x , then $F'(x) = f(x)$ since

$$\mathbb{P}(\{X < x + \delta x\}) - \mathbb{P}(\{X < x\}) \approx f_X(x) \delta x$$

for small $\delta x > 0$. A cumulative distribution function is monotonically increasing, and satisfies $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$. A cumulative distribution function F is continuous from the right, i.e., $F(x) = \lim_{h \rightarrow 0+} F(x + h)$. Furthermore, we have $\mathbb{P}(X < c) = \lim_{x \rightarrow c-} F(x)$.

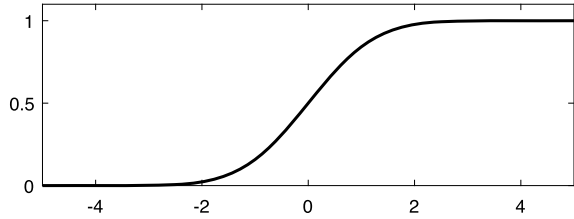
Example 4.8 The probability density function $f(x)$ of the standard normal distribution is given by

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

and a standard normal random variable X has the cumulative distribution function

$$N(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt .$$

Fig. 4.5 Cumulative distribution function of the standard normal distribution



(See Fig. 4.5.)

Definition 4.4 (*Fourier Transform*) For an integrable function $f(x)$ on $\mathbb{R}$, its *Fourier transform* $\hat{f}$ is defined by $\hat{f}(t) = \int_{-\infty}^{\infty} f(x)e^{-itx} dx$, $t \in \mathbb{R}$, where $i^2 = -1$. It is continuous on the real line.

Definition 4.5 (*Characteristic Function*) The *characteristic function* ϕ_X of a random variable X is defined by

$$\phi_X(t) = \mathbb{E}[e^{itX}], \quad -\infty < t < \infty,$$

where $i = \sqrt{-1}$. Note that X need not be continuous for the expectation to exist. For example, the characteristic function of a binomial distribution $B(n, k)$ defined by $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ is given by $\phi(t) = (pe^{it} + (1-p))^n$. For a continuous random variable X with its pdf f_X the characteristic function is the Fourier transform of f_X , i.e.,

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx, \quad -\infty < t < \infty.$$

(In probability theory it is more common to use e^{itx} instead of e^{-itx} .)

Theorem 4.1 For $t \in \mathbb{R}$ the following facts are known:

- (i) $\phi_X(0) = 1$,
- (ii) $|\phi_X(t)| \leq 1$,
- (iii) $\phi_X(t)$ is continuous,
- (iv) $\phi_{a+bX}(t) = e^{iat} \phi_X(bt)$,
- (v) $\phi'_X(0) = i\mu$,
- (vi) $\phi''_X(0) = -\mathbb{E}[X^2] = -\mu^2 - \sigma^2$, and
- (vii) the Fourier inversion formula holds, i.e.,

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \phi_X(t) dt.$$

Remark 4.2 (i) Since we have a formal power series expansion

$$e^{itX} = \sum_{n=0}^{\infty} \frac{1}{n!} i^n t^n X^n ,$$

the power series expansion of the characteristic function is given by

$$\phi_X(t) = \mathbb{E}[e^{itX}] = \sum_{n=0}^{\infty} \frac{i^n \mathbb{E}[X^n]}{n!} t^n$$

under the assumption that the series converges and the equality makes sense.

- (ii) The characteristic function of a random variable completely determines the distribution of X . More precisely, if X and Y have the same characteristic function, then their probability distributions are identical due to the Fourier inversion formula.

Example 4.9 If $X = Z$ is normally distributed with mean 0 and variance 1, then by the result in Example 4.6 the characteristic function is given by

$$\phi_Z(t) = \sum_{n=0}^{\infty} \frac{i^{2n} \mathbb{E}[Z^{2n}]}{(2n)!} t^{2n} = \sum_{n=0}^{\infty} \frac{(-1)^n \mathbb{E}[Z^{2n}]}{(2n)!} t^{2n} . \quad (4.1)$$

On the other hand,

$$\phi_Z(t) = \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{-t^2/2} = \sum_{n=0}^{\infty} \frac{1}{n!} \frac{(-1)^n}{2^n} t^{2n} , \quad (4.2)$$

where the second equality is given by the result in Exercise 4.8. By comparing the coefficients of the right-hand sides of (4.1) and (4.2), we obtain

$$\mathbb{E}[Z^{2n}] = \frac{(2n)!}{n! 2^n} = 1 \times 3 \times 5 \times \cdots \times (2n-1) .$$

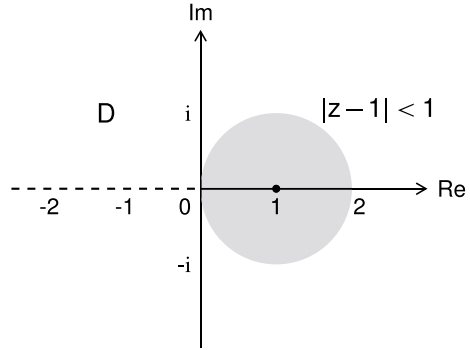
By taking complex conjugates in deriving (4.2), we also have

$$\int_{-\infty}^{\infty} e^{-itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = e^{-t^2/2} ,$$

which implies that the Fourier transform of the standard normal density function is itself up to constant multiple. If a normal variable Y has mean μ and variance σ^2 , then $\phi_Y(t) = e^{it\mu} e^{-\sigma^2 t^2/2}$. (See Example 4.22.)

Remark 4.3 Let X_n , $n \geq 1$, and X be random variables with their cumulative distribution functions F_{X_n} and F_X satisfying

Fig. 4.6 A disk $|z - 1| < 1$ for the power series expansion of $\log z$



$$\lim_{n \rightarrow \infty} \phi_{X_n}(t) = \phi_X(t), \quad -\infty < t < \infty.$$

Then $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$ at the points x where F_X is continuous.

Remark 4.4 For $z \in D = \mathbb{C} \setminus \{z \in \mathbb{C} : z \text{ is real and } \operatorname{Re} z \leq 0\}$, define

$$\log z = \log |z| + i \operatorname{Arg}(z), \quad -\pi < \operatorname{Arg}(z) < \pi.$$

(See Fig. 4.6.) Then for $|z - 1| < 1$ we have the Taylor series expansion

$$\log z = (z - 1) - \frac{1}{2}(z - 1)^2 + \dots.$$

Since

$$\begin{aligned} \phi_X(t) &= \phi_X(0) + \phi'_X(0)t + \frac{\phi''_X(0)}{2}t^2 + \dots \\ &= 1 + i\mu t - \frac{\mu^2 + \sigma^2}{2}t^2 + \dots, \end{aligned}$$

we have

$$\begin{aligned} \log \phi_X(t) &= \left(i\mu t - \frac{\mu^2 + \sigma^2}{2}t^2 + \dots \right) - \frac{1}{2} \left(i\mu t - \frac{\mu^2 + \sigma^2}{2}t^2 + \dots \right)^2 + \dots \\ &= i\mu t - \frac{\sigma^2}{2}t^2 + \dots, \end{aligned}$$

and hence

$$\frac{\log \phi_X(t) - i\mu t}{t^2} = -\frac{\sigma^2}{2} + \dots.$$

Theorem 4.2 *If $X_n, n \geq 1$, are normal random variables with mean μ_n and variance σ_n^2 and if X_n converges to X in L^2 , then X is also normally distributed with mean $\mu = \lim_{n \rightarrow \infty} \mu_n$ and variance $\sigma^2 = \lim_{n \rightarrow \infty} \sigma_n^2$ under the assumption that the limits exist.*

Proof Since the characteristic functions of X_n converge everywhere to a limit that is also a characteristic function of a normal variable, the sequence X_n itself converges to a normal variable. $\square$

Definition 4.6 (*Moment Generating Function*) The *moment generating function* of a random variable X is defined for u in a neighborhood of 0 by

$$M_X(u) = \mathbb{E}[e^{uX}]$$

if the expectation exists. Depending on X , the expectation may not exist for large real values of u . Note that $M_X(0) = 1$ and

$$M_X(u) = 1 + m_1 u + \frac{m_2}{2!} u^2 + \frac{m_3}{3!} u^3 + \cdots + \frac{m_k}{k!} u^k + \cdots$$

where $m_k = \mathbb{E}[X^k]$ and the power series converges. If the k th moment m_k grows too fast then the power series may not converge to a finite limit. If $X \geq 0$, the moment generating function is essentially the Laplace transform of the probability density function. Note that u need not be real. If u is purely imaginary, then we obtain the characteristic function in Definition 4.5.

4.3 Independent Random Variables

Definition 4.7 (*Independence of Events*) If measurable sets (or events) $A_1, \dots, A_n$ satisfy

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \times \cdots \times \mathbb{P}(A_n),$$

then $A_1, \dots, A_n$ are said to be *independent*.

If two events A_1 and A_2 are independent, then the pairs A_1 and A_2^c , A_1^c and A_2 , A_1^c and A_2^c are independent, where B^c denotes the complement of a set B .

Definition 4.8 (*Independence of σ -Algebras*) Let $\mathcal{F}_i, 1 \leq i \leq n$, be σ -algebras on a set Ω . If $A_1, \dots, A_n$ are independent for any choice of $A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n$, then we say that $\mathcal{F}_1, \dots, \mathcal{F}_n$ are *independent*.

Definition 4.9 (*Independence of Random Variables*) Let $X_i : \Omega \rightarrow \mathbb{R}$, $1 \leq i \leq n$, be random variables. If $\sigma(X_i)$ are independent sub- σ -algebras, then X_i are said to be *independent*. (See Definition 3.7.) Equivalently, if $X_i^{-1}(B_i)$, $1 \leq i \leq n$, are independent for arbitrary Borel measurable subsets $B_i \subset \mathbb{R}$, then X_i are independent. Random variables X_γ , $\gamma \in \Gamma$, are said to be independent if $X_{\gamma_1}, \dots, X_{\gamma_n}$ are independent for any $n \geq 1$ and $\gamma_1, \dots, \gamma_n \in \Gamma$.

Definition 4.10 (*Independence of a Random Variable and a σ -Algebra*) Given a random variable X and a sub- σ -algebra $\mathcal{F}$, if $\sigma(X)$ and $\mathcal{F}$ are independent, then X and $\mathcal{F}$ are said to be *independent*.

Lemma 4.1 *Suppose that random variables X and Y are independent. For arbitrary Borel measurable functions $f : \mathbb{R} \rightarrow \mathbb{R}$ and $g : \mathbb{R} \rightarrow \mathbb{R}$, the random variables $f(X)$ and $g(Y)$ are independent.*

Proof Use $\sigma(f(X)) \subset \sigma(X)$ and $\sigma(g(Y)) \subset \sigma(Y)$. □

Example 4.10 Suppose that random variables X and Y are independent.

- (i) For constants α and β we take $f(x) = \alpha x + \beta$. Then we see that $\alpha X + \beta$ and Y are independent.
- (ii) $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$ are independent.
- (iii) Take $f(x) = \max\{x, 0\}$. Then $X^+ = \max\{X, 0\}$ and Y are independent.

Theorem 4.3 *Suppose that $X_1, \dots, X_n$ are independent random variables.*

- (i) *For an arbitrary Borel measurable $B_i \subset \mathbb{R}$ we have*

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \times \dots \times \mathbb{P}(X_n \in B_n) .$$

- (ii) *If $-\infty < \mathbb{E}[X_i] < \infty$ for every i , then*

$$\mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \times \dots \times \mathbb{E}[X_n] .$$

- (iii) *If $\mathbb{E}[X_i^2] < \infty$ for every i , then*

$$\text{Var}(X_1 + \dots + X_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) ,$$

where Var denotes variance.

Proof For the sake of notational convenience we prove only the case $n = 2$.

- (i) Use the fact

$$\begin{aligned} \mathbb{P}(X_1 \in B_1, X_2 \in B_2) &= \mathbb{P}(X_1^{-1}(B_1) \cap X_2^{-1}(B_2)) \\ &= \mathbb{P}(X_1^{-1}(B_1)) \mathbb{P}(X_2^{-1}(B_2)) . \end{aligned}$$

(ii) In part (i) we take $X_1 = \sum_i \alpha_i \mathbb{1}_{A_i}$, $X_2 = \sum_i \beta_i \mathbb{1}_{B_i}$, and obtain

$$\mathbb{E}[X_1 X_2] = \sum_{i,j} \alpha_i \beta_j \mathbb{P}(A_i \cap B_j) = \sum_i \alpha_i \mathbb{P}(A_i) \sum_j \beta_j \mathbb{P}(B_j) = \mathbb{E}[X_1] \mathbb{E}[X_2] .$$

For the general case, approximate the given random variables by sequences of simple functions and take the limits.

(iii) Note that

$$\begin{aligned} & \text{Var}(X_1 + X_2) \\ &= \mathbb{E}[(X_1 - \mathbb{E}[X_1]) + (X_2 - \mathbb{E}[X_2])]^2 \\ &= \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] + \mathbb{E}[(X_2 - \mathbb{E}[X_2])^2] + 2\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] . \end{aligned}$$

By Example 4.10(ii) and part (ii) of the above theorem,

$$\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])] = \mathbb{E}[X_1 - \mathbb{E}[X_1]] \mathbb{E}[X_2 - \mathbb{E}[X_2]] = 0 .$$

Thus $\text{Var}(X_1 + X_2) = \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] + \mathbb{E}[(X_2 - \mathbb{E}[X_2])^2]$. □

Definition 4.11 (*Joint Probability*) For random variables X_i , $1 \leq i \leq n$, defined on a probability space $(\Omega, \mathbb{P})$, define a *random vector* $(X_1, \dots, X_n) : \Omega \rightarrow \mathbb{R}^n$ by $(X_1, \dots, X_n)(\omega) = (X_1(\omega), \dots, X_n(\omega))$, $\omega \in \Omega$. Its *joint probability density function* $f_{X_1, \dots, X_n} : \mathbb{R}^n \rightarrow \mathbb{R}$ is given by

$$\int \cdots \int_B f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n = \mathbb{P}((X_1, \dots, X_n)^{-1}(B)) ,$$

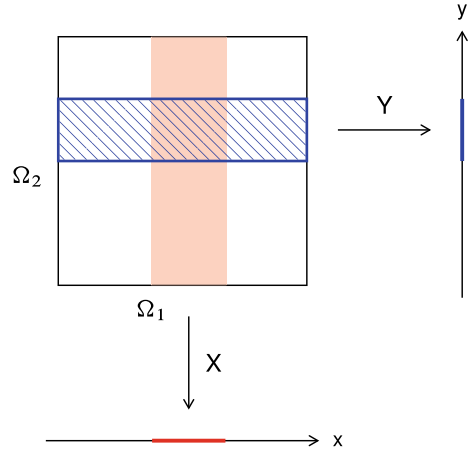
where $B \subset \mathbb{R}^n$ is an arbitrary Borel measurable subset. The *characteristic function* of a random vector $\mathbf{X} = (X_1, \dots, X_n)$ is defined by the n -dimensional Fourier transform

$$\phi_{\mathbf{X}}(\mathbf{t}) = \mathbb{E}[e^{i\mathbf{t} \cdot \mathbf{X}}] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i\mathbf{t} \cdot \mathbf{x}} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n ,$$

where $\mathbf{t} \cdot \mathbf{x} = t_1 x_1 + \cdots + t_n x_n$ and $\mathbf{t} \cdot \mathbf{X} = t_1 X_1 + \cdots + t_n X_n$.

Example 4.11 (*Product probability*) Let Z_i be a random variable on a probability space $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ for $i = 1, 2$. Consider the product measure space $\Omega_1 \times \Omega_2$ with the product probability measure $\mathbb{P}_1 \times \mathbb{P}_2$ and define $X, Y : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ by $X(\omega_1, \omega_2) = Z_1(\omega_1)$ and $Y(\omega_1, \omega_2) = Z_2(\omega_2)$. Then X and Y are independent. (See Fig. 4.7 where the product measure space is represented by a rectangle and X and Y are given by projections. Inverse images are shown as vertical and horizontal strips.)

Fig. 4.7 Independent random variables X and Y are viewed as projections



Definition 4.12 (*Jointly Normal Distribution*) The random variables $X_1, \dots, X_n$ are *jointly normal* with a covariance matrix Σ if their joint pdf is given by

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{1/2}} \exp \left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right),$$

where

$$\begin{aligned} \Sigma &= \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})^t (\mathbf{X} - \boldsymbol{\mu})] \\ &= \begin{bmatrix} \text{Cov}(X_1, X_1) & \cdots & \text{Cov}(X_1, X_n) \\ \vdots & & \vdots \\ \text{Cov}(X_n, X_1) & \cdots & \text{Cov}(X_n, X_n) \end{bmatrix} \end{aligned}$$

and

$$\mathbf{x} = (x_1, \dots, x_n), \quad \boldsymbol{\mu} = (\mu_1, \dots, \mu_n) \quad \text{and} \quad \mathbf{X} = (X_1, \dots, X_n).$$

In this case, we write $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$. Note that $\mathbb{E}[X_i] = \mu_i$, $1 \leq i \leq n$, and that Σ is symmetric and nonnegative definite.

Remark 4.5 (i) Here is a generalization of the jointly normal distribution. The random variables $X_1, \dots, X_n$ are said to be jointly normal in the extended sense if there exists a (not necessarily invertible) symmetric and nonnegative definite $n \times n$ matrix $C = (c_{jk})$ and a vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ satisfying

$$\phi_{\mathbf{X}}(x_1, \dots, x_n) = \exp \left(-\frac{1}{2} \sum_{j,k=1}^n c_{jk} x_j x_k + i \sum_{j=1}^n \mu_j x_j \right). \quad (4.3)$$

This condition (4.3) is satisfied if we choose $C = \Sigma^{-1}$ in Definition 4.12. It is equivalent to the statement that any linear combination of $X_1, \dots, X_n$ is normal. When $X_1, \dots, X_n$ are jointly normal, and X_1 and X_j are uncorrelated for $2 \leq j \leq n$, then X_1 is independent of $X_2, \dots, X_n$.

- (ii) Suppose that X_n is normal for $n \geq 1$ and that X_n converges to X in L^2 . Then X is normal. For the proof, consult Oksendal (2003).

Theorem 4.4 (Independence) *Let X and Y be random variables with their pdf's f_X and f_Y , respectively. Let $f_{X,Y}$ denote the joint pdf of the random vector (X, Y) defined in Definition 4.11. Then $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ if and only if X and Y are independent.*

Proof First, if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, then

$$X^{-1}(B_1) \cap Y^{-1}(B_2) = (X, Y)^{-1}(B_1 \times B_2)$$

for any Borel subsets B_1, B_2 . Hence

$$\begin{aligned} \mathbb{P}(X^{-1}(B_1) \cap Y^{-1}(B_2)) &= \iint_{B_1 \times B_2} f_{X,Y}(x, y) \, dx \, dy \\ &= \int_{B_2} \left(\int_{B_1} f_X(x) \, dx \right) f_Y(y) \, dy \\ &= \mathbb{P}(X^{-1}(B_1)) \mathbb{P}(Y^{-1}(B_2)) . \end{aligned}$$

Next, suppose that X and Y are independent. For any pair of real numbers a, b , let $A = (-\infty, a) \times (-\infty, b)$. Then, by definition,

$$\mathbb{P}((X, Y) \in A) = \iint_A f_{X,Y}(x, y) \, dx \, dy .$$

By the independence,

$$\begin{aligned} \mathbb{P}((X, Y) \in A) &= \mathbb{P}(X \in (-\infty, a)) \mathbb{P}(Y \in (-\infty, b)) \\ &= \int_{-\infty}^a f_X(x) \, dx \int_{-\infty}^b f_Y(y) \, dy \\ &= \int_{-\infty}^a \int_{-\infty}^b f_X(x) f_Y(y) \, dx \, dy . \end{aligned}$$

Hence $f_{X,Y}(x, y) = f_X(x)f_Y(y)$. □

Definition 4.13 (Convolution) *Let f and g be integrable functions on $\mathbb{R}$. Then their convolution $f * g$ is defined by*

$$(f * g)(x) = \int_{-\infty}^{\infty} f(x - y)g(y) \, dy .$$

It can be shown that $f * g = g * f$. Furthermore, for integrable functions f, g, h we have $f * (g * h) = (f * g) * h$.

Theorem 4.5 (Sum of Random Variables) *Let X and Y be independent random variables with probability density functions f_X and f_Y , respectively. Then the probability density function of $X + Y$, denoted by f_{X+Y} , is given by the convolution, i.e.,*

$$f_{X+Y} = f_X * f_Y .$$

Proof First, we present a heuristic proof. Let $F_X(x) = \mathbb{P}(X \leq x)$ and let $-\infty < \cdots < y_i < y_{i+1} < \cdots < \infty$ be a partition of $(-\infty, \infty)$ such that $\sup_i \delta_i$ is sufficiently small, where $\delta_i = y_{i+1} - y_i > 0$. Then

$$\begin{aligned} \mathbb{P}(X + Y \leq a) &\approx \sum_i \mathbb{P}(X \leq a - y_i | y_i \leq Y < y_{i+1}) \times \mathbb{P}(y_i \leq Y < y_{i+1}) \\ &= \sum_i \mathbb{P}(X \leq a - y_i) \times \mathbb{P}(y_i \leq Y < y_{i+1}) \\ &\approx \sum_i F_X(a - y_i) f_Y(y_i) \delta y_i . \end{aligned}$$

(See Fig. 4.8.) By letting $\sup_i \delta_i \rightarrow 0$, we obtain

$$\mathbb{P}(X + Y \leq a) = \int_{-\infty}^{\infty} F_X(a - y) f_Y(y) dy .$$

Taking the derivatives of both sides with respect to a , we have

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a - y) f_Y(y) dy = f_X * f_Y(a) .$$

We used the fact that F_X is sufficiently smooth.

A more mathematically rigorous proof goes as follows: Since the joint probability density function of (X, Y) is equal to $f_X(x) f_Y(y)$, we have

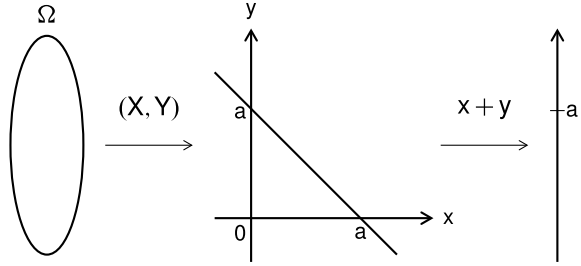
$$\mathbb{P}(X + Y \leq a) = \int_{-\infty}^{\infty} \int_{-\infty}^{a-y} f_X(x) f_Y(y) dx dy .$$

Differentiating both sides with respect to a , we obtain

$$f_{X+Y}(a) = \int_{-\infty}^{\infty} f_X(a - y) f_Y(y) dy = f_X * f_Y(a) .$$

□

Fig. 4.8 Sum of two independent random variables



Remark 4.6 Let ϕ_X , ϕ_Y and ϕ_{X+Y} denote the characteristic functions of X , Y and $X + Y$, respectively. If X and Y are independent, then $\phi_{X+Y}(t) = \phi_X(t) \phi_Y(t)$. For, if we let f_X , f_Y and f_{X+Y} denote the probability density functions of X , Y and $X + Y$, respectively, then

$$\phi_{X+Y}(t) = \widehat{f_{X+Y}} = \widehat{f_X * f_Y} = \widehat{f_X} \widehat{f_Y} = \phi_X(t) \phi_Y(t)$$

by Theorem 4.5. This can be shown directly since

$$\phi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}] \mathbb{E}[e^{itY}] = \phi_X(t) \phi_Y(t) .$$

Example 4.12 Let U_1 , U_2 be independent and uniformly distributed in $[0, 1]$. Then the pdf f_V of $V = U_1 + U_2$ is given by

$$f_V(x) = \begin{cases} x, & x \in [0, 1] , \\ 2 - x, & x \in (1, 2] , \\ 0, & x \notin [0, 2] . \end{cases}$$

We can prove this fact by computing the convolution $f * f$, where $f(x) = 1$ on $[0, 1]$ and $f(x) = 0$ elsewhere.

Example 4.13 Let X_1 , X_2 be independent normally distributed with means μ_1 and μ_2 and variances σ_1 and σ_2 , respectively. Then $X_1 + X_2$ is also normally distributed with mean $\mu_1 + \mu_2$ and variance $\sigma_1 + \sigma_2$. To prove this it suffices to compute the convolution of two normal density functions f_{X_1} and f_{X_2} . For notational simplicity, consider the case $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$. Then

$$\begin{aligned} (f_{X_1} * f_{X_2})(z) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-(z-t)^2/2} e^{-t^2/2} dt \\ &= \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{\infty} e^{-(t-z/2)^2} dt \\ &= \frac{1}{2\pi} e^{-z^2/4} \int_{-\infty}^{\infty} e^{-t^2} dt \\ &= \frac{1}{2\pi} e^{-z^2/4} \sqrt{\pi} = \frac{1}{\sqrt{2\pi} \times 2} e^{-z^2/(2 \times 2)} , \end{aligned}$$

which implies that $X_1 + X_2$ is also normal with mean 0 and variance 2. In general, if X_i are independent normal variables with mean μ_i and variance σ_i^2 for $1 \leq i \leq n$, then $S_n = X_1 + \cdots + X_n$ is normally distributed with mean $\mu_1 + \cdots + \mu_n$ and variance $\sigma_1^2 + \cdots + \sigma_n^2$.

Remark 4.7 If X and Y are independent, then the characteristic function of a random vector (X, Y) satisfies

$$\phi_{X,Y}(s, t) = \mathbb{E} [e^{i(sX+tY)}] = \mathbb{E} [e^{isX} e^{itY}] = \mathbb{E} [e^{isX}] \mathbb{E} [e^{itY}] = \phi_X(s) \phi_Y(t) .$$

To show that the converse also holds, we assume that $\phi_{X,Y}(s, t) = \phi_X(s) \phi_Y(t)$. Let $X \times Y : \Omega \times \Omega \rightarrow \mathbb{R}^2$ be the direct product defined by $X \times Y(\omega_1, \omega_2) = (X(\omega_1), Y(\omega_2))$, then

$$\begin{aligned} \phi_{X \times Y}(s, t) &= \mathbb{E} [e^{i(sX+tY)}] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{i(sx+ty)} f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} e^{isx} f_X(x) dx \int_{-\infty}^{\infty} e^{ity} f_Y(y) dy \quad (\text{Fubini's theorem}) \\ &= \phi_X(s) \phi_Y(t) . \end{aligned}$$

(For notational simplicity we considered the case that X and Y are continuous and have probability density functions.) Hence the joint probability distribution of X and Y satisfies $df_{X,Y}(x, y) = df_X(x) df_Y(y)$, and X and Y are independent.

4.4 Change of Variables

Theorem 4.6 (Composition) *If $h(x)$ is a differentiable and monotone function, and if a random variable X has a continuous probability density function $f_X(x)$, then the probability density function of $Y = h(X)$ is given by*

$$f_Y(y) = \frac{1}{|h'(h^{-1}(y))|} f_X(h^{-1}(y)) ,$$

where h' denotes the derivative of h . (See Fig. 4.9.)

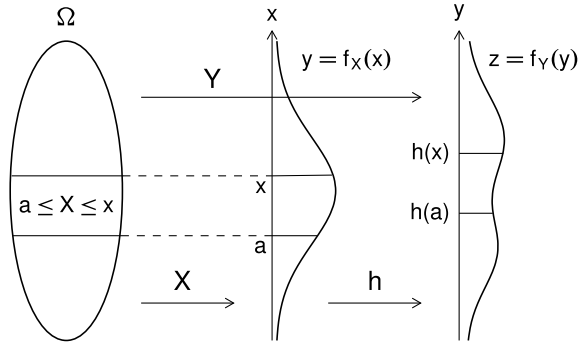
Proof First, consider the case when h is monotonically increasing. Since

$$\Pr(a \leq X \leq x) = \Pr(h(a) \leq Y \leq h(x)) ,$$

we have

$$\int_a^x f_X(x) dx = \int_{h(a)}^{h(x)} f_Y(y) dy .$$

Fig. 4.9 A new probability density function obtained by change of variables



By differentiating with respect to x , we have $f_X(x) = f_Y(h(x))h'(x)$. Now substitute $x = h^{-1}(y)$. For the case when h is decreasing, note that

$$\int_a^x f_X(x) dx = \int_{h(x)}^{h(a)} f_Y(y) dy = - \int_{h(a)}^{h(x)} f_Y(y) dy .$$

□

Example 4.14 If a random variable X has a probability density function $f_X(x)$, let us find the probability density function $f_Y(y)$ of $Y = aX + b$ where $a > 0$. Note that for arbitrary y we have

$$\int_{-\infty}^y f_Y(y) dy = \mathbb{P}(aX + b \leq y) = \mathbb{P}\left(X \leq \frac{y-b}{a}\right) = \int_{-\infty}^{(y-b)/a} f_X(x) dx .$$

By differentiating with respect to y , we obtain $f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$. For a standard normal variable X , the pdf of $Y = aX + b$ is given by

$$f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) .$$

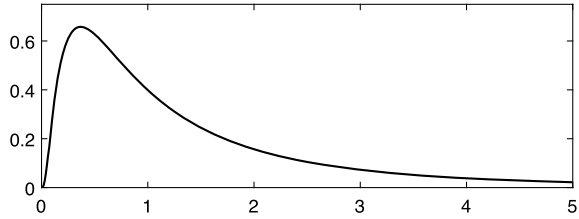
Example 4.15 Let U be uniformly distributed in $[0, 1]$, i.e.,

$$f_U(x) = \begin{cases} 1, & u \in (0, 1] , \\ 0, & u \notin (0, 1] . \end{cases}$$

Consider $Y = \sqrt{U}$, i.e., $h(x) = \sqrt{x}$ in Theorem 4.6. Then $f_Y(y) = 2y$, $0 \leq y \leq 1$. (See also Exercise 4.4.)

Example 4.16 Let X be uniformly distributed in $(0, 1]$, and define $Y = -\log X$ by choosing $h(x) = -\ln x$. Since $h'(x) = -\frac{1}{x}$ and $x = e^{-y}$,

Fig. 4.10 The probability density function of $\exp X$ when $X \sim N(0, 1)$



$$f_Y(y) = e^{-y}, \quad y \geq 0.$$

Example 4.17 (*Lognormal distribution*) If X is normally distributed with mean μ and variance σ^2 , then $Y = e^X$ has the probability density function given by

$$f_Y(y) = \frac{1}{\sigma y \sqrt{2\pi}} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right), \quad y > 0.$$

It can be shown that $\mathbb{E}[Y] = e^{\mu + \frac{1}{2}\sigma^2}$ and $\text{Var}(Y) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$. (For the proof see Exercise 4.7.) See Fig. 4.10 for the graph for $\mu = 0, \sigma = 1$.

Example 4.18 (*Geometric Brownian motion*) This is a special case of Example 4.17. Suppose that an asset price S_t at time $t \geq 0$ is given by

$$S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t},$$

where a Brownian motion $W_t, t \geq 0$, is normally distributed with mean 0 and variance t . (See Chap. 7 for the definition of a Brownian motion.) Put $Y = S_t$. Since $\log Y = \log S_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t$, we see that $X = \log Y$ is normally distributed with mean $\log S_0 + (\mu - \frac{1}{2}\sigma^2)t$ and variance $\sigma^2 t$. Hence the pdf of $Y = e^X$ is given by

$$f_Y(y) = \frac{1}{\sigma y \sqrt{2\pi t}} \exp\left(-\frac{(\log y - \log S_0 - (\mu - \frac{1}{2}\sigma^2)t)^2}{2\sigma^2 t}\right),$$

and hence $\mathbb{E}[S_t] = e^{\log S_0 + (\mu - \frac{1}{2}\sigma^2)t + \frac{1}{2}\sigma^2 t} = S_0 e^{\mu t}$, $\text{Var}(S_t) = S_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1)$ by Example 4.17. This result will be used in Sect. 16.2.

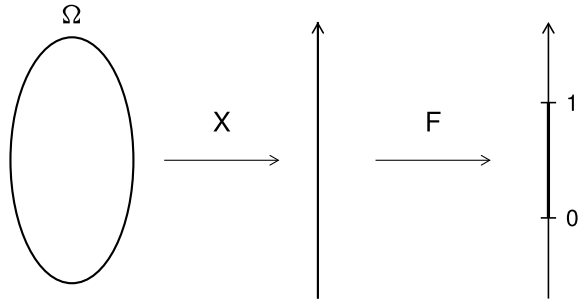
Remark 4.8 1. If X has a pdf $f_X(x)$, and if $y = h(x)$ is differentiable but not necessarily one-to-one, then $Y = h(X)$ has its pdf given by

$$f_Y(y) = \sum_{h(z)=y} \frac{1}{|h'(z)|} f_X(z),$$

where the summation is over z satisfying $h(z) = y$.

2. For the two-dimensional version, see Exercise 4.15.

Fig. 4.11 The composite random variable $F(X)$



Example 4.19 (*Square of the standard normal variable*) Suppose that X is the standard normal variable, i.e., $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Let us find the pdf of $Y = X^2$ on $[0, \infty)$. Since $h(x) = x^2$, we have $h^{-1}(y) = \pm\sqrt{y}$ for $y \geq 0$ and

$$f_Y(y) = \frac{1}{|-2\sqrt{y}|} f_X(-\sqrt{y}) + \frac{1}{|2\sqrt{y}|} f_X(\sqrt{y}) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}.$$

For $y < 0$, clearly $f_Y(y) = 0$.

Theorem 4.7 (Uniform Distribution) *Let X be a random variable with its cumulative distribution function F . Assume that F is invertible. Then $F(X)$ is uniformly distributed in $[0, 1]$. (See Fig. 4.11.)*

Proof Since $\Pr(F(X) \leq x) = \Pr(X \leq F^{-1}(x)) = F(F^{-1}(x)) = x$, we have $\Pr(a \leq F(X) \leq b) = \Pr(F(X) \leq b) - \Pr(F(X) \leq a) = b - a$. $\square$

For computer experiments, see Simulations 4.1 and 4.2. See also Simulation 4.3.

4.5 The Law of Large Numbers

In this section the symbol μ denotes an average not a measure.

Theorem 4.8 (Weak Law of Large Numbers) *Let X_n , $n \geq 1$, be a sequence of independent and identically distributed random variables with $\mathbb{E}[X_n] = \mu$ and $\text{Var}(X_n) = \sigma^2$. Put $S_n = X_1 + \cdots + X_n$. Then $\frac{1}{n}S_n$ converges to μ in probability as $n \rightarrow \infty$. In other words, for every $\varepsilon > 0$ we have*

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0.$$

Proof Since $\frac{S_n}{n} - \mu = \frac{1}{n}(X_1 + \cdots + X_n) - \mu$, the characteristic function is equal to $e^{-i\mu t} (\phi_{X_1}(\frac{t}{n}))^n$, and its logarithm is equal to $t(\log \phi_{X_1}(\frac{t}{n}) - i\mu \frac{t}{n})/(\frac{t}{n})$, which

converges to 0 as $n \rightarrow \infty$. Hence the characteristic function of $\frac{S_n}{n} - \mu$ converges to 1, which is the characteristic function of the constant random variable $X = 0$. The cdf of X is given by $F_X(x) = 1, x \geq 0$, and $F_X(x) = 0, x < 0$. Choose $\varepsilon > 0$. By continuity, we have $\lim_{n \rightarrow \infty} \Pr\left(\frac{S_n}{n} - \mu \leq -\varepsilon\right) = F_X(-\varepsilon) = 0$ and $\lim_{n \rightarrow \infty} \Pr\left(\frac{S_n}{n} - \mu \leq \varepsilon\right) = F_X(\varepsilon) = 1$. $\square$

Remark 4.9 With an additional condition that X_n are square-integrable, the proof of Theorem 4.8 is simple. Using Chebyshev's inequality for $p = 2$, we obtain

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{j=1}^n (X_j - \mu)\right| > \varepsilon\right) \leq \frac{1}{n^2 \varepsilon^2} \sum_{j=1}^n \mathbb{E}[(X_j - \mu)^2] = \frac{\sigma^2}{n \varepsilon^2}$$

and let $n \rightarrow \infty$, completing the proof.

Theorem 4.9 (Strong Law of Large Numbers) *Let $X_n, n \geq 1$, be a sequence of independent and identically distributed random variables with $\mathbb{E}[X_n] = \mu$. Put $S_n = X_1 + \dots + X_n$. Then $\frac{1}{n} S_n$ converges to μ almost surely as $n \rightarrow \infty$.*

Proof Consult Folland (1999). $\square$

Let $A_1, A_2, A_3, \dots$ be a sequence of events in a probability space $(\Omega, \mathbb{P})$. Let $\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$. Note that $\omega \in \limsup A_n$ if and only if $\omega \in A_n$ for infinitely many n .

Theorem 4.10 (Borel–Cantelli Lemma) *Let $A_n, n \geq 1$, be a sequence of events.*

- (i) *If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\limsup A_n) = 0$.*
- (ii) *If A_n are independent and if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(\limsup A_n) = 1$.*

Proof (i) For each $k \geq 1$, we have

$$\mathbb{P}(\limsup A_n) \leq \mathbb{P}\left(\bigcup_{n \geq k} A_n\right) \leq \sum_{n=k}^{\infty} \mathbb{P}(A_n) \rightarrow 0.$$

- (ii) If $\mathbb{P}(A_n) = 1$ for infinitely many n , then it is obvious that the statement is true. Assume that $\mathbb{P}(A_n) < 1$ for every $n \geq N$ for some N . Note that $\prod_n (1 - a_n) = 0$ if and only if $\sum_n a_n = \infty$ where $0 \leq a_n < 1$ for every n . For each k , we have

$$\mathbb{P}\left(\Omega \setminus \bigcup_{n=k}^{\infty} A_n\right) = \mathbb{P}\left(\bigcap_{n=k}^{\infty} (\Omega \setminus A_n)\right) = \prod_{n=k}^{\infty} (1 - \mathbb{P}(A_n)) = 0.$$

We use the independence of the complements of $A_1, A_2, A_3, \dots$ for the second equality. Hence $\mathbb{P}(\bigcup_{n=k}^{\infty} A_n) = 1$ for every k . $\square$

For example, if a fair coin is tossed n times, then the probability of the event A_n that heads comes up every time is 2^{-n} . Since $\sum_1^{\infty} \mathbb{P}(A_n) < \infty$, we conclude that tails will eventually come up with probability 1.

4.6 The Central Limit Theorem

Definition 4.14 (*Convergence in Distribution*) A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ converges in distribution to a random variable X if

$$\lim_{n \rightarrow \infty} \mathbb{P}\{X_n \leq x\} = \mathbb{P}\{X \leq x\}$$

at all points where the cumulative distribution function of X is continuous. If X_n converges to X in distribution, we write $X_n \xrightarrow{D} X$. It is known that if $\{X_n\}_{n=1}^{\infty}$ converges almost surely to X as $n \rightarrow \infty$, then $X_n \xrightarrow{D} X$.

Theorem 4.11 (Central Limit Theorem) Let $\{X_n\}_{n=1}^{\infty}$ be independent, identically distributed L^2 -random variables with $\mathbb{E}[X_n] = \mu$ and $\text{Var}(X_n) = \sigma^2$ for $n \geq 1$. Put $S_n = X_1 + \cdots + X_n$. Then

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = N(x), \quad -\infty < x < \infty$$

where $N(x)$ is the cumulative distribution function for the standard normal distribution defined in Example 4.8. In other words, $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ converges in distribution to a standard normal variable.

Proof Here is a sketch of the proof. Put $S_n^* = (S_n - n\mu)/(\sigma\sqrt{n})$. Then

$$\phi_{S_n^*}(t) = \exp\left(\frac{-in\mu t}{\sigma\sqrt{n}}\right) \phi_{S_n}\left(\frac{t}{\sigma\sqrt{n}}\right) = \exp\left(\frac{-in\mu t}{\sigma\sqrt{n}}\right) \left(\phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.$$

The logarithm of the leftmost side is

$$n \left(\log \phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right) - i\mu \frac{t}{\sigma\sqrt{n}} \right) = \frac{t^2}{\sigma^2} \frac{\log \phi_{X_1}\left(\frac{t}{\sigma\sqrt{n}}\right) - i\mu \frac{t}{\sigma\sqrt{n}}}{\left(\frac{t}{\sigma\sqrt{n}}\right)^2},$$

which converges to $\frac{t^2}{\sigma^2} \left(-\frac{\sigma^2}{2}\right) = -\frac{t^2}{2}$ as in Remark 4.4. Hence $\lim_{n \rightarrow \infty} \phi_{S_n^*} = e^{-t^2/2}$, which is the characteristic function of the standard normal distribution. (Consult Example 4.9.) $\square$

The Central Limit Theorem states that if we take n samples $X_1, \dots, X_n$ with average μ and standard deviation σ , not necessarily normally distributed, then S_n/n is approximately normally distributed with average μ and standard deviation $\sigma/\sqrt{n}$ for sufficiently large n . (See Fig. 4.15.) This idea is behind the Monte Carlo integration method for option pricing.

4.7 Statistical Ideas

Definition 4.15 (*Covariance*) The covariance between random variables X and Y is defined by $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$.

Theorem 4.12 For random variables X, Y, Z, X_i, Y_j the following facts hold:

- (i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$.
- (ii) $\text{Cov}(X, X) = \text{Var}(X)$.
- (iii) $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.
- (iv) $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$.
- (v) $\text{Cov}(\sum_i X_i, \sum_j Y_j) = \sum_i \sum_j \text{Cov}(X_i, Y_j)$.
- (vi) For a constant a , $\text{Cov}(X, aY) = a \text{Cov}(X, Y)$.
- (vii) For a constant b , $\text{Cov}(X, b) = 0$.
- (viii) If X and Y are independent then $\text{Cov}(X, Y) = 0$ by (iii).

Definition 4.16 (*Correlation*) The correlation coefficient $\rho(X, Y)$ for random variables X and Y with $\text{Var}(X) > 0$ and $\text{Var}(Y) > 0$ is defined by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

By the Cauchy–Schwarz inequality we have $-1 \leq \rho(X, Y) \leq 1$.

- Example 4.20**
- (i) Given a random variable X with $\text{Var}(X) > 0$, and constants $a \neq 0$ and b , we define $Y = aX + b$. Then $\rho(X, Y) = 1$ for $a > 0$ and $\rho(X, Y) = -1$ for $a < 0$ since $\text{Cov}(X, Y) = a\text{Cov}(X, X) + \text{Cov}(X, b) = a\text{Var}(X)$ and $\text{Var}(Y) = \text{Var}(aX + b) = \text{Var}(aX) = a^2\text{Var}(X)$.
 - (ii) Let U be a uniformly distributed random variable in $[0, 1]$. Then U and $1 - U$ are negatively correlated since $\text{Cov}(U, 1 - U) = \text{Cov}(U, 1) - \text{Cov}(U, U) = 0 - \text{Var}(U) = -\frac{1}{12}$ by Example 4.3.
 - (iii) If Z is a standard normal variable, then $\text{Cov}(Z, -Z) = -\text{Var}(Z) = -1$.

- Remark 4.10**
- (i) For $X_1, X_2 \in L^2(\Omega, \mathbb{P})$ on a probability space $(\Omega, \mathbb{P})$, note that $\text{Cov}(X_1, X_2) = 0$ if and only if $X_1 - \mu_1 \mathbb{1}$ and $X_2 - \mu_2 \mathbb{1}$ are orthogonal where $\mu_i = \mathbb{E}[X_i]$, $i = 1, 2$, and $\mathbb{1}$ is the constant function equal to 1 on Ω .
 - (ii) Define $Y = X_2 - \frac{\mathbb{E}[X_2 X_1]}{\mathbb{E}[X_1^2]} X_1$. Then $\mathbb{E}[Y X_1] = 0$ since $\frac{\mathbb{E}[X_2 X_1]}{\mathbb{E}[X_1^2]} X_1$ is the component of X_2 in the direction of X_1 .
 - (iii) Let X_1, X_2 be random variables with mean μ_i and variance $\sigma_i^2 > 0$, $i = 1, 2$, and the correlation coefficient ρ . Define $Y = X_2 - \rho \frac{\sigma_2}{\sigma_1} X_1$. Then $\text{Cov}(Y, X_1) = 0$ since $\mathbb{E}[Y] = \mu_2 - \rho \frac{\sigma_2}{\sigma_1} \mu_1$ and since

$$\begin{aligned}
\text{Cov}(Y, X_1) &= \mathbb{E}[(X_2 - \rho \frac{\sigma_2}{\sigma_1} X_1) X_1] - (\mu_2 - \rho \frac{\sigma_2}{\sigma_1} \mu_1) \mu_1 \\
&= \mathbb{E}[X_1 X_2] - \rho \frac{\sigma_2}{\sigma_1} \mathbb{E}[X_1^2] - (\mu_2 - \rho \frac{\sigma_2}{\sigma_1} \mu_1) \mu_1 \\
&= (\rho \sigma_1 \sigma_2 + \mu_1 \mu_2) - \rho \frac{\sigma_2}{\sigma_1} (\sigma_1^2 + \mu_1^2) - (\mu_2 - \rho \frac{\sigma_2}{\sigma_1} \mu_1) \mu_1 = 0 .
\end{aligned}$$

Remark 4.11 If jointly normal random variables $X_1, \dots, X_n$ are uncorrelated, i.e., $\text{Cov}(X_i, X_j) = 0$ for $i \neq j$, then Σ is a diagonal matrix $\text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ where $\sigma_i^2 = \text{Var}(X_i)$, and

$$f(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi} \sigma_i} \exp\left(-\frac{1}{2\sigma_i^2} (x_i - \mu_i)^2\right) .$$

Since the joint pdf is the product of individual densities for X_i , we conclude that $X_1, \dots, X_n$ are independent. (See Theorem 4.4.)

Remark 4.12 For jointly normal variables $X_1, \dots, X_n$ and an $m \times n$ matrix $A = (a_{ij})$, it can be shown that the linear combinations $Y_i = \sum_{j=1}^n a_{ij} X_j$, $1 \leq i \leq m$, are jointly normal. Since independent normal variables are jointly normal, their linear combinations are jointly normal.

Theorem 4.13 Using independent standard normal variables Z_1 and Z_2 , we can construct a pair of correlated normal variables X_1 and X_2 with a given correlation ρ through the linear transformation given by

$$\begin{cases} X_1 = Z_1 \\ X_2 = \rho Z_1 + \sqrt{1 - \rho^2} Z_2 \end{cases}$$

Proof Note that

$$\text{Var}(X_2) = \rho^2 \text{Var}(Z_1) + (1 - \rho^2) \text{Var}(Z_2) + 2\rho \sqrt{1 - \rho^2} \text{Cov}(Z_1, Z_2) = 1 ,$$

and the correlation $\text{Corr}(X_1, X_2)$ is equal to

$$\text{Corr}(Z_1, \rho Z_1 + \sqrt{1 - \rho^2} Z_2) = \rho \text{Corr}(Z_1, Z_1) + \sqrt{1 - \rho^2} \text{Corr}(Z_1, Z_2) = \rho .$$

Now use Remark 4.12 to prove that X_1 and X_2 are normal variables. □

Example 4.21 By Theorem 4.13 we construct correlated normal variables X_1 and X_2 with $\rho(X_1, X_2) = -0.5$, using independent standard normal variables Z_1 and Z_2 . Figure 4.12 shows (Z_1, Z_2) and (X_1, X_2) in the left and right panels, respectively. Consult Simulation 4.5.

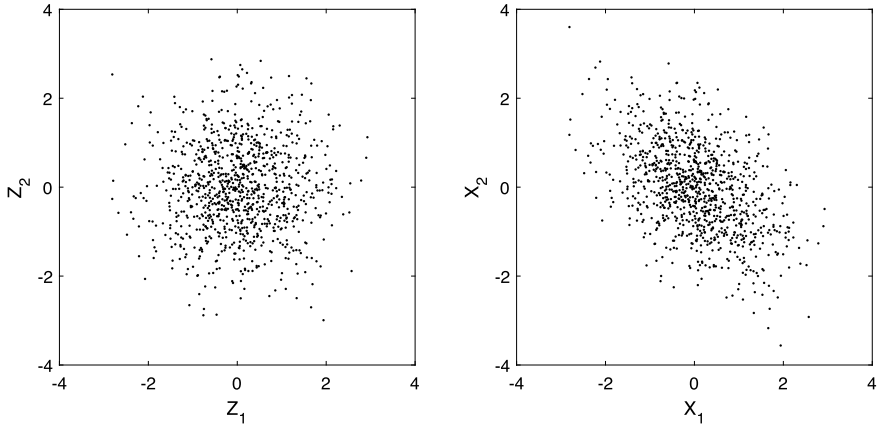


Fig. 4.12 Scatter plots for independent standard normal variables Z_1 and Z_2 (left), and negatively correlated normal variables X_1 and X_2 (right)

Definition 4.17 (*Skewness and Kurtosis*) Let X be a random variable with mean μ and variance σ^2 .

- (i) The *skewness* of X is defined by $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$.
- (ii) The *kurtosis* of X is defined by $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$.

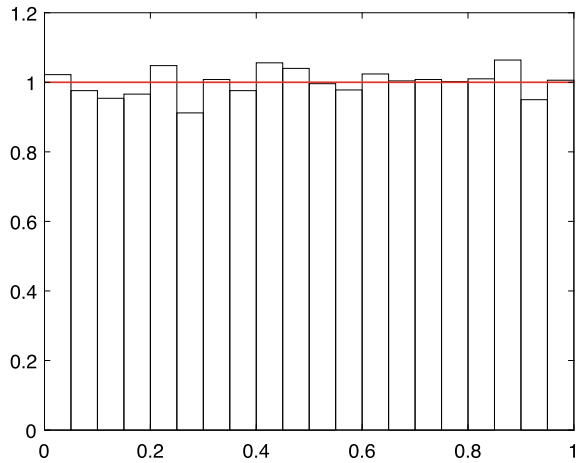
Example 4.22 A normal variable with mean μ and variance σ^2 has skewness 0 and kurtosis 3. (See Example 4.9.)

4.8 Computer Experiments

Simulation 4.1 (Uniform Distribution) Let $F_X(x)$ be the cdf of a random variable X . We plot the histogram of the sample values of the random variable $F_X(X)$, and check that $F_X(X)$ is uniformly distributed in $(0, 1)$. (See Theorem 4.7.) In the following experiment we divide the unit interval into N subintervals of length $\frac{1}{N}$, and choose a standard normal variable X . The MATLAB commands for X and F_X are given by `randn` and `normcdf`, respectively. See Fig. 4.13.

```
N = 20; % number of bins
Sample = 10^4; % number of samples
width = 1/N
bin = zeros(1,N); % a zero matrix of size 1 x N

for i = 1:Sample
```

Fig. 4.13 Uniform distribution of $F_X(X)$ 

```

j_bin = ceil(normcdf(randn)/width);
bin(j_bin) = bin(j_bin)+1;
end

x = 0:width:1;
y = ones(1,N+1); % the constant function 1
mid_points = 0+width/2:width:1; % middle points of bins
bar(mid_points,bin/Sample/width, 1, 'w')
hold on
plot(x,y)

```

We may use the command `histogram` instead of `bar` as follows:

```

FX = normcdf(randn(1,Sample));
histogram(FX,N,'Normalization','pdf','FaceColor','w')

```

Simulation 4.2 (Equally Spaced Points in the Unit Interval) Let $F_X(x)$ be the cdf of X . Generate N random values $x_1, \dots, x_N$ of X , sort them in ascending order, and obtain $y_1 < y_2 < \dots < y_N$. Since the values $F_X(x_1), \dots, F_X(x_N)$ are approximately uniformly distributed for sufficiently large N by Theorem 4.7, so are $F_X(y_1), \dots, F_X(y_N)$ since they are a rearrangement of $F_X(x_1), \dots, F_X(x_N)$. Note that $F_X(y_1) < F_X(y_2) < \dots < F_X(y_N)$ since F_X is monotonically increasing. Hence the sequence is close to the equally spaced points $\frac{1}{N+1} < \frac{2}{N+1} < \dots < \frac{N}{N+1}$, and hence $F_X(y_i)$ is close to $\frac{i}{N+1}$ for each i . Therefore, if we plot the points $(F_X(y_i), \frac{i}{N+1})$ they are concentrated along the diagonal of the unit square. For a scatter plot of one hundred such points for the standard normal distribution, see the left plot in Fig. 4.14.

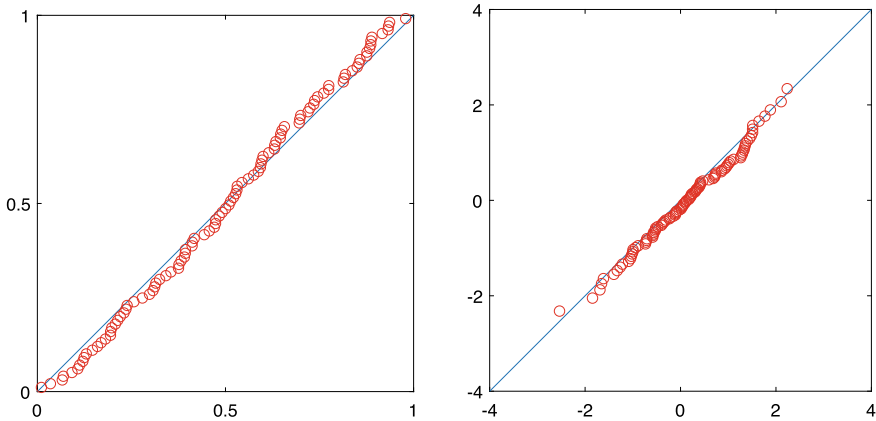


Fig. 4.14 Equally spaced points along the diagonal (*left*) and a quantile-quantile plot for the standard normal distribution (*right*)

```

N = 100; % number of points
X = randn(N,1);
Y = sort(X);

U = 1/(N+1):1/(N+1):N/(N+1);
V = normcdf(Y); % the normal cdf

x = 0:0.01:1;
plot(x,x)
hold on
plot(V,U,'o')

```

Simulation 4.3 (Quantile-Quantile Plot) Let $F_X(x)$ be the cdf of X . Generate N random values $x_1, \dots, x_N$ of X , sort them in ascending order, and obtain $y_1 < y_2 < \dots < y_N$. Since the values $F_X(y_1) < F_X(y_2) < \dots < F_X(y_N)$ are more or less uniformly distributed, they are close to $\frac{1}{N+1}, \frac{2}{N+1}, \dots, \frac{N}{N+1}$, respectively, i.e., $F_X(y_i)$ is close to $\frac{i}{N+1}$ for each i . Therefore, if we plot the points $(y_i, F_X^{-1}(\frac{i}{N+1}))$, then they are scattered along the straight line $y = x$, $-\infty < x < \infty$. For a scatter plot of one hundred such points for the standard normal distribution, see the right plot in Fig. 4.14.

```

N = 100; % number of points
X = randn(N,1);
Y = sort(X);

```

```

U = 1/(N+1):1/(N+1):N/(N+1);
V = norminv(U); % inverse of the normal cdf

x = -4:0.1:4;
plot(x,x)
hold on
plot(Y,V,'o')

```

Simulation 4.4 (The Central Limit Theorem) We check the Central Limit Theorem for the uniform distribution. Take a random vector $(X_1, \dots, X_n)$ where X_i are i.i.d. and uniformly distributed in $(0, 1)$. Note that $\mathbb{E}[X_i] = \frac{1}{2} (= \mu)$ and $\text{Var}(X_i) = \frac{1}{12} (= \sigma^2)$. Generate independently M such n -dimensional vectors and consider the distribution of M values of $S_n/n = (X_1 + \dots + X_n)/n$. For sufficiently large n , the distribution is close to the normal distribution $N(\mu, \sigma^2/n)$. In the simulation we choose $n = 2, 5, 20$ and 100 . For $n = 2$ the theoretical pdf of $(X_1 + X_2)/2$ is given by a triangle, which is the convolution of the pdf of the uniform distribution. To check the convergence to normal distribution see Fig. 4.15 where the pdf of the normal distribution is plotted in red.

```

mu = 1/2; % expectation
sigma2 = 1/12; % variance
n = 100; % size of each sample in the fourth plot
M = 10^6; % number of samples of size n

Aver = zeros(M,1);
num_bins = 60; % number of bins

X = rand(M,n); % unif distribution
for i = 1:M
    Aver(i,1) = mean(X(i,:));
end
histogram(Aver,num_bins,'Normalization','pdf');
hold on

% the probability density function of the normal distribution
x = 0:0.01:1;
s = sqrt(sigma2/n); % standard deviation
f = exp(-(x-mu).^2/(2*s^2))/(s*sqrt(2*pi));
plot(x,f,'Color','red')

```

Simulation 4.5 (Correlated Normal Variables) We plot the ordered pairs (Z_1, Z_2) and (X_1, X_2) as in Example 4.21.

```

N = 1000; % number of points
Z1 = randn(N,1); Z2 = randn(N,1);
CorrelationZ1Z2 = corr(Z1,Z2)

```

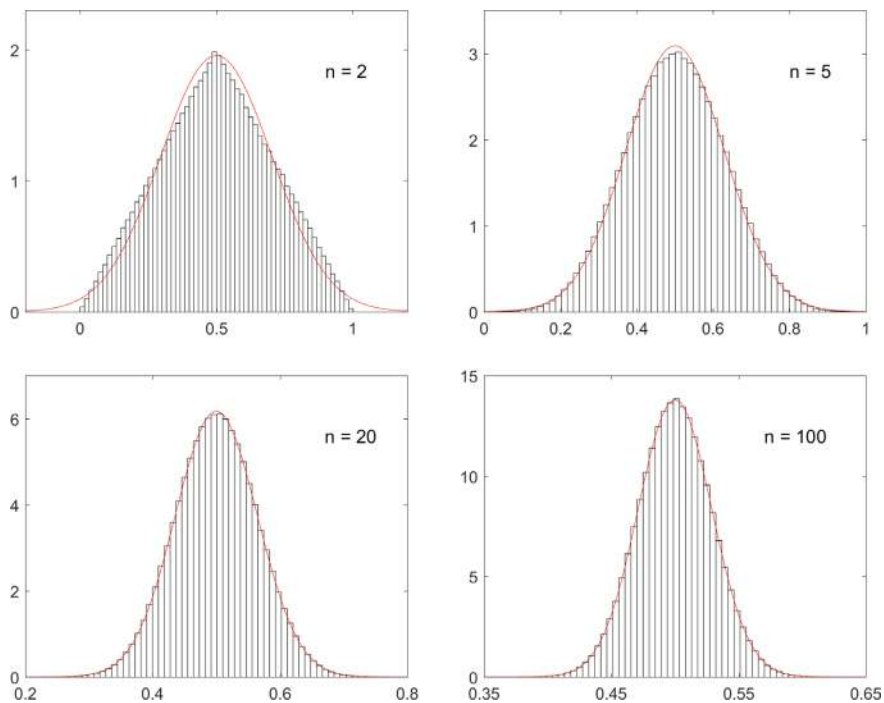


Fig. 4.15 The Central Limit Theorem: distribution of sample averages $(U_1 + \dots + U_n)/n$, $n = 2, 5, 20, 100$, for the uniform distribution $U(0, 1)$

```
rho = -0.5;
X1 = Z1;
X2 = rho*Z1 + sqrt(1-rho^2)*Z2;
CorrelationX1X2 = corr(X1,X2)
```

```
figure(1)
plot(Z1,Z2,'.');
hold off;
```

```
figure(2)
plot(X1,X2,'.')
```

We have the following output:

```
CorrelationZ1Z2 = 0.0168
CorrelationX1X2 = -0.4730
```

Exercises

4.1 Let $(\Omega, \mathbb{P})$ be a probability space, and let $X \geq 0$ be a random variable. Show that $\mathbb{E}[X] < \infty$ if and only if $\sum_{j=0}^{\infty} \mathbb{P}(\{X \geq j\}) < \infty$. (See also Exercise 3.6.)

4.2 Let X be a random variable taking values ± 1 with equal probability. Show that if we choose $\phi(x) = x^2$ in Jensen's inequality then we have a strict inequality.

4.3 For a random variable X satisfying $\mathbb{P}(X = a) = 1$ find the cumulative distribution function.

4.4 Let U be a uniformly distributed random variable in $[0, 1]$. Show that $\mathbb{E}[e^{\sqrt{U}}] = 2$. (Hint: See Example 4.15.)

4.5 Let N denote the cdf of the standard normal distribution. Prove that $N(\alpha) + N(-\alpha) = 1$.

4.6 Let X be a random variable with its cumulative distribution function $F(x)$. Assume that F is invertible. Show that if U is uniformly distributed in $[0, 1]$, then the cumulative distribution function of $F^{-1}(U)$ is F . In other words, $F^{-1}(U)$ and X are identically distributed. See Fig. 4.16.

4.7 Let X be a normal variable with mean μ and variance σ^2 . Show that the mean and the variance of e^X are given by $e^{\mu + \sigma^2/2}$ and $e^{2\mu + \sigma^2}(e^{\sigma^2} - 1)$, respectively. For example, for a standard normal variable Z we have $\mathbb{E}[e^Z] = \sqrt{e}$.

4.8 Let Z be a standard normal variable and take $u = a + ib \in \mathbb{C}$, $a, b \in \mathbb{R}$. Show that $\mathbb{E}[e^{uZ}] = e^{u^2/2}$. (For example, by taking $u = i\theta$, we have $\mathbb{E}[e^{i\theta Z}] = e^{-\theta^2/2}$, which is a special case of Exercise 4.11 with $u = i\sqrt{t}$. In general, for $X \sim N(\mu, \sigma^2)$ we have $\mathbb{E}[e^{uX}] = e^{\mu u + \frac{1}{2}\sigma^2 u^2}$.) Prove that $\mathbb{E}[Z^{2n}] = (2n)!/(n!2^n)$.

4.9 Let Z denote a standard normal variable. For a real constant α , show that $e^{\alpha Z} \in L^p$ for every $1 \leq p < \infty$.

Fig. 4.16 The composite random variable $F^{-1}(U)$

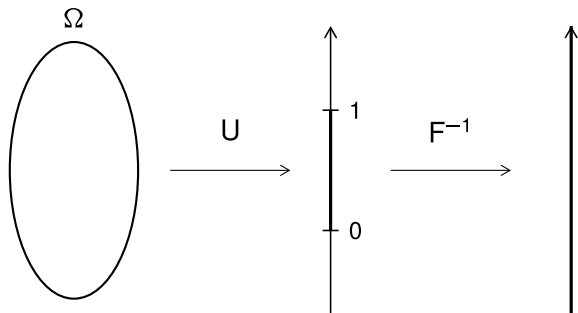
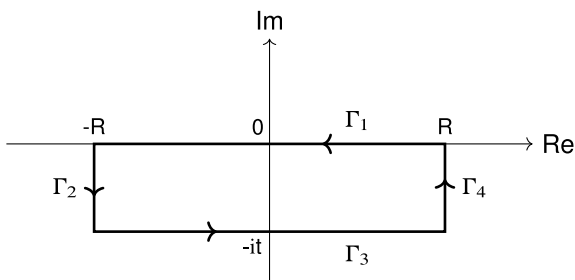


Fig. 4.17 A contour for $f(z) = \exp(-z^2/(2t))$ for $z \in \mathbb{C}$



4.10 Let Z denote a standard normal variable. Show that $\mathbb{E}[e^{\alpha Z^2}] = \frac{1}{\sqrt{1-2\alpha}}$ for $-\infty < \alpha < \frac{1}{2}$. For example, $\mathbb{E}[e^{-Z^2}] = \frac{1}{\sqrt{3}}$. (See also Appendix A.8.)

4.11 Let X be a normal variable with mean 0 and variance $t > 0$. Using the contour integral in complex analysis, show that $\mathbb{E}[e^{iX}] = e^{-t/2}$. (Hint: Consider $f(z) = \exp(-\frac{z^2}{2t})$, $z \in \mathbb{C}$, and a closed curve $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$ given in Fig. 4.17.)

4.12 Let $U_1, \dots, U_{12}$ be independent random variables uniformly distributed in the unit interval. Compute the kurtosis of $Y = U_1 + \dots + U_{12} - 6$.

4.13 If X is uniformly distributed in $[-1, 1]$, what is the probability distribution of $Y = X^2$? (Hint: Note that $f_X = \frac{1}{2} \times \mathbb{1}_{[-1,1]}$. For $y \notin (0, 1]$, $f_Y(y) = 0$.)

4.14 Let X and Y be independent standard normal variables. Show that $\tan^{-1}(\frac{Y}{X})$ is uniformly distributed in $(-\frac{\pi}{2}, \frac{\pi}{2})$.

4.15 Suppose that (X, Y) is a continuous two-dimensional random vector with probability density function $f_{X,Y}(x, y)$. Prove that, if $(V, W) = \Phi(X, Y)$ for some bijective differentiable mapping $\Phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, then

$$f_{V,W}(v, w) = f_{X,Y}(\Phi^{-1}(v, w)) \frac{1}{|J_\Phi(\Phi^{-1}(v, w))|},$$

where J_Φ denotes the Jacobian of Φ .

4.16 Find random variables X, Y which are not independent but $\text{Cov}(X, Y) = 0$.

4.17 Find a pair of standard normal variables with correlation $-1 \leq \rho \leq 1$. (Hint: First, generate a pair of independent standard normal variables Z_1, Z_2 . Define X_1, X_2 by linear combinations of Z_1, Z_2 .)

4.18 (Uncorrelated normal variables) Let X, Y be random variables with the joint density function

$$f_{X,Y}(x, y) = \begin{cases} \frac{2|x+y|}{\sqrt{2\pi}} \exp\left(-\frac{(2|x+y|)^2}{2}\right), & y \geq -|x|, \\ 0, & y < -|x|. \end{cases}$$

(See Fig. 4.18.) Prove that X and Y are uncorrelated standard normal random variables. Show that they are not independent.

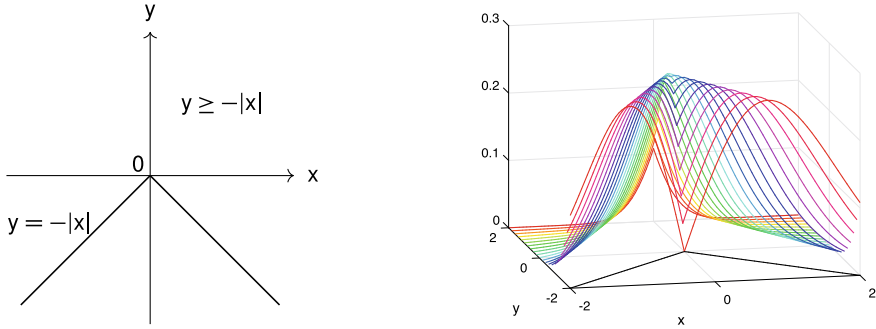


Fig. 4.18 Support of the joint pdf (left) of uncorrelated normal variables X, Y in Exercise 4.18 and the graph of the joint pdf (right)

4.19 Let $X \geq 0$ be a square-integrable random variable. Prove the following statements:

- (i) If X has continuous distribution, then $\mathbb{E}[X^2] = 2 \int_0^\infty t \mathbb{P}(X > t) dt$.
- (ii) If X has discrete distribution, then $\mathbb{E}[X^2] = \sum_{k=1}^\infty (2k+1) \mathbb{P}(X > k)$.

4.20 (Product of Random Variables) Let X, Y be independent random variables with their probability density functions f_X and f_Y , respectively. Prove that the probability density function of XY is given by

$$f_{XY}(a) = \int_{-\infty}^{\infty} \frac{1}{|y|} f_X\left(\frac{a}{y}\right) f_Y(y) dy .$$

4.21 Suppose that U and V are independent and uniformly distributed in $[0, 1]$. We consider the random variable UV .

- (i) Prove that $f_{UV}(a) = -\log a$ for $0 < a \leq 1$, and $f_{UV}(a) = 0$ elsewhere.
- (ii) Find the cumulative distribution function for UV .

For a numerical confirmation for cdf, use the following code:

```
M = 10^4;
bins = 20;
U = rand(M,1); V = rand(M,1);
X = U.*V;
histogram(X,bins,'Normalization','cdf')
hold on
u = 0:0.002:1;
cdf = u - log(u).*u;
axis equal
plot(u,cdf,'r');
```

See Fig. 4.19, where we use $M = 10^4$ samples and 20 bins.

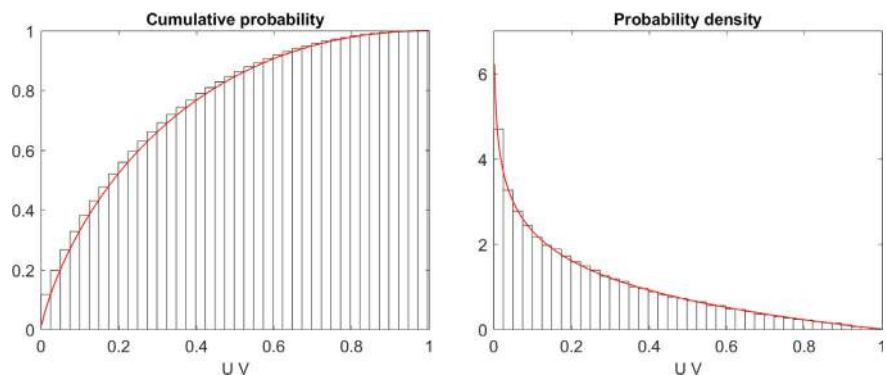


Fig. 4.19 The pdf and cdf of the product of two independent uniformly distributed random variables U and V

4.22 Let $Z \sim N(0, 1)$. Compute $\mathbb{E}[|Z|]$. (For a numerical confirmation of the answer see Appendix A.8.)

Chapter 5

Conditional Expectation



Conditional expectation is an essential mathematical tool in stating option pricing formulas. The option price at time t is given by the conditional expectation based on all the available information up to time t . The concept of conditional expectation will be developed in three stages. First, we define the conditional expectation on a given event, $\mathbb{E}[X|A]$, and next define the conditional expectation with respect to a sub- σ -algebra, $\mathbb{E}[X|\mathcal{G}]$, then finally the conditional expectation with respect to a random variable, $\mathbb{E}[X|Y]$. Recall that a sub- σ -algebra represents a collection of currently available information. In this chapter we present an intuitive introduction to conditional expectation.

5.1 Conditional Expectation Given an Event

For measurable subsets A and B , $\mathbb{P}(A) > 0$, define the *conditional probability* of B given A by

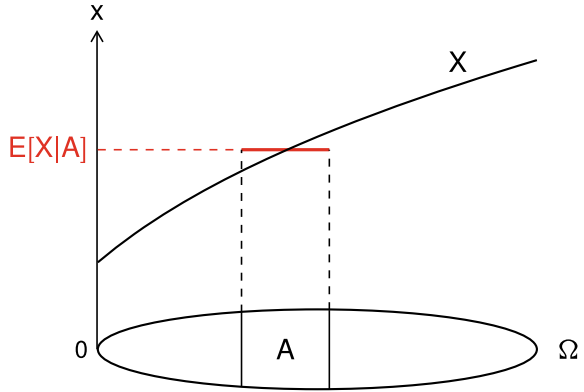
$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} .$$

(See also Definition 3.4.)

Definition 5.1 (*Conditional Expectation*) Consider a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an event $A \in \mathcal{F}$ of positive probability. The conditional expectation of X given A , denoted by $\mathbb{E}[X|A]$, is defined by

$$\mathbb{E}[X|A] = \frac{\int_A X \, d\mathbb{P}}{\mathbb{P}(A)} = \frac{\mathbb{E}[\mathbb{1}_A X]}{\mathbb{P}(A)} .$$

Fig. 5.1 Conditional expectation of X given an event A



In other words, $\mathbb{E}[X|A]$ is the average of X on a set A . (See Fig. 5.1.) If a random variable X is given by $X = \mathbb{1}_B$ for a measurable set B , we have

$$\mathbb{E}[X|A] = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B|A), \quad (5.1)$$

which is an extension of the concept of conditional probability.

5.2 Conditional Expectation with Respect to a σ -Algebra

As a motivation, we first consider a sample space Ω which is a disjoint union $\Omega = A_1 \cup \dots \cup A_n$ of measurable subsets $A_1, \dots, A_n$ of positive measure. Let $\mathcal{G}$ be the σ -algebra generated by $A_1, \dots, A_n$, then $\mathcal{G}$ is a sub- σ -algebra of $\mathcal{F}$. Note that if $B \in \mathcal{G}$ and $\mathbb{P}(B) > 0$, then B is a union of at least one subset from the collection $A_1, \dots, A_n$. If a subset B of A_i has positive measure, then $B = A_i$. Sometimes $\mathcal{G}$ is denoted by $\sigma(\{A_1, \dots, A_n\})$. In this case, a new random variable $E[X|\mathcal{G}]$ is defined by

$$\mathbb{E}[X|\mathcal{G}](\omega) = \mathbb{E}[X|A_i], \quad \omega \in A_i,$$

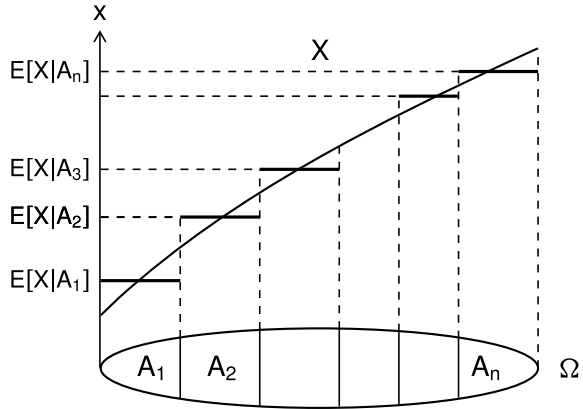
where $\mathbb{E}[X|A_i]$ is given by (5.1). In other words,

$$\mathbb{E}[X|\mathcal{G}] = \sum_{i=1}^n \mathbb{E}[X|A_i] \mathbb{1}_{A_i}.$$

(See Fig. 5.2.) Note that $E[X|\mathcal{G}]$ is $\mathcal{G}$ -measurable and for $B \in \mathcal{G}$ we have

$$\int_B \mathbb{E}[X|\mathcal{G}] d\mathbb{P} = \int_B X d\mathbb{P}.$$

Fig. 5.2 Conditional expectation when a probability space is decomposed into subsets of positive measure



The idea can be easily extended to the case when the sample space is a countable union of pairwise disjoint subsets.

Now consider a general sub- σ -algebra $\mathcal{G}$ where the probability space need not be a disjoint union of countably many subsets. In this case, even if we define the value of the function $\mathbb{E}[X|\mathcal{G}]$ on a sufficiently small size subset $A \in \mathcal{G}$ it is necessary that we define on a further smaller size subset $B \in \mathcal{G}$ such that $B \subset A$, and with the property that

$$\frac{1}{\mathbb{P}(B)} \int_B \mathbb{E}[X|\mathcal{G}] d\mathbb{P} = \frac{1}{\mathbb{P}(B)} \int_B X d\mathbb{P} = \mathbb{E}[X|B].$$

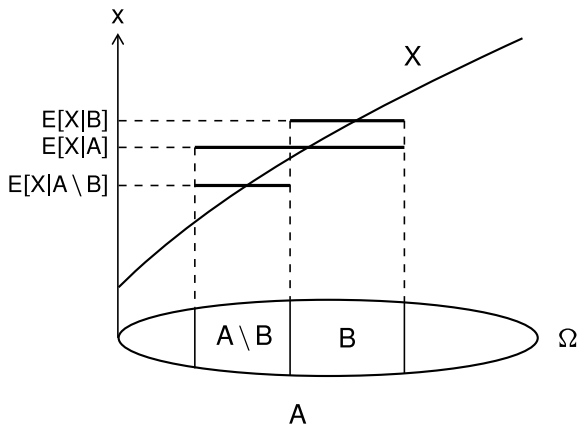
More precisely, if $A = B \cup (A - B)$, then the average of the averages $\mathbb{E}[X|B]$ and $\mathbb{E}[X|A - B]$ on B and $A - B$, respectively, should be equal to $\mathbb{E}[X|A]$. (See Fig. 5.3.) If there exists a $C \subset B$ such that $C \in \mathcal{G}$ and $\mathbb{P}(C) > 0$, then this procedure should be repeated. In general, such a procedure continues forever and does not come to an end, and hence we have to define $\mathbb{E}[X|\mathcal{G}]$ in an abstract way in what follows.

Definition 5.2 (*Conditional Expectation*) Let $\mathcal{F}$ be a σ -algebra on Ω , and $\mathcal{G}$ a sub- σ -algebra of $\mathcal{F}$. If X is an $\mathcal{F}$ -measurable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and if there exists a $\mathcal{G}$ -measurable random variable Y such that

$$\int_A Y d\mathbb{P} = \int_A X d\mathbb{P}$$

for every $A \in \mathcal{G}$, then Y is called the *conditional expectation* of X with respect to $\mathcal{G}$ and denoted by $\mathbb{E}[X|\mathcal{G}]$.

Fig. 5.3 Comparison of $\mathbb{E}[X|A]$, $\mathbb{E}[X|B]$, $\mathbb{E}[X|A \setminus B]$ where $B \subset A$



Theorem 5.1 *The conditional expectation Y in Definition 5.2 exists and is unique.*

Proof Given a random variable X , let $X^+ = \max\{X, 0\}$, $X^- = \max\{-X, 0\}$, then $X = X^+ - X^-$. For a sub- σ -algebra $\mathcal{G}$ define $\mathbb{Q} : \mathcal{G} \rightarrow [0, \infty)$ by

$$\mathbb{Q}(A) = \int_A X^+ d\mathbb{P}, \quad A \in \mathcal{G}.$$

Then $\mathbb{Q}$ is a measure on $\mathcal{G}$. Since $\mathbb{Q}$ is absolutely continuous with respect to the measure $\mathbb{P}|_{\mathcal{G}}$, the restriction of $\mathbb{P} : \mathcal{F} \rightarrow [0, \infty)$ to $\mathcal{G}$, by the Radon–Nikodym theorem there exists a $\mathcal{G}$ -measurable random variable Y^+ satisfying

$$\int_A X^+ d\mathbb{P} = \mathbb{Q}(A) = \int_A Y^+ d\mathbb{P}, \quad A \in \mathcal{G}.$$

Similarly for X^- there exists a $\mathcal{G}$ -measurable random variable Y^- satisfying

$$\int_A X^- d\mathbb{P} = \int_A Y^- d\mathbb{P}, \quad A \in \mathcal{G}.$$

Now let $Y = Y^+ - Y^-$, then Y is $\mathcal{G}$ -measurable and satisfies

$$\int_A X d\mathbb{P} = \int_A Y d\mathbb{P}, \quad A \in \mathcal{G}.$$

To show the uniqueness, suppose that there exist $\mathcal{G}$ -measurable random variables Y_1 and Y_2 satisfying

$$\int_A X d\mathbb{P} = \int_A Y_1 d\mathbb{P} = \int_A Y_2 d\mathbb{P}$$

for every $A \in \mathcal{G}$. Hence $\int_A (Y_1 - Y_2) d\mathbb{P} = 0$ for every $A \in \mathcal{G}$, thus $Y_1 - Y_2 = 0$, i.e., $Y_1 = Y_2$ modulo $\mathbb{P}$ -measure zero subsets. $\square$

Example 5.1 (*Binary Expansion*) Let $\Omega = \prod_1^\infty \{0, 1\}$, a convenient choice for the fair coin tossing problem, identified with the unit interval via the binary expansion of real numbers. (See also Example 3.2.) For a *cylinder subset*

$$[a_1, \dots, a_n] = \{(x_1 x_2 x_3 \dots) : x_i = a_i, 1 \leq i \leq n\}$$

define its size by $\mu([a_1, \dots, a_n]) = 2^{-n}$. Then the set function μ is extended to a probability measure, also denoted by μ , defined on the σ -algebra generated by the cylinder sets. (If we regard μ as a measure on the unit interval, it is Lebesgue measure.)

Now let $\mathcal{F}_n$ be the sub- σ -algebra generated by cylinder sets $[a_1, \dots, a_n]$. (From the perspective of the unit interval, $\mathcal{F}_n$ is the sub- σ -algebra generated by the intervals $[(i-1) \times 2^{-n}, i \times 2^{-n})$, $1 \leq i \leq 2^n$.) An $\mathcal{F}$ -measurable function is of the form $\xi(x_1, x_2, x_3, \dots)$, $x_i \in \{0, 1\}$, $i \geq 1$, and an $\mathcal{F}_n$ -measurable function is of the form $\eta(x_1, \dots, x_n)$, $x_i \in \{0, 1\}$, which does not depend on the variables $x_{n+1}, x_{n+2}, \dots$. Hence $\mathbb{E}[\xi | \mathcal{F}_n] = \eta(x_1, \dots, x_n)$ for some function $\eta : \prod_1^n \{0, 1\} \rightarrow \mathbb{R}$. With abuse of notation, we may write

$$\mathbb{E}[\xi | \mathcal{F}_n](x_1, \dots, x_n) = \int \cdots \int \xi(x_1, \dots, x_n, x_{n+1}, \dots) dx_{n+1} \cdots dx_\infty,$$

where the integration is done over $\prod_{n+1}^\infty \{0, 1\}$.

Theorem 5.2 (Conditional Expectation) *For a given measure space $(\Omega, \mathcal{F}, \mathbb{P})$, let $\mathcal{G}$ and $\mathcal{H}$ be sub- σ -algebras of a σ -algebra $\mathcal{F}$. Then the following facts hold:*

- (i) (*Linearity*) $\mathbb{E}[aX + bY | \mathcal{G}] = a \mathbb{E}[X | \mathcal{G}] + b \mathbb{E}[Y | \mathcal{G}]$ for any constants a, b .
- (ii) (*Average of average*) $\mathbb{E}[\mathbb{E}[X | \mathcal{G}]] = \mathbb{E}[X]$.
- (iii) (*Taking out what is known*) If X is $\mathcal{G}$ -measurable and XY is integrable, then $\mathbb{E}[XY | \mathcal{G}] = X \mathbb{E}[Y | \mathcal{G}]$.
- (iv) (*Independence*) If X and $\mathcal{G}$ are independent, then $\mathbb{E}[X | \mathcal{G}] = \mathbb{E}[X]$.
- (v) (*Tower property*) If $\mathcal{H} \subset \mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{H}] = \mathbb{E}[X | \mathcal{H}]$. Hence, if $\mathcal{H} = \mathcal{G}$, the conditional expectation is a projection, i.e., $\mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{G}] = \mathbb{E}[X | \mathcal{G}]$.
- (vi) (*Positivity*) If $X \geq 0$, then $\mathbb{E}[X | \mathcal{G}] \geq 0$.

Proof (i) Use the fact that for every $A \in \mathcal{G}$

$$\begin{aligned} \int_A (a \mathbb{E}[X | \mathcal{G}] + b \mathbb{E}[Y | \mathcal{G}]) d\mathbb{P} &= a \int_A \mathbb{E}[X | \mathcal{G}] d\mathbb{P} + b \int_A \mathbb{E}[Y | \mathcal{G}] d\mathbb{P} \\ &= a \int_A X d\mathbb{P} + b \int_A Y d\mathbb{P} \\ &= \int_A (aX + bY) d\mathbb{P}. \end{aligned}$$

- (ii) For a proof choose $A = \Omega$ in the definition of conditional expectation. For another proof take $\mathcal{H} = \{\emptyset, \Omega\}$ in item (v) and use $\mathbb{E}[\cdot | \mathcal{H}] = \mathbb{E}[\cdot]$.
- (iii) First, consider the case when $X = \mathbb{1}_A$ for $A \in \mathcal{G}$. For any $B \in \mathcal{G}$ we have

$$\int_B \mathbb{1}_A \mathbb{E}[Y | \mathcal{G}] d\mathbb{P} = \int_{A \cap B} \mathbb{E}[Y | \mathcal{G}] d\mathbb{P} = \int_{A \cap B} Y d\mathbb{P} = \int_B \mathbb{1}_A Y d\mathbb{P},$$

and hence $\mathbb{1}_A \mathbb{E}[Y | \mathcal{G}] = \mathbb{E}[\mathbb{1}_A Y | \mathcal{G}]$, which holds for simple functions by (i). Next, use the fact that a measurable function can be approximated by a convergent sequence of simple functions.

- (iv) Since X and $\mathcal{G}$ are independent, if we let $0 = a_0 < a_1 < a_2 < \dots$ be a partition of the real line that is the range of X , then for any $A \in \mathcal{G}$ we have

$$\mathbb{P}(X^{-1}([a_i, a_{i+1}]) \cap A) = \mathbb{P}(X^{-1}([a_i, a_{i+1}])) \mathbb{P}(A).$$

Hence

$$\begin{aligned} \int_A X d\mathbb{P} &= \lim \sum_i a_i \mathbb{P}(X^{-1}([a_i, a_{i+1}]) \cap A) \\ &= \lim \sum_i a_i \mathbb{P}(X^{-1}([a_i, a_{i+1}])) \mathbb{P}(A) \\ &= \mathbb{E}[X] \mathbb{P}(A) \\ &= \int_A \mathbb{E}[X] d\mathbb{P}. \end{aligned}$$

- (v) Note that for any $B \in \mathcal{G}$ we have

$$\int_B \mathbb{E}[X | \mathcal{G}] d\mathbb{P} = \int_B X d\mathbb{P}$$

and that for any $B \in \mathcal{H}$ we have

$$\int_B \mathbb{E}[X | \mathcal{H}] d\mathbb{P} = \int_B X d\mathbb{P}.$$

Thus if $\mathcal{H} \subset \mathcal{G}$ then for any $B \in \mathcal{H}$ we have

$$\int_B \mathbb{E}[X | \mathcal{G}] d\mathbb{P} = \int_B \mathbb{E}[X | \mathcal{H}] d\mathbb{P}.$$

By the definition of conditional expectation we conclude

$$\mathbb{E}[\mathbb{E}[X | \mathcal{G}] | \mathcal{H}] = \mathbb{E}[X | \mathcal{H}].$$

- (vi) For each $n \geq 1$, let $A_n = \{\omega : \mathbb{E}[X|\mathcal{G}](\omega) \leq -\frac{1}{n}\}$. If $X \geq 0$, then $A_n \in \mathcal{G}$. Put $A = \bigcup_{n=1}^{\infty} A_n$. Then $A = \{\mathbb{E}[X|\mathcal{G}] < 0\}$ and

$$0 \leq \int_{A_n} X \, d\mathbb{P} = \int_{A_n} \mathbb{E}[X|\mathcal{G}] \, \mathbb{P} \leq -\frac{1}{n} \mathbb{P}(A_n)$$

and $\mathbb{P}(A_n) = 0$. By Fact 3.1 (ii) we conclude that $\mathbb{P}(A) = 0$. $\square$

The following is Jensen's inequality for conditional expectation.

Theorem 5.3 *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub- σ -algebra of $\mathcal{F}$, and let $\phi : \mathbb{R} \rightarrow \mathbb{R}$ be convex. If X and $\phi \circ X$ are integrable, then*

$$\phi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\phi(X)|\mathcal{G}].$$

For a concave function the inequality is in the opposite direction.

Proof Note that $\phi(x_0) = \sup\{ax_0 + b : ay + b \leq \phi(y) \text{ for all } y \in \mathbb{R}\}$ for each x_0 . Since $aX + b \leq \phi(X)$ for a, b such that $ay + b \leq \phi(y)$ for $y \in \mathbb{R}$, we have $a\mathbb{E}[X|\mathcal{G}] + b = \mathbb{E}[aX + b|\mathcal{G}] \leq \mathbb{E}[\phi(X)|\mathcal{G}]$. Now use the fact that the supremum of the left-hand side is equal to $\phi(\mathbb{E}[X|\mathcal{G}])$. $\square$

Corollary 5.1 *Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G}$ be a sub- σ -algebra of $\mathcal{F}$. If $|X|^p$ is integrable for $1 \leq p < \infty$, then*

$$\|\mathbb{E}[X|\mathcal{G}]\|_p \leq \|X\|_p.$$

That is, the linear transformation $\mathbb{E}[\cdot|\mathcal{G}] : L^p(\Omega, \mathcal{F}) \rightarrow L^p(\Omega, \mathcal{G}) \subset L^p(\Omega, \mathcal{F})$ has norm 1, and hence is continuous.

Proof Take a convex function $\phi(x) = |x|^p$. By Jensen's inequality we obtain $|\mathbb{E}[X|\mathcal{G}]|^p \leq \mathbb{E}[|X|^p|\mathcal{G}]$ and $\mathbb{E}[|\mathbb{E}[X|\mathcal{G}]|^p] \leq \mathbb{E}[\mathbb{E}[|X|^p|\mathcal{G}]] = \mathbb{E}[|X|^p]$. Now take the p th root of the both sides. It is clear that the norm is less than or equal to 1. Since $\mathbb{E}[1|\mathcal{G}] = 1$ we conclude that the norm is equal to 1. $\square$

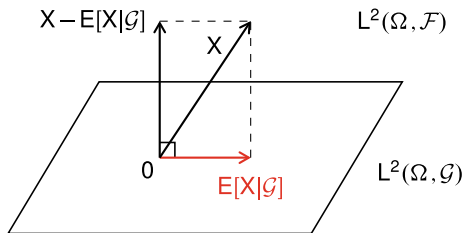
Recall that if $\mathcal{G}$ is a sub- σ -algebra of $\mathcal{F}$, then $L^2(\Omega, \mathcal{G})$ is a subspace of $L^2(\Omega, \mathcal{F})$. Define an inner product for $X, Y \in L^2(\Omega, \mathcal{F})$ by $\mathbb{E}[XY]$, which is equal to $\int_{\Omega} XY \, d\mathbb{P}$.

Theorem 5.4 *Let $\mathcal{G}$ be a sub- σ -algebra of $\mathcal{F}$. Then the linear transformation*

$$\mathbb{E}[\cdot|\mathcal{G}] : L^2(\Omega, \mathcal{F}) \rightarrow L^2(\Omega, \mathcal{G}) \subset L^2(\Omega, \mathcal{F})$$

is an orthogonal projection onto $L^2(\Omega, \mathcal{G})$. (See Fig. 5.4.)

Fig. 5.4 Conditional expectation as an orthogonal projection



Proof To show that $\mathbb{E}[X|\mathcal{G}]$ and $X - \mathbb{E}[X|\mathcal{G}]$ are orthogonal, we note that

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X|\mathcal{G}](X - \mathbb{E}[X|\mathcal{G}])] &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]X] - \mathbb{E}[\mathbb{E}[X|\mathcal{G}]\mathbb{E}[X|\mathcal{G}]] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]X] - \mathbb{E}[\mathbb{E}[\mathbb{E}[X|\mathcal{G}]X|\mathcal{G}]] \\ &= \mathbb{E}[\mathbb{E}[X|\mathcal{G}]X] - \mathbb{E}[\mathbb{E}[X|\mathcal{G}]X] \\ &= 0. \end{aligned}$$

The second equality holds since $\mathbb{E}[X|\mathcal{G}]$ is $\mathcal{G}$ -measurable and $\mathbb{E}[X|\mathcal{G}]\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]X|\mathcal{G}]$ by Theorem 5.2(iii). For the third equality use Theorem 5.2(ii). $\square$

Corollary 5.2 Let $\mathcal{G}$ be a sub- σ -algebra of $\mathcal{F}$. Let X be an $\mathcal{F}$ -measurable random variable such that $\mathbb{E}[X^2] < \infty$. Then

$$\mathbb{E}[\mathbb{E}[X|\mathcal{G}]^2] \leq \mathbb{E}[X^2].$$

Proof Either we use Theorem 5.4, or apply Jensen's inequality $\mathbb{E}[X|\mathcal{G}]^2 \leq \mathbb{E}[X^2|\mathcal{G}]$, which implies $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]^2] \leq \mathbb{E}[\mathbb{E}[X^2|\mathcal{G}]] = \mathbb{E}[X^2]$. $\square$

Remark 5.1 If $\mathbb{E}[X^2] < \infty$, then

$$\mathbb{E}[X - \mathbb{E}[X|\mathcal{G}]]^2 = \min_{Y \in L^2(\Omega, \mathcal{G}, \mathbb{P})} \mathbb{E}[X - Y]^2.$$

Hence $\mathbb{E}[X|\mathcal{G}]$ is the best approximation based on the information contained in $\mathcal{G}$ from the viewpoint of the least squares method in $L^2(\Omega, \mathcal{F})$.

Example 5.2 A probability measure space $(\Omega, \mathcal{F}, \mathbb{P})$ is given by $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $\mathcal{F} = \sigma(\{\omega_1\}, \{\omega_2\}, \{\omega_3\})$ with a probability measure $\mathbb{P}$ defined by $\mathbb{P}(\{\omega_i\}) = \frac{1}{3}$, $1 \leq i \leq 3$. If a random variable X is defined by $X(\omega_1) = x_1$, $X(\omega_2) = x_2$, $X(\omega_3) = x_3$, and if $\mathcal{G} = \{\emptyset, \Omega, \{\omega_1, \omega_2\}, \{\omega_3\}\}$, then

$$\mathbb{E}[X|\mathcal{G}](\omega) = \begin{cases} \frac{1}{2}(x_1 + x_2), & \omega \in \{\omega_1, \omega_2\} \\ x_3, & \omega \in \{\omega_3\}. \end{cases}$$

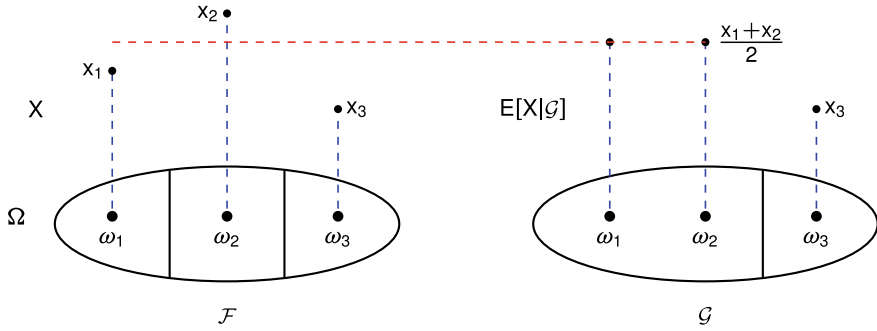


Fig. 5.5 Conditional expectation with respect to a sub- σ -algebra $\mathcal{G}$ on $\Omega = \{\omega_1, \omega_2, \omega_3\}$

(See Fig. 5.5. Note that if a random variable Y is measurable with respect to $\mathcal{G}$, then $Y(\omega_1) = Y(\omega_2)$.) Since

$$X(\omega) - \mathbb{E}[X|\mathcal{G}](\omega) = \begin{cases} \frac{1}{2}(x_1 - x_2), & \omega = \omega_1 \\ \frac{1}{2}(-x_1 + x_2), & \omega = \omega_2 \\ 0, & \omega = \omega_3 \end{cases}$$

the expectation of the product of $\mathbb{E}[X|\mathcal{G}]$ and $X - \mathbb{E}[X|\mathcal{G}]$ is zero, and they are orthogonal to each other. For more examples, see Brzeźniak and Zastawniak (1999).

5.3 Conditional Expectation with Respect to a Random Variable

Definition 5.3 (*Conditional Expectation*) On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we are given $\mathcal{F}$ -measurable random variables X and Y . We assume that X is integrable. Define the conditional expectation of X with respect to Y , denoted by $\mathbb{E}[X|Y]$, by $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$. Then $\mathbb{E}[X|Y]$ is measurable with respect to $\sigma(Y)$, and satisfies

$$\int_{\{a \leq Y \leq b\}} \mathbb{E}[X|Y] d\mathbb{P} = \int_{\{a \leq Y \leq b\}} X d\mathbb{P}$$

for an arbitrary interval $[a, b] \subset \mathbb{R}$.

Remark 5.2 The conditional expectation of a random variable given an event is a constant. However, the conditional expectation with respect to a sub- σ -algebra or a random variable is again a random variable.

Theorem 5.5 *If the conditions given in Definition 5.3 are satisfied, then $\mathbb{E}[X|Y]$ is a function of Y . More precisely, there exists a measurable function $f : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ such that $\mathbb{E}[X|Y] = f \circ Y$.*

Proof We sketch the proof. First, consider the case when Y is a simple function of the form $Y = \sum_{i=1}^n y_i \times \mathbb{1}_{A_i}$ where y_i are distinct and the subsets $A_i = Y^{-1}(y_i)$ are pairwise disjoint. Then for $\omega \in A_i$ we have $\mathbb{E}[X|Y](\omega) = \mathbb{E}[X|A_i]$. Hence $\mathbb{E}[X|Y] = \sum_{i=1}^n \mathbb{E}[X|A_i] \times \mathbb{1}_{A_i}$. Thus if we take a measurable function f such that $f(y_i) = \mathbb{E}[X|A_i]$ for every i , then $\mathbb{E}[X|Y] = f \circ Y$.

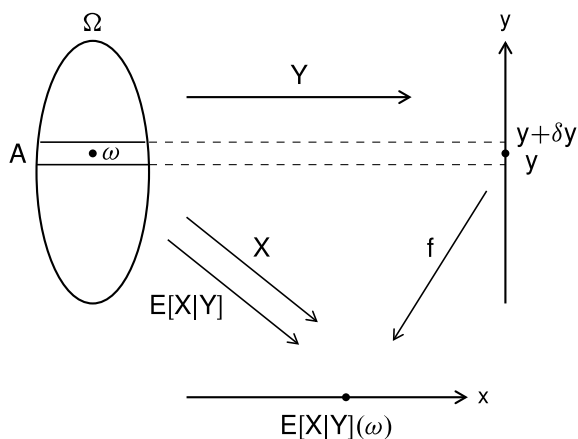
For the general case when Y is continuous, choose a sequence of partitions $-\infty < \cdots < y_{i-1}^{(n)} < y_i^{(n)} < y_{i+1}^{(n)} < \cdots < +\infty$ such that $\delta y_i^{(n)} = y_{i+1}^{(n)} - y_i^{(n)} \rightarrow 0$ as $n \rightarrow \infty$, and approximate Y by a sequence of simple functions

$$Y_n = \sum_{i=1}^n y_i^{(n)} \times \mathbb{1}_{A_i^{(n)}}$$

where $A_i^{(n)} = Y^{-1}([y_i^{(n)}, y_{i+1}^{(n)}))$, and find $\mathbb{E}[X|Y_n] = f_n \circ Y_n$, and finally take the limit of f_n as $n \rightarrow \infty$. $\square$

Remark 5.3 In Fig. 5.6 we are given two random variables X and Y . We assume that the subset $A = \{y \leq Y \leq y + \delta y\}$ has sufficiently small measure, in other words, $\delta y \approx 0$. If $\omega \in A$, then $\mathbb{E}[X|Y](\omega)$ is approximated by $\mathbb{E}[X|A]$. The inverse images under Y of short intervals are represented by horizontal strips in the sample space Ω , and they generate the σ -algebra $\sigma(Y)$. In the limiting case we regard the widths of these strips as being equal to 0, and the strips are treated as line segments. Since $\mathbb{E}[X|Y]$ is measurable with respect to $\sigma(Y)$, $\mathbb{E}[X|Y]$ is constant on each strip. Since each strip is identified with a point y on the real line via Y , we can define f naturally. Usually, instead of $f(y)$ we use the notation $\mathbb{E}[X|Y = y]$ and regard it as a function of y . Strictly speaking, $\mathbb{E}[X|Y = y]$ is a function defined on Ω satisfying $\mathbb{E}[X|Y = y] = \mathbb{E}[X|Y]|_{\{\omega: Y(\omega)=y\}}$.

Fig. 5.6 Conditional expectation of X with respect to Y



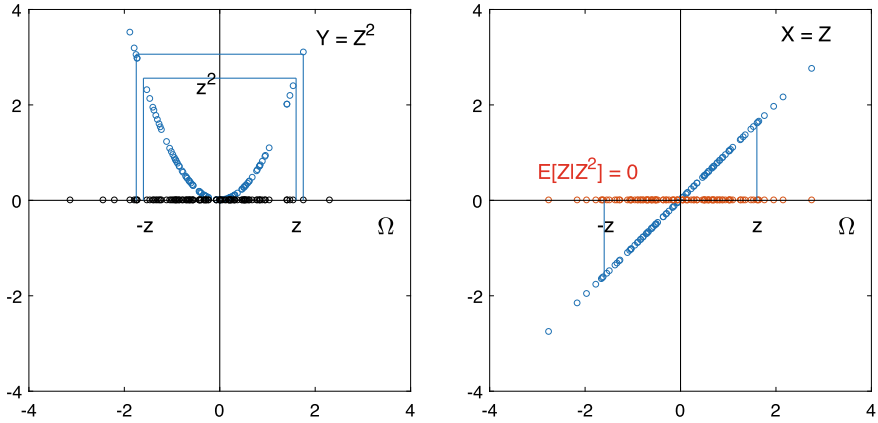


Fig. 5.7 Scatter plot of sample values of (Z^2, Z) marked by blue circles (*left*), and the plots of (Z, Z) marked by blue circles and $(Z, \mathbb{E}[Z|Z^2]=0)$ marked by red circles on symmetric sets $\{z, -z\}$ (*right*)

Example 5.3 (*Conditional Expectation Given a Random Variable*) We consider a standard normal variable Z , regarded as a function $X : \Omega \rightarrow \mathbb{R}^1$ defined by $X(\omega) = \omega$ on a probability space $\Omega = (\mathbb{R}^1, \phi(x)dx)$, where ϕ is the standard normal density function. (The probabilistic structure of Ω is represented by the scatter plot of $(Z, 0)$ marked by 100 black circles in the left plot of Fig. 5.7. The sample points are more concentrated near 0.) Since the inverse image of z^2 , $z \geq 0$, under $Y = Z^2$ equals $\{-z, z\}$, the σ -algebra $\sigma(Z^2)$ is generated by symmetric subsets $J = [-z - \delta, -z] \cup [z, z + \delta]$, $z \geq 0$, $\delta > 0$. (In the left plot the inverse image of an interval under $Y = Z^2$ is shown to be symmetric.) The average of Z on J equals zero, and hence $\mathbb{E}[Z|Z^2] = 0$. (In the right the averages of X on 50 subsets of the form $\{-z, z\}$ are marked by red circles.) See also Simulation 5.2.

5.4 Conditional Variance

Definition 5.4 (*Conditional Variance*) The *conditional variance* of X given Y is defined by $\text{Var}(X|Y) = \mathbb{E}[(X - \mathbb{E}[X|Y])^2|Y]$.

Theorem 5.6 *Let X and Y denote random variables. Then*

$$\text{Var}(X|Y) = \mathbb{E}[X^2|Y] - \mathbb{E}[X|Y]^2$$

and

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y]) .$$

Proof Note that $\text{Var}(X|Y) = \mathbb{E}[(X - \mathbb{E}[X|Y])^2|Y] = \mathbb{E}[X^2|Y] - 2\mathbb{E}[X|Y]^2 + \mathbb{E}[X|Y]^2 = \mathbb{E}[X^2|Y] - \mathbb{E}[X|Y]^2$. Hence

$$\mathbb{E}[\text{Var}(X|Y)] = \mathbb{E}[\mathbb{E}[X^2|Y]] - \mathbb{E}[\mathbb{E}[X|Y]^2] = \mathbb{E}[X^2] - \mathbb{E}[\mathbb{E}[X|Y]^2] .$$

On the other hand,

$$\text{Var}(\mathbb{E}[X|Y]) = \mathbb{E}[\mathbb{E}[X|Y]^2] - \mathbb{E}[\mathbb{E}[X|Y]]^2 = \mathbb{E}[\mathbb{E}[X|Y]^2] - \mathbb{E}[X]^2 .$$

Therefore, $\mathbb{E}[\text{Var}(X|Y)] + \text{Var}(\mathbb{E}[X|Y]) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. □

5.5 Computer Experiments

Simulation 5.1 (Monty Hall Problem) In a game show there are three closed doors: only one door has a car behind it, and behind the others, goats. A guest chooses one door, and if the car is behind it, they win it. Suppose that the guest picks a door, say Door 1. Before checking what is behind Door 1, the host opens another door, say Door 3, and shows there is a goat and asks if the guest wants to switch the doors, i.e., to choose Door 2 instead of Door 1. Does it make sense for the guest to choose Door 2? The host knows what is behind the doors, and opens a door with a goat. Prove that the guest should switch the doors to have a better chance of winning the prize.

If the guest sticks with the first choice, then the probability of having chosen the door with a car is $\frac{1}{3}$. On the other hand, if the door is switched then the probability increases to $\frac{2}{3}$. Here is why: Let C denote the door behind which there is a car, X the first choice made by the guest, H the door opened by the host. Since

$$\mathbb{P}(C = 2, X = 1, H = 3) = \mathbb{P}(C = 2|X = 1, H = 3)\mathbb{P}(H = 3|X = 1)\mathbb{P}(X = 1) ,$$

and since

$$\mathbb{P}(C = 2, X = 1, H = 3) = \mathbb{P}(H = 3|C = 2, X = 1)\mathbb{P}(C = 2|X = 1)\mathbb{P}(X = 1) ,$$

we have

$$\begin{aligned} & \mathbb{P}(C = 2|X = 1, H = 3)\mathbb{P}(H = 3|X = 1) \\ &= \mathbb{P}(H = 3|C = 2, X = 1)\mathbb{P}(C = 2|X = 1) , \end{aligned}$$

and finally

$$\begin{aligned}\mathbb{P}(C = 2|X = 1, H = 3) &= \frac{\mathbb{P}(H = 3|C = 2, X = 1)\mathbb{P}(C = 2|X = 1)}{\mathbb{P}(H = 3|X = 1)} \\ &= \frac{1 \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{2}{3}.\end{aligned}$$

```

N = 1000; % number of trials
prize = randi([0,2],N,1);
% N random integers from the set {0,1,2} for the location of the car
% Doors are numbered 0,1,2.

first = randi([0,2],N,1);
% N random integers from the set {0,1,2} representing the first choice made

second = zeros(N,1);
% the second choice to be made by switching

for i = 1:N
    if first(i) == prize(i)
        k=randi([1,2],1,1); % Randomly choose 1 or 2.
        % The host opens the door with number mod((3-k) + first(i),3)
        % out of two remaining doors without prize.
        second(i) = mod( k + first(i),3); % The guest switches the doors.
    else
        second(i) = prize(i);
        % Switching the doors always wins a prize in this case.
    end
end
n = 0; % number of right answers;
for i = 1:N
    if prize(i) == second(i)
        n = n + 1;
    end
end
fprintf('Probability of winning if door is switched = %f\n',n/N)

```


Simulation 5.2 (Conditional Expectation) Plot the inverse images of the interval $[z, z + \delta]$ under $Y = Z^2$. For the output see Fig. 5.7.

```
M = 100; % number of samples
X = randn(M,1);
Y = X.^2;
z = 1.6;
figure(1)
scatter(X,Y,15);
hold on
scatter(X,zeros(M,1));
text(z,0,'z');
text(-z,0,'-z');
text(3.3,-0.5,'\Omega');
text(2,3.5,'Y = Z^2');
text(0,z^2,'z^2');
hold off

M0 = M/2;
X0 = randn(M0,1);
X = [X0; -X0];
figure(2)
scatter(X,X);
hold on
scatter(X,zeros(M,1));
text(2.1,3.5,'X = Z');
text(-3,0.7,'E[Z|Z^2] = 0', 'Color', 'r');
```

Exercises

5.1 Let $\Omega = \{a, b, c\}$ and let $X : \Omega \rightarrow \mathbb{R}^1$ be a random variable defined by $X(a) = 0, X(b) = X(c) = 1$.

- (i) What is the σ -algebra generated by X ?
- (ii) If Y is defined by $Y(a) = 0, Y(b) = 1, Y(c) = 2$, then what is $\mathbb{E}[X|Y]$?
- (iii) Prove or disprove $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$.

5.2 Let $\Omega = \{a, b, c, d\}$, let $\mathcal{F}$ be the σ -algebra consisting of all the subsets of Ω , and let $\mathcal{G}$ be the sub- σ -algebra generated by $\{a, b\}$ and $\{c, d\}$.

- (i) List all of the measurable subsets belonging to $\mathcal{G}$.
- (ii) Let $Z : \Omega \rightarrow \mathbb{R}^1$ be the function defined by $Z(a) = Z(b) = Z(c) = Z(d) = 1$. Is Z measurable with respect to $\mathcal{G}$?

- (iii) Suppose that $Y : \Omega \rightarrow \mathbb{R}^1$ is a $\mathcal{G}$ -measurable function such that $Y(a) = 5$. What are the possible values of $Y(b)$?
- (iv) Let $X : \Omega \rightarrow \mathbb{R}^1$ be a random variable defined by $X(a) = 0, X(b) = X(c) = 3$ and $X(d) = 1$. What is the σ -algebra generated by X ?
- (v) Let X be the random variable given in (iv). Let $W : \Omega \rightarrow \mathbb{R}^1$ be a random variable defined by $W(a) = 10, W(b) = W(c) = W(d) = 20$. Find $\mathbb{E}[W|X]$.

5.3 Let $\Omega = \{uu, ud, du, dd\}$ and let $\mathcal{F}$ be the σ -algebra consisting of all the subsets of Ω . Let $X : \Omega \rightarrow \mathbb{R}^1$ be a random variable defined by $X(uu) = 5, X(ud) = X(du) = 3$ and $X(dd) = 1$.

- (i) What is the sub- σ -algebra generated by X ?
- (ii) Let $Y : \Omega \rightarrow \mathbb{R}^1$ be a random variable defined by $Y(uu) = 1, Y(ud) = Y(du) = Y(dd) = -1$. Find $\mathbb{E}[Y|X]$.

5.4 Let $\Omega = [0, 1]$ with $\mathbb{P}$ the Lebesgue measure on $[0, 1]$. Suppose that X and Y are random variables on $(\Omega, \mathbb{P})$, and $Y(\omega) = \omega(1 - \omega)$. Find $\mathbb{E}[X|Y]$. (We assume that all the integrals under consideration exist.)

5.5 Show that Example 5.2 can be interpreted as a linear algebra problem. (Hint: Note that $L^2(\Omega, \mathcal{F}, \mathbb{P})$ can be identified with $\mathbb{R}^3$, and $L^2(\Omega, \mathcal{G}, \mathbb{P})$ with a subspace $\{(a, a, b) : a, b \in \mathbb{R}\}$.)

5.6 Suppose that X and Y have the joint density $f_{X,Y}(x, y) = x + y$ for any $x, y \in [0, 1]$ and $f_{X,Y}(x, y) = 0$ otherwise. Find $\mathbb{E}[X|Y]$.

5.7 (Bayes' Theorem) Let $B_1, \dots, B_n$ be a partition of a probability space $(\Omega, \mathbb{P})$ such that $\mathbb{P}(B_i) > 0, 1 \leq i \leq n$, and take $A \subset \Omega, \mathbb{P}(A) > 0$. Prove that

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{k=1}^n \mathbb{P}(A|B_k)\mathbb{P}(B_k)} \quad \text{for } 1 \leq i \leq n.$$

(Hint: See Fig. 5.8 where Ω is represented by the unit square and A is the region enclosed by thick line segments in the right.)

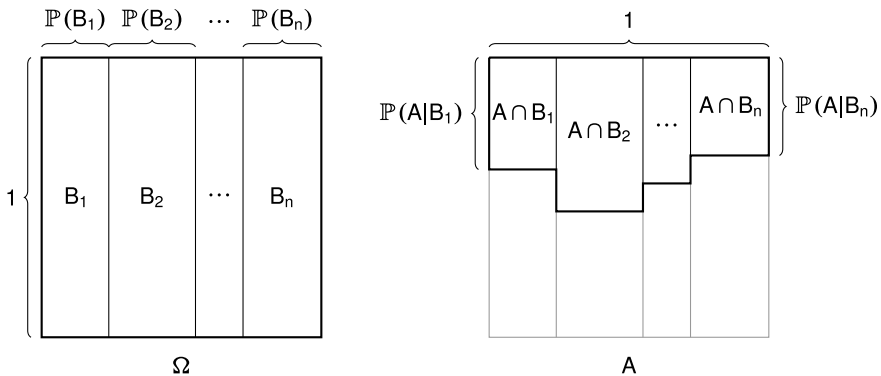


Fig. 5.8 A diagram for Bayes' theorem

5.8 Consider a test of a certain disease caused by a virus which infects one out of a thousand people on average. The test method is effective, but not perfect. For a person who has already been infected there is a 95% chance of detection, and for a healthy person there is a 1% probability of false alarm. Suppose that a person, who is not yet known to be infected or not, has a positive test result. What is the probability of his being truly infected?

Part III
Stochastic Processes and Brownian Motion

Chapter 6

Stochastic Processes



We collect and analyze sequential data from nature and society, especially from the financial market, in the form of numerical sequences, and try to predict what will happen next. The observed data is regarded as a realization of a sequence of random variables, called a stochastic process. Due to uncertainty of payoff in financial investment before maturity date, the theory of stochastic processes has become the language of mathematical finance. In this chapter we introduce basic concepts of stochastic processes such as martingales and the first hitting time.

6.1 Stochastic Processes

Definition 6.1 (*Filtration*) Let Ω be a measurable space with a σ -algebra $\mathcal{F}$. Consider a collection of sub- σ -algebras $\{\mathcal{F}_t\}_{t \in I}$ of $\mathcal{F}$, indexed by $I \subset \mathbb{R}$. (For example, $I = \{0, 1, \dots, n\}$, $I = \mathbb{N}$, $I = [0, T]$ and $I = [0, \infty)$). The parameter or index t represents time in general.) If $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for $s, t \in I$ such that $s \leq t$, then $\{\mathcal{F}_t\}_{t \in I}$ is called a *filtration*. Unless stated otherwise, we assume that 0 is the smallest element in I and $\mathcal{F}_0 = \{\emptyset, \Omega\}$.

A filtration is interpreted as a monotone increment of information as time passes by. As an example, consider the game of twenty questions. One of the players of the game is chosen as ‘answerer’, and the remaining players ask him/her up to twenty questions. The answerer chooses an object that is not revealed to other players until they correctly guess the answer. The answerer can answer ‘Yes’ or ‘No’, and after twenty questions are asked, the players present their guess. To model the game mathematically, we define the sample space Ω as the set of all objects in the world and after each question is asked we increase the information on the object $\omega \in \Omega$. Each question X_n is something like ‘Is ω alive?’, or ‘Is ω visible?’, and it may be regarded as a random variable $X_n : \Omega \rightarrow \{0, 1\}$ depending on the correctness of the guess. Depending on whether the answer to the question is correct or incorrect, we

have $X_n = 0$ or 1 , respectively. A wise player would ask questions in such a way that $\{\omega : X_n(\omega) = 0\}$ and $\{\omega : X_n(\omega) = 1\}$ have comparable probabilities. After n questions we obtain information $X_i = 0$ or 1 , $1 \leq i \leq n$, which defines a partition of Ω , or equivalently a σ -algebra $\mathcal{F}_n = \sigma(\{X_1, \dots, X_n\})$. As n increases, $\mathcal{F}_n$ also increases, and we see that $\mathcal{F}_n$, $1 \leq n \leq 20$, is a filtration. If we don't ask a wise question, i.e., if we don't define X_n wisely, then some subsets among 2^n subsets are empty and some others are too big, containing many objects, thus making the game difficult to win. If we ask twenty reasonable questions, the sample space is partitioned into $2^{20} \approx 1,000,000$ subsets, and each subset contains one and only one object in the world, corresponding to almost all the words in a dictionary, and we can find the correct answer.

Definition 6.2 (*Stochastic Process*)

- (i) A *stochastic process* is a sequence of random variables $X_t : \Omega \rightarrow \mathbb{R}$ parameterized by time t belonging to an index set $I \subset \mathbb{R}$. In other words, when a stochastic process $X : I \times \Omega \rightarrow \mathbb{R}$ is given, $X(t, \cdot) = X_t : \Omega \rightarrow \mathbb{R}$ is a measurable mapping for each $t \in I$.
- (ii) If I is a discrete set, then a process $\{X_t\}_{t \in I}$ is called a discrete time stochastic process, and if I is an interval then $\{X_t\}_{t \in I}$ is called a continuous time stochastic process.
- (iii) For each $\omega \in \Omega$ the mapping $t \mapsto X_t(\omega)$ is called a *sample path*.
- (iv) If almost all sample paths of a continuous time process are continuous, then we call the process a *continuous process*.
- (v) The filtration $\mathcal{F}_t$ generated by a process X_t , i.e., $\mathcal{F}_t = \sigma(\{X_s : 0 \leq s \leq t\})$ is called a *natural filtration* for X_t .

Definition 6.3 (*Adapted Process*) Consider a filtration $\{\mathcal{F}_t\}_{t \in I}$ and a stochastic process $\{X_t\}_{t \in I}$. If X_t is measurable with respect to $\mathcal{F}_t$ for every t , then $\{X_t\}_{t \in I}$ is said to be *adapted* to the filtration $\{\mathcal{F}_t\}_{t \in I}$.

Definition 6.4 (*Markov Property*) Let $\{X_t\}$ be a stochastic process and let $\mathcal{F}_t = \sigma(\{X_u : 0 \leq u \leq t\})$ be the sub- σ -algebra determined by the history of the process up to time t , i.e., $\mathcal{F}_t$ is generated by $X_u^{-1}(B)$ where $0 \leq u \leq t$ and B is an arbitrary Borel subset of $\mathbb{R}$. We say that $\{X_t\}$ has the *Markov property* if for every $0 \leq s \leq t$ the conditional probability of X_t given $\mathcal{F}_s$ is the same as the conditional probability of X_t given X_s , i.e.,

$$\mathbb{P}(X_t \leq y | \mathcal{F}_s) = \mathbb{P}(X_t \leq y | X_s)$$

with probability 1. In other words, if the present state of the process, X_s , is known, then its future movement is independent of its past history $\mathcal{F}_s$. A stochastic process with the Markov property is called a *Markov process*. The *transition probability* of a Markov process $\{X_t\}$ is defined by

$$\mathbb{P}(X_t \in (y, y + dy) | X_s = x),$$

which is the conditional probability of the process at time t to be found in a very short interval $(y, y + dy)$ given the condition that $X_s = x$.

Example 6.1 (Random Walk) Let $Z_n, n \geq 1$, be a sequence of independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(Z_n = 1) = p$ and $\mathbb{P}(Z_n = -1) = 1 - p$ for some $0 \leq p \leq 1$. A one-dimensional *random walk* $X_0, X_1, X_2, \dots$ is defined by $X_0 = 0, X_n = Z_1 + \dots + Z_n, n \geq 1$. Let $\mathcal{F}_0$ be the trivial σ -algebra $\{\emptyset, \Omega\}$ and let $\mathcal{F}_n = \sigma(Z_1, \dots, Z_n)$ be the sub- σ -algebra generated by $Z_1, \dots, Z_n$. If $p = \frac{1}{2}$, the process $\{X_n\}_{n \geq 0}$ is called a *symmetric random walk*.

Example 6.2 (Coin Tossing) Take

$$\Omega = \prod_{n=1}^{\infty} \{0, 1\} = \{\omega = \omega_1 \omega_2 \omega_3 \dots \mid \omega_n = 0, 1\},$$

which is the set of all possible outcomes of coin tossing with the symbols 0 and 1 representing heads and tails, respectively, and may be identified with the unit interval via the binary representation $\omega = \sum_{n=1}^{\infty} \omega_n 2^{-n}$. Define $Z_n(\omega) = (-1)^{\omega_n}$. Then $\mathcal{F}_n$ is the σ -algebra generated by the blocks or the cylinder sets of length n defined by $[a_1, \dots, a_n] = \{\omega : \omega_1 = a_1, \dots, \omega_n = a_n\}$ whose probability is equal to $p^k(1-p)^{n-k}$, where k is the number of 0's among $a_1, \dots, a_n$. Then by the Kolmogorov Extension Theorem we obtain a probability measure on arbitrary measurable subsets of Ω . If we regard Ω as the unit interval, then $\mathcal{F}_n$ is generated by subintervals of the form $[\frac{i-1}{2^n}, \frac{i}{2^n})$, $1 \leq i \leq 2^n$. If $p = \frac{1}{2}$, we have a fair coin tossing and the probability measure on $[0, 1]$ is Lebesgue measure. If $p \neq \frac{1}{2}$, the coin is biased and the corresponding measure on $[0, 1]$ is singular. For more information on singular measures, consult (Choe, 2005).

6.2 Predictable Processes

We cannot receive information that influences the financial market and act on it simultaneously. There should be some time lag, however short, before we decide on an investment strategy. This idea is reflected in the definition of predictability or previsibility of a stochastic process.

Definition 6.5 (Predictable Process)

- (i) (Discrete time) If a process $\{X_n\}_{n=0}^{\infty}$ is adapted to a filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$, and if X_n is measurable with respect to $\mathcal{F}_{n-1}$ for every $n \geq 1$, then $\{X_n\}_{n=0}^{\infty}$ is said to be *predictable*.

- (ii) (Continuous time) If a process $\{X_t\}_{t \geq 0}$ is adapted to a filtration $\{\mathcal{F}_t\}_{t \geq 0}$, and if X_t is measurable with respect to a sub- σ -algebra $\mathcal{F}_{t-}$ defined by

$$\mathcal{F}_{t-} = \sigma\left(\bigcup_{0 \leq s < t} \mathcal{F}_s\right)$$

for every t , then $\{X_t\}_{t \geq 0}$ is said to be *predictable*.

Example 6.3 (i) (Discrete time) Let $\mathcal{S} = \{s_1, s_2, \dots\}$ be a set of symbols and let $\Omega = \prod_{n=1}^{\infty} \mathcal{S}$ be the set of all infinite sequences of the symbols in $\mathcal{S}$. Let $\mathcal{F}_n$ be a σ -algebra on Ω generated by the cylinder sets of the form

$$[a_1, \dots, a_n] = \{(s_{i_1}, s_{i_2}, \dots) \in \Omega : s_{i_k} = a_k, 1 \leq k \leq n\}.$$

Note that $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$ if and only if X_n is of the form

$$X_n(s_{i_1}, \dots, s_{i_n}, \dots) = X_n(s_{i_1}, \dots, s_{i_n})$$

as a function defined on $\prod_{i=1}^n \mathcal{S}$, i.e., its values depend only on the first n coordinates. Also note that X_n is predictable if and only if X_n is of the form

$$X_n(s_{i_1}, \dots, s_{i_{n-1}}, \dots) = X_n(s_{i_1}, \dots, s_{i_{n-1}})$$

as a function defined on $\prod_{i=1}^{n-1} \mathcal{S}$, i.e., its values depend only on the first $n - 1$ coordinates.

- (ii) (Continuous time) It is known that a continuous process is predictable. For example, a Brownian motion is predictable.

6.3 Martingales

Definition 6.6 (*Martingale*) Suppose that a stochastic process $\{X_t\}_{t \in I}$ is adapted to a filtration $\{\mathcal{F}_t\}_{t \in I}$, and that X_t is integrable for every t , i.e., $\mathbb{E}[|X_t|] < \infty$. If

$$X_s = \mathbb{E}[X_t | \mathcal{F}_s]$$

for arbitrary $s \leq t$, then $\{X_t\}_{t \in I}$ is called a *martingale* with respect to $\{\mathcal{F}_t\}_{t \in I}$. If $X_s \leq \mathbb{E}[X_t | \mathcal{F}_s]$ for $s \leq t$, then it is called a *submartingale*, and if $X_s \geq \mathbb{E}[X_t | \mathcal{F}_s]$ for $s \leq t$, then it is a *supermartingale*.

A martingale is both a submartingale and a supermartingale. A stochastic process that is both a submartingale and a supermartingale is a martingale.

Remark 6.1 If a martingale $\{X_n\}_{n \geq 0}$ is *increasing*, i.e., $X_n \leq X_{n+1}$ for $n \geq 0$, then it is constant. For the proof, first note that

$$\mathbb{E}[X_n - X_{n-1} | \mathcal{F}_{n-1}] = \mathbb{E}[X_n | \mathcal{F}_{n-1}] - \mathbb{E}[X_{n-1} | \mathcal{F}_{n-1}] = X_{n-1} - X_{n-1} = 0.$$

Since $X_n - X_{n-1} \geq 0$, we have $X_n - X_{n-1} = 0$, and hence

$$X_n = X_{n-1} = \cdots = X_1 = X_0.$$

Since $X_0 = \mathbb{E}[X_1 | \mathcal{F}_0]$, X_n is the constant X_0 for $n \geq 0$.

Theorem 6.1 If $\{X_t\}_{t \geq 0}$ is a martingale with respect to a filtration $\{\mathcal{F}_t\}_{t \geq 0}$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$, then $\mathbb{E}[X_t] = \mathbb{E}[X_0]$ for every t .

Proof For $0 \leq s \leq t$, we have

$$\mathbb{E}[X_t] = \mathbb{E}[X_t | \mathcal{F}_0] = \mathbb{E}[\mathbb{E}[X_t | \mathcal{F}_s] | \mathcal{F}_0] = \mathbb{E}[X_s | \mathcal{F}_0] = \mathbb{E}[X_s].$$

Now take $s = 0$. □

Example 6.4 Let $X_n, n \geq 0$, be the symmetric random walk in Example 6.1. Since, for $j = k \pm 1$,

$$\begin{aligned} \mathbb{P}(X_{n+1} = j | X_n = k) &= \mathbb{P}(k + Z_{n+1} = j | X_n = k) \\ &= \mathbb{P}(Z_{n+1} = j - k | X_n = k) \\ &= \mathbb{P}(Z_{n+1} = j - k) = \frac{1}{2}, \end{aligned}$$

we have

$$\mathbb{E}[X_{n+1} | X_n = k] = \mathbb{E}[k + Z_{n+1} | X_n = k] = k + \mathbb{E}[Z_{n+1} | X_n = k] = k.$$

Hence $\mathbb{E}[X_{n+1} | X_n = k] = X_n$ on the subset $\{\omega \in \Omega : X_n = k\}$ for every k . Thus $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n$ and $\{X_n\}_{n \geq 0}$ is a martingale. See Fig. 6.1 where the probability space Ω is partitioned into the subsets $\{\omega \in \Omega : X_n = k\}$, $-n \leq k \leq n$, for $n \geq 1$. Note that the averages of X_{n+1} over those subsets are equal to the values of X_n .

Remark 6.2 The name ‘martingale’ comes from an old French gambling game called ‘martingale’. We bet one dollar that the first flip of a coin comes up heads. If it does come up with heads then we stop, having won one dollar. Otherwise, we bet two dollars that the second toss comes up heads. If it does, then the net gain is one dollar, and we stop. Otherwise, we have lost three dollars and this time we bet $2^2 = 4$ dollars that the third flip is a head. And so on. If the first $n - 1$ flips all come up tails, then we have lost $\sum_{j=0}^{n-1} 2^j = (2^n - 1)$ dollars and we bet 2^n dollars on the n th toss. Since the coin will eventually come up heads with probability 1, we are

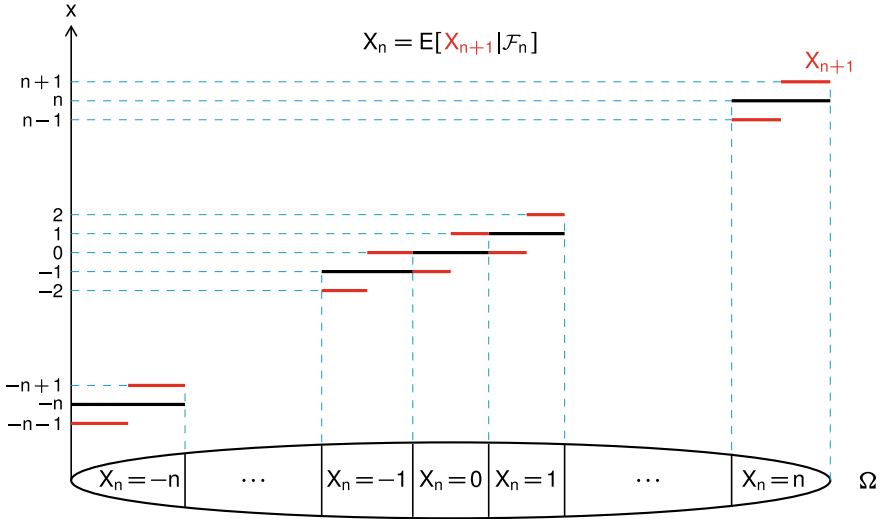


Fig. 6.1 Martingale property of the symmetric random walk

guaranteed to win one dollar. This works if we have an infinite amount of funds. See also Example 6.10.

Theorem 6.2 Let X_t be a martingale. If ϕ is a convex function such that $\phi(X_t)$ is integrable for every t , then $\phi(X_t)$ is a submartingale.

Proof By Theorem 5.3, $\mathbb{E}[\phi(X_t)|\mathcal{F}_s] \geq \phi(\mathbb{E}[X_t|\mathcal{F}_s])$, $s < t$, almost surely for a stochastic process X_t . If X_t is a martingale, then $\phi(\mathbb{E}[X_t|\mathcal{F}_s]) = \phi(X_s)$, $s < t$, almost surely. $\square$

Example 6.5 Let X_t be a martingale. Then $|X_t|$, X_t^2 , $X^+ = \max\{X, 0\}$ and e^X are submartingales since $|x|$, x^2 , $\max\{x, 0\}$ and e^x are convex functions. (We assume that all the integrals under consideration exist.)

Example 6.6 Let X_n be the symmetric random walk in Example 6.1. Then $X_n^2 - n$ is a martingale. For the proof, note that

$$\begin{aligned}
 \mathbb{E}[X_{n+1}^2 - (n+1)|\mathcal{F}_n] &= \mathbb{E}[(X_n + Z_{n+1})^2 - (n+1)|\mathcal{F}_n] \\
 &= \mathbb{E}[X_n^2 + 2X_n Z_{n+1} + Z_{n+1}^2 - (n+1)|\mathcal{F}_n] \\
 &= \mathbb{E}[X_n^2 + 2X_n Z_{n+1} + 1 - (n+1)|\mathcal{F}_n] \\
 &= \mathbb{E}[X_n^2 + 2X_n Z_{n+1} - n|\mathcal{F}_n] \\
 &= \mathbb{E}[X_n^2|\mathcal{F}_n] + \mathbb{E}[2X_n Z_{n+1}|\mathcal{F}_n] - \mathbb{E}[n|\mathcal{F}_n] \\
 &= X_n^2 + 2X_n \mathbb{E}[Z_{n+1}|\mathcal{F}_n] - n \\
 &= X_n^2 - n.
 \end{aligned}$$

For the equality just before the last one, we take out what is known at time n , and for the last equality we use

$$\mathbb{E}[Z_{n+1}|\mathcal{F}_n] = \mathbb{E}[Z_{n+1}] = 0$$

since Z_{n+1} and $\mathcal{F}_n$ are independent.

Remark 6.3 The continuous limit of a symmetric random walk is Brownian motion W_t . Hence the fact that $X_n^2 - n$ is a martingale corresponds to the fact that $W_t^2 - t$ is a martingale. See Theorem 7.8.

Definition 6.7 (*Quadratic Variation*) Let $\{X_t\}_{0 \leq t \leq T}$ be a continuous martingale. The *quadratic variation* process of X_t , denoted by $[X, X]_t$ or $[X]_t$, is defined by

$$[X, X]_t = \lim_{\max_j |t_j - t_{j-1}| \rightarrow 0} \sum_{j=1}^n |X_{t_j} - X_{t_{j-1}}|^2,$$

where the convergence is in probability. (For more information, consult (Oksendal, 2003).) Using the idea from the identity

$$||\mathbf{v} + \mathbf{w}||^2 = ||\mathbf{v}||^2 + ||\mathbf{w}||^2 + 2\mathbf{v} \cdot \mathbf{w}$$

for $\mathbf{v}, \mathbf{w} \in \mathbb{R}^k$, we define the *quadratic covariation* process $[X, Y]_t$ of two continuous square-integrable martingales X_t and Y_t by

$$[X, Y]_t = \frac{1}{2} ([X + Y, X + Y]_t - [X, X]_t - [Y, Y]_t).$$

It is known that $X_t^2 - [X]_t$ is a martingale.

6.4 Stopping Time

Consider the situation when an investor wants to reduce the risk of stock market crash by selling some shares of a stock, whose price at time t is denoted by S_t , when its value falls below a level b , $b < S_0$, i.e., the shares will be sold at time

$$\tau = \min\{n > 0 : S_{t_n} \leq b\}$$

if the stock price is observed at discrete times $0 < t_1 < t_2 < \dots$. If $\{n \geq 1 : S_{t_n} \leq b\} = \emptyset$, then $\tau = +\infty$ by convention. Since the decision to trade depends on information obtained by the time t_n , which is represented by $\mathcal{F}_n = \sigma(\{S_{t_k} : 0 \leq k \leq n\})$, we require the condition that $\{\tau = n\}$ is $\mathcal{F}_n$ -measurable.

Definition 6.8 (*Stopping Time*)

- (i) (Discrete time) Let $\{\mathcal{F}_n\}_{n \geq 0}$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A random variable $\tau : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{+\infty\}$ satisfying

$$\{\omega \in \Omega : \tau(\omega) \leq n\} \in \mathcal{F}_n, \quad n \geq 1,$$

is called a *stopping time*. This condition is equivalent to $\{\tau = n\} \in \mathcal{F}_n$ for every $n \geq 1$ since

$$\{\tau = n\} = \{\tau \leq n\} \cap \{\tau \leq n-1\}^c \in \mathcal{F}_n$$

and

$$\{\tau \leq n\} = \{\tau = 1\} \cup \dots \cup \{\tau = n\} \in \mathcal{F}_n.$$

- (ii) (Continuous time) Let $\{\mathcal{F}_t\}_{t \geq 0}$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A random variable $\tau : \Omega \rightarrow [0, \infty]$ satisfying

$$\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t \quad \text{for every } t \geq 0$$

is called a *stopping time*. (If we consider a finite maturity date T in financial applications, the filtration in the above may be defined only on $0 \leq t \leq T$.)

Remark 6.4 (*Stopped Process*) Let τ be a stopping time, and let $\{X_n\}_{n \geq 1}$ be a stochastic process. Here are some standard notations: $a \wedge b = \min\{a, b\}$, $(\tau \wedge n)(\omega) = \tau(\omega) \wedge n$ and

$$X_{\tau \wedge n} = X_n \mathbb{1}_{\{n < \tau\}} + X_\tau \mathbb{1}_{\{n \geq \tau\}}.$$

If X_n is adapted to a filtration $\mathcal{F}_n$, then $X_{\tau \wedge n}$ is also adapted to $\mathcal{F}_n$ since for a Borel subset $B \subset \mathbb{R}$ we have

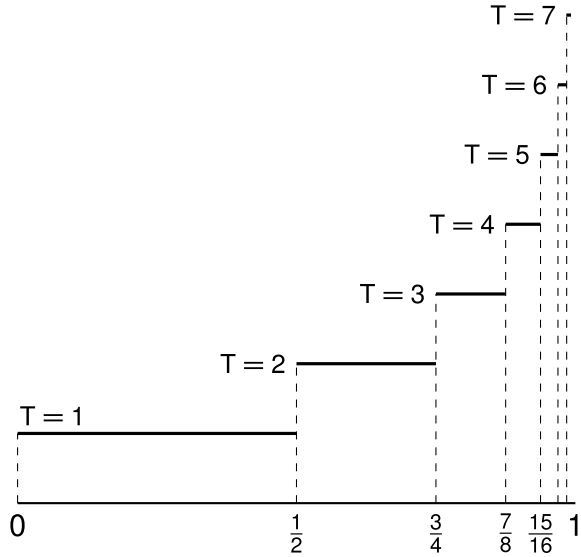
$$\{X_{\tau \wedge n} \in B\} = \{X_n \in B, \tau > n\} \cup \left(\bigcup_{k=1}^n \{X_k \in B, \tau = k\} \right).$$

Example 6.7 (*Binary Expansion*) Let $\Omega = [0, 1)$ be the unit interval with its filtration $\{\mathcal{F}_n\}_{n \geq 0}$ defined by

$$\mathcal{F}_n = \sigma \left(\left\{ \left[\frac{i}{2^n}, \frac{i+1}{2^n} \right) : 0 \leq i \leq 2^n - 1 \right\} \right).$$

Note that $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Define $T : \Omega \rightarrow \{0, 1, 2, \dots\}$ by $T(\omega) = k$ if $\omega = \omega_1 \omega_2 \omega_3 \dots$ and $k = \min\{i \geq 1 : \omega_i = 0\}$. In other words, T is the first time that the digit in the binary expansion of ω is equal to 0 for the first time. For example, $T = 1$ on $[0, \frac{1}{2})$. See Fig. 6.2. There are points with two different binary expansions such as

Fig. 6.2 The first time T that the digit '0' appears in the binary expansion



$\omega = 0111 \dots = 1000 \dots$, but those points form a subset of probability zero. Except at such points, T is uniquely defined, and hence it is well-defined at almost every point. Since $\{\omega : T \leq n\}$ is $\mathcal{F}_n$ -measurable, T is a stopping time.

Definition 6.9 (*The First Hitting Time*) The first hitting time τ_B of a Borel set $B \subset \mathbb{R}$ by an adapted process $\{X_n\}_{n \geq 0}$ is defined by

$$\tau_B(\omega) = \min\{n \geq 0 : X_n(\omega) \in B\}.$$

It is a stopping time since for every $n \geq 0$ we have

$$\{\omega : \tau_B(\omega) \leq n\} = \bigcup_{k \leq n} \{\omega : X_k(\omega) \in B\} \in \mathcal{F}_n.$$

For the continuous time case, we define the first hitting time τ_B by

$$\tau_B(\omega) = \inf\{t \geq 0 : X_t(\omega) \in B\}.$$

Example 6.8 Consider a random walker starting from $0 \in \mathbb{Z}$. Toss a coin, and he moves to the right or left by the distance 1 depending on the outcome 0 and 1 (for heads and tails) until he arrives at $-5 \in \mathbb{Z}$ or $5 \in \mathbb{Z}$. Let X_n denote the position of the random walker at time n . Note that $X_0 = 0$. Define

$$\tau(\omega) = \min\{n : X_n(\omega) = -5 \text{ or } 5\}$$

where $\omega \in \Omega = \prod_{n=1}^{\infty} \{0, 1\}$, which is identified with $[0, 1]$ as in Example 6.7. As before, the filtration is given by $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Note that

$$\{\tau = 1\} = \{\tau = 2\} = \{\tau = 3\} = \{\tau = 4\} = \emptyset$$

and $\tau(\omega) = 5$ if and only if $\omega = 00000 \dots$ or $\omega = 11111 \dots$, i.e.,

$$\omega \in \left[0, \frac{1}{2^5}\right] \cup \left[1 - \frac{1}{2^5}, 1\right].$$

Then τ is a stopping time. To see why, note that

$$\{\tau = n\} = \left(\bigcap_{k=1}^{n-1} \{-5 < X_k < 5\} \right) \cap \{X_n = \pm 5\},$$

which is an intersection of two subsets belonging to $\mathcal{F}_{n-1}$ and $\mathcal{F}_n$.

Theorem 6.3 (*Stopped Martingale*) Let τ be a stopping time, and let $\{X_t\}$ be a martingale. Then $\{X_{t \wedge \tau}\}$ is also a martingale.

Proof Consult Rogers and Williams (2000a), Rogers and Williams (2000b). $\square$

Example 6.9 As a corollary of Theorem 6.3, we have $\mathbb{E}[X_{\tau \wedge n}] = \mathbb{E}[X_0]$. Recall that the symmetric random walk $\{X_n\}$ on $\mathbb{Z}$ starting from 0 moving to the next position by the distance ± 1 is a martingale. Let τ be the stopping time defined by $\tau = \inf\{n : X_n = 1\}$. Since $X_\tau = 1$, we have $\mathbb{E}[X_{\tau \wedge n}] = \mathbb{E}[X_0] = 0 < 1 = \mathbb{E}[X_\tau]$.

Example 6.10 (*Martingale*) A fair coin is tossed repeatedly, and the outcome is given by $Z_i = \pm 1, i \geq 1$, representing a win or a loss. A strategy, or a betting system, is defined by $\beta_n = 2^{n-1}$ dollars if $Z_1 = \dots = Z_{n-1} = -1$ ($n-1$ consecutive losses), and $\beta_n = 0$, otherwise. Note that the n th bet is placed after the $n-1$ observations. Let

$$Y_n = Z_1 + 2Z_2 + \dots + 2^{n-1}Z_n.$$

Define a stopping time $\tau = \min\{n : Y_n = 1\}$. Then $\tau < \infty$ almost surely, and τ is the total number of wins if we play until we collect one dollar, and $Y_{\tau \wedge n}$ is the total amount of winnings after n plays. It is a martingale. Note that for $n-1$ consecutive losses from the beginning, we have

$$Y_\tau(\omega) = -1 - 2 - \dots - 2^{n-2} + 2^{n-1} = 1,$$

and hence $Y_\tau = 1$ almost surely. This game was called ‘martingale’, and is the origin of the mathematical term.

Definition 6.10 (*Local Martingale*) Let $\{X_t\}_{0 \leq t < \infty}$ be a (continuous) stochastic process. Suppose that there exists a nondecreasing sequence of stopping times $\tau_n, n \geq 1$, such that $\{X_{t \wedge \tau_n}\}_{0 \leq t < \infty}$ is a martingale for every $n \geq 1$ and $\lim_{n \rightarrow \infty} \tau_n = \infty$ almost surely. Then we call $\{X_t\}_{0 \leq t < \infty}$ a (continuous) *local martingale*.

Remark 6.5 (i) A martingale $\{M_t\}_{0 \leq t < \infty}$ is a local martingale. For, if we take a localizing sequence $\tau_n = n, n \geq 1$, then $\{M_{t \wedge \tau_n}\}_{0 \leq t < \infty}$ is a martingale for every $n \geq 1$ by Theorem 6.3.

(ii) $\mathbb{E}[X_t]$ need not exist for a local martingale X_t .

6.5 Computer Experiments

Simulation 6.1 (Estimation of Exponential) Let $N = \min\{n \geq 2 : U_1 \leq U_2 \leq \dots \leq U_{n-1} > U_n\}$ where $U_i, i \geq 1$, are independent and uniformly distributed in $[0, 1]$. We check the fact that the expectation of N equals $e = 2.7183 \dots$ (For a proof, consult Exercise 6.5.)

```
num = 10^7; % number of iterations
N = ones(num,1);
for j=1:num
    count = 2;
    U1 = rand;
    U2 = rand;
    while (U1 <= U2)
        U1 = U2;
        U2 = rand;
        count = count + 1;
    end
    N(j) = count;
end
ave = mean(N)
```

Simulation 6.2 (Estimation of Exponential) Let $N = \min\{n : U_1 + \dots + U_n > 1\}$ where $U_i, i \geq 1$, are independent and uniformly distributed in $[0, 1]$. We check the fact that the expectation of N equals $e = 2.7183 \dots$ (For a proof, consult Exercise 6.6.)

```
num = 10^7; % number of iterations
N = zeros(num,1);
for j=1:num
    count = 0;
    total = 0;
    while(total <= 1)
```

```

        total = total + rand;
        count = count + 1;
    end
    N(j) = count;
end
ave = mean(N)

```

Exercises

6.1 Let $s < t < u$. Show that $P(X_u \leq y \mid X_s, X_t) = P(X_u \leq y \mid X_t)$ for a Markov process $\{X_t\}$.

6.2 Let $Z_1, Z_2, Z_3, \dots$ be independent and identically distributed random variables such that $\Pr(Z_n = 1) = p > \frac{1}{2}$ and $\Pr(Z_n = -1) = q = 1 - p$. Using the notations in Example 6.1, let $X_0 = 0$, $X_n = Z_1 + \dots + Z_n$, $n \geq 1$, be a random walk. Prove that X_n is not a martingale.

6.3 Let $M_0, M_1, M_2, \dots$ be a martingale. Show that the stochastic process $\exp(M_0), \exp(M_1), \exp(M_2), \dots$ is a submartingale. (We assume that all the integrals under consideration exist.)

6.4 Suppose that $|X_t| \leq Y$ for every t where $Y \geq 0$ and $\mathbb{E}[Y] < \infty$. Show that $\{X_t\}_t$ be uniformly integrable.

6.5 Let $N = \min\{n \geq 2 : U_1 \leq U_2 \leq \dots \leq U_{n-1} > U_n\}$ where $U_i, i \geq 1$, are independent and uniformly distributed in $[0, 1]$.

- (i) Prove that $\mathbb{P}\{N > n\} = \frac{1}{n!}, n \geq 1$.
- (ii) Show that $\mathbb{E}[N] = e$.

6.6 Let $U_i, i \geq 1$, be independent and uniformly distributed in $[0, 1]$. Define $N = \min\{n : U_1 + \dots + U_n > 1\}$. Show that $\mathbb{E}[N] = e$.

6.7 Given a stochastic process $\{X_t\}_{t \geq 0}$, let $\mathcal{G}_t$ be the σ -algebra generated by $\{X_s : 0 \leq s \leq t\}$. Prove that if X_t is a martingale with respect to some filtration $\{\mathcal{H}_t\}_{t \geq 0}$, then X_t is also a martingale with respect to the filtration $\{\mathcal{G}_t\}_{t \geq 0}$.

6.8 Assume that $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Show that if a discrete time martingale $\{X_n\}$ is predictable, then it is constant.

6.9 If τ_1 and τ_2 are stopping times with respect to a filtration $\{\mathcal{F}_n\}_{n \geq 0}$, then we define $(\tau_1 \vee \tau_2)(\omega) = \max\{\tau_1(\omega), \tau_2(\omega)\}$. Show that $\tau_1 \vee \tau_2$ is also a stopping time.

Chapter 7

Brownian Motion



In 1827, while studying pollen under a microscope, Scottish botanist Robert Brown (1773–1858) discovered continuous random movements of minute particles suspended in water. The theoretical properties of this motion, now called Brownian motion, were studied by French mathematician Louis Bachelier (1870–1946) in 1900 to compute option prices. Albert Einstein (1879–1955) also investigated Brownian motion in 1905. Later, an American mathematician Norbert Wiener (1894–1964) gave a rigorous framework. For the English translation of Bachelier’s work, see Davis and Etheridge (2006). While the Brownian motions found in nature should have bounded speed, the mathematical model allows unbounded speed of transition from one location to another with small but positive probability. Therefore, when we need to distinguish the mathematical definition of Brownian motion from physical Brownian motion, we may use the terminology Wiener process instead of Brownian motion. On most occasions in this book we use the terminology Brownian motion. To denote a Brownian motion we use two notations $W(t)$ and W_t interchangeably. In this chapter we introduce the basic properties of Brownian motion needed in mathematical finance, including the Law of Large Numbers for Brownian motion. We are more interested in computer simulations of Brownian motion than a deeper theoretical exposition. For a general introduction to Brownian motion and various equivalent definitions, consult (Billingsley 2012; Durrett 2019; Karatzas and Shreve 1998).

7.1 Brownian Motion as a Stochastic Process

We consider only one-dimensional Brownian motion in this chapter. Brownian motion as an idealized physical phenomena can be described as an axiomatic mathematical system. Assume that at time $t = 0$ a particle is located at $0 \in \mathbb{R}^1$, and as

time passes by it continuously moves in the positive or negative direction at random. The probability that the particle is found in the interval $[a, b]$ at time $t > 0$ is given by

$$\int_a^b \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)} dx .$$

Now, we present an axiomatic definition of a Brownian motion, and in the next section we introduce a sample space consisting of continuous curves $\omega(t)$ defined on $[0, \infty)$ such that $\omega(0) = 0$, and call the associated stochastic process W_t , $W_t(\omega) = \omega(t)$ for $t \geq 0$, a Brownian motion.

Definition 7.1 (*Brownian Motion*) A stochastic process W_t , $t \geq 0$, is called a Brownian motion if it has the following properties:

- (i) $W_0 = 0$ and $t \mapsto W_t$, $t \geq 0$, is continuous with probability 1.
- (ii) If $0 \leq t_0 \leq t_1 \leq \dots \leq t_n$, then

$$W(t_1) - W(t_0), \dots, W(t_n) - W(t_{n-1})$$

are independent.

- (iii) For $0 \leq s \leq t$ the increment $W_t - W_s$ is normally distributed with mean 0 and variance $t - s$.

The existence of such processes will be proved in Sect. 7.2.

Remark 7.1 The transition probability density of moving from a fixed point x at time $s \geq 0$ to another point y at time t , $s < t$, is given by

$$p(t - s; x, y) = \frac{1}{\sqrt{2\pi(t - s)}} \exp\left(-\frac{(x - y)^2}{2(t - s)}\right), \quad (7.1)$$

which has variance $t - s$ and mean x .

In Fig. 7.1 twenty sample paths of Brownian motion are given together with the parabola $t = W^2$ (left) and the probability density functions for W_t , $t = 1, \dots, 10$, which spread out over time (right). Consult Simulation 7.1.

In Fig. 7.2 sample paths of Brownian motion are given for $0 \leq t \leq 10^{-8}$ (left) and for $0 \leq t \leq 10^8$ (right), respectively. Compare the scalings. We observe that Brownian motion has huge increments over a very short time period (left) and that its values are almost zero compared with time t for large t (right). See Remark 7.9 and Corollary 7.2, respectively.

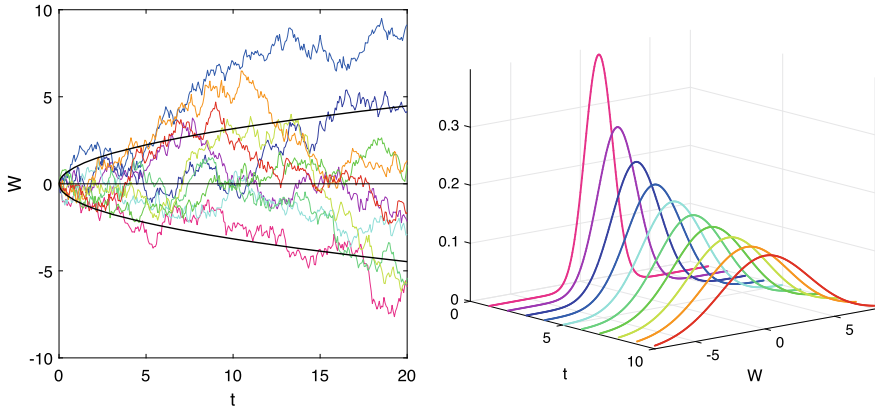


Fig. 7.1 Sample paths of Brownian motion together with the parabola $t = W^2$ (left) and the probability density functions for W_t as t increases (right)

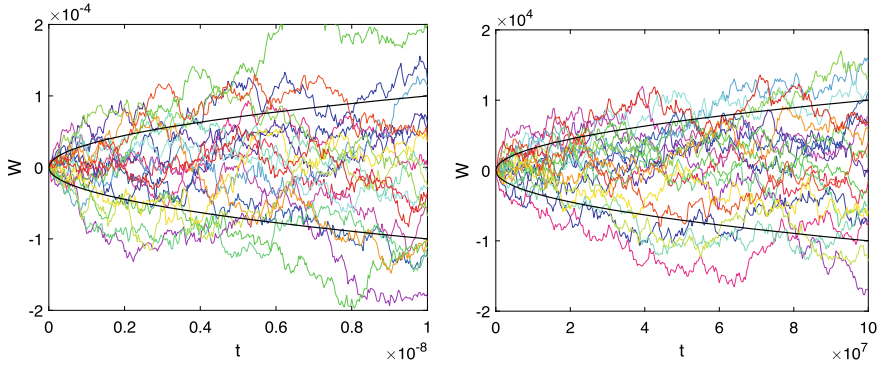


Fig. 7.2 Sample paths of Brownian motion for $0 \leq t \leq 10^{-8}$ and for $0 \leq t \leq 10^8$

Theorem 7.1 For $0 = t_0 < t_1 < \dots < t_n$ and for intervals $I_1, \dots, I_n \subset \mathbb{R}$, put $\delta t_j = t_j - t_{j-1}$. Then

$$\begin{aligned} & \mathbb{P}(W_{t_1} \in I_1, \dots, W_{t_n} \in I_n) \\ &= \int_{I_1} \dots \int_{I_n} p(\delta t_1; 0, x_1) p(\delta t_2; x_1, x_2) \dots p(\delta t_n; x_{n-1}, x_n) dx_n \dots dx_1. \end{aligned}$$

Theorem 7.2 (Invariance Under Time Translation) Given a Brownian motion $\{W_t\}_{t \geq 0}$ and a constant $a \geq 0$ put

$$B_t = W_{t+a} - W_a.$$

Then $\{B_t\}_{t \geq 0}$ is also a Brownian motion.

Proof First, note that $B_0 = 0$ and $t \mapsto B_t$ is continuous with probability 1. Since the increment of B_t satisfies $B_t - B_s = W_{t+a} - W_{s+a}$, which is also an increment of W_t , the rest of the conditions in Definition 7.1 also hold. $\square$

Theorem 7.3 (i) $\mathbb{E}[W_t W_s] = \min\{t, s\}$. As a special case, $\mathbb{E}[W_t^2] = t$.
(ii) $\mathbb{E}[(W_t - W_s)^2] = |t - s|$.

Proof (i) Assume that $s \leq t$. Then we have

$$\begin{aligned} \mathbb{E}[W_t W_s] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x y p(s, 0, x) p(t - s, x, y) dx dy \\ &= \int_{-\infty}^{\infty} x p(s, 0, x) \left(\int_{-\infty}^{\infty} y p(t - s, x, y) dy \right) dx \\ &= \int_{-\infty}^{\infty} x p(s, 0, x) x dx \\ &= \int_{-\infty}^{\infty} x^2 p(s, 0, x) dx \\ &= s . \end{aligned}$$

(ii) Note that

$$\mathbb{E}[W_t^2 - 2W_t W_s + W_s^2] = t - 2\min\{t, s\} + s = |t - s| .$$

Or, by Theorem 7.2, since $W_t - W_s$ and W_{t-s} are identically distributed for $s \leq t$, Part (i) implies that $\mathbb{E}[(W_t - W_s)^2] = \mathbb{E}[W_{t-s}^2] = t - s$. $\square$

Remark 7.2 Either by direct computation, or by comparing the coefficients of Taylor expansions of the equality in Lemma 7.1 we have $\mathbb{E}[W_t^n] = 0$ for n odd, $\mathbb{E}[W_t^2] = t$ and $\mathbb{E}[W_t^4] = 3t^2$.

Definition 7.2 (Quadratic Variation) Let $0 = t_0 < t_1 < \dots < t_n = T$ be a partition of the interval $[0, T]$ such that $\lim_{n \rightarrow \infty} \max_i \delta_i t = 0$, where $\delta_i t = t_{i+1} - t_i$. The quadratic variation of Brownian motion is defined by

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} |W(t_{i+1}) - W(t_i)|^2 ,$$

where the convergence is in the L^2 -sense. Put $\delta_i W = W_{t_{i+1}} - W_{t_i}$. Note that $\mathbb{E}[\sum_i |\delta_i W|^2] = \sum_i \delta_i t = T$.

Theorem 7.4 *The quadratic variation of Brownian motion is given by*

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}|^2 = T ,$$

where the convergence is in the L^2 -sense.

Proof Since $(\delta_i W)^2$ and $(\delta_j W)^2$ are independent for $i \neq j$, we have

$$\mathbb{E}[(\delta_i W)^2 (\delta_j W)^2] = \mathbb{E}[(\delta_i W)^2] \mathbb{E}[(\delta_j W)^2] .$$

By Remark 7.2, $\mathbb{E}[(\delta_i W)^4] = 3(\delta_i t)^2$ and $\mathbb{E}[(\delta_i W)^2] = \delta_i t$. Hence

$$\begin{aligned} & \mathbb{E} \left[\left| \sum_{i=0}^{n-1} (\delta_i W)^2 - T \right|^2 \right] \\ &= \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \mathbb{E}[(\delta_i W)^2 (\delta_j W)^2] - 2T \sum_{i=0}^{n-1} \mathbb{E}[(\delta_i W)^2] + T^2 \\ &= \sum_{i=0}^{n-1} \mathbb{E}[(\delta_i W)^4] + 2 \sum_{i < j} \mathbb{E}[(\delta_i W)^2 (\delta_j W)^2] - 2T \sum_{i=0}^{n-1} \mathbb{E}[(\delta_i W)^2] + T^2 \\ &= \sum_{i=0}^{n-1} 3(\delta_i t)^2 + 2 \sum_{i < j} (\delta_i t) (\delta_j t) - 2T \sum_{i=0}^{n-1} \delta_i t + T^2 \\ &= 2 \sum_{i=0}^{n-1} (\delta_i t)^2 + \left(\sum_{i=0}^{n-1} \delta_i t \right)^2 - 2T \sum_{i=0}^{n-1} \delta_i t + T^2 \\ &= 2 \sum_{i=0}^{n-1} (\delta_i t)^2 + T^2 - 2T^2 + T^2 \leq 2 \max_i \delta_i t \sum_{i=0}^{n-1} \delta_i t = 2 \max_i \delta_i t \times T \rightarrow 0 \end{aligned}$$

as $n \rightarrow \infty$, and the proof is complete. $\square$

Remark 7.3 If we take $\delta_i t = \delta t = \frac{T}{n}$ for every i , then

$$\mathbb{E} \left[\left| \sum_{i=0}^{n-1} (\delta_i W)^2 - T \right|^2 \right] = 2n \frac{T^2}{n^2} = \frac{2T^2}{n} .$$

Corollary 7.1 (The First Order Variation) *The first order variation of Brownian motion diverges. In other words, with probability 1 we have*

$$\lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}| = \infty .$$

Proof For a continuous sample path ω belonging to a set Ω_1 of measure 1 we put $C_n = \max_{0 \leq i \leq n-1} |W_{t_{i+1}} - W_{t_i}|$. Since ω is a uniformly continuous function defined on a compact set $[0, T]$, we have $\lim_{n \rightarrow \infty} C_n = 0$ and

$$\sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}|^2 \leq C_n \sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}|.$$

The left-hand side converges to T for almost every ω belonging to a set Ω_2 of measure 1 along a subsequence $\{n_k\}_{k=1}^\infty$ by Theorem 7.4. Hence $\sum_{i=0}^{n-1} |W_{t_{i+1}} - W_{t_i}|$ increases to infinity with probability 1. $\square$

Lemma 7.1 For a real constant θ we have

$$\mathbb{E}[e^{\theta W_t}] = e^{\frac{1}{2}\theta^2 t}, \quad t \geq 0.$$

Proof Note that

$$\begin{aligned} \mathbb{E}[e^{W_t}] &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^x e^{-\frac{x^2}{2t}} dx \\ &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2 - t^2}{2t}} dx \\ &= e^{\frac{t}{2}} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2t}} dy = e^{\frac{1}{2}t}. \end{aligned}$$

Since θW_t and $W_{\theta^2 t}$ have the same distribution, we have $\mathbb{E}[e^{\theta W_t}] = \mathbb{E}[e^{W_{\theta^2 t}}] = e^{\frac{1}{2}\theta^2 t}$. $\square$

Example 7.1 By taking $t = 1$ and $\theta = 1$, we see that $\mathbb{E}[e^Z] = \sqrt{e}$, where $Z \sim N(0, 1)$.

Remark 7.4 By replacing θ by $-\theta$ in Lemma 7.1, we have $\mathbb{E}[e^{-\theta W_t}] = e^{\frac{1}{2}\theta^2 t}$, which can be proved also by the fact that θW_t and $-\theta W_t$ are identically distributed.

Example 7.2 In the geometric Brownian motion model the stock price S_t satisfies $S_t = S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}$. Since $\mathbb{E}[e^{-\frac{1}{2}\theta^2 t + \theta W_t}] = e^{-\frac{1}{2}\theta^2 t} \mathbb{E}[e^{\theta W_t}] = 1$, we have $\mathbb{E}[S_t] = S_0 e^{\mu t}$, i.e., the stock price increases exponentially on average regardless of the volatility σ as time increases. In Fig. 7.3 twenty sample paths of S_t are plotted together with the curve $y = S_0 e^{\mu t}$ for $\mu = 0.2$ and $\sigma = 0.1$. Compare it with Fig. 7.4 for $\mu = 0.2$ and $\sigma = 0.3$. Consult Simulation 7.2. See also Remark 7.10 for a less intuitive result.

Lemma 7.2 If $i = \sqrt{-1}$, then $\mathbb{E}[e^{i\theta W_t}] = e^{-\frac{1}{2}\theta^2 t}$ for any real constant θ .

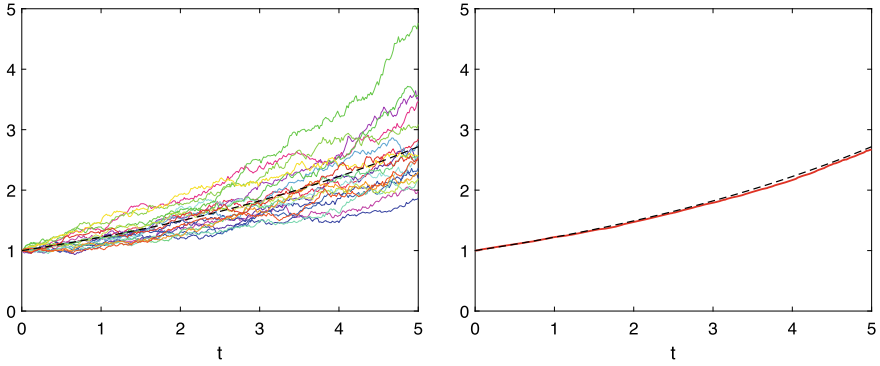


Fig. 7.3 Exponential growth of the geometric Brownian motion: The sample average of S_t (in red on the right) is close to the exponential curve $y = S_0 e^{\mu t}$ (dashed in black) for $\mu = 0.2$ and $\sigma = 0.1$

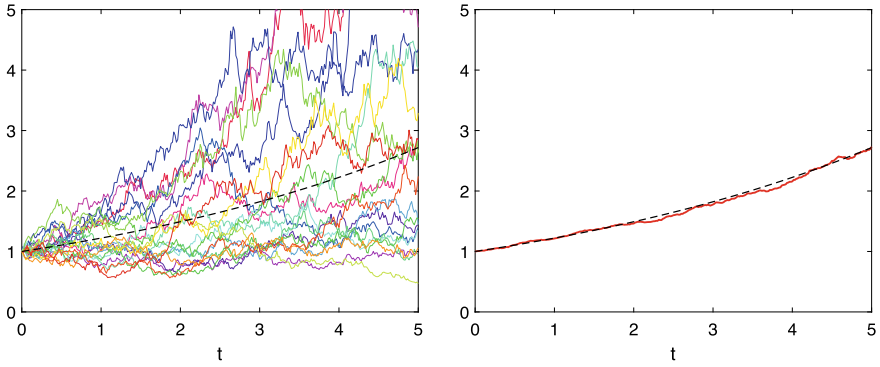


Fig. 7.4 Exponential growth of the geometric Brownian motion: The sample average of S_t (in red on the right) is close to the exponential curve $y = S_0 e^{\mu t}$ (dashed in black) for $\mu = 0.2$ and $\sigma = 0.3$

Proof Recall that the Fourier transform of the probability density function of the standard normal variable is given by the probability density function itself up to a constant multiple. (See Example 4.9.) Hence

$$\begin{aligned} \mathbb{E}[e^{iW_t}] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)} e^{ix} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} e^{i\sqrt{t}y} dy \\ &= e^{-(\sqrt{t})^2/2} = e^{-t/2}. \end{aligned}$$

Since θW_t and $W_{\theta^2 t}$ are identically distributed, we have $\mathbb{E}[e^{i\theta W_t}] = \mathbb{E}[e^{iW_{\theta^2 t}}] = e^{-\frac{1}{2}\theta^2 t}$. $\square$

Remark 7.5 The probability distribution of W_t is symmetric around 0 and hence $\mathbb{E}[\sin W_t] = 0$. Thus $\mathbb{E}[e^{iW_t}] = \mathbb{E}[\cos W_t] + i\mathbb{E}[\sin W_t] = \mathbb{E}[\cos W_t]$ and we obtain a real number even when the function under integration is complex-valued in the above. As a by-product we have $\mathbb{E}[\cos W_t] = e^{-t/2}$. (See Remark 7.11 and Exercise 7.1. See also Exercise 10.3.) By taking $t = 1$ we have $\mathbb{E}[\cos Z] = \frac{1}{\sqrt{e}}$, where Z is a standard normal variable.

Theorem 7.5 Let $0 \leq s < t$. Under the conditions that $W_s = x$ and $W_t = z$ the distribution of $W_{(s+t)/2}$ is normal with expectation $\frac{1}{2}(x + z)$ and variance $\frac{1}{4}(t - s)$. That is, the conditional probability density function $f(y)$ is given by

$$f(y) = \frac{1}{\sqrt{2\pi \frac{t-s}{4}}} \exp\left(-\frac{(y - \frac{x+z}{2})^2}{2\frac{t-s}{4}}\right).$$

Proof We sketch the proof for $s = 0$, in which case we have $x = 0$. To compute the conditional density function, we compute first the conditional expectation $\mathbb{E}[\phi(W_{t/2}) | W_t = z]$ for an arbitrary function $\phi : \mathbb{R} \rightarrow \mathbb{R}$. For arbitrarily small $\delta z > 0$ let $A = \{\omega : W_t(\omega) \in [z, z + \delta z]\}$. Then

$$\begin{aligned} & \frac{1}{\mathbb{P}(A)} \int_A \phi(W_{t/2}) d\mathbb{P} \\ &= \frac{\int_z^{z+\delta z} \int_{-\infty}^{\infty} \phi(y) p(\frac{t}{2}; 0, y) p(\frac{t}{2}; y, u) dy du}{\int_z^{z+\delta z} p(t; 0, u) du} \\ &\approx \frac{\int_{-\infty}^{\infty} \phi(y) p(\frac{t}{2}; 0, y) p(\frac{t}{2}; y, z) dy \delta z}{p(t; 0, z) \delta z} \\ &= \int_{-\infty}^{\infty} \phi(y) \frac{p(\frac{t}{2}; 0, y) p(\frac{t}{2}; y, z)}{p(t; 0, z)} dy. \end{aligned}$$

Thus the conditional probability density of $W_{t/2}$ given $W_t = z$ is obtained by

$$\begin{aligned} & \frac{p(\frac{t}{2}; 0, y) p(\frac{t}{2}; y, z)}{p(t; 0, z)} \\ &= \frac{1}{\sqrt{2\pi \frac{t}{2}}} \exp\left(-\frac{y^2}{2\frac{t}{2}}\right) \frac{1}{\sqrt{2\pi \frac{t}{2}}} \exp\left(-\frac{(z-y)^2}{2\frac{t}{2}}\right) \sqrt{2\pi t} \exp\left(-\frac{z^2}{2t}\right) \\ &= \frac{\sqrt{2}}{\sqrt{\pi t}} \exp\left(-\frac{2(y - \frac{1}{2}z)^2}{t}\right). \end{aligned}$$

7.2 Construction of Brownian Motion

In Theorem 7.6 presented in the next paragraph, we construct a sample Brownian path satisfying $W_T = b$ using the above result. First, we find the value at $t = \frac{1}{2}T$, then find the values at $t = \frac{1}{4}T, \frac{3}{4}T$, and so on by adding midpoints. We let $\delta t = \frac{1}{2^n}T$ for some n . Connect the points $(0, 0), (\delta t, W(\delta t)), \dots, (k\delta t, W(k\delta t)), \dots, (T, W(T))$ by line segments, and obtain a piecewise curve which converges to a Brownian sample path. See also Sect. 12.3.

Theorem 7.6 (Brownian Motion Conditional on $W_T = b$) *By using successive approximation we construct a sample path of Brownian motion over an interval $[0, T]$. For the sake of notational simplicity we consider the case $T = 1$. We will define piecewise linear functions $W^{(n)}$, $n \geq 1$, recursively. Given $n \geq 1$, consider the points $\frac{k}{2^n}$, $0 \leq k \leq 2^n$, and define $W^{(n)}$ using $W^{(n-1)}$ at those points depending on whether k is even or odd.*

(i) *If k is even, then $k = 2i$, and define*

$$W^{(n)}\left(\frac{k}{2^n}\right) = W^{(n)}\left(\frac{i}{2^{n-1}}\right) = W^{(n-1)}\left(\frac{i}{2^{n-1}}\right).$$

(ii) *If k is odd, then $k = 2i + 1$, and define*

$$W^{(n)}\left(\frac{k}{2^n}\right) = \frac{1}{2} \left[W^{(n)}\left(\frac{2i}{2^n}\right) + W^{(n)}\left(\frac{2i+2}{2^n}\right) \right] + 2^{-\frac{n+1}{2}} Z_{2i+1}^{(n)},$$

where $Z_{2i+1}^{(n)}$, $1 \leq 2i + 1 \leq 2^n$, are independent standard normal variables, i.e.,

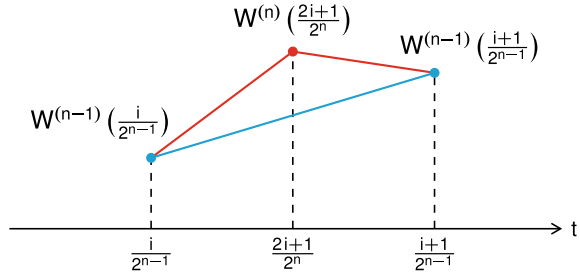
$$W^{(n)}\left(\frac{2i+1}{2^n}\right) = \frac{1}{2} \left[W^{(n-1)}\left(\frac{i}{2^{n-1}}\right) + W^{(n-1)}\left(\frac{i+1}{2^{n-1}}\right) \right] + 2^{-\frac{n+1}{2}} Z_{2i+1}^{(n)}.$$

(See Fig. 7.5.) Finally, we connect the points $(\frac{k}{2^n}, W^{(n)}(\frac{k}{2^n}))$ to obtain a piecewise linear continuous function $W^{(n)}$ on the whole interval. Then $W^{(n)}$, $n \geq 1$, converges uniformly to a continuous function with probability 1.

Proof Let $(\Omega, \mathbb{P})$ denote the probability space under consideration, and put

$$M_n(\omega) = \max\{|Z_k^{(n)}(\omega)| : 0 < k < 2^n, k \text{ is odd}\}$$

Fig. 7.5 Construction of a sample path of Brownian motion



and $A_n = \{M_n > n\}$. Take $\alpha > 0$. Since

$$\begin{aligned} \mathbb{P}\left(|Z_k^{(n)}| > \alpha\right) &= 2 \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-z^2/2} dz \\ &< \sqrt{\frac{2}{\pi}} \int_{\alpha}^{\infty} \frac{\alpha}{z} e^{-z^2/2} dz = \sqrt{\frac{2}{\pi}} \frac{1}{\alpha} e^{-\alpha^2/2}, \end{aligned}$$

we have

$$\mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_k \{|Z_k^{(n)}| > n\}\right) \leq 2^{n-1} \mathbb{P}(|Z_1^{(n)}| > n) \leq \sqrt{\frac{2}{\pi}} \frac{2^{n-1}}{n} e^{-n^2/2},$$

and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$.

By the Borel–Cantelli Lemma (Theorem 4.10) there exists a set Ω_1 such that $\mathbb{P}(\Omega_1) = 1$ and for $\omega \in \Omega_1$ there exists $N = N(\omega)$ satisfying $M_n(\omega) \leq n$ for $n \geq N$. Then, for $\omega \in \Omega_1$ and $n \geq N$, we have

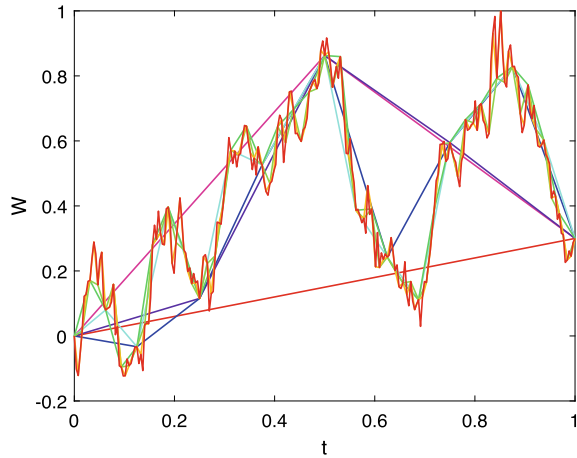
$$\sup_{0 \leq t \leq 1} |W^{(n+1)}(t) - W^{(n)}(t)| = 2^{-\frac{n+1}{2}} M_n \leq 2^{-\frac{n+1}{2}} n,$$

and hence, for $n, m \geq N$,

$$\sup_{0 \leq t \leq 1} |W^{(m)}(t) - W^{(n)}(t)| = \sum_{j=n}^{\infty} 2^{-\frac{j+1}{2}} j.$$

Since the infinite sum in the above converges monotonically to 0 as $n \rightarrow \infty$, we conclude that $W^{(n)}(t)$ converges uniformly on $[0, 1]$. $\square$

Fig. 7.6 A sample path of Brownian motion constructed by the bisection method



Fact 7.1 Let $W(t)$ be the limit of $W^{(n)}(t)$ in the preceding construction.

(i) The increments

$$W^{(n)}\left(\frac{k}{2^n}\right) - W^{(n)}\left(\frac{k-1}{2^n}\right)$$

are independent and normally distributed with mean 0 and variance $\frac{1}{2^n}$.

(ii) If $0 \leq t_0 < t_1 < \cdots < t_n \leq 1$, then the increments $W(t_{i+1}) - W(t_i)$, $0 \leq i \leq n-1$, are independent and normally distributed with mean 0 and variance $t_{i+1} - t_i$. Thus $W(t)$ is a Brownian motion. (See Fig. 7.6.)

Proof Consult (Karatzas and Shreve 1998). □

Remark 7.6 Sometimes a generalized Brownian motion $a + W_t$ is called a Brownian motion conditional on $W_0 = a$. (See also Theorem 7.2.)

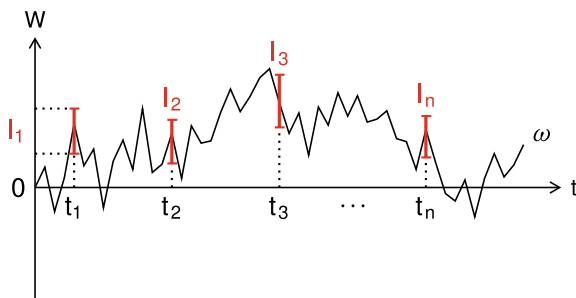
Definition 7.3 (*Cylinder Subset*) Let Ω be the set of functions $\omega : [0, \infty) \rightarrow \mathbb{R}^1$, called sample paths, satisfying the condition $\omega(0) = 0$. To endow a measurable structure on Ω we consider cylinder subsets which are building blocks of all measurable subsets. More precisely, for arbitrary time points $0 = t_0 < t_1 < \cdots < t_n$ and arbitrary intervals $I_1, \dots, I_n \subset \mathbb{R}$, we define a *cylinder subset* by

$$C(t_1, \dots, t_n; I_1, \dots, I_n) = \{\omega \in \Omega \mid \omega(t_1) \in I_1, \dots, \omega(t_n) \in I_n\}.$$

An infinitely long cylinder $\{(x, y) \mid x^2 + y^2 = 1\} \times \mathbb{R}^1 \subset \mathbb{R}^3$ is a set of points whose third coordinates are arbitrary.¹ (See Fig. 7.7 for a graphical representation of a cylinder subset where a sample path passes through the gates represented by the intervals.)

¹ A cylinder subset for a Brownian motion is regarded as a subset of the infinite Cartesian product $\prod_{t \geq 0} \mathbb{R}_t^1$ of the real line $\mathbb{R}_t^1 = \mathbb{R}^1$, $t \geq 0$, whose points are arbitrary except at finitely many values $0 = t_0 < t_1 < \cdots < t_n$.

Fig. 7.7 A cylinder subset $C(t_1, \dots, t_n; I_1, \dots, I_n)$ for Brownian motion



The σ -algebra generated by the cylinder subsets $C(t_1, \dots, t_n; I_1, \dots, I_n)$ for $n \geq 1$ and $0 = t_0 < t_1 < \dots < t_n \leq t$, is denoted by $\mathcal{F}_t$. Note that $\{\mathcal{F}_t\}_{t \geq 0}$ is a filtration and that a σ -algebra $\mathcal{F}$ is generated by $\bigcup_{t \geq 0} \mathcal{F}_t$. Then we obtain a filtered measurable space $(\Omega, \{\mathcal{F}_t\}_{t \geq 0}, \mathcal{F})$, which is called a sample space of the Brownian motion.

Remark 7.7 The following are basic properties of the cylinder subsets.

- (i) $C(t; I_1) \cup C(t; I_2) = C(t; I_1 \cup I_2)$.
- (ii) $C(t; I_1) \cap C(t; I_2) = C(t; I_1 \cap I_2)$.
- (iii) For $t_1 < t_2$ we have $C(t_1; I_1) \cap C(t_2; I_2) = C(t_1, t_2; I_1, I_2)$.

Now we define a measure on Ω : Put $\delta t_j = t_j - t_{j-1}$ and define the size $\mathbb{P}_0$ of a cylinder subset $C(t_1, \dots, t_n; I_1, \dots, I_n)$ by

$$\begin{aligned} & \mathbb{P}_0(C(t_1, \dots, t_n; I_1, \dots, I_n)) \\ &= \int_{I_1} \cdots \int_{I_n} p(\delta t_1; 0, x_1) p(\delta t_2; x_1, x_2) \cdots p(\delta t_n; x_{n-1}, x_n) dx_n \cdots dx_1, \end{aligned}$$

where $p(\delta t; x, y)$ is the transition probability density of the normal distribution defined in (7.1). By the Kolmogorov Extension Theorem, $\mathbb{P}_0$ is extended to a probability measure $\mathbb{P}$ defined on $\mathcal{F}$. To be more precise, we first consider the algebra $\mathcal{F}_0$ generated by finite unions of cylinder subsets and define an additive set function $\mathbb{P}_0 : \mathcal{F}_0 \rightarrow [0, 1]$, and next we extend the domain of $\mathbb{P}_0$ to the whole σ -algebra $\mathcal{F}$ and obtain a countably additive set function, i.e., a measure, $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$.

Remark 7.8 We may regard a Brownian motion as a sequence of random variables W_t , $t \geq 0$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$ by $W_t(\omega) = \omega(t)$. Since $W_t^{-1}(I) = C(t; I)$ for a Borel subset $I \subset \mathbb{R}$, we have $\sigma(W_t) \subset \mathcal{F}_t$. In other words, $\{W_t\}_{t \geq 0}$ is adapted to $\{\mathcal{F}_t\}_{t \geq 0}$. (See Fig. 7.8.)

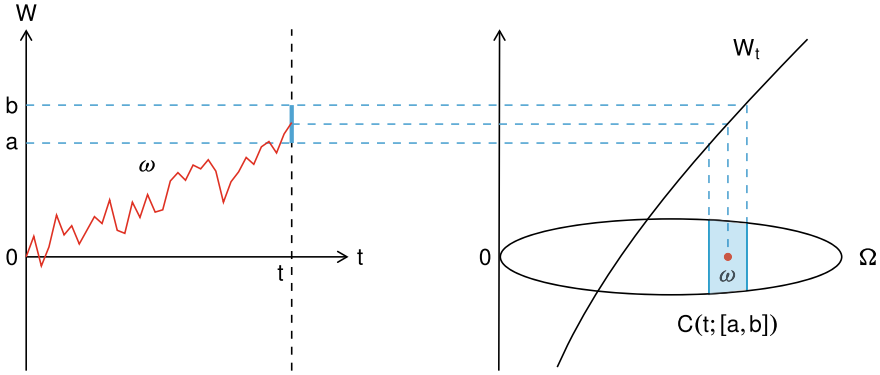


Fig. 7.8 Brownian motion and a cylinder subset

7.3 Properties of Brownian Motion

Remark 7.9 (Nondifferentiability of Brownian motion) From the equation

$$\frac{\delta W_t}{\delta t} = \frac{W_{t+\delta t} - W_t}{\delta t} \sim \frac{1}{\sqrt{\delta t}} N(0, 1)$$

we observe that as the increment $\delta t > 0$ converges to 0 the rate $\frac{\delta W_t}{\delta t}$ can be arbitrarily large. This is the reason why almost every sample path of Brownian motion is not differentiable at every t . Therefore we use short line segments in plotting sample paths of Brownian motion, and should not use any smoothing technique. (See Fig. 7.2.)

Theorem 7.7 (Time Inversion) *Let $W_t, t \geq 0$, be a Brownian motion. The stochastic process X_t defined by*

$$X_t = \begin{cases} 0, & t = 0 \\ t W_{1/t}, & t > 0 \end{cases}$$

is also a Brownian motion.

Proof Consult (Mörters and Peres 2010). □

Corollary 7.2 (Law of Large Numbers for Brownian Motion) *Let $W_t, t \geq 0$, be a Brownian motion. Then*

$$\lim_{t \rightarrow \infty} \frac{W_t}{t} = 0$$

almost surely.

Proof Note that

$$\lim_{t \rightarrow \infty} \frac{W_t}{t} = \lim_{u \rightarrow 0+} u W_{1/u} = \lim_{u \rightarrow 0+} X_u = 0$$

almost surely since X_u defined by $X_u = u W_{1/u}$, $u > 0$, and $X_0 = 0$, is a Brownian motion by Theorem 7.7. $\square$

For a simulation of Corollary 7.2, see Simulation 7.4 and Fig. 7.9. For a related result, see Exercise 11.7.

Remark 7.10 In Example 7.2, for any constants μ and σ we showed that

$$\mathbb{E}[S_t] = \mathbb{E}[S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}] = S_0 e^{\mu t}, \quad t \geq 0.$$

However, if $\mu - \frac{1}{2}\sigma^2 < 0$, then

$$\begin{aligned} \lim_{t \rightarrow \infty} S_0 e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t} &= \lim_{t \rightarrow \infty} S_0 e^{t[(\mu - \frac{1}{2}\sigma^2) + \sigma \frac{1}{t} W_t]} \\ &= \lim_{t \rightarrow \infty} S_0 e^{t(\mu - \frac{1}{2}\sigma^2)} = 0 \end{aligned}$$

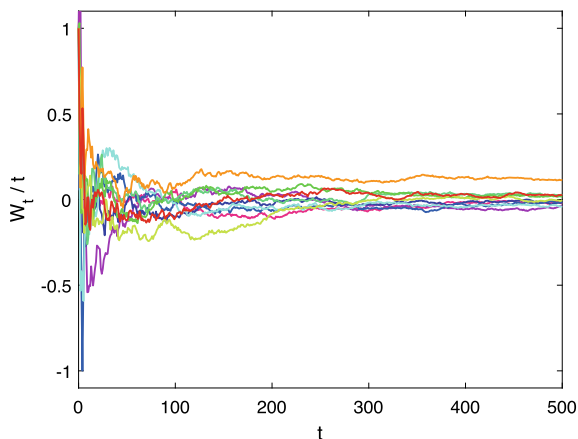
almost surely by the Law of Large Numbers. Since

$$\lim_{t \rightarrow \infty} \mathbb{E}[S_t] = S_0 e^{\mu t} = \begin{cases} \infty, & \mu > 0 \\ S_0, & \mu = 0 \\ 0, & \mu < 0 \end{cases}$$

we have

$$0 = \mathbb{E}\left[\lim_{t \rightarrow \infty} S_t\right] \neq \lim_{t \rightarrow \infty} \mathbb{E}[S_t]$$

Fig. 7.9 The Law of Large Numbers for Brownian motion



for $\mu \geq 0$. If $\mu = 0$ then S_t is a martingale and $\mathbb{E}[S_t] = S_0$ for every $t \geq 0$ by Theorem 6.1, which is in agreement with the above result. Note that even in this case we have $\lim_{t \rightarrow \infty} S_t = 0$.

See Figs. 7.3 and 7.4 for the simulations for $\mu = 0.2$ and $\mu - \frac{1}{2}\sigma^2 > 0$ where individual sample paths tend to increase as $t \rightarrow \infty$ and the average of sample paths grows exponentially. See also Fig. 7.10 for $\mu > 0$ and $\mu - \frac{1}{2}\sigma^2 < 0$ where individual sample paths converge to 0 while their average grows exponentially.

Finally, for $\mu = 0$ presented in Fig. 7.11, the martingale property makes the sample paths stay around the constant average S_0 for every t .

In Table 7.1 the limiting behaviors of geometric Brownian motion are classified by the signs of μ and $\mu - \frac{1}{2}\sigma^2$ under the assumption that $\sigma > 0$.

Fig. 7.10 The Law of Large Numbers for geometric Brownian motion for $\mu > 0$ and $\mu - \frac{1}{2}\sigma^2 < 0$ with exponentially increasing average

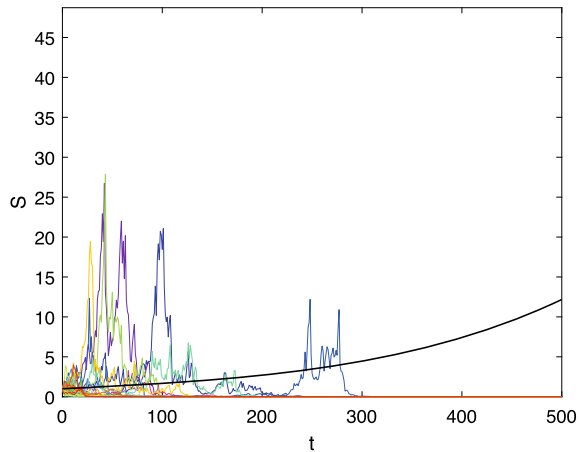


Fig. 7.11 The Law of Large Numbers for geometric Brownian motion and the martingale property for $\mu = 0$

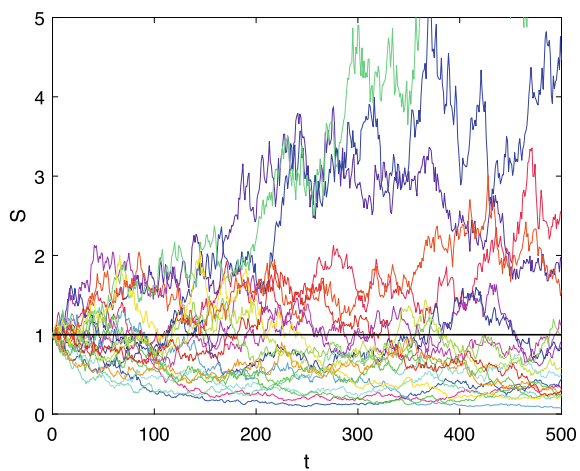


Table 7.1 Limiting behaviors of geometric Brownian motion depending on the signs of μ and $\mu - \frac{1}{2}\sigma^2$

	$\mu > 0$	$\mu = 0$	$\mu < 0$
$\mu - \frac{1}{2}\sigma^2 > 0$	$S_t \rightarrow +\infty$ a.s. $\mathbb{E}[S_t] \rightarrow +\infty$	Not applicable	Not applicable
$\mu - \frac{1}{2}\sigma^2 = 0$	$S_t = e^{\sigma W_t}$ $\mathbb{E}[S_t] \rightarrow +\infty$	Not applicable	Not applicable
$\mu - \frac{1}{2}\sigma^2 < 0$	$S_t \rightarrow 0$ a.s. $\mathbb{E}[S_t] \rightarrow +\infty$	$S_t \rightarrow 0$ a.s. $\mathbb{E}[S_t] = S_0$	$S_t \rightarrow 0$ a.s. $\mathbb{E}[S_t] \rightarrow 0$

7.4 Random Walk to the Horizon

An observer on a platform watches a man staggering away from the point $(0, 0, 0)$ and trying to move along a straight road represented by the x -axis. Suppose that at time $t = 0$ the random walker starts moving forward with constant speed $v > 0$ meters per second while his sideways movement is assumed to be a constant multiple of a Brownian motion W_t where t is measured in seconds. (Thus his position at time t is $(vt, cW_t, 0)$ for some $c > 0$. See Fig. 7.12.) What is the limiting behavior of the random walk from the perspective of the observer? In other words, if P is the centrally projected point onto the yz -plane of the random walker, what is the trajectory of P ? (For an interesting tale related to the given problem see Chap. 2 in Bruce (2002).)

To find $P = (0, y, z)$, solve the equation

$$u(-a, 0, h) + (1 - u)(vt, cW_t, 0) = (0, y, z)$$

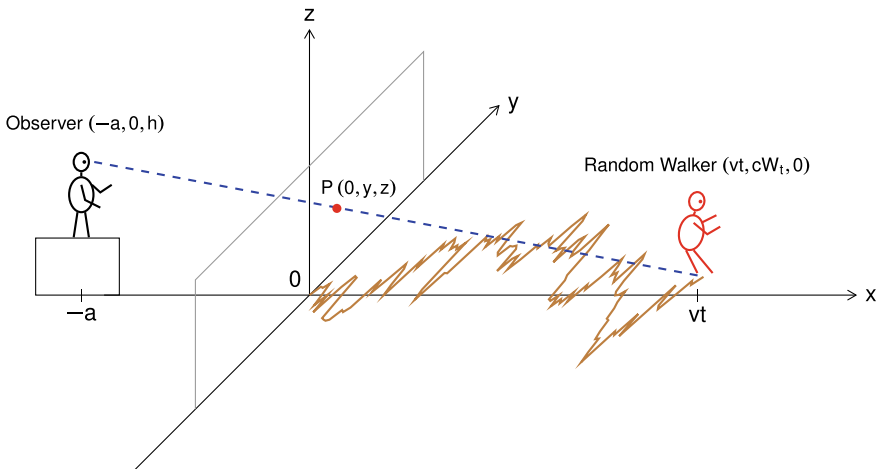
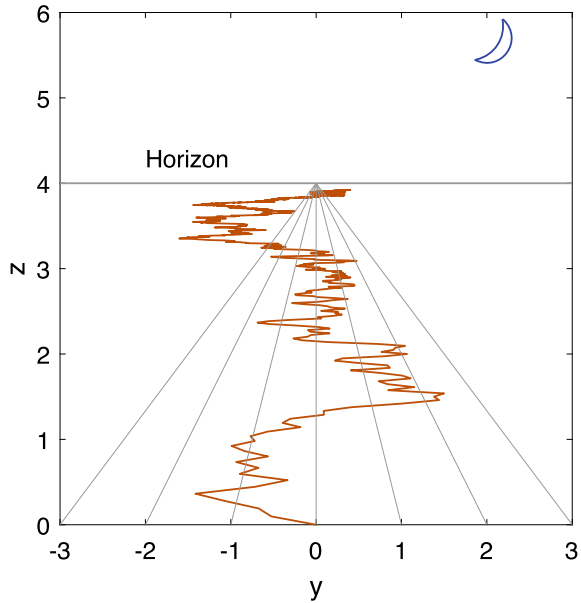
**Fig. 7.12** While an observer watches on a platform, a random walker staggers away

Fig. 7.13 The projected trajectory of a random walker converges to a point on the horizon



and obtain $u = bt/(a + bt)$. Then

$$y = \frac{ac}{a + bt} W_t \quad \text{and} \quad z = \frac{hbt}{a + bt}.$$

Thus $\lim_{t \rightarrow +\infty} (y, z) = (0, h)$ by the Law of Large Numbers for Brownian motion. To the observer, the random walker vanishes to a point on the horizon $z = h$ and all parallel lines, given by $z = 0$ and $y = y_0$ for y_0 constant, converge to the same vanishing point.

See Fig. 7.13, where we take $a = 5$, $v = 0.5$, $c = 1$ and $h = 4$ for the simulation. The straight lines given by $(x, y, x) = (x, y_0, 0)$, $0 \leq x < \infty$, are projected to the line segments on the vertical plane $x = 0$ defined by $\frac{y}{y_0} + \frac{z}{4} = 1$ for $y_0 \neq 0$ and $(0, z)$, $0 \leq z < 4$ for $y_0 = 0$. To see why, note that the projection is a subset of the intersection of the unique plane determined by the observer's viewpoint $(-a, 0, h)$ and the straight line $(x, y, x) = (x, y_0, 0)$, $0 \leq x < \infty$, and the vertical plane $x = 0$. For a computer experiment, see Simulation 7.5.

7.5 Brownian Motion and Martingales

In this section we present some of the most important examples of martingales defined by a Brownian motion $\{W_t\}_{t \geq 0}$.

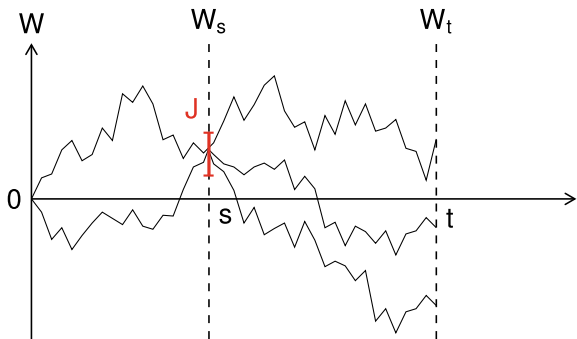
Before we present a formal proof of the martingale property of a Brownian motion in Theorem 7.8, we provide an intuitive view as follows: Let $\dots < a_{k-1} < a_k < a_{k+1} < \dots$, and $J_k = (a_k, a_{k+1}]$. Fix $0 \leq s < t$. Note that the cylinder sets $C(s; J_k) = \{\omega : W_s(\omega) \in J_k\}$ partition the set of all Brownian paths Ω . In Fig. 7.14 the interval J_k is plotted along the line with the time coordinate equal to s . Let J be one such interval. If $a_{k+1} - a_k$ is close to zero, then $\mathbb{E}[W_t | \mathcal{F}_s]$ is approximated by a number $\mathbb{E}[W_t | C(s; J)]$ on $C(s; J)$. Assume that J is of the form $J = (a, a + \delta x]$ for some a and $\delta x > 0$ and that δx is sufficiently small. Then

$$\begin{aligned} \mathbb{E}[W_t | C(s; J)] &= \frac{\int_{C(s; J)} W_t d\mathbb{P}}{\mathbb{P}(C(s; J))} \\ &= \frac{\int_{-\infty}^{\infty} \int_J x_2 p(s; 0, x_1) p(t-s; x_1, x_2) dx_1 dx_2}{\int_J p(s; 0, x_1) dx_1} \\ &\approx \frac{\int_{-\infty}^{\infty} x_2 p(s; 0, a) p(t-s; a, x_2) \delta x dx_2}{p(s; 0, a) \delta x} \\ &= \int_{-\infty}^{\infty} x_2 p(t-s; a, x_2) dx_2 = a, \end{aligned}$$

where the symbol ‘ $\approx$ ’ means ‘being approximately equal to’ for sufficiently small $\delta x > 0$ and the difference decreases to 0 as $\delta x \rightarrow 0$, and $p(u; x, y)$ denotes the transition probability density from x to y in time u . Since a is the representative value of W_s on $C(s; J)$, we observe that on each subset $C(s; J)$ two random variables $\mathbb{E}[W_t | \mathcal{F}_s]$ and W_s are sufficiently close to each other. In fact, $\mathbb{E}[W_t | \mathcal{F}_s] = W_s$, i.e., W_t is a martingale.

Here we present some examples of martingales.

Fig. 7.14 Conditional expectation of W_t on a cylinder set $C(s; J)$



Theorem 7.8 *Let $\{W_t\}_{t \geq 0}$ denote a Brownian motion. The following stochastic processes are martingales:*

- (i) $\{W_t\}_{t \geq 0}$
- (ii) $\{W_t^2 - t\}_{t \geq 0}$
- (iii) $\{e^{-\frac{1}{2}\theta^2 t + \theta W_t}\}_{t \geq 0}$
- (iv) $\{e^{\frac{1}{2}t} \cos W_t\}_{t \geq 0}$
- (v) $\{W_t^3 - 3tW_t\}_{t \geq 0}$

Proof Take $0 \leq s < t$.

- (i) Since $W_t - W_s$ and $\mathcal{F}_s$ are independent, we have $\mathbb{E}[W_t - W_s | \mathcal{F}_s] = \mathbb{E}[W_t - W_s] = 0$, and hence $\mathbb{E}[W_t | \mathcal{F}_s] = \mathbb{E}[W_s | \mathcal{F}_s] = W_s$.
- (ii) To show that $\mathbb{E}[W_t^2 - t | \mathcal{F}_s] = W_s^2 - s$ we note that

$$\begin{aligned} \mathbb{E}[W_t^2 | \mathcal{F}_s] &= \mathbb{E}[(W_t - W_s)^2 + 2W_s W_t - W_s^2 | \mathcal{F}_s] \\ &= \mathbb{E}[(W_t - W_s)^2 | \mathcal{F}_s] + 2W_s \mathbb{E}[W_t | \mathcal{F}_s] - \mathbb{E}[W_s^2 | \mathcal{F}_s] \\ &= \mathbb{E}[(W_t - W_s)^2] + 2W_s W_s - W_s^2 \\ &= \mathbb{E}[W_{t-s}^2] + W_s^2 = t - s + W_s^2. \end{aligned}$$

- (iii) Since $W_t - W_s$ is independent of $\mathcal{F}_s$ we have

$$\mathbb{E}[e^{\theta(W_t - W_s)} | \mathcal{F}_s] = \mathbb{E}[e^{\theta(W_t - W_s)}] = \mathbb{E}[e^{\theta W_{t-s}}] = e^{\frac{1}{2}\theta^2(t-s)}.$$

Hence $e^{-\theta W_s} \mathbb{E}[e^{\theta W_t} | \mathcal{F}_s] = e^{\frac{1}{2}\theta^2(t-s)}$ and $\mathbb{E}[e^{\theta W_t - \frac{1}{2}\theta^2 t} | \mathcal{F}_s] = e^{-\frac{1}{2}\theta^2 s} e^{\theta W_s}$.

- (iv) Let $z = \alpha + i\beta \in \mathbb{C}$, $\alpha, \beta \in \mathbb{R}$ and $i^2 = -1$, be a complex number, and let $\Re(w)$ denote the real part of a complex number w . Then

$$\begin{aligned} \mathbb{E}[\cos W_t | \mathcal{F}_s] &= \mathbb{E}[\Re(e^{iW_t}) | \mathcal{F}_s] = \Re(\mathbb{E}[e^{iW_t} | \mathcal{F}_s]) \\ &= \Re(\mathbb{E}[e^{iW_s} e^{i(W_t - W_s)} | \mathcal{F}_s]) \\ &= \Re(e^{iW_s} \mathbb{E}[e^{i(W_t - W_s)} | \mathcal{F}_s]) \\ &= \Re(e^{iW_s} \mathbb{E}[e^{i(W_t - W_s)}]) = \Re(e^{iW_s} \mathbb{E}[e^{iW_{t-s}}]) . \end{aligned}$$

By Lemma 7.2 we have

$$\mathbb{E}[\cos W_t | \mathcal{F}_s] = \Re(e^{iW_s} e^{-(t-s)/2}) = e^{-(t-s)/2} \cos W_s$$

and $\mathbb{E}[e^{\frac{1}{2}t} \cos W_t | \mathcal{F}_s] = e^{\frac{1}{2}s} \cos W_s$.

- (v) Since $W_t - W_s$ is independent of $\mathcal{F}_s$, $(W_t - W_s)^2$ and $(W_t - W_s)^3$ are also independent of $\mathcal{F}_s$. Hence we have

$$\begin{aligned}
 \mathbb{E}[W_t^3 | \mathcal{F}_s] - W_s^3 &= \mathbb{E}[(W_t - W_s + W_s)^3 - W_s^3 | \mathcal{F}_s] \\
 &= \mathbb{E}[(W_t - W_s)^3 + 3(W_t - W_s)^2 W_s + 3(W_t - W_s) W_s^2 | \mathcal{F}_s] \\
 &= \mathbb{E}[(W_t - W_s)^3] + 3W_s \mathbb{E}[(W_t - W_s)^2] + 3W_s^2 \mathbb{E}[W_t - W_s] \\
 &= 0 + 3W_s(t - s) + 3W_s^2 \times 0 \\
 &= \mathbb{E}[3t W_t | \mathcal{F}_s] - 3s W_s .
 \end{aligned}$$

$$\text{Hence } \mathbb{E}[W_t^3 - 3t W_t | \mathcal{F}_s] = W_s^3 - 3s W_s .$$

□

Remark 7.11 As a by-product of the proof of Theorem 7.8 (iv), we obtain

$$\mathbb{E}[\cos W_t | \mathcal{F}_s] = e^{-(t-s)/2} \cos W_s .$$

For $s = 0$ the fact that $\mathbb{E}[\cos W_t] = e^{-t/2}$ has already been mentioned in Remark 7.5.

The following facts, due to Paul Lévy,² characterize Brownian motions using the concept of martingale.

Theorem 7.9 *Given a stochastic process $\{X_t\}_{t \geq 0}$ and the filtration $\mathcal{F}_t = \sigma(\{X_s : 0 \leq s \leq t\})$, the process X_t is a Brownian motion if and only if all of the following conditions hold:*

- (i) $X_0 = 0$ with probability 1.
- (ii) A sample path $t \mapsto X_t$ is continuous with probability 1.
- (iii) $\{X_t\}_{t \geq 0}$ is a martingale with respect to $\{\mathcal{F}_t\}_{t \geq 0}$.
- (iv) $\{X_t^2 - t\}_{t \geq 0}$ is a martingale with respect to $\{\mathcal{F}_t\}_{t \geq 0}$.

We have another version, which will be proved.

Theorem 7.10 *Let $\{M_t\}_{t \geq 0}$ be a martingale with respect to a filtration $\{\mathcal{F}_t\}_{t \geq 0}$ and a probability measure $\mathbb{Q}$. If M_t is continuous, $M_0 = 0$ and $[M, M]_t = t$, then it is a Brownian motion.*

Proof Let $F(t, x) = \exp(\frac{1}{2}\lambda^2 t + i\lambda x)$ for some $\lambda \in \mathbb{R}$ where $i^2 = -1$. Then

$$\begin{aligned}
 dF(t, M_t) &= \frac{1}{2}\lambda^2 F(t, M_t) dt + i\lambda F(t, M_t) dM_t - \frac{1}{2}\lambda^2 F(t, M_t) d[M, M]_t \\
 &= i\lambda F(t, M_t) dM_t .
 \end{aligned}$$

² Paul Pierre Lévy (1886–1971) was a French mathematician who made fundamental contributions to probability theory.

Hence $F(t, M_t) = \exp(\frac{1}{2}\lambda^2 t + i\lambda M_t)$ is a martingale. Since, for $s \leq t$,

$$\mathbb{E} \left[\exp \left(\frac{1}{2}\lambda^2 t + i\lambda M_t \right) \middle| \mathcal{F}_s \right] = \exp \left(\frac{1}{2}\lambda^2 s + i\lambda M_s \right),$$

the characteristic function of $M_t - M_s$ is given by

$$\mathbb{E} \left[\exp (i\lambda (M_t - M_s)) \middle| \mathcal{F}_s \right] = \exp \left(-\frac{1}{2}\lambda^2 (t - s) \right).$$

Hence $M_t - M_s, s \leq t$, is normally distributed with mean 0 and variance $t - s$ under $\mathbb{Q}$. As a by-product we obtain

$$\mathbb{E} \left[e^{i\lambda M_t} \middle| \mathcal{F}_s \right] = e^{-\frac{1}{2}\lambda^2 (t-s)} e^{i\lambda M_s}, \quad s < t.$$

Next, suppose that $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$. Then, for real constants $\lambda_1, \dots, \lambda_n$,

$$\begin{aligned} & \mathbb{E} \left[e^{i\lambda_1 M_{t_1} + i\lambda_2 (M_{t_2} - M_{t_1}) + \dots + i\lambda_n (M_{t_n} - M_{t_{n-1}})} \right] \\ &= \mathbb{E} \left[e^{i(\lambda_1 - \lambda_2) M_{t_1} + i(\lambda_2 - \lambda_3) M_{t_2} + \dots + i(\lambda_{n-1} - \lambda_n) M_{t_{n-1}} + i\lambda_n M_{t_n}} \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[e^{i(\lambda_1 - \lambda_2) M_{t_1} + i(\lambda_2 - \lambda_3) M_{t_2} + \dots + i(\lambda_{n-1} - \lambda_n) M_{t_{n-1}} + i\lambda_n M_{t_n}} \middle| \mathcal{F}_{t_{n-1}} \right] \right] \\ &= \mathbb{E} \left[e^{i(\lambda_1 - \lambda_2) M_{t_1} + i(\lambda_2 - \lambda_3) M_{t_2} + \dots + i(\lambda_{n-1} - \lambda_n) M_{t_{n-1}}} \mathbb{E} \left[e^{i\lambda_n M_{t_n}} \middle| \mathcal{F}_{t_{n-1}} \right] \right] \\ &= \mathbb{E} \left[e^{i(\lambda_1 - \lambda_2) M_{t_1} + i(\lambda_2 - \lambda_3) M_{t_2} + \dots + i(\lambda_{n-1} - \lambda_n) M_{t_{n-1}}} e^{-\frac{1}{2}\lambda_n^2 (t_n - t_{n-1})} e^{i\lambda_n M_{t_{n-1}}} \right] \\ &= e^{-\frac{1}{2}\lambda_n^2 (t_n - t_{n-1})} \mathbb{E} \left[e^{i(\lambda_1 - \lambda_2) M_{t_1} + i(\lambda_2 - \lambda_3) M_{t_2} + \dots + i\lambda_{n-1} M_{t_{n-1}}} \right] \\ &= e^{-\frac{1}{2}\lambda_n^2 (t_n - t_{n-1})} \mathbb{E} \left[\mathbb{E} \left[e^{i(\lambda_1 - \lambda_2) M_{t_1} + i(\lambda_2 - \lambda_3) M_{t_2} + \dots + i\lambda_{n-1} M_{t_{n-1}}} \middle| \mathcal{F}_{t_{n-2}} \right] \right] \\ &\quad \vdots \\ &= e^{-\frac{1}{2}\lambda_n^2 (t_n - t_{n-1})} e^{-\frac{1}{2}\lambda_{n-1}^2 (t_{n-1} - t_{n-2})} \times \dots \times e^{-\frac{1}{2}\lambda_1^2 t_1}, \end{aligned}$$

thus the increments are independent. (See Remark 4.7.) □

7.6 Computer Experiments

Simulation 7.1 (Sample Paths of Brownian Motion) We plot 100 sample paths of Brownian motion. We take the time increment³ $\delta t = \frac{T}{N}$, where $T = 20$, $N = 200$. See Fig. 7.1.

³ ‘Forever—is composed of Nows—’, Emily Dickinson (1830–1886).

```

T = 20;
N = 200; % number of time steps
dt = T/N;
time = 0:dt:T;
M = 30; % number of sample paths

W = zeros(M,N+1); % dW and W are matrices.
dW = sqrt(dt)*randn(M,N);
for i=1:N
    W(:,i+1) = W(:,i) + dW(:,i); % Note W(j,1) = 0.0.
end

for j = 1:M
    plot(time,W(j,1:N+1),'b');
hold on;
end

```

Simulation 7.2 (Average of Geometric Brownian Motion) We check the formula $\mathbb{E}[S_t] = S_0 e^{\mu t}$ given in Example 7.2. See Figs. 7.3 and 7.4.

```

mu = 0.25; sigma = 0.3;
N = 100; % number of time steps
T = 5; dt = T/N;

time = 0:dt:T;
num = 200; % number of sample paths
W = zeros(num,N + 1);
dW = zeros(num,N);

S = zeros(num,N + 1);
S0 = 1;
S([1:num],1) = S0;

for i = 1:num
    dW(i,1:N) = sqrt(dt)*randn(1,N);
end

for i = 1:num
    for j=1:N
        S(i,j+1) = S(i,j) + mu*S(i,j)*dt + sigma*S(i,j)*dW(i,j);
    end
end

ave = zeros(1,N+1);
for j = 1: N+1

```

```

        ave(j) = mean(S(:,j));
    end

    plot(time,ave);
    hold on
    t = 0:0.01:T;
    plot(t,S0*exp(mu*t),'r')

```

Simulation 7.3 (Brownian Motion with Boundary Condition) Using the bisection method we generate a sample path of Brownian motion with a given condition at the final time T using conditional expectation. For an output see Fig. 7.6.

```

T = 1; % length of time interval
n = 8; % number of bisections
dt = T/2^n; % length of time step
time = 0:dt:T; % partition of the time interval
M = 2^n + 1;
W = zeros(1,M); % a sample path of Brownian motion
W(1) = 0; % initial condition
W(M) = 0.3; % condition on Brownian motion W at time T=dt*(M-1)

for i=1:n
    Increment = 2^(n-i+1);
    for j=1:Increment:2^n
        index1 = j;
        index2 = j+Increment;
        t1 = time(index1);
        t2 = time(index2);
        W1 = W(index1);
        W2 = W(index2);
        ave = (W1 + W2)/2;
        var = (t2 - t1)/4;
        ind_mid = (index1 + index2)/2;
        W(ind_mid)=random('normal',ave,sqrt(var)); %conditional expectation
    end
end

for i=0:n
    t_value = zeros(1,2^i+1);
    W_value = zeros(1,2^i+1);
    for k=1:2^i+1
        t_value(k)=(k-1)*dt*2^(n-i);
        W_value(k) = W((k-1)*2^(n-i)+1);
    end
end

```

```

plot(t_value,W_value,'-', 'color', hsv2rgb([1-i/n 1 1]));
hold on;
pause(0.3); % Pause between successive stages.
end

```

Simulation 7.4 (Law of Large Numbers for Brownian Motion) We plot 10 sample paths of $\frac{1}{t}W_t$, $0 \leq t \leq 500$, to simulate the Law of Large Numbers for Brownian motion (Corollary 7.2). For an output see Fig. 7.9.

```

N = 200; % number of time steps
T = 500; dt = T/N; time = 0:dt:T;
M = 10; % number of sample paths
W = zeros(M,N+1); % dW and W are matrices.
dW = sqrt(dt)*randn(M,N);
for i=1:N
    W(:,i+1) = W(:,i) + dW(:,i); % Note W(j,1) = 0.0.
end

X = zeros(M,N+1);
X(:,1) = 1;
for i = 2:N+1
    X(:,i) = W(:,i)/((i-1)*dt);
end

for j = 1:M
    plot(time,X(j,1:N+1));
hold on;
end

```

Simulation 7.5 (Random Walk to the Horizon)

```

a = 5; v = 0.5; c = 1; h = 4;
N = 2000; % number of time steps
T = 400; % Choose a large number.
dt = T/N; t = 0:dt:T;
W = zeros(1,N+1); % Brownian motion
dW = sqrt(dt)*randn(1,N); % increments of Brownian motion
for i=1:N
    W(1,i+1) = W(1,i) + dW(1,i); % Note W(j,1) = 0.0.
end
plot(a*c*W(1,1:N+1)./(a+v*t),h*v*t./(a+v*t));

```


Exercises

7.1 Here is another proof of Remark 7.5: For a Brownian motion W_t , define $f(\alpha) = \mathbb{E}[\cos(\alpha W_t)]$, $-\infty < \alpha < \infty$, and show that $f'(\alpha) = -\alpha t f(\alpha)$ and $\mathbb{E}[\cos W_t] = f(1) = e^{-\frac{1}{2}t}$.

7.2 Suppose that X is normally distributed with mean μ and variance σ^2 . Compute $\mathbb{E}[e^{\theta X}]$ and $\mathbb{E}[X^4]$.

7.3 Show that $\{W_t^3 - 3tW_t\}_{t \geq 0}$ is a martingale.

7.4 Prove that $\{W_t^3\}_{t \geq 0}$ is not a martingale even though $\mathbb{E}[W_t^3] = 0$ for every $t \geq 0$. (Hint: Use $\mathbb{E}[(W_t - W_s)^3 | \mathcal{F}_s] = 0$.)

7.5 Compute (i) $\mathbb{E}[(W_t - W_s)^2]$, (ii) $\mathbb{E}[e^{-W_t}]$ and (iii) $\mathbb{E}[e^{2W_t}]$.

7.6 Given a Brownian motion W_t , $0 \leq t \leq 1$, define a stochastic process $X_t = W_t - tW_1$, which is called a Brownian bridge. Then clearly $X_0 = X_1 = 0$ and $\mathbb{E}[X_t] = 0$. Show that $\mathbb{E}[X_t^2] = t - t^2$.

7.7 Compute the following conditional expectations:

- (i) $\mathbb{E}[W_t^2 | W_s = x]$, $0 \leq s < t$.
- (ii) $\mathbb{E}[W_t^2 | W_r = x, W_s = y]$, $0 \leq r < s < t$.
- (iii) $\mathbb{E}[e^{-W_t} | W_s = x]$, $0 \leq s < t$.
- (iv) $\mathbb{E}[W_s | W_t = 0]$, $0 \leq s < t$.

7.8 For a real constant θ show that $\mathbb{E}[W_t e^{\theta W_t}] = \theta t e^{\frac{1}{2}\theta^2 t}$.

7.9 For a real constant θ show that $\mathbb{E}[W_t^2 e^{\theta W_t}] = (t + \theta^2 t^2) e^{\frac{1}{2}\theta^2 t}$.

7.10 Let $\{W_t\}_{t \geq 0}$ be a Brownian motion. Show that the processes $\{X_t\}_{t \geq 0}$ in the following are also Brownian motions.

- (i) $X_t = \alpha W_{t/\alpha^2}$ for $\alpha \neq 0$.
- (ii) $X_t = a^{-1/2} W_{at}$ for $a > 0$.
- (iii) $X_t = W_{T+t} - W_T$ for $T > 0$.
- (iv) $X_t = \begin{cases} W_t, & t \leq T, \\ 2W_T - W_t, & t > T, \end{cases}$ for $T > 0$.

7.11 Prove that $\mathbb{E}[W_s | W_t] = \frac{s}{t} W_t$, $0 \leq s < t$. In other words, $\mathbb{E}[W_s | W_t = y] = \frac{s}{t} y$.

7.12 Show that $\mathbb{E}[W_s | W_t] = \frac{s}{t} W_t$ for $0 \leq s < t$ using the following idea: Define a process $\{X_u\}_{u \geq 0}$ by $X_0 = 0$ and $X_u = uW_{1/u}$ for $u > 0$. Then X_u is a Brownian motion.

7.13 Find a constant a for which $X_t = e^{at} \cos W_t$, $t \geq 0$, is a martingale. (See also Theorem 7.8.)

7.14 Let W_t , $t \geq 0$, be a Brownian motion. Show that $\Pr(W_T > 0, W_{2T} > 0) = \Pr(W_T < 0, W_{2T} < 0) = \frac{3}{8}$ and $\Pr(W_T > 0, W_{2T} < 0) = \Pr(W_T < 0, W_{2T} > 0) = \frac{1}{8}$ for every $T > 0$.

7.15 Compute $\mathbb{E}[|W_t|]$. (Hint: Use Exercise 4.22.)

Chapter 8

The Reflection Principle of Brownian Motion



We investigate the reflection properties of Brownian motion, and as a motivation we start the chapter with discrete random walks. Presented in this chapter are the formulas for the maximum of Brownian motion with drift. Also shown are the scatter plot and the probability density function of Brownian motion and its maximum. The results in this chapter will be used for the pricing of barrier options in Sect. 18.2. For the sake of simplicity of exposition we consider only one barrier problems. For more information consult Durrett (2019).

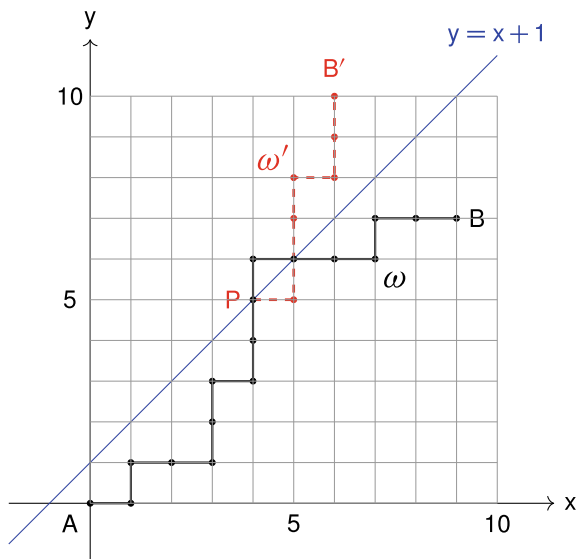
8.1 The Reflection of Discrete Random Walks

Before we analyze a Brownian motion, we consider first a simpler case of discrete time random walk on a lattice for a combinatorial problem. Suppose that nine people have five dollar bills and seven people have ten dollar bills, and they are in line at a ticket counter which has no money and charges five dollars in cash for admission. If a ten dollar bill is presented and there is no change, the line stops. What is the probability the ticket seller was always able to give change?

8.1.1 Two One-Directional Random Walks

Consider the following random walks in the first quadrant of the two-dimensional integer lattice: The movements are random, but they are in nonnegative directions, i.e., in each of the x - and y -directions there is no going backward. We have queues of length 16 with $\binom{9+7}{7}$ possible configurations. To have a valid configuration it has to be assured that for each person with a 10 dollar bill at least one another person with a 5 dollar bill has to be in front of him. We consider paths of length 16 starting

Fig. 8.1 A sample path ω of the two dimensional random walk arising from a sequence of payments at a ticket office (solid in black), and the reflected path ω' after ω hits a barrier $y = x + 1$ (dotted in red)



at $A = (0, 0)$ and ending at $B = (9, 7)$. The paths consist of $(1, 0)$ -steps in the x -direction representing persons with 5 dollar bills and $(0, 1)$ -steps in the y -direction representing persons with 10 dollar bills. Since for each 10 dollar bill a person with a 5 dollar bill has to be in front, the valid paths are those having at each point a y -height less than or equal to the x -length. (See Fig. 8.1).

The number of all possible configurations is the number of all paths of length 16 from $A = (0, 0)$ to $B = (9, 7)$. The number of valid configurations is the number of paths of length 16 from A to B which do not touch the line $y = x + 1$.

The number of paths from A to B of length 16 is equal to $\binom{16}{7}$ since we can select nine $(1, 0)$ -steps in the x -direction and seven $(0, 1)$ -steps in the y -direction. The number of invalid configurations can be calculated using the reflection principle: A path is invalid if it touches or crosses the line $y = x + 1$. If a path touches the line for the first time at a point on the line $y = x + 1$, then the point is marked by P . Now, we reflect the latter part of the path from P to B at the line $y = x + 1$, obtaining a path from A via P to $B' = (6, 9)$. The first part of the path from A to P is the same as the original path.

Each invalid path ω from A to B has to touch the line $y = x + 1$ at least once and corresponds bijectively to the reflected path ω' from A via P to B' , where the first part of the path from A to P is identical to the original path ω . Since this is a bijection, the number of invalid paths from A to B is the number of all paths from A to B' , which equals $\binom{10+6}{6}$. (We use the fact that any path from A to B' has to cross $y = x + 1$, and corresponds to an invalid path from A to B .) The resulting probability is equal to

$$\frac{\binom{16}{7} - \binom{16}{6}}{\binom{16}{7}} = \frac{3}{10}.$$

In general, if there are k people with five dollar bills and ℓ people with ten dollar bills, where $k > \ell$, then $B = (k, \ell)$ and $B' = (\ell - 1, k + 1)$. Hence the probability that the ticket seller is always able to make change equals

$$\frac{\binom{k+\ell}{\ell} - \binom{k+\ell}{\ell-1}}{\binom{k+\ell}{\ell}} = \frac{k+1-\ell}{k+1}.$$

8.1.2 One-Dimensional Random Walks

Now we consider one-dimensional random walks on the integer lattice $\mathbb{Z}$ to solve the same problem: Find the probability that every person in the queue gets a ticket without waiting because of the seller's temporary inability to give them change. Let $\omega_0, \omega_1, \dots, \omega_{16}$ represent the number of five dollar bills at the counter. Note that $\omega_0 = 0$. The sequence ω_i corresponds to the position of a random walker on the integers starting from 0. At time i , $0 \leq i < 16$, the walker is at position $\omega_i \in \mathbb{Z}$. If a person in the queue buys a ticket with a five dollar bill, the random walker moves to $\omega_i = \omega_{i-1} + 1$. If a person in the queue buys a ticket with a ten dollar bill, the random walker moves to $\omega_i = \omega_{i-1} - 1$. In this method we reflect the paths with respect to the horizontal line $y = -1$, and proceed as in Sect. 8.1.1 obtaining the same answer. (See Fig. 8.2). Consult Durrett (2019) for more information.

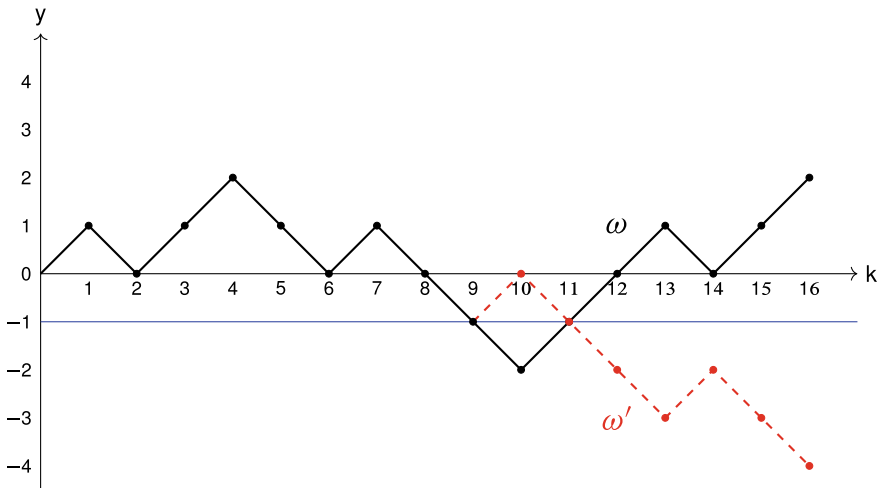


Fig. 8.2 A sample path of ω the random walk arising from a sequence of payments at a ticket office (solid in black), and the reflected path ω' after ω hits a barrier $y = -1$ (dotted in red)

8.2 The Reflection Property of Brownian Motion

Take $m > 0$ (the symbol m is chosen for ‘maximum’) and let $\{W_t\}_{t \geq 0}$ be a Brownian motion with respect to a probability measure $\mathbb{P}$. (The problem is symmetric with respect to m . For $m < 0$, we obtain corresponding equivalent results.) Let τ denote the *first hitting time* of a Brownian particle, i.e.,

$$\tau = \inf\{t \geq 0 : W_t = m\}.$$

See Fig. 8.3, where a sample Brownian path hits the level $m = 3$ and stops there. Consult Simulation 8.1.

In Fig. 8.4 we consider a Brownian path, Path A, and its reflection, Path B, after Path A hits a barrier of height m at $\tau < T$. Since a Brownian particle at the position m at time $\tau < T$ has the equal probability of being above or below m , we observe the following:

$$\mathbb{P}\{W_T < m, \tau < T\} = \mathbb{P}\{W_T > m, \tau < T\} \quad (8.1)$$

and

$$\mathbb{P}\{\tau < T\} = \mathbb{P}\{W_T < m, \tau < T\} + \mathbb{P}\{W_T > m, \tau < T\}. \quad (8.2)$$

Substituting (8.1) into (8.2), we have

$$\mathbb{P}\{\tau < T\} = 2\mathbb{P}\{W_T > m, \tau < T\}.$$

Note that if $W_T > m$ then $\tau < T$, i.e.,

$$\mathbb{P}\{W_T > m, \tau < T\} = \mathbb{P}\{W_T > m\}.$$

Fig. 8.3 A Brownian path is stopped when it hits a barrier $m = 3$

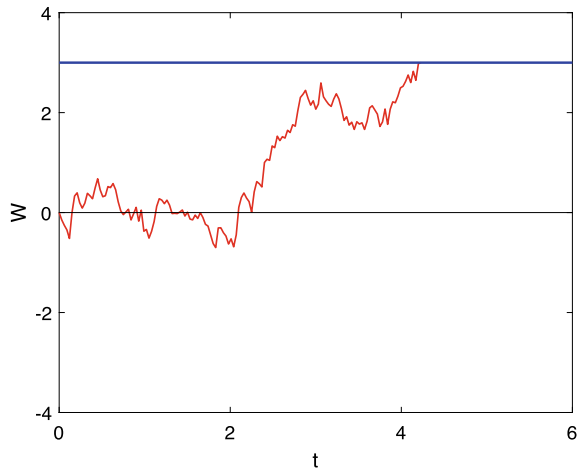
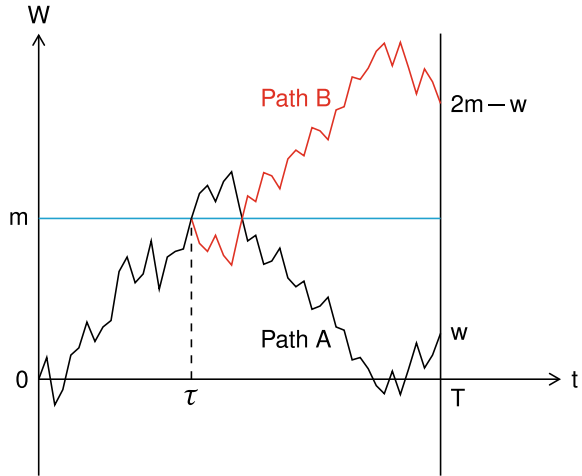


Fig. 8.4 A Brownian path is reflected after it hits a barrier



Hence

$$\mathbb{P}\{\tau < T\} = 2 \mathbb{P}\{W_T > m\}. \quad (8.3)$$

Thus

$$\begin{aligned} \mathbb{P}\{W_T > m | \tau < T\} &= \frac{\mathbb{P}\{W_T > m, \tau < T\}}{\mathbb{P}\{\tau < T\}} \\ &= \frac{1}{2}. \end{aligned}$$

In other words, once a Brownian particle is at the position m at time τ before T , then there is a fifty-fifty chance of being above or below m at T . Let $N(x)$ denote the standard normal distribution function. Then we obtain the following result.

Lemma 8.1 *The first hitting time τ satisfies*

$$\mathbb{P}\{\tau < T\} = 2 N\left(-\frac{m}{\sqrt{T}}\right).$$

Hence the probability density function of τ is given by

$$f_\tau(T) = \frac{m}{T\sqrt{2\pi T}} \exp\left(-\frac{m^2}{2T}\right).$$

Proof Note that (8.3) implies that

$$\begin{aligned}\mathbb{P}\{\tau < T\} &= \frac{2}{\sqrt{2\pi T}} \int_m^\infty \exp\left(-\frac{x^2}{2T}\right) dx \\ &= \frac{2}{\sqrt{2\pi}} \int_{m/\sqrt{T}}^\infty \exp\left(-\frac{x^2}{2}\right) dx .\end{aligned}$$

To obtain the pdf of τ , we take the derivative of $\mathbb{P}\{\tau < T\}$ with respect to T . □

8.3 The Maximum of Brownian Motion

Define

$$M_T = \max\{W_t : 0 \leq t \leq T\} .$$

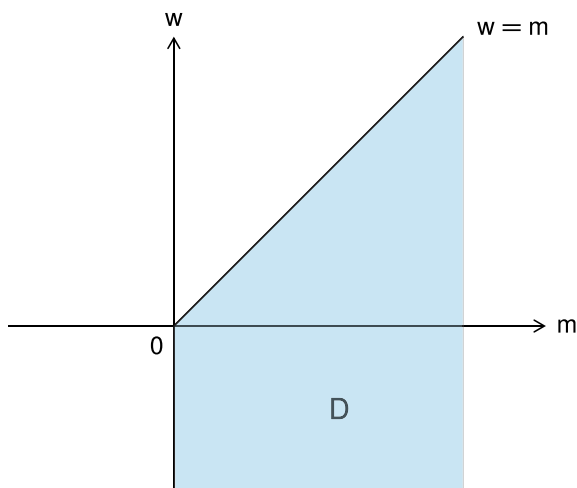
Hence the values of the ordered pair (M_T, W_T) are distributed in the set

$$D = \{(m, w) : w \leq m, m \geq 0\} .$$

(See Fig. 8.5). Note that the condition $\tau < T$ is equivalent to the condition $M_T > m$. Let $w < m$, where w will denote a value assumed by W_T . In Fig. 8.4 we observe that $W_T = w$ for Path A if and only if $W_T = 2m - w$ for Path B. Hence we make the following observations:

$$\mathbb{P}\{M_T > m, W_T < w\} = \mathbb{P}\{M_T > m, W_T > 2m - w\} \quad (8.4)$$

Fig. 8.5 Range of (M_T, W_T)



and

$$\mathbb{P}\{W_T < w\} = \mathbb{P}\{M_T > m, W_T < w\} + \mathbb{P}\{M_T < m, W_T < w\}. \quad (8.5)$$

Substituting (8.4) into (8.5), we have

$$\mathbb{P}\{M_T < m, W_T < w\} = \mathbb{P}\{W_T < w\} - \mathbb{P}\{M_T > m, W_T > 2m - w\}.$$

Since $w < m$, the condition $W_T > 2m - w$ implies that $M_T > m$. Hence

$$\mathbb{P}\{M_T < m, W_T < w\} = \mathbb{P}\{W_T < w\} - \mathbb{P}\{W_T > 2m - w\}. \quad (8.6)$$

Lemma 8.2 (i) Let $w \leq m$ and $m \geq 0$. The probability distribution of W_T to be below a barrier m is given by

$$\mathbb{P}\{M_T < m, W_T < w\} = N\left(\frac{w}{\sqrt{T}}\right) - N\left(-\frac{2m - w}{\sqrt{T}}\right).$$

(ii) For $w \leq m$ and $m \geq 0$, the joint probability density of M_T and W_T is given by

$$f_{M_T, W_T}(m, w) = \frac{2(2m - w)}{T\sqrt{2\pi T}} \exp\left(-\frac{(2m - w)^2}{2T}\right).$$

Proof (i) Use (8.6).

(ii) It suffices to take the second order partial derivative of the cumulative probability distribution

$$\mathbb{P}\{M_T < m, W_T < w\}$$

with respect to w and m . More precisely, we have

$$\frac{\partial}{\partial m} N\left(\frac{w}{\sqrt{T}}\right) = 0$$

and

$$\begin{aligned} & \frac{\partial^2}{\partial w \partial m} N\left(-\frac{2m - w}{\sqrt{T}}\right) \\ &= \frac{\partial}{\partial w} \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(-\frac{2m - w}{\sqrt{T}}\right)^2\right) \left(-\frac{2}{\sqrt{T}}\right) \right] \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(-\frac{2m - w}{\sqrt{T}}\right)^2\right) \frac{\partial}{\partial w} \left[-\frac{1}{2} \left(-\frac{2m - w}{\sqrt{T}}\right)^2 \right] \left(-\frac{2}{\sqrt{T}}\right). \end{aligned}$$

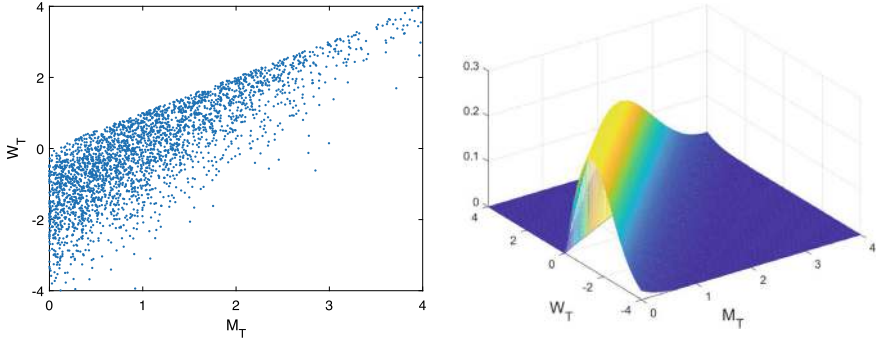


Fig. 8.6 A scatter plot of (M_T, W_T) (left) and $f_{M_T, W_T}(m, w)$ (right) for $T = 2$

Now note that

$$\begin{aligned} \frac{\partial}{\partial w} \left[-\frac{1}{2} \left(-\frac{2m-w}{\sqrt{T}} \right)^2 \right] &= \left(-\frac{1}{2} \right) 2 \left(-\frac{2m-w}{\sqrt{T}} \right) \frac{1}{\sqrt{T}} \\ &= \frac{2m-w}{T}, \end{aligned}$$

which completes the proof. $\square$

See Fig. 8.6 for a scatter plot of (M_T, W_T) and the probability density function $f_{M_T, W_T}(m, w)$, $0 \leq m$ and $w \leq m$. For the simulation consult Simulation 8.2.

8.4 The Maximum of Brownian Motion with Drift

Let $\tilde{W}_t$, $0 \leq t \leq T$, be a Brownian motion with respect to a probability measure $\tilde{\mathbb{P}}$. For a real constant θ let $\hat{W}_t$ be the Brownian motion with drift θ per unit time. More precisely, define

$$\hat{W}_t = \tilde{W}_t + \theta t, \quad 0 \leq t \leq T.$$

Define the maximum of $\hat{W}_t$ by

$$\hat{M}_T = \max_{0 \leq t \leq T} \hat{W}_t.$$

Since $\hat{M}_0 = 0$, we have $\hat{M}_T \geq 0$ and $\hat{M}_T \geq \hat{W}_T$.

Theorem 8.1 *The joint probability density function $\tilde{f}_{\hat{M}_T, \hat{W}_T}$ of $(\hat{M}_T, \hat{W}_T)$ with respect to $\tilde{\mathbb{P}}$ is given by*

$$\tilde{f}_{\hat{M}_T, \hat{W}_T}(m, w) = \frac{2(2m - w)}{T\sqrt{2\pi T}} \exp\left(\theta w - \frac{1}{2}\theta^2 T - \frac{1}{2T}(2m - w)^2\right)$$

in

$$D = \{(m, w) : w \leq m, m \geq 0\},$$

and

$$\tilde{f}_{\hat{M}_T, \hat{W}_T} = 0$$

elsewhere.

Proof Define

$$\hat{Z}_t = e^{-\theta \hat{W}_t - \frac{1}{2}\theta^2 t} = e^{-\theta \hat{W}_t + \frac{1}{2}\theta^2 t}, \quad 0 \leq t \leq T,$$

which is a $\tilde{\mathbb{P}}$ -martingale. Note that there exists a probability measure $\hat{\mathbb{P}}$ such that

$$d\hat{\mathbb{P}} = \hat{Z}_T d\tilde{\mathbb{P}}$$

for which $\hat{W}_t$ is a Brownian motion. (See Theorem 11.1, whose proof does not depend on the results in Chap. 8.) Then

$$\hat{f}_{\hat{M}_T, \hat{W}_T}(m, w) = \frac{2(2m - w)}{T\sqrt{2\pi T}} e^{-\frac{1}{2T}(2m - w)^2}$$

on D , and 0 elsewhere by Lemma 8.2. Now the probability density function of $(\hat{M}_T, \hat{W}_T)$ with respect to $\tilde{\mathbb{P}}$ is given by

$$\begin{aligned} \tilde{\mathbb{P}}\{\hat{M}_T \leq m, \hat{W}_T \leq w\} &= \mathbb{E}^{\tilde{\mathbb{P}}}\left[\mathbb{1}_{\{\hat{M}_T \leq m, \hat{W}_T \leq w\}}\right] \\ &= \mathbb{E}^{\hat{\mathbb{P}}}\left[\frac{1}{\hat{Z}_T} \mathbb{1}_{\{\hat{M}_T \leq m, \hat{W}_T \leq w\}}\right] \\ &= \mathbb{E}^{\hat{\mathbb{P}}}\left[e^{\theta \hat{W}_T - \frac{1}{2}\theta^2 T} \mathbb{1}_{\{\hat{M}_T \leq m, \hat{W}_T \leq w\}}\right] \\ &= \int_{-\infty}^w \int_{-\infty}^m e^{\theta w - \frac{1}{2}\theta^2 T} \hat{f}_{\hat{M}_T, \hat{W}_T}(m, w) dm dw. \end{aligned}$$

By differentiating with respect to m and w , we have

$$\tilde{f}_{\hat{M}_T, \hat{W}_T}(m, w) = e^{\theta w - \frac{1}{2}\theta^2 T} \hat{f}_{\hat{M}_T, \hat{W}_T}(m, w).$$

□

Corollary 8.1 *We have*

$$\tilde{f}_{\hat{M}_T}(m) = \begin{cases} \frac{2}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(m-\theta T)^2} - 2\theta e^{2\theta m} N\left(\frac{-m-\theta T}{\sqrt{T}}\right), & m \geq 0, \\ 0, & m < 0. \end{cases}$$

Proof For $m < 0$, it is clear that

$$\tilde{f}_{\hat{M}_T}(m) = 0.$$

For $m \geq 0$ note that

$$\begin{aligned} \tilde{f}_{\hat{M}_T}(m) &= \int_{-\infty}^{\infty} \tilde{f}_{\hat{M}_T, \hat{W}_T}(m, w) dw \\ &= \int_{-\infty}^m \frac{2(2m-w)}{T\sqrt{2\pi T}} e^{\theta w - \frac{1}{2}\theta^2 T - \frac{1}{2T}(2m-w)^2} dw \\ &= e^{2\theta m} \frac{2}{T\sqrt{2\pi T}} \int_{-\infty}^m (2m-w) e^{-\frac{1}{2T}(w-2m-\theta T)^2} dw \\ &= e^{2\theta m} \frac{2}{T\sqrt{2\pi}} \int_{-\infty}^{(-m-\theta T)/\sqrt{T}} (-\sqrt{T}u - \theta T) e^{-\frac{1}{2}u^2} du \\ &= -\frac{2e^{2\theta m}}{\sqrt{T}\sqrt{2\pi}} \int_{-\infty}^{(-m-\theta T)/\sqrt{T}} u e^{-\frac{1}{2}u^2} du \\ &\quad - \frac{2\theta e^{2\theta m}}{\sqrt{2\pi}} \int_{-\infty}^{(-m-\theta T)/\sqrt{T}} e^{-\frac{1}{2}u^2} du \\ &= \frac{2e^{2\theta m}}{\sqrt{T}\sqrt{2\pi}} e^{-(m+\theta T)^2/2T} - 2\theta e^{2\theta m} N\left(\frac{-m-\theta T}{\sqrt{T}}\right), \end{aligned}$$

where we used the identity

$$\theta w - \frac{1}{2}\theta^2 T - \frac{1}{2T}(2m-w)^2 = -\frac{1}{2T}(w-2m-\theta T)^2 + 2\theta m$$

in the third equality, and we used the substitution

$$u = \frac{w-2m-\theta T}{\sqrt{T}}$$

in the fourth equality. □

Corollary 8.2 *For $m \geq 0$, we have*

$$\tilde{\mathbb{P}}\{\hat{M}_T \leq m\} = N\left(\frac{m-\theta T}{\sqrt{T}}\right) - e^{2\theta m} N\left(\frac{-m-\theta T}{\sqrt{T}}\right).$$

Proof Note that

$$\begin{aligned}
 & \widetilde{\mathbb{P}}\{\widehat{M}_T \leq m\} \\
 &= \int_{-\infty}^m \widetilde{f}_{\widehat{M}_T}(\mu) d\mu \\
 &= \int_0^m \left[\frac{2}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(\mu-\theta T)^2} - 2\theta e^{2\theta\mu} N\left(\frac{-\mu-\theta T}{\sqrt{T}}\right) \right] d\mu \\
 &= \int_0^m \frac{2}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(\mu-\theta T)^2} d\mu - \int_0^m 2\theta e^{2\theta\mu} N\left(\frac{-\mu-\theta T}{\sqrt{T}}\right) d\mu \\
 &= I_1 - I_2,
 \end{aligned}$$

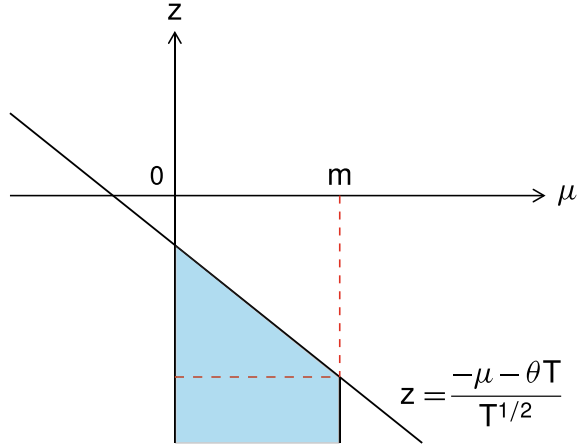
where

$$\begin{aligned}
 I_1 &= \int_0^m \frac{2}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(\mu-\theta T)^2} d\mu \\
 &= 2 \int_{-\theta\sqrt{T}}^{(m-\theta T)/\sqrt{T}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
 &= 2 \left\{ N\left(\frac{m-\theta T}{\sqrt{T}}\right) - N(-\theta\sqrt{T}) \right\}
 \end{aligned}$$

and

$$\begin{aligned}
 I_2 &= \int_0^m 2\theta e^{2\theta\mu} \int_{-\infty}^{(-\mu-\theta T)/\sqrt{T}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz d\mu \\
 &= \int_{-\infty}^{(-m-\theta T)/\sqrt{T}} \int_0^m 2\theta e^{2\theta\mu} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} d\mu dz \\
 &\quad + \int_{(-m-\theta T)/\sqrt{T}}^{-\theta\sqrt{T}} \int_0^{-\sqrt{T}z-\theta T} 2\theta e^{2\theta\mu} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} d\mu dz \\
 &= (e^{2\theta m} - 1) N\left(\frac{-m-\theta T}{\sqrt{T}}\right) \\
 &\quad + \int_{(-m-\theta T)/\sqrt{T}}^{-\theta\sqrt{T}} (e^{2\theta(-\sqrt{T}z-\theta T)} - 1) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \quad (\text{See Fig. 8.7.}) \\
 &= (e^{2\theta m} - 1) N\left(\frac{-m-\theta T}{\sqrt{T}}\right) \\
 &\quad + \int_{(-m-\theta T)/\sqrt{T}}^{-\theta\sqrt{T}} \frac{1}{\sqrt{2\pi}} e^{2\theta(-\sqrt{T}z-\theta T)-\frac{1}{2}z^2} dz \\
 &\quad - \left\{ N(-\theta\sqrt{T}) - N\left(\frac{-m-\theta T}{\sqrt{T}}\right) \right\}.
 \end{aligned}$$

Fig. 8.7 Domain of integration for the double integral I_2



Since

$$2\theta(-\sqrt{T}z - \theta T) - \frac{1}{2}z^2 = -\frac{1}{2}(z + 2\theta\sqrt{T})^2,$$

we have

$$\begin{aligned} I_2 = & (e^{2\theta m} - 1)N\left(\frac{-m - \theta T}{\sqrt{T}}\right) + \left\{N(\theta\sqrt{T}) - N\left(\frac{-m + \theta T}{\sqrt{T}}\right)\right\} \\ & - \left\{N(-\theta\sqrt{T}) - N\left(\frac{-m - \theta T}{\sqrt{T}}\right)\right\}. \end{aligned}$$

Hence $I_1 - I_2$ is equal to

$$\begin{aligned} & 2N\left(\frac{m - \theta T}{\sqrt{T}}\right) - N(-\theta\sqrt{T}) - e^{2\theta m}N\left(\frac{-m - \theta T}{\sqrt{T}}\right) - N(\theta\sqrt{T}) \\ & \quad + N\left(\frac{-m + \theta T}{\sqrt{T}}\right) \\ = & 2N\left(\frac{m - \theta T}{\sqrt{T}}\right) - e^{2\theta m}N\left(\frac{-m - \theta T}{\sqrt{T}}\right) - 1 + N\left(\frac{-m + \theta T}{\sqrt{T}}\right) \\ = & 2N\left(\frac{m - \theta T}{\sqrt{T}}\right) - e^{2\theta m}N\left(\frac{-m - \theta T}{\sqrt{T}}\right) - N\left(\frac{m - \theta T}{\sqrt{T}}\right) \\ = & N\left(\frac{m - \theta T}{\sqrt{T}}\right) - e^{2\theta m}N\left(\frac{-m - \theta T}{\sqrt{T}}\right), \end{aligned}$$

where we used the fact $N(-x) + N(x) = 1$. □

8.5 Computer Experiments

Simulation 8.1 (The First Hitting Time) We plot a sample Brownian path hitting a barrier $m = 3$. In the beginning of the program we set

$$\text{hitting_time} = N + 1,$$

where N is the number of subintervals in the partition of the time interval $[0, T]$ just in case that the generated sample Brownian path does not hit the level $m = 3$ before or at T . See Fig. 8.3.

```
T = 6;
m = 3;
N = 200;
hitting_time = N+1;
dt = T/N;
t = 0:dt:T;
W = zeros(1,N+1);
W(1,1) = 0;
dW = sqrt(dt)*randn(1,N);

for j = 1:N;
    W(1,j+1) = W(1,j) + dW(1,j);
    if ( (W(1,j+1) >= m) )
        W(1,j+1) = m;
        hitting_time = j+1;
        break;
    end
end
plot(0:dt:(hitting_time-1)*dt,W(1,1:hitting_time));
```

Simulation 8.2 (The Maximum of Brownian Motion) For M_T and W_T in Lemma 8.2 the scatter plot of (M_T, W_T) and the joint probability density function $f_{M_T, W_T}(m, w)$ are given. For the output see Fig. 8.6.

```
L = 3000; % number of sample Brownian paths
T = 2;
mu = 0.1;
sigma = 0.3;
N = 3000; % number of time steps
dt = T/N;
t = 0:dt:T;
W = zeros(L,N+1);
```

```

dW = sqrt(dt)*randn(L,N);

M_T = zeros(L,1);

for i = 1:N
    W(:,i+1) = W(:,i) + dW(:,i);
end
W_T(:,1) = W(:,N+1);

for k = 1:L
    M_T(k,1) = max(W(k,:));
end

figure(1)
scatter(M_T(:,1),W_T(:,1),'.')
xlabel('M_T'); ylabel('W_T');
set(gca,'xlim',[0,4],'ylim',[-4,4]);

figure(2)
[m,w] = meshgrid(0:0.02:4,-4:0.04:4);
Z = 2/T/sqrt(2*pi*T)*(2*m-w).*exp(-(2*m-w).^2/2/T);
Z(m < w) = 0;
mesh(m,w,Z)
xlabel('M_T'); ylabel('W_T');

```

Exercises

8.1 (Bertrand's ballot problem) In an election where candidate A receives k votes and candidate B receives ℓ votes with $k > \ell$, what is the probability that A will be strictly ahead of B throughout the count?

Part IV

Itô Calculus

Chapter 9

The Itô Integral



In this chapter we define the Itô integral of a stochastic process and investigate its properties. To define a Riemann–Stieltjes type integral $\int_0^T f(t) d\alpha(t)$ using a function $\alpha : [0, T] \rightarrow \mathbb{R}$ as an integrator, we need the condition that the variation of α is bounded. However, a sample path of a Brownian motion is of unbounded variation since the growth rate of δW is approximately equal to $\sqrt{\delta t}$, which is very large compared with δt as δt decreases to 0. Therefore a Brownian sample path cannot be used as an integrator in a definition of a Riemann–Stieltjes type integral. K. Itô's idea is to take a suitable average over all possible Brownian paths. This idea will be explained gradually since it requires a considerable amount of preparation. Also given is a computer simulation of the Itô integral. For an elementary introduction to the Itô integral, see Roberts (2009) and Wiersema (2008).

9.1 Definition of the Itô Integral

If a stochastic process $\{f_t(\omega)\}_{t \geq 0}$ is measurable with respect to the filtration $\{\mathcal{F}_t\}$ for every $t \geq 0$, then the Itô integral of a stochastic process $f_t(\omega) = f(t, \omega)$ is integrated with respect to an integrator given by a Brownian motion W_t , and the resulting integral is a random variable. More precisely, if we let $t_i = \frac{T}{n}i$ we have

$$\int_0^T f_t(\omega) dW_t(\omega) = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f_{t_i}(\omega) (W_{t_{i+1}}(\omega) - W_{t_i}(\omega)) ,$$

where the limit is defined in $L^2(\Omega, \mathbb{P})$ space. In other words,

$$\mathbb{E} \left[\left| \sum_{i=0}^{n-1} f_{t_i} (W_{t_{i+1}} - W_{t_i}) - \int_0^T f_t dW_t \right|^2 \right] \rightarrow 0$$

as $n \rightarrow \infty$.

Note that in the Itô integral, to approximate $f_t(\omega)$ on a subinterval $[t_i, t_{i+1}]$ we use the value at the left endpoint t_i . This means the Itô integral possesses many useful properties for financial mathematics since we do not know at time t_i what would happen in the future $t_i < t \leq t_{i+1}$.

Let $f_t(\omega) = f(t, \omega)$ be a stochastic process that is adapted to a filtration $\{\mathcal{F}_t\}_{0 \leq t < \infty}$ such that

$$\mathbb{E} \left[\int_0^\infty |f_t(\omega)|^2 dt \right] < \infty .$$

Here $f_t(\omega)$ is regarded as a function $f : t \mapsto f(t, \omega)$, and is continuous for almost every ω , and

$$\int_0^\infty |f_t(\omega)|^2 dt$$

is a Riemann integral for almost every ω where the continuity holds.

Let V denote the collection of all such stochastic processes. It is a vector space with a norm defined by

$$\|f\|_V^2 = \mathbb{E} \left[\int_0^\infty |f_t(\omega)|^2 dt \right] .$$

On the right-hand side we write f_t instead of f to emphasize the fact that it is an integral with respect to t .

Definition 9.1 (*Simple Process*) Let $0 = t_0 < t_1 < \dots < t_n = T$ be a partition of the time interval $[0, T]$. Suppose that for $0 \leq i \leq n-1$ there exist $\mathcal{F}_{t_i}$ -measurable random variables $\zeta_i \in L^2(\Omega)$ such that a stochastic process $\{f_t\}$ is of the form

$$f_t(\omega) = \sum_{i=0}^{n-1} \zeta_i(\omega) \mathbb{1}_{[t_i, t_{i+1})}(t) .$$

Then f_t is called a *simple stochastic process*. (See Fig. 9.1.) The set of all simple stochastic processes is denoted by $\mathcal{H}_0^2$. Since $f \in \mathcal{H}_0^2$ satisfies

$$\|f\|_V^2 = \sum_{i=0}^{n-1} \mathbb{E} [|\zeta_i|^2] (t_{i+1} - t_i) < \infty ,$$

we have $f \in V$. In other words, $\mathcal{H}_0^2$ is a subspace of V .

Definition 9.2 (*Itô Integral*) For a simple process

$$f_t(\omega) = \sum_{i=0}^{n-1} \zeta_i(\omega) \mathbb{1}_{[t_i, t_{i+1})}(t) \in \mathcal{H}_0^2$$

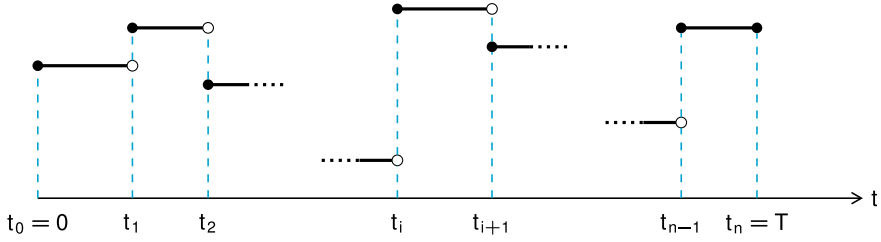


Fig. 9.1 A sample path of a simple process

we define its *Itô integral* $I(f)$ by

$$I(f)(\omega) = \sum_{i=0}^{n-1} \zeta_i(\omega) (W_{t_{i+1}}(\omega) - W_{t_i}(\omega)) .$$

Remark 9.1 (i) An Itô integral $I(f)$ is a random variable.

(ii) To avoid a subscript to a subscript as in W_{t_k} , which is rather difficult to read, we sometimes write $W(t_k)$. Similarly, sometimes $f(t)$ may denote a stochastic process f_t , not a deterministic function of t .

(iii) From time to time we write $\delta W_k = W_{t_{k+1}} - W_{t_k}$ and $\delta t_k = t_{k+1} - t_k$.

(iv) A simple process

$$f_t = \sum_{i=0}^{n-2} \zeta_i \mathbb{1}_{[t_i, t_{i+1})}(t) + \zeta_{n-1} \mathbb{1}_{[t_{n-1}, t_n]}(t)$$

is regarded as being defined over $[0, \infty)$ by taking $f_t = 0$ for $t \geq T$.

9.2 Itô Isometry

Theorem 9.1 (Itô Isometry) *If $f \in \mathcal{H}_0^2$, then $I(f) \in L^2(\Omega, \mathbb{P})$ and*

$$I : \mathcal{H}_0^2 \rightarrow L^2(\Omega, \mathbb{P})$$

is a norm-preserving linear transformation. In other words,

$$\mathbb{E} [|I(f)|^2] = \mathbb{E} \left[\int_0^\infty |f_t|^2 dt \right] .$$

Proof The proof that I is a linear transformation is trivial, and omitted. If a step stochastic process f_t is of the form

$$f_t(\omega) = \sum_{i=0}^{n-1} \zeta_i(\omega) \mathbb{1}_{[t_i, t_{i+1})}(t) \in \mathcal{H}_0^2,$$

then

$$\begin{aligned} |I(f)|^2 &= \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \zeta_j \zeta_k \delta W_j \delta W_k \\ &= \sum_{k=0}^{n-1} |\zeta_k|^2 (\delta W_k)^2 + 2 \sum_{j < k} \zeta_j \zeta_k \delta W_j \delta W_k. \end{aligned}$$

Since ζ_k and δW_k are independent and $\mathbb{E}[(\delta W_k)^2] = \delta t_k$, we have

$$\mathbb{E}[|\zeta_k|^2 (\delta W_k)^2] = \mathbb{E}[|\zeta_k|^2] \mathbb{E}[(\delta W_k)^2] = \mathbb{E}[|\zeta_k|^2] \delta t_k.$$

If $j < k$, then $\zeta_j \zeta_k \delta W_j$ and δW_k are independent, and hence

$$\mathbb{E}[\zeta_j \zeta_k \delta W_j \delta W_k] = \mathbb{E}[\zeta_j \zeta_k \delta W_j] \mathbb{E}[\delta W_k] = 0.$$

Thus

$$\mathbb{E}[|I(f)|^2] = \sum_{k=0}^{n-1} \mathbb{E}[|\zeta_k|^2] \delta t_k < \infty. \quad (9.1)$$

Therefore, $I(f) \in L^2(\Omega)$.

On the other hand, since

$$|f_t|^2 = \sum_{j=0}^{n-1} \sum_{k=0}^{n-1} \zeta_j \zeta_k \mathbb{1}_{[t_j, t_{j+1})}(t) \mathbb{1}_{[t_k, t_{k+1})}(t) = \sum_{k=0}^{n-1} |\zeta_k|^2 \mathbb{1}_{[t_k, t_{k+1})}(t),$$

we have

$$\int_0^\infty |f_t|^2 dt = \sum_{k=0}^{n-1} |\zeta_k|^2 \Delta t_k$$

and finally obtain

$$\mathbb{E} \left[\int_0^\infty |f_t|^2 dt \right] = \sum_{k=0}^{n-1} \mathbb{E}[|\zeta_k|^2] \Delta t_k. \quad (9.2)$$

Now we use (9.1) and (9.2) to complete the proof. $\square$

Definition 9.3 (*Extension of Itô Integral*) Since the transformation $I : \mathcal{H}_0^2 \rightarrow L^2(\Omega, \mathbb{P})$ is linear and continuous, and since $L^2(\Omega, \mathbb{P})$ is a complete metric space, by Corollary B.1 the domain of I can be extended from $\mathcal{H}_0^2$ to a set, denoted by $\mathcal{H}^2$, which contains $\mathcal{H}_0^2$ as a dense subset. (When we emphasize the finite time interval $[0, T]$, we write $\mathcal{H}_T^2$ in place of $\mathcal{H}^2$.) Then the integral $I(f) \in L^2(\Omega, \mathbb{P})$ of a stochastic process $f(t, \omega)$ which is approximated by a sequence of simple processes $f_n(t, \omega)$ is determined by $I(f) = \lim_{n \rightarrow \infty} I(f_n)$, where the limit is taken in the L^2 -sense. In other words,

$$\lim_{n \rightarrow \infty} \mathbb{E} [|I(f) - I(f_n)|^2] = 0 .$$

Then $I(f)$ is called the *Itô integral* and is denoted by

$$\int_0^\infty f_t \, dW_t .$$

Remark 9.2 (i) It is clear that Theorem 9.1 holds for $f \in \mathcal{H}^2$.

(ii) To approximate $f_t \in \mathcal{H}^2$ we use a simple process

$$f_n(t, \omega) = \sum_{i=0}^{n-1} f_{t_i}(\omega) \mathbb{1}_{[t_i, t_{i+1})}(t) \in \mathcal{H}_0^2 ,$$

and $I(f)$ is the L^2 -limit of

$$I(f_n) = \sum_{i=0}^{n-1} f_{t_i} (W_{t_{i+1}} - W_{t_i}) .$$

(See Fig. 9.2.) Note that we choose the *left endpoint* t_i in each interval $[t_i, t_{i+1})$.

This rule makes the Itô integral a martingale. See Theorem 9.4.

(iii) Considering the product measure space $\Omega \times [0, \infty)$ and Fubini's theorem we have

$$\mathbb{E} \left[\int_0^\infty |f_t|^2 \, dt \right] = \int_0^\infty \mathbb{E}[|f_t|^2] \, dt .$$

Theorem 9.2 (Itô Isometry) *For stochastic processes $f, g \in \mathcal{H}^2$ we have*

$$\mathbb{E}[I(f)I(g)] = \mathbb{E} \left[\int_0^\infty f_t g_t \, dt \right] .$$

Using the inner products defined on $\mathcal{H}^2$ and $L^2(\Omega, \mathbb{P})$, we observe that the equation is equivalent to the relation

$$(I(f), I(g))_{L^2} = (f, g)_{\mathcal{H}^2} .$$

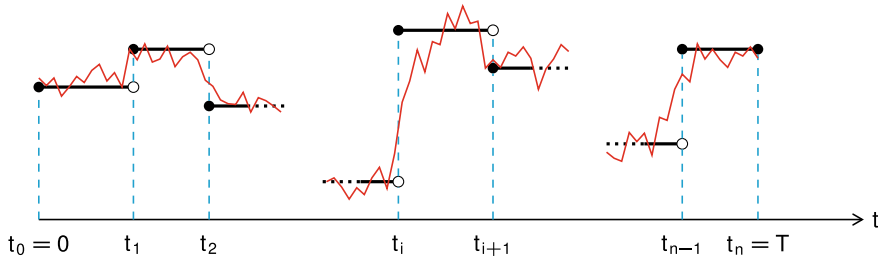


Fig. 9.2 Approximation of a general stochastic process by a simple process

Proof Using the inner products on the spaces $\mathcal{H}^2$ and $L^2(\Omega, \mathbb{P})$, we obtain

$$\begin{aligned}
 & (f, f)_{\mathcal{H}^2} + (g, g)_{\mathcal{H}^2} + 2(f, g)_{\mathcal{H}^2} \\
 &= (f + g, f + g)_{\mathcal{H}^2} \\
 &= (I(f + g), I(f + g))_{L^2} \quad (\text{by Theorem 9.1}) \\
 &= (I(f) + I(g), I(f) + I(g))_{L^2} \\
 &= (I(f), I(f))_{L^2} + (I(g), I(g))_{L^2} + 2(I(f), I(g))_{L^2} \\
 &= (f, f)_{\mathcal{H}^2} + (g, g)_{\mathcal{H}^2} + 2(I(f), I(g))_{L^2} . \quad (\text{by Theorem 9.1})
 \end{aligned}$$

Hence $(f, g)_{\mathcal{H}^2} = (I(f), I(g))_{L^2}$. □

Definition 9.4 If a stochastic process $\{f_t\}_{t \geq 0}$ is adapted to a filtration $\{\mathcal{F}_t\}_{t \geq 0}$, then $\{\mathbb{1}_{[0, T]}(t)f_t\}_{t \geq 0}$ is also adapted to $\{\mathcal{F}_t\}_{t \geq 0}$. For arbitrary $T > 0$ we define the Itô integral on a finite interval $[0, T]$ by

$$\int_0^T f_t dW_t = \int_0^\infty \mathbb{1}_{[0, T]}(t) f_t dW_t .$$

For example, if we take $f_t = 1$, then

$$\int_0^T dW_t = W_T - W_0 = W_T .$$

As we have seen in Theorem 9.2, $I : \mathcal{H}^2 \rightarrow L^2(\Omega, \mathbb{P})$ is norm-preserving. Since $\mathcal{H}_0^2$ is dense in $\mathcal{H}^2$, most properties of I on $\mathcal{H}_0^2$ are extended to $\mathcal{H}^2$. In summary, we have the following result.

Theorem 9.3 For stochastic processes f_t, g_t , the following holds:

(i) For a, b constant,

$$\int_0^t (af_u + bg_u) dW_u = a \int_0^t f_u dW_u + b \int_0^t g_u dW_u .$$

(ii) For every $t \geq 0$,

$$\mathbb{E} \left[\left| \int_0^t f_u dW_u \right|^2 \right] = \mathbb{E} \left[\int_0^t |f_u|^2 du \right] .$$

Corollary 9.1 Let $f(t)$, $a \leq t \leq b$, be a deterministic function in $L^2[a, b]$. Then the Itô integral $\int_a^b f(t) dW_t$ is a normally distributed random variable with mean 0 and variance

$$\|f\|^2 = \int_a^b |f(t)|^2 dt .$$

Proof The statement holds true if f is a step function by Theorem 9.1. For an arbitrary $f \in L^2[a, b]$, we take a sequence of step functions f_n converging to f in $L^2[a, b]$ and apply Theorem 9.2. $\square$

9.3 Examples of the Itô Integral

9.3.1 A Correction Term for the Itô Integral

If a function $g(x)$ is differentiable on $[0, T]$ and $g(0) = 0$, then the Riemann integral satisfies

$$\int_0^T g(t) g'(t) dt = \frac{1}{2} g(T)^2 ,$$

however, for the Itô integral we have

$$\int_0^T W_t dW_t = \frac{1}{2} W_T^2 - \frac{1}{2} T . \quad (9.3)$$

(For a computational experiment, see Fig. 9.3, where a single sample path of the numerically obtained integral and the exact solution are compared. In Fig. 9.4 we generate multiple sample paths.)

To prove (9.3), we take $f(t) = \mathbb{1}_{[0, T]}(t) W(t) \in \mathcal{H}^2$. Then

$$\int_0^T W_t dW_t = \int_0^\infty f_t dW_t .$$

Take $t_i = i \times \frac{T}{n}$, and partition the interval $[0, T]$ by $0 = t_0 < t_1 < \dots < t_n = T$. Define

$$f_n(t) = \sum_{i=0}^{n-1} \mathbb{1}_{[t_i, t_{i+1})}(t) W(t_i) \in \mathcal{H}_0^2 .$$

Fig. 9.3 A sample path of the Itô integral $\int_0^t W_s dW_s$ (solid) and a sample path of the exact solution $\frac{1}{2}W_t^2 - \frac{1}{2}t$ (dotted) in Sect. 9.3.1

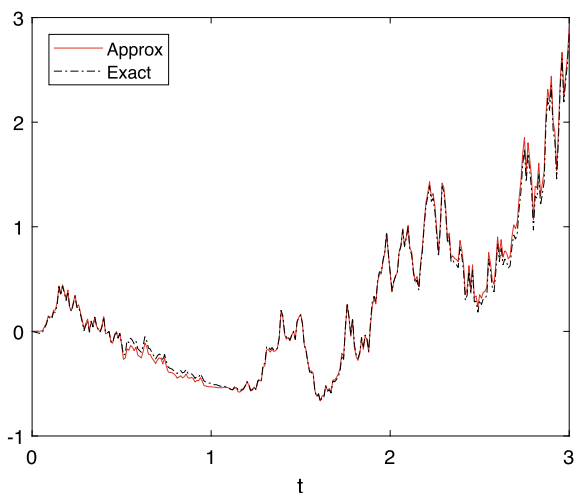
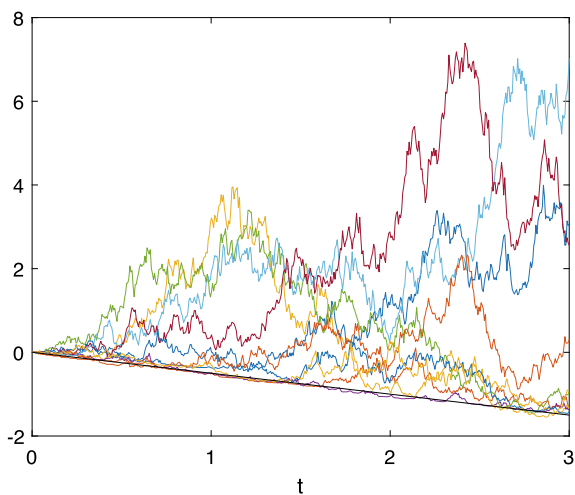


Fig. 9.4 Sample paths of the stochastic process $\int_0^t W_s dW_s$, $0 \leq t \leq 3$, in Sect. 9.3.1



Since

$$\begin{aligned}
 \mathbb{E} \left[\int_0^\infty |f(t) - f_n(t)|^2 dt \right] &= \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \mathbb{E} [|W(t) - W(t_i)|^2] dt \\
 &= \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} (t - t_i) dt \\
 &= \frac{1}{2} \sum_{i=0}^{n-1} (t_{i+1} - t_i)^2 \\
 &= \frac{1}{2} n \left(\frac{T}{n} \right)^2 = \frac{T^2}{2n} \rightarrow 0,
 \end{aligned}$$

we see that $f_n(t)$ converges to $f(t)$ as $n \rightarrow \infty$ by Theorem 9.2. Now we have

$$\begin{aligned} I(f_n) &= \sum_{i=0}^{n-1} W(t_i) \{W(t_{i+1}) - W(t_i)\} \\ &= \frac{1}{2} \sum_{i=0}^{n-1} \{W(t_{i+1})^2 - W(t_i)^2\} - \frac{1}{2} \sum_{i=0}^{n-1} \{W(t_{i+1}) - W(t_i)\}^2 \\ &= \frac{1}{2} W(T)^2 - \frac{1}{2} \sum_{i=0}^{n-1} \{W(t_{i+1}) - W(t_i)\}^2, \end{aligned}$$

which converges to $\frac{1}{2} W(T)^2 - \frac{1}{2} T$ in the L^2 -sense by Theorem 7.4.

In Simulation 9.3 where we take $T = 3$ and consider the speed of convergence of the L^2 -norm of the difference

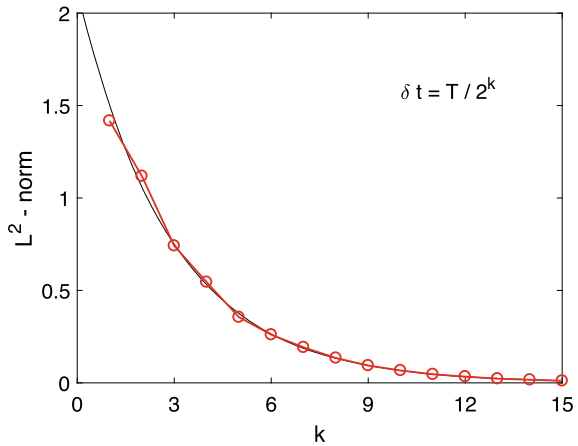
$$I(f_n) - \left(\frac{1}{2} W_T^2 - \frac{1}{2} T \right) = -\frac{1}{2} \sum_{i=0}^{n-1} \{W(t_{i+1}) - W(t_i)\}^2 + \frac{1}{2} T,$$

where $\delta t = \frac{T}{n}$, $n = 2^k$, $1 \leq k \leq 15$. Then, by Remark 7.3, we have

$$\left\| I(f_n) - \left(\frac{1}{2} W_T^2 - \frac{1}{2} T \right) \right\|_2 = \frac{1}{2} \sqrt{\frac{2T^2}{n}} = 2^{-(k+1)/2} T.$$

(See Fig. 9.5 where the numerical estimate of the L^2 -norm of the difference is shown (red circles) together with the graph of $y = 2^{-(x+1)/2} T$, $0 \leq x \leq 15$.)

Fig. 9.5 Speed of L^2 -convergence of $I(f_n) = \sum W_i \delta W_i$ to $\frac{1}{2} W_T^2 - \frac{1}{2} T$ in Sect. 9.3.1



9.3.2 The Itô Integral of Deterministic Functions

The Itô isometry in Theorem 9.1 implies that the following Itô integrals

$$X = \int_0^T t \, dW_t \quad \text{and} \quad Y = \int_0^T (T - t) \, dW_t$$

are normally distributed with mean 0 and variance equal to

$$\int_0^T t^2 \, dt = \int_0^T (T - t)^2 \, dt = \frac{T^3}{3}.$$

That is, $X \sim N(0, \frac{T^3}{3})$ and $Y \sim N(0, \frac{T^3}{3})$. See Fig. 9.6 where the normalized histograms of X and Y are presented. Also plotted is the pdf $\frac{1}{\sqrt{2\pi\sigma}} \exp(-\frac{x^2}{2\sigma^2})$ of the normal distribution $N(0, \sigma^2)$ where $\sigma^2 = \frac{T^3}{3}$.

Since

$$X + Y = \int_0^T t \, dW_t + \int_0^T (T - t) \, dW_t = \int_0^T T \, dW_t = T W_T \sim N(0, T^3),$$

we see that $\text{Cov}(X, Y) = \frac{1}{6}T^3$. (For more information on covariance, see Sect. 4.7.) We may prove it directly by applying Theorem 9.2 since

$$\text{Cov}(X, Y) = \mathbb{E}[XY] = \int_0^T t(T - t) \, dt = \frac{1}{6}T^3.$$

Note that the correlation of X and Y equals $\frac{1}{2}$. (See Fig. 9.7 for the scatter plot of the pair (X, Y) and also Simulation 9.4.)

9.4 The Martingale Property of the Itô Integral

Theorem 9.4 *Let $(\Omega, \mathbb{P})$ be a probability space with a filtration $\{\mathcal{F}_t : t \geq 0\}$ and let $\{W_t\}_{t \geq 0}$ be a $(\mathbb{P}, \{\mathcal{F}_t : t \geq 0\})$ -Brownian motion. Let $\{g(t, \omega) : t \geq 0\}$ be an $\mathcal{F}_t$ -predictable process such that $\mathbb{E}[g^2(t, \omega)] < \infty$, then the stochastic process*

$$M_t = \int_0^t g(s, \omega) \, dW_s$$

is a martingale.

Proof First, we show that the statement holds for $\mathcal{H}_0^2$. By linearity, it suffices to show the claim for a simple process of the form

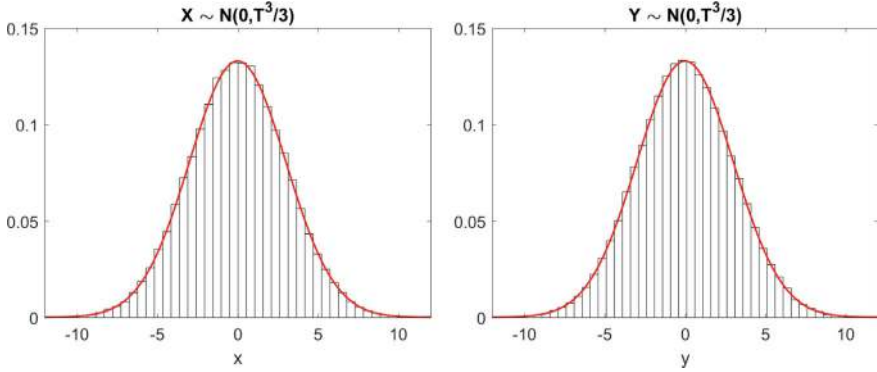
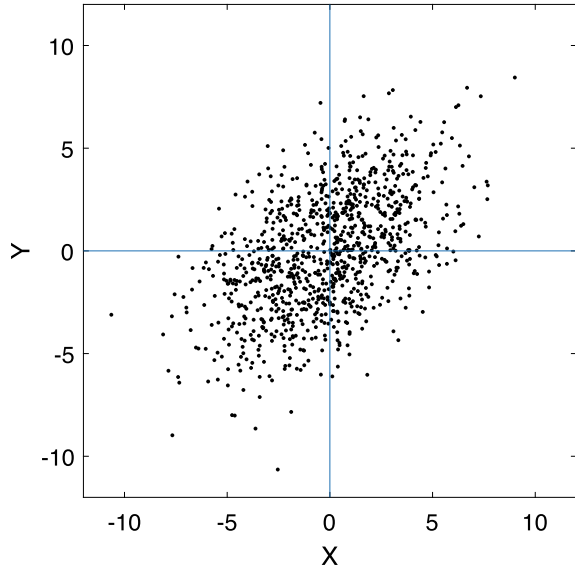


Fig. 9.6 For $T = 3$, normalized histograms of the probability distributions of random variables $X = \int_0^T t dW_t$ (left) and $Y = \int_0^T (T-t) dW_t$ (right) in Sect. 9.3.2 plotted together with the probability density function of the normal distribution $N(0, T^3/3)$ (in red)

Fig. 9.7 Scatter plot of (X, Y) where $X = \int_0^T t dW_t$ and $Y = \int_0^T (T-t) dW_t$ for $T = 3$ in Sect. 9.3.2



$$g_s(\omega) = \zeta(\omega) \mathbb{1}_{[a,b)}(s)$$

for every $0 \leq a < b$ where ζ is $\mathcal{F}_a$ -measurable. Then

$$M_t = \begin{cases} 0, & 0 \leq t \leq a, \\ \zeta(W_t - W_a), & a \leq t \leq b, \\ \zeta(W_b - W_a), & b \leq t. \end{cases}$$

Now it is straightforward to check that $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$ for $0 \leq s \leq t$. Finally, we use the fact that the linear transformation defined by taking conditional expectation $g \mapsto \mathbb{E}[g|\mathcal{F}_t]$, $g \in \mathcal{H}_0^2$, is continuous. $\square$

Corollary 9.2 *The Itô integral has mean zero, that is,*

$$\mathbb{E} \left[\int_0^t f(s, \omega) dW_s \right] = 0$$

where the expectation represents the Lebesgue integral on Ω .

Proof Put

$$M_t = \int_0^t f(s, \omega) dW_s .$$

Then

$$M_0 = \int_0^0 f(s, \omega) dW_s = 0 .$$

Since $\{M_t\}$ is a martingale, $\mathbb{E}[M_t]$ is constant, and thus $\mathbb{E}[M_t] = \mathbb{E}[M_0] = 0$. $\square$

Example 9.1 From the result in Sect. 9.3.1 we have

$$0 = \mathbb{E} \left[\int_0^T W_t dW_t \right] = \frac{1}{2} \mathbb{E} [W_T^2] - \frac{1}{2} T ,$$

and hence $\mathbb{E}[W_T^2] = T$, which coincides with the fact that the variance of W_T is equal to T .

In Fig. 9.8 obtained from Simulation 9.5 we check Corollary 9.2 for an arbitrary stochastic process $f(t, W_t)$, $0 \leq t \leq T$. More specifically, we show that the average of M sample paths of $\int_0^t f(s, W_s) dW_s$ at each fixed t is close to zero in agreement with Corollary 9.2. We take $M = 5$, which is too small to apply the Central Limit Theorem. However, we plot just a few sample paths for the visual clarity. For a better numerical result in the final conclusion, we need to choose a larger number for M .

For a computer experiment, we choose $f(t, W_t) = \exp(W_t)$, $T = 1$ and $N = 10$. Let $\delta t = \frac{T}{N} = 0.1$, and $t_i = i \times \delta t$, $i = 0, 1, \dots, N$. Note that $t_0 = 0$ and $t_N = T$. In the first panel we plot M increments of a Brownian motion, $dW(t_i)$, at $t = t_i$, $i = 0, 1, \dots, N - 1$. In the second panel, we plot the Brownian motion W by adding the increments dW , i.e., $W(0) = 0$ and $W(t_i) = \sum_{k=0}^{i-1} dW(t_k)$, $i = 1, \dots, N$. The third panel shows a step process on $0 \leq t \leq T$ defined by $f(t_i, W(t_i))$, $t_i \leq t < t_{i+1}$, $i = 0, 1, \dots, N - 2$, and $f(t_{N-1}, W(t_{N-1}))$, $t_{N-1} \leq t \leq t_N$. The fourth panel shows the product $f(t_i, W(t_i))dW(t_i)$ at $t = t_i$, $i = 0, 1, \dots, N - 1$. In the fifth panel we plot the sample paths of a process $\int_0^t f(u, W_u) dW_u$ evaluated at $t = t_i$, $i = 1, \dots, N$. In fact, we use $\sum_{k=0}^{i-1} f(t_k, W(t_k))dW(t_k)$ as an approximation of $\int_0^{t_i} f(u, W_u) dW_u$, $i = 1, \dots, N$. In the sixth and final panel we plot the average of

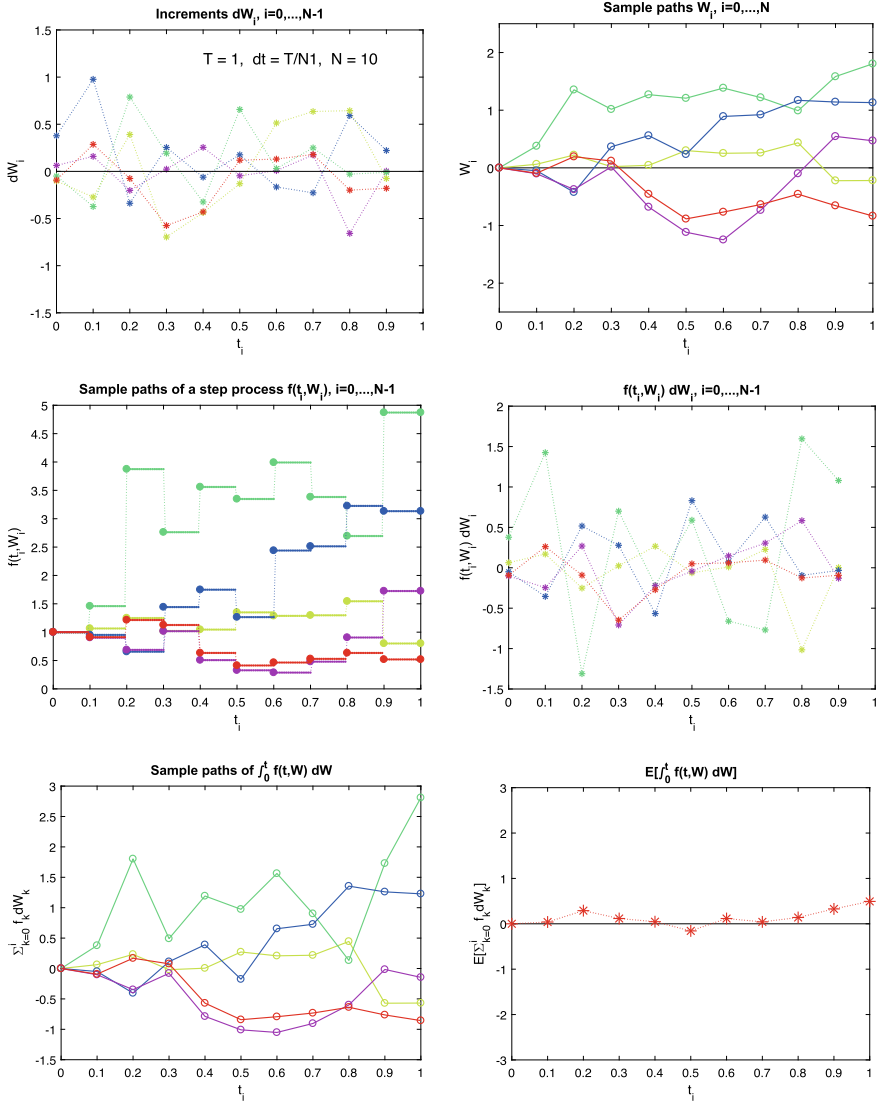


Fig. 9.8 The Itô integral $\int_0^t f_s dW_s$ has zero mean for $f(t, W_t) = \exp(W_t)$, $0 \leq t \leq 1$: five samples of increments dW_i of a Brownian motion, $0 \leq i \leq N-1$; sample paths of a Brownian motion $W(t)$ observed at time $t = t_i$, $0 \leq i \leq N$; sample paths of a step process defined by $f(t_i, W(t_i))$, $t_i \leq t < t_{i+1}$, $0 \leq i \leq N-1$; $f(t_i, W_i) dW_i$, $0 \leq i \leq N-1$; sample paths of $\int_0^t f(u, W_u) dW_u$ observed at $t = t_i$, $0 \leq i \leq N$; and a numerical demonstration of $\mathbb{E}[\int_0^t f(s, W_s) dW_s] = 0$ at $t = t_i$, $0 \leq i \leq N$ (from top left to bottom right)

$$\sum_{k=0}^{i-1} f(t_k, W(t_k)) dW(t_k)$$

over M sample paths for each $i = 0, 1, \dots, N$.

9.5 Stochastic Integrals with Respect to a Martingale

For $n = 0, 1, 2, \dots$, let M_n be a discrete time martingale and H_n a discrete time predictable process. If H_n is bounded, or if both H_n and M_n are L^2 -integrable, then $X_0 = 0$ and

$$X_n = \sum_{k=1}^n H_k (M_k - M_{k-1})$$

for $n \geq 1$, defines a discrete time martingale, and we may informally write

$$X_n = \int H_k dM_k .$$

If M_t is a continuous time martingale, as in the special case when $M_t = W_t$, using M_t as an integrator, we can define a stochastic integral

$$X_t = \int_0^t f(s, M_s) dM_s$$

and derive the properties corresponding to the standard Itô integral. First, we define the stochastic integral for a simple process

$$f_t(\omega) = \sum_{i=0}^{n-1} \zeta_i(\omega) \mathbb{1}_{[t_i, t_{i+1})}(t)$$

by

$$\int_0^T f_s dM_s = \sum_{i=0}^{n-1} \zeta_i (M_{t_{i+1}} - M_{t_i}) ,$$

then extend it to the general case by taking the limit and obtain X_t . It is known that $\{X_t\}_{t \geq 0}$ is also a martingale. More precisely, take a predictable process $f(t, \omega) \in L^2([0, T] \times \Omega)$ such that

$$\mathbb{E} \left[\int_0^T |f_t|^2 d\langle M \rangle_t \right] < \infty .$$

Then

$$X_t = \int_0^t f_s \, dM_s, \quad 0 \leq t \leq T,$$

is a martingale and the following equality holds:

$$\mathbb{E}[X_t^2] = \mathbb{E} \left[\int_0^t |f_s|^2 \, d\langle M \rangle_s \right].$$

Furthermore,

$$\langle X \rangle_t = \int_0^t |f_s|^2 \, d\langle M \rangle_s.$$

Example 9.2 Let W_t be a Brownian motion, and take an adapted process $g(s, \omega) \in L^2([0, T], \Omega)$. Consider the martingale

$$M_t = \int_0^t g_s \, dW_s, \quad 0 \leq t \leq T.$$

Then

$$\langle M \rangle_t = \int_0^t |g_s|^2 \, ds.$$

Hence $f \in L^2_{\text{pred}}([0, T] \times \Omega)$ with the condition that

$$\mathbb{E} \left[\int_0^T |f_t|^2 \, d\langle M \rangle_t \right] < \infty$$

if and only if

$$\mathbb{E} \left[\int_0^T |f_t|^2 |g_t|^2 \, dt \right] < \infty.$$

Furthermore, we have

$$\int_0^T f_t \, dM_t = \int_0^T f_t g_t \, dW_t.$$

Theorem 9.5 Let $\{X_t\}_{t \geq 0}$ be a continuous martingale. Then the process

$$\int_0^t f(u, X_u) dX_u, \quad 0 \leq t < \infty,$$

is a martingale.

Proof Here is a sketch of the proof. For $0 \leq s < t$, we have

$$\begin{aligned} & \mathbb{E} \left[\int_0^t f(u, X_u) dX_u \middle| \mathcal{F}_s \right] \\ &= \mathbb{E} \left[\int_0^s f(u, X_u) dX_u \middle| \mathcal{F}_s \right] + \mathbb{E} \left[\int_s^t f(u, X_u) dX_u \middle| \mathcal{F}_s \right] \end{aligned}$$

and the second term on the right is approximated by

$$\begin{aligned} \mathbb{E} \left[\int_s^t f(u, X_u) dX_u \middle| \mathcal{F}_s \right] &\approx \mathbb{E} \left[\sum_{i=0}^{n-1} f(u_i, X_{u_i})(X_{u_{i+1}} - X_{u_i}) \middle| \mathcal{F}_s \right] \\ &= \sum_{i=0}^{n-1} \mathbb{E} \left[f(u_i, X_{u_i})(X_{u_{i+1}} - X_{u_i}) \middle| \mathcal{F}_s \right] \\ &= \sum_{i=0}^{n-1} \mathbb{E} \left[\mathbb{E} \left[f(u_i, X_{u_i})(X_{u_{i+1}} - X_{u_i}) \middle| \mathcal{F}_{u_i} \right] \middle| \mathcal{F}_s \right] \\ &= \sum_{i=0}^{n-1} \mathbb{E} \left[f(u_i, X_{u_i})(X_{u_i} - X_{u_i}) \middle| \mathcal{F}_s \right] \\ &= 0 \end{aligned}$$

where $s = u_0 < \cdots < u_i < u_{i+1} < \cdots < u_n = t$. □

For more details see Kuo (2006).

9.6 The Martingale Representation Theorem

Let $\{\mathcal{F}_t\}_{t \geq 0}$ be a filtration on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\{M_t\}_{t \geq 0}$ be a martingale adapted to $\{\mathcal{F}_t\}$. If $\mathbb{E}[M_t^2] < \infty$ for every $t \geq 0$, then $\{M_t\}_{t \geq 0}$ is called a *square-integrable martingale*.

The following is a special case of the Martingale Representation Theorem when the martingale under consideration is given by Brownian motion.

Theorem 9.6 (Martingale Representation Theorem) *For a $\mathbb{P}$ -Brownian motion $\{W_t\}_{t \geq 0}$ let $\{\mathcal{F}_t\}_{t \geq 0}$ be the filtration generated by $\{W_t\}$. If a square-integrable process $\{M_t\}_{t \geq 0}$ is adapted to $\{\mathcal{F}_t\}_{t \geq 0}$, and it is a $\mathbb{P}$ -martingale, then there exists an $\{\mathcal{F}_t\}$ -predictable process $\{\alpha_t\}_{t \geq 0}$ satisfying*

$$M_t = M_0 + \int_0^t \alpha_s dW_s$$

with probability 1.

Remark 9.3 Theorem 9.6 is the converse of Theorem 9.4. Its conclusion can be written as $dM_t = \alpha_t dW_t$. In other words, if M_t is a martingale, then dM_t has no dt -term, and it looks as if M_t were differentiable with respect to W_t with its derivative equal to α_t .

Proof Since M_0 is $\mathcal{F}_0$ -measurable, it is constant with probability 1. It suffices to prove the statement for $M_0 = 0$ since we can prove the above result for N_t using $N_t = M_t - M_0$ instead of M_t if necessary. By the properties of a martingale, we can easily see that $\mathbb{E}[M_t] = 0$ and $\mathbb{E}[M_T | \mathcal{F}_t] = M_t$.

Let V be the set of square-integrable stochastic processes adapted to the given filtration. For $X_t, Y_t \in V$ define

$$(X_t, Y_t)_V = \mathbb{E} \left[\int_0^T X_s Y_s ds \right],$$

then $(\cdot, \cdot)_V$ is an inner product on a Hilbert space V . If we let

$$L_0^2(\Omega, \mathcal{F}_T, \mathbb{P}) = \{Z \in L^2(\Omega, \mathcal{F}_T, \mathbb{P}) : \mathbb{E}[Z] = 0\},$$

then the linear transformation $I : V \rightarrow L_0^2(\Omega, \mathcal{F}_T, \mathbb{P})$ defined by

$$I(\{X_t\}_{t \geq 0}) = \int_0^T X_s dW_s$$

preserves distance. Now it remains to show that I is surjective. Then there would exist $\{\alpha_t\}_{0 \leq t \leq T} \in V$ such that $M_T = \int_0^T \alpha_s dW_s$. Hence

$$M_t = \mathbb{E}[M_T | \mathcal{F}_t] = \mathbb{E} \left[\int_0^T \alpha_s dW_s \middle| \mathcal{F}_t \right] = \int_0^t \alpha_s dW_s$$

and the proof is complete. For the details consult (Korn and Korn 2001; Oksendal 2003). $\square$

Example 9.3 Let $X_t = \int_0^t a(s) dW_s$, where $a(t)$ is a deterministic function. Then X_t is a martingale. For, if we apply the Itô formula to X_t^2 , then

$$d(X_t^2) = X_t dX_t + X_t dX_t + dX_t dX_t = 2X_t dX_t + a(t)^2 dt$$

and $d(X_t^2) - a(t)^2 dt = 2X_t dX_t$. Hence $X_t^2 - \int_0^t a(s)^2 ds$ is a martingale.

9.7 Computer Experiments

Simulation 9.1 (Itô Integral) We generate a single sample path of $\int_0^t W_s dW_s$ and compare it with the theoretical formula $\frac{1}{2}W_t^2 - \frac{1}{2}t$. Recall that in the definition of the Itô integral we take the left endpoint from the subinterval $[t_i, t_{i+1}]$ to evaluate the integrand. In the following program the partition of the time interval $[0, T]$ is represented by an array $t = 0:dt:T$ of length $N + 1$ in such a way that

$$t(1) = t_0 = 0, \dots, t(i+1) = t_i, \dots, t(N+1) = t_N = T.$$

The array `Integral` represents the partial sum $\sum_{j=0}^i W_{t_j} dW_{t_j}$, $0 \leq i \leq N-1$, as the time progresses while `Exact` is the exact formula $\frac{1}{2}W_t^2 - \frac{1}{2}t$. For the output see Fig. 9.3.

```
T= 3.0; N = 300; dt = T/N; t = 0:dt:T;
dW = sqrt(dt)*randn(1,N);
W = zeros(1,N+1);
Integral = zeros(1,N+1);
Exact = zeros(1,N+1);

for i = 1:N
    W(i+1) = W(i) + dW(i);
    Integral(i+1) = Integral(i) + W(i)*dW(i); % Take the left endpoint.
    Exact(i+1)=W(i+1)^2/2 - i*dt/2;
end
plot(t,Integral,'r-',t,Exact,'k-.');
xlabel('t');
hlegend=legend('approx','exact');
```

Simulation 9.2 (Itô Integral) We generate ten sample paths of the stochastic process $\int_0^t W_s dW_s$, $0 \leq t \leq T$. See Fig. 9.4, where the straight line $y = -\frac{1}{2}t$ is also plotted to check whether $\frac{1}{2}W_t^2 - \frac{1}{2}t \geq -\frac{1}{2}t$. Observe that the sample paths might go below the straight line due to numerical discretization of the time axis.

```
T = 3.0; M = 10; N = 500; dt = T/N; t = 0:dt:T;
dW = sqrt(dt)*randn(M,N);
W = [zeros(M,1), cumsum(dW,2)];
Integral = [zeros(M,1), cumsum(W(:,1:N).*dW,2)];
plot(t,Integral)
hold on
plot(t,-0.5*t)
xlabel('t');
```

Simulation 9.3 (Convergence) For the Itô integral $\int_0^T W_t dW_t$ in Sect. 9.3.1 we check the L^2 -convergence of the partial sum $\sum_{i=0}^{N-1} W_{t_i} dW_{t_i}$ to the limit $\frac{1}{2}W_T^2 - \frac{1}{2}T$. In our simulation we compute the square root of $\mathbb{E} \left[\left| \sum_{i=0}^{N-1} W_{t_i} \delta W_{t_i} - \left(\frac{1}{2}W_T^2 - \frac{1}{2}T \right) \right|^2 \right]$. The average is computed using $M = 1000$ samples.

```
T = 3;
M = 10^3; % number of Brownian sample paths
K = 15;
L2_norm = zeros(K,1); % L2-norm
for k = 1:K
    N = 2^k; % [0,T] is partitioned into N subintervals.
    dt = T/N;
    W = zeros(M,N+1);
    dW = sqrt(dt)*randn(M,N);
    W(1:M,2:N+1) = cumsum(dW,2);
    Integrand = W(1:M,1:N).* dW;
    Ito_Integral = cumsum(Integrand(1:M,:),2);
    Exact = 0.5*(W(:,N+1).^2-T*ones(M,1));
    differ = Ito_Integral(:,N) - Exact;
    L2_norm(k,1) = sqrt(mean(differ.^2));
end

x = 0:0.1:K;
plot(x, 2.^(-(x+1)/2)*T,'k-');
hold on;
plot(L2_norm,'or-');
```

See Fig. 9.5.

Simulation 9.4 (Normal Distribution of Itô Integral) First, we plot the normalized histograms of X and Y in Sect. 9.3.2 to check whether they are normally distributed with variance $\frac{T^3}{3}$. We choose $T = 3$ and take $M = 10^5$ Brownian sample paths. The time interval $[0, T]$ is equally partitioned into $N = 1000$ subintervals.

```
T = 3;
M = 10^5; % number of Brownian sample paths
N = 1000; % [0,T] is partitioned into N subintervals.
dt = T/N; t = zeros(M,N+1);
for k = 1:N+1
    t(:,k) = (k-1)*dt*ones(M,1);
end
dW = sqrt(dt)*randn(M,N);
Integrand1 = t(:,1:N).* dW;
Integrand2 = (T-t(:,1:N)).* dW;
Ito_Integ1 = cumsum(Integrand1(1:M,:),2);
```

```

Ito_Integ2 = cumsum(Integrand2(1:M,:),2);
Variance = T^3/3

figure(1)
histogram(Ito_Integ1(:,N),50,'Normalization','pdf','FaceColor','w');
hold on
x = -12:0.1:12;
plot(x,1/sqrt(2*pi*Variance)*exp(-x.^2/(2*Variance)),'r-');

figure(2)
histogram(Ito_Integ2(:,N),50,'Normalization','pdf','FaceColor','w');

```

See Fig. 9.6. Now we compute the covariance and correlation of X and Y .

```

Variance_Exact = T^3/3 % Variance_Exact
Covariance_Exact = T^3/6 % Covariance_Exact
Correlation_Exact = 0.5 % Correlation_Exact
Var1 = var(Ito_Integ1(:,N)) % variance
Var2 = var(Ito_Integ2(:,N)) % variance
Covariance = cov(Ito_Integ1(:,N),Ito_Integ2(:,N))
Correlation = corr(Ito_Integ1(:,N),Ito_Integ2(:,N))

```

Here is the output.

```

Variance_Exact =
           9
Covariance_Exact =
      4.5000
Correlation_Exact =
      0.5000
Var1 =
      8.9956
Var2 =
      9.0760
Covariance =
      8.9956      4.5232
      4.5232      9.0760
Correlation =
      0.5006

```

Note that the numerically obtained covariances and correlation are close to the exact values. In the following we scatter plot the sample values of the pair (X, Y) .

```

figure(1)
scatter(Ito_Integ1(:,N),Ito_Integ2(:,N),'k.')
xlabel('X');
ylabel('Y');

```

See Fig. 9.7.

Simulation 9.5 (Expectation of Itô Integral is Zero) We check the validity of Corollary 9.2 for the stochastic process $f(t, W_t) = \exp(W_t)$, $0 \leq t \leq T$.

```
clear all; clc; clf;
T = 1; % indexed by i, t_0, t_1, ..., t_i, ..., t_N
M = 5; % number of Brownian sample paths
N = 10; % [0,T] is partitioned into N subintervals.
dt = T/N
sqrt(dt)
dW = sqrt(dt)*randn(M,N);
W = [zeros(M,1), cumsum(dW,2)];
for k = 1:N+1
    t(:,k) = (k-1)*dt*ones(M,1);
end

f = exp(W);
f = [f(:,1:N), f(:,N)];
fdW = f(:,1:N).*dW;
Ito_Integ = [zeros(M,1), cumsum(fdW(1:M,:),2)];
for k = 1:N+1
    Expectation(1,k) = mean(Ito_Integ(1:M,k));
end

figure(1) % dW
for i = 1:M
    plot(t(i,1:N), dW(i,1:N), '*:', 'Color', hsv2rgb([i/M 1 1]));
    hold on
end
plot(0:dt:T, zeros(1,N+1), 'k');
xlabel('t_i');
ylabel('dW_i');
title('Increments dW_i, i=0,...,N-1');
hold off

figure(2) % W
plot(0:dt:T, zeros(1,N+1), 'k');
hold on
for i = 1:M
    plot(t(i,:), W(i,:), 'o-', 'Color', hsv2rgb([i/M 1 1]));
    hold on
end
ylabel('W_i');
title('Sample paths W_i, i=0,...,N');
```

```

hold off

figure(3) % a stochastic process f(t,W) at t_i
for i = 1:M
X = t(i,1:N+1);
Xq = 0:0.002:1;
V = f(i,1:N+1);
Yq = interp1(X,V,Xq,'previous');
plot(X,V,'o',Xq,Yq,':.','Color',hsv2rgb([i/M 1 1]));
hold on
end
ylabel('f(t_i,W_i)');
title('Sample paths of a step process f(t_i,W_i), i=0,...,N-1');
hold off

figure(4) % f(t,W)dW at t_i
for i = 1:M
plot(t(i,1:N),fdW(i,1:N),'*','Color',hsv2rgb([i/M 1 1]));
hold on
end
ylabel('f(t_i,W_i) dW_i');
title('f(t_i,W_i) dW_i, i=0,...,N-1');
hold off

figure(5) % integral f(t,W)dW
for i = 1:M
plot(t(i,1:N+1),Ito_Integ(i,1:N+1),'o-','Color',hsv2rgb([i/M 1 1]));
hold on
end
ylabel('\Sigma_{k=0}^{i} f_k dW_k');
title('Sample paths of \int_0^t f(t,W) dW');
hold off

figure(6) % Expectation of integral fdW
plot(0:dt:T,zeros(1,N+1),'k');
hold on
plot(0:dt:N*dt,Expectation(1:N+1),'r*');
ylabel('E[\Sigma_{k=0}^{i} f_k dW_k]');
title('E[\int_0^t f(t,W) dW]');
hold off

```

See Fig. 9.8. See also Fig. 10.5.

Exercises

9.1 Without applying Corollary 9.2, show that $\mathbb{E} \left[\int_0^T (W_t + 1)^2 dW_t \right] = 0$.

9.2 Compute $\int_0^T e^{W_t} dW_t$.

9.3 Show that $\int_0^t s dW_s = t W_t - \int_0^t W_s ds$ from the definition of the Itô integral.

9.4 Show that $\int_0^t W_s^2 dW_s = \frac{1}{3} W_t^3 - \int_0^t W_s ds$ from the definition of the Itô integral. (Hint: Use the identity $b^2(a - b) = \frac{1}{3}(b^3 - a^3) - \frac{1}{3}(b - a)^3 - b(b^2 - a^2)$.)

9.5 Show that $X_t = t^2 W_t - 2 \int_0^t s W_s ds$ is a martingale.

9.6 For a continuous deterministic function $f(t)$ define $X_t = \int_0^t f(s) dW_s$. Show that $\text{Cov}(X_t, X_{t+u}) = \int_0^t f(s)^2 ds$, $u \geq 0$.

9.7 Let $X_t = e^{W_t^2}$, $t \geq 0$. Show that $\mathbb{E}[X_t^2] = \frac{1}{\sqrt{1-4t}}$, $0 \leq t < \frac{1}{4}$, and hence $X_t \notin \mathcal{H}^2$. (Thus the identity in Theorem 9.1 does not hold.)

9.8 Suppose that $X_t = f(t, W_t)$ is a martingale for some sufficiently smooth $f(t, x)$. Show that f satisfies the differential equation $f_t + \frac{1}{2} f_{xx} = 0$.

9.9 Does there exist $X_t = f(t, W_t)$ such that $dX_t = X_t dW_t$?

9.10 Let $dX_t = a_t dt + dW_t$ where a_t is a bounded process. Show that X_t is not a martingale unless $a_t = 0$.

9.11 Fix $a > 0$. Define $f(t, x) = \mathbb{1}_{[0, a]}(t)$ for $(t, x) \in [0, \infty) \times (-\infty, \infty)$.

(i) Compute $\int_0^t f(s, W_s) dW_s$.

(ii) Prove that the stochastic process X_t defined by $X_t = W_t$, $0 \leq t \leq a$, and $X_t = W_a$, $t > a$, is a martingale.

(iii) Find α_t such that $dX_t = \alpha_t dW_t$.

9.12 Show that the Martingale Representation Theorem holds true for a martingale $W_t^3 - 3t W_t$. (See Theorem 7.8 (v).)

9.13 Find a deterministic function $g(t)$ for which $e^{W_t + g(t)}$, $t \geq 0$, is a martingale.

9.14 Show that $X_t = (W_t + \theta t)e^{-\theta W_t - \frac{1}{2}\theta^2 t}$ is a martingale.

9.15 Let $dX_t = a_t dt + dW_t$ where a_t is a bounded process. Define

$$L_t = \exp \left(- \int_0^t a_s dW_s - \frac{1}{2} \int_0^t a_s^2 ds \right).$$

Show that $X_t L_t$ is a martingale.

9.16 Let $X_t = t^2 W_t - 2 \int_0^t s W_s ds$. Show that X_t is a martingale.

9.17 Find a constant a for which $X_t = e^{at} \cos W_t$ is a martingale.

9.18 Show that $d(W_t^6) = 15W_t^4 dt + 6W_t^5 dW_t$. Verify that $\mathbb{E}[W_t^6] = 15t^3$.

Chapter 10

The Itô Formula



In this chapter we present the Itô formula, or the Itô lemma, which is the most frequently used fundamental fact in stochastic calculus. It approximates a function of both time and Brownian motion W_t in a style similar to the Taylor series expansion of a function of two variables except that the closeness of approximation is measured in terms of the probabilistic distribution of the increment of Brownian motion δW . Since $\mathbb{E}[\delta W] = 0$ and $\mathbb{E}[(\delta W)^2] = \delta t$, we may write $|\delta W| \approx \sqrt{\delta t}$. Thus, even when $\delta t \approx 0$ we cannot ignore $(\delta W)^2$ completely. Itô's lemma or Itô's formula regards $(\delta W)^2$ as δt in a mathematically rigorous way.

As an application, we let S_t denote the price of a risky asset such as stock. The standard model for asset price movement is given by the geometric Brownian motion. More precisely, if δS denotes the increment of S_t then we have an approximate stochastic differential equation $\delta S_t = \mu S_t \delta t + \sigma S_t \delta W_t$, where δW represents uncertainty or risk in financial investment.

10.1 Motivation for the Itô Formula

The idea underlying the Itô formula is a probabilistic interpretation of the second order Taylor expansion. Given a sufficiently smooth function $f(t, x)$, consider the increment of the curve $t \mapsto f(t, W_t)$ using the second order Taylor expansion:

$$\begin{aligned} \delta f &= f(t + \delta t, W_{t+\delta t}) - f(t, W_t) \\ &= f_t \delta t + f_x \delta W_t + \frac{1}{2} \{ f_{tt} (\delta t)^2 + 2 f_{tx} \delta t \delta W_t + f_{xx} (\delta W_t)^2 \} . \end{aligned}$$

See Fig. 10.1. If $\delta t > 0$ is close to 0, then $(\delta t)^2$ is far smaller than δt . We note that δW_t is normally distributed with mean 0 and variance δt and it can be large, albeit with very small probability, which is troublesome. We adopt the practical rule of

Fig. 10.1 Graph of a function and Itô's lemma

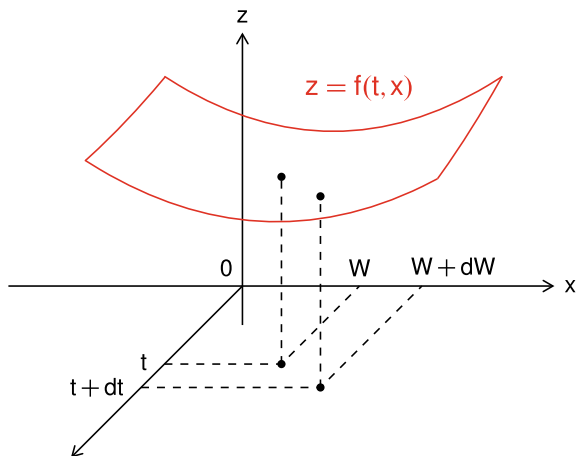


Table 10.1 Multiplication rule for Itô calculus

$\times$	dt	dW
dt	0	0
dW	0	dt

thumb $(\delta t)^2 = 0$, $(\delta W)^2 = \delta t$, and take $\delta W \times \delta t = \pm(\delta t)^{3/2} = 0$, i.e., we ignore the δt -terms of order greater than 1. See Table 10.1, which is written in differential notation.

According to the new viewpoint we have

$$\delta f = \left(f_t + \frac{1}{2}f_{xx}\right)\delta t + f_x \delta W_t.$$

Before we prove the Itô formula in Sect. 10.2, we present a more detailed analysis of the rule $(dW_t)^2 = dt$, which is the essential idea in the proof of the formula. Take a partition $0 = t_0 < t_1 < \dots < t_n = T$ of a given time interval $[0, T]$ and let $\delta t_i = t_{i+1} - t_i$ and $\delta W_i = W_{t_{i+1}} - W_{t_i}$, $0 \leq i \leq n-1$. Suppose we want to express $f(T, W_T) - f(0, W_0)$ as a sum of small increments

$$(\delta f)_i = f(t_{i+1}, W_{t_{i+1}}) - f(t_i, W_{t_i}),$$

i.e., we consider

$$f(T, W_T) - f(0, W_0) = \sum_{i=1}^{n-1} (\delta f)_i.$$

Since δW_i can be considerably large compared to δt_i , we ignore the terms corresponding to $(\delta t_i)^2$ and $(\delta t_i)(\delta W_i)$ in the second order Taylor expansion, then the

approximation of $(\delta f)_i$ by the second order Taylor expansion is given by

$$(\delta f)_i^{\text{Taylor}} = f_t \delta t_i + f_x \delta W_i + \frac{1}{2} f_{xx} (\delta W_i)^2,$$

where the partial derivatives are evaluated at the point (t_i, W_{t_i}) . It is known to be a good approximation for sufficiently small δt and δW . Compare it with the approximation by the Itô formula

$$(\delta f)_i^{\text{Itô}} = (f_t + \frac{1}{2} f_{xx}) \delta t_i + f_x \delta W_i.$$

Observe that the essential difference between two approximations is given by

$$\varepsilon_i = \frac{1}{2} f_{xx} (\delta W_i)^2 - \frac{1}{2} f_{xx} \delta t_i = \frac{1}{2} f_{xx} ((\delta W_i)^2 - \delta t_i).$$

If this is small enough in some sense, then we can say that the approximation by the Itô formula is as good as Taylor approximation.

As a test we choose $f(t, x) = x^2$ and check the approximation errors for $\delta t \approx 0$ and $\delta W \approx 0$ resulting from two methods. We choose $i = 0$, and $t_i = 0$, $W_{t_i} = 0$, and plot the error of the second order Taylor approximation

$$f(t_i + \delta t_i, W_{t_i} + \delta W_i) - f(t_i, W_{t_i}) - (\delta f)_i^{\text{Taylor}}$$

and the error of the Itô approximation

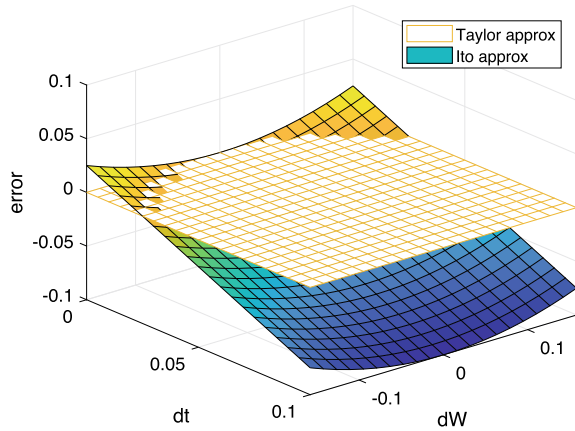
$$f(t_i + \delta t_i, W_{t_i} + \delta W_i) - f(t_i, W_{t_i}) - (\delta f)_i^{\text{Itô}}$$

over the region $0 \leq \delta t_i \leq T$, $-1.5\sqrt{T} \leq \delta W_i \leq 1.5\sqrt{T}$, $T = 0.01$. See Fig. 10.2 where the approximation by the Itô formula is better and the corresponding error is close to zero. The second order Taylor expansion has negligible error so that the graph is almost flat with height 0 while the Itô formula does not approximate the given function well enough. Therefore we need to give a new interpretation for Itô approximation, which is nothing but taking the average error in a suitable sense over all possible values of δW_i so that the average of ε_i would be close to 0 since δW_i is normally distributed with average 0 and variance δt_i .

In the following discussion we will compute the L^2 -norm of ε_i to estimate its closeness to 0, and estimate the total error $\varepsilon_0 + \dots + \varepsilon_{n-1}$.

For $i < j$, we have

Fig. 10.2 Approximation errors of $f(t, x) = x^2$ by the second order Taylor expansion and the Itô formula



$$\begin{aligned}
 \mathbb{E}[\varepsilon_i \varepsilon_j] &= \mathbb{E} \left[\frac{1}{4} f_{xx}(t_i, W_{t_i}) \{\delta t_i - (\delta W_i)^2\} f_{xx}(t_j, W_{t_j}) \{\delta t_j - (\delta W_j)^2\} \right] \\
 &= \mathbb{E} \left[\frac{1}{4} f_{xx}(t_i, W_{t_i}) \{\delta t_i - (\delta W_i)^2\} f_{xx}(t_j, W_{t_j}) \right] \mathbb{E} [\delta t_j - (\delta W_j)^2] \\
 &= 0
 \end{aligned}$$

by independence of

$$f_{xx}(t_i, W_{t_i}) \{\delta t_i - (\delta W_i)^2\} f_{xx}(t_j, W_{t_j})$$

and $\delta t_j - (\delta W_j)^2$, and by the fact that $\mathbb{E}[\delta t_j - (\delta W_j)^2] = 0$.

For $i = j$ we have

$$\begin{aligned}
 \mathbb{E}[\varepsilon_i^2] &= \mathbb{E} \left[\frac{1}{4} (f_{xx}(t_i, W_{t_i}))^2 \{\delta t_i - (\delta W_i)^2\}^2 \right] \\
 &\leq \frac{C^2}{4} \mathbb{E} [(\delta t_i)^2 - 2 \delta t_i (\delta W_i)^2 + (\delta W_i)^4] \\
 &= \frac{C^2}{4} \{(\delta t_i)^2 - 2(\delta t_i)^2 + 3(\delta t_i)^2\} = \frac{C^2}{2} (\delta t_i)^2
 \end{aligned}$$

where

$$C = \max\{|f_{xx}(t, x)| : t \geq 0, -\infty < x < \infty\} < \infty. \quad (10.1)$$

Hence

$$\mathbb{E}[(\varepsilon_0 + \cdots + \varepsilon_{n-1})^2] = \mathbb{E}[\varepsilon_0^2] + \cdots + \mathbb{E}[\varepsilon_{n-1}^2] \leq n \times \frac{C^2}{2} \left(\frac{T}{n}\right)^2 = \frac{C^2 T}{2} \delta t$$

if we take $\delta t = \frac{T}{n}$, $t_i = i \times \delta t$, $0 \leq i \leq n$. Thus the L^2 -norm of the total error satisfies

$$(\mathbb{E}[(\varepsilon_0 + \cdots + \varepsilon_{n-1})^2])^{\frac{1}{2}} \leq C \left(\frac{T}{2}\right)^{\frac{1}{2}} (\delta t)^{\frac{1}{2}}$$

which gives a bound for the speed of convergence to 0 of the total error in approximating $f(T, W_T) - f(0, 0)$ as $\delta t \rightarrow 0$.

Example 10.1 Take $f(t, x) = x^2$ as a test function. Since $f_t = 0$, $f_x = 2x$ and $f_{tt} = 0$, $f_{xx} = 2$, $f_{tx} = 0$, we have

$$\varepsilon_i = (\delta W_i)^2 - \delta t_i,$$

and $C = 2$ where C is defined by (10.1). Thus

$$\varepsilon_0 + \cdots + \varepsilon_{n-1} = \sum_{i=0}^{n-1} ((\delta W_i)^2 - \delta t_i) = \sum_{i=0}^{n-1} (\delta W_i)^2 - T$$

and

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{i=0}^{n-1} (\delta W_i)^2 - T \right)^2 \right] \\ &= \mathbb{E} \left[\left(\sum_{i=0}^{n-1} (\delta W_i)^2 \right)^2 - 2T \sum_{i=0}^{n-1} (\delta W_i)^2 + T^2 \right] \\ &= \mathbb{E} \left[\sum_{i=0}^{n-1} (\delta W_i)^4 + \sum_{i \neq j} (\delta W_i)^2 (\delta W_j)^2 - 2T \sum_{i=0}^{n-1} (\delta W_i)^2 + T^2 \right] \\ &= \sum_{i=0}^{n-1} 3(\delta t_i)^2 + \sum_{i \neq j} (\delta t_i)(\delta t_j) - 2T \sum_{i=0}^{n-1} \delta t_i + T^2 \\ &= \sum_{i=0}^{n-1} 2(\delta t_i)^2 + \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (\delta t_i)(\delta t_j) - 2T \sum_{i=0}^{n-1} \delta t_i + T^2 \\ &= \sum_{i=0}^{n-1} 2(\delta t_i)^2 + \sum_{i=0}^{n-1} \delta t_i \sum_{j=0}^{n-1} \delta t_j - 2T \sum_{i=0}^{n-1} \delta t_i + T^2 \\ &= \sum_{i=0}^{n-1} 2(\delta t_i)^2, \end{aligned} \tag{10.2}$$

where we used $\sum_{i=0}^{n-1} \delta t_i = T$ for the last equality. We choose $T = 3$, $n = 2^k$, $\delta t = \frac{T}{n} = 2^{-k}T$, $1 \leq k \leq 15$, and compute the L^2 -norm of $\sum_{i=0}^{n-1} (\delta W_i)^2 - T$, theoretically equal $(2n(\delta t)^2)^{1/2} = 2^{-(k-1)/2}T$ by (10.2). In the left plot in Fig. 10.3 we plot both numerically and theoretically obtained L^2 -norms as functions of k .

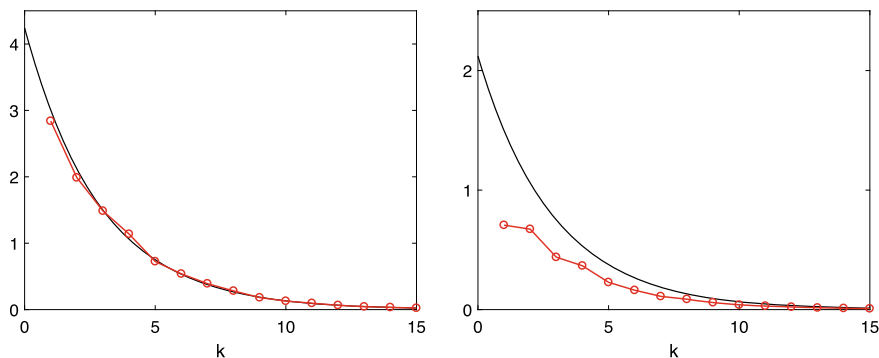


Fig. 10.3 Speed of L^2 -convergence in Itô formula for $f(t, x) = x^2$ (left) and $f(t, x) = \sin x$ (right)

Example 10.2 Take $f(t, x) = \sin x$ as a test function. Since $f_t = 0$, $f_x = \cos x$ and $f_{tt} = 0$, $f_{xx} = -\sin x$, $f_{tx} = 0$, we have

$$\varepsilon_i = \frac{1}{2} (\sin W_i) (\delta t_i - (\delta W_i)^2) ,$$

and $C = 1$. Thus $\mathbb{E}[(\varepsilon_1 + \cdots + \varepsilon_n)^2] \leq \frac{T}{2} \delta t$. In the right plot in Fig. 10.3 we choose $T = 3$, $n = 2^k$, $\delta t = 2^{-k} T$, $1 \leq k \leq 15$, and compute the L^2 -norm of $\varepsilon_1 + \cdots + \varepsilon_n$ which is bounded by $(\frac{T}{2} \delta t)^{1/2} = 2^{-(k+1)/2} T$. Consult Simulation 10.1.

10.2 The Itô Formula: Basic Form

The following fact is called the Itô formula or the Itô lemma.

Theorem 10.1 (Itô) *Let $f(t, x) : [0, \infty) \times \mathbb{R}^1 \rightarrow \mathbb{R}^1$ be a sufficiently smooth function such that its partial derivatives f_t , f_x , f_{xx} exist and are continuous. Fix a constant $0 < T < \infty$. Then almost surely the following equation holds:*

$$\begin{aligned} & f(T, W_T) - f(0, W_0) \\ &= \int_0^T f_t(s, W_s) ds + \int_0^T f_x(s, W_s) dW_s + \frac{1}{2} \int_0^T f_{xx}(s, W_s) ds , \end{aligned}$$

where the first and the third integrals are Riemann integrals for each fixed sample Brownian path.

Before we present a proof, it is necessary to give some comments on this very important theorem.

Remark 10.1 (i) The formula in the theorem is written in integral notation, and it can also be written in differential notation as follows:

$$df(t, W_t) = f_t(t, W_t) dt + f_x(t, W_t) dW_t + \frac{1}{2} f_{xx}(t, W_t) dt$$

or simply,

$$df = \left(f_t + \frac{1}{2} f_{xx} \right) dt + f_x dW_t .$$

(ii) For $f(t, x) = x^2$ in Example 10.1 a straightforward computation yields

$$\begin{aligned} f(t_{i+1}, W_{t_{i+1}}) - f(t_i, W_{t_i}) &= (W_{t_i} + \delta W_i)^2 - W_{t_i}^2 \\ &= 2W_{t_i} \delta W_i + (\delta W_i)^2 \end{aligned}$$

and

$$\begin{aligned} W_T^2 - W_0^2 &= f(T, W_T) - f(0, W_0) \\ &= \sum_{i=0}^{n-1} 2W_{t_i} \delta W_i + \sum_{i=0}^{n-1} (\delta W_i)^2 , \end{aligned}$$

where the first sum converges to

$$\int_0^T 2W_t dW_t$$

as $n \rightarrow \infty$ in the L^2 -sense and the second sum converges to T . On the other hand, since

$$(\delta f)_i^{\text{Itô}} = 2W_{t_i} \delta W_i + \delta t_i ,$$

we have

$$f(T, W_T) - f(0, W_0) \approx \sum_{i=0}^{n-1} (\delta f)_i^{\text{Itô}} = \sum_{i=0}^{n-1} 2W_{t_i} \delta W_i + T$$

in agreement with the fact that

$$\sum_{i=0}^{n-1} (\delta W_i)^2$$

converges to T in the L^2 -sense.

Proof The essential idea of the proof is already presented in Sect. 10.1. Now we will proceed with mathematical rigour. First, consider the case when there exists a constant $C > 0$ such that for every $(t, x) \in [0, \infty) \times \mathbb{R}^1$ we have

$$|f(t, x)| \leq C, \quad |f_t(t, x)| \leq C, \quad |f_x(t, x)| \leq C, \quad |f_{xx}(t, x)| \leq C.$$

The unbounded case will be considered at the end of the proof. Let $0 = t_0 < t_1 < \dots < t_n = T$ be a partition of $[0, T]$, and put $\delta W_i = W_{t_{i+1}} - W_{t_i}$, $\delta t_i = t_{i+1} - t_i$. We assume that $\max_{1 \leq i \leq n} \delta t_i \rightarrow 0$ as $n \rightarrow \infty$.

On each subinterval $[t_i, t_{i+1}]$ by Taylor's theorem there exist $t_i^* \in [t_i, t_{i+1}]$ and $W_i^* \in [W_{t_i}(\omega), W_{t_{i+1}}(\omega)]$ such that the following holds:

$$\begin{aligned} & f(T, W_T(\omega)) - f(0, W_0(\omega)) \\ &= \sum_{i=0}^{n-1} (f(t_{i+1}, W_{t_{i+1}}(\omega)) - f(t_i, W_{t_i}(\omega))) \\ &= \sum_{i=0}^{n-1} (f(t_{i+1}, W_{t_{i+1}}) - f(t_i, W_{t_{i+1}})) + \sum_{i=0}^{n-1} (f(t_i, W_{t_{i+1}}) - f(t_i, W_{t_i})) \\ &= \sum_{i=0}^{n-1} f_t(t_i^*, W_{t_{i+1}}) \delta t_i + \sum_{i=0}^{n-1} f_x(t_i, W_i^*) \delta W_i + \frac{1}{2} \sum_{i=0}^{n-1} f_{xx}(t_i, W_i^*) (\delta W_i)^2 \\ &= \sum_{i=0}^{n-1} f_t(t_i^*, W_{t_{i+1}}) \delta t_i + \sum_{i=0}^{n-1} f_x(t_i, W_i^*) \delta W_i + \frac{1}{2} \sum_{i=0}^{n-1} f_{xx}(t_i, W_{t_i}) \delta t_i \\ &\quad + \frac{1}{2} \sum_{i=0}^{n-1} f_{xx}(t_i, W_{t_i}) ((\delta W_i)^2 - \delta t_i) + \frac{1}{2} \sum_{i=0}^{n-1} [f_{xx}(t_i, W_i^*) - f_{xx}(t_i, W_{t_i})] (\delta W_i)^2 \\ &= \Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4 + \Sigma_5, \end{aligned}$$

where the symbol ω is omitted for the sake of notational simplicity if there is no danger of confusion. Note that if $W_{t_{i+1}}(\omega) < W_{t_i}(\omega)$, then $W_i^* \in [W_{t_{i+1}}(\omega), W_{t_i}(\omega)]$.

First, we consider Σ_1 and Σ_3 , which are Riemann integrals of continuous functions for almost every ω . By the definition of the Riemann integral their limits for almost every ω are given by

$$\begin{aligned} \lim_{n \rightarrow \infty} \Sigma_1 &= \int_0^T f_t(t, W_t(\omega)) dt, \\ \lim_{n \rightarrow \infty} \Sigma_3 &= \int_0^T f_{xx}(t, W_t(\omega)) dt. \end{aligned}$$

For $\Sigma_2, \Sigma_4, \Sigma_5$ we first prove the L^2 -convergence of the limits

$$\begin{aligned}\lim_{n \rightarrow \infty} \Sigma_2 &= \int_0^T f_x(t, W_t) dW_t, \\ \lim_{n \rightarrow \infty} \Sigma_4 &= 0, \\ \lim_{n \rightarrow \infty} \Sigma_5 &= 0,\end{aligned}$$

and show that there exist subsequences which converge for almost every ω .

As for Σ_2 we use the boundedness condition $|f_x(t, x)| \leq C$. Now, if we let

$$g(t, \omega) = f_x(t, W_t(\omega)) \in \mathcal{H}^2,$$

and define a simple stochastic process by

$$g_n(t, \omega) = \sum_{i=0}^{n-1} f_x(t_i, W_{t_i}(\omega)) \mathbb{1}_{[t_i, t_{i+1})}(t) \in \mathcal{H}_{\text{simple}}^2,$$

then $g_n(t, \omega)$ approximates g . Note that for every $0 \leq t \leq T$

$$\lim_{n \rightarrow \infty} |g_n(t, \omega) - g(t, \omega)|^2 = 0$$

for almost every ω by continuity.

Since

$$|g_n(t) - g(t)|^2 \leq 4C^2,$$

by the Lebesgue Dominated Convergence Theorem we have

$$\lim_{n \rightarrow \infty} \int_0^T |g_n(t, \omega) - g(t, \omega)|^2 dt = 0$$

for almost every ω . Furthermore, since

$$\int_0^T |g_n(t, \omega) - g(t, \omega)|^2 dt \leq 4TC^2,$$

we apply the Lebesgue Dominated Convergence Theorem once more to obtain

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\int_0^T |g_n(t) - g(t)|^2 dt \right] = 0,$$

where $\int_0^T |g_n(t, \omega) - g(t, \omega)|^2 dt, n \geq 1$, is regarded as a bounded sequence of functions of ω with an upper bound for all n given by $4TC^2$, and $\mathbb{E}$ denotes the Lebesgue integral on a probability space $(\Omega, \mathbb{P})$. Hence g_n converges to g in the L^2 -sense in M_T^2 , and hence

$$I(g_n) = \int_0^T g_n dW = \sum_{i=0}^{n-1} f_x(t_i, W_{t_i}) \delta W_i$$

converges to

$$I(g) = \int_0^T f_x(t, W_t) dW_t ,$$

where $I(\cdot)$ denotes the Itô integral of a function inside the parentheses.

Now we consider Σ_4 . Recall that, for $i < j$, a function depending on $W_{t_i}, \delta W_i, W_{t_j}$ and another function depending on δW_j are independent. In other words, for any measurable functions $\phi_1(\cdot, \cdot, \cdot)$ and $\phi_2(\cdot)$ we have

$$\mathbb{E} [\phi_1(W_{t_i}, \delta W_i, W_{t_j}) \phi_2(\delta W_j)] = \mathbb{E} [\phi_1(W_{t_i}, \delta W_i, W_{t_j})] \mathbb{E} [\phi_2(\delta W_j)] .$$

Now since f_{xx} is bounded by C we have

$$\begin{aligned} & \mathbb{E} \left[\left| \sum_{i=0}^{n-1} f_{xx}(t_i, W_{t_i}) ((\delta W_i)^2 - \delta t_i) \right|^2 \right] \\ &= \mathbb{E} \left[\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} f_{xx}(t_i, W_{t_i}) ((\delta W_i)^2 - \delta t_i) f_{xx}(t_j, W_{t_j}) ((\delta W_j)^2 - \delta t_j) \right] \\ &= \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \mathbb{E} [f_{xx}(t_i, W_{t_i}) ((\delta W_i)^2 - \delta t_i) f_{xx}(t_j, W_{t_j}) ((\delta W_j)^2 - \delta t_j)] \\ &= \sum_{i < j} \mathbb{E} [f_{xx}(t_i, W_{t_i}) ((\delta W_i)^2 - \delta t_i) f_{xx}(t_j, W_{t_j})] \mathbb{E} [((\delta W_j)^2 - \delta t_j)] \\ &\quad + \sum_{j < i} \mathbb{E} [((\delta W_i)^2 - \delta t_i)] \mathbb{E} [f_{xx}(t_i, W_{t_i}) f_{xx}(t_j, W_{t_j}) ((\delta W_j)^2 - \delta t_j)] \\ &\quad + \sum_{i=0}^{n-1} \mathbb{E} [f_{xx}(t_i, W_{t_i})^2 |(\delta W_i)^2 - \delta t_i|^2] \\ &= 0 + 0 + \sum_{i=0}^{n-1} \mathbb{E} [f_{xx}(t_i, W_{t_i})^2 |(\delta W_i)^2 - \delta t_i|^2] \\ &= \sum_{i=0}^{n-1} \mathbb{E} [f_{xx}(t_i, W_{t_i})^2] \mathbb{E} [|(\delta W_i)^2 - \delta t_i|^2] \\ &\leq C^2 \sum_{i=0}^{n-1} \mathbb{E} [|(\delta W_i)^2 - \delta t_i|^2] \\ &= C^2 \sum_{i=0}^{n-1} (\mathbb{E} [(\delta W_i)^4] - 2 \mathbb{E}[(\delta W_i)^2] \delta t_i + (\delta t_i)^2) \end{aligned}$$

$$\begin{aligned}
&= C^2 \sum_{i=0}^{n-1} (3 - 2 + 1)(\delta t_i)^2 \\
&\leq 2C^2 \max_{1 \leq i \leq n} \delta t_i \sum_{i=0}^{n-1} \delta t_i \\
&= 2C^2 T \max_{1 \leq i \leq n} \delta t_i \rightarrow 0.
\end{aligned}$$

The third equality holds since, for $i < j$, by independence we have

$$\begin{aligned}
&\mathbb{E} [f_{xx}(t_i, W_{t_i})((\delta W_i)^2 - \delta t_i) f_{xx}(t_j, W_{t_j})((\delta W_j)^2 - \delta t_j)] \\
&= \mathbb{E} [f_{xx}(t_i, W_{t_i})((\delta W_i)^2 - \delta t_i) f_{xx}(t_j, W_{t_j})] \mathbb{E} [(\delta W_j)^2 - \delta t_j]
\end{aligned}$$

and the fourth equality by $\mathbb{E}[(\delta W_j)^2 - \delta t_j] = 0$. The fifth equality holds since $|f_{xx}(t_i, W_{t_i})|^2$ and $|(\delta W_i)^2 - \delta t_i|^2$ are independent. Hence

$$\sum_{i=0}^{n-1} f_{xx}(t_i, W_{t_i}) ((\delta W_i)^2 - \delta t_i) \rightarrow 0$$

in the L^2 -sense.

As for Σ_5 we have

$$\lim_{n \rightarrow \infty} \max_i |f_{xx}(t_i, W_i^*) - f_{xx}(t_i, W_{t_i})| = 0$$

for almost every ω by continuity. Since $\sum_{i=0}^{n-1} (\delta W_i)^2 \rightarrow T$ in the L^2 -sense, by Fact 3.8 there exists a subsequence $\{n_k\}_{k=1}^\infty$ such that

$$\lim_{n_k \rightarrow \infty} \sum_{i=0}^{n_k-1} (\delta W_i(\omega))^2 = T$$

for almost every ω . Hence for $\{n_k\}_{k=1}^\infty$ for almost every ω we have

$$\begin{aligned}
&\left| \sum_{i=0}^{n_k} (f_{xx}(t_i, W_i^*) - f_{xx}(t_i, W_{t_i}(\omega))) (\delta W_i(\omega))^2 \right| \\
&\leq \max_i |f_{xx}(t_i, W_i^*) - f_{xx}(t_i, W_{t_i}(\omega))| \sum_{i=0}^{n_k-1} (\delta W_i(\omega))^2 \rightarrow 0
\end{aligned}$$

as $k \rightarrow \infty$. Since Σ_2 converges in the L^2 -sense along $\{n_k\}$, we can show the convergence for almost every ω by using a subsequence of $\{n_k\}$.

Now, we show that the same conclusion holds without the condition on boundedness of partial derivatives. Define

$$f_n(t, x) = \mathbb{I}_{[-n, n]}(x) f(t, x) ,$$

where $\mathbb{I}_{[-n, n]}$ is the indicator of the interval $[-n, n]$. Since f_n and its derivatives are bounded, we may apply the previously proved result and obtain

$$\begin{aligned} & f_n(T, W_T) - f_n(0, W_0) \\ &= \int_0^T (f_n)_t(s, W_s) ds + \int_0^T (f_n)_x(s, W_s) dW_s + \frac{1}{2} \int_0^T (f_n)_{xx}(s, W_s) ds . \end{aligned}$$

Now define a monotonically increasing sequence of subsets

$$B_n = \{\omega : \sup_{0 \leq t \leq T} |W_t(\omega)| \leq n\} .$$

Since $f = f_n$ on B_n , the Itô formula holds on B_n for f . Since $\mathbb{P}(B_n) \rightarrow 1$, the Itô formula for f holds for almost every ω . $\square$

In the following example we obtain stochastic differential equations.

- Example 10.3** (i) For $X_t = W_t$, take $f(t, x) = x$. Since $f_x = 1$ and $f_{xx} = 0$, we have $dX_t = dW_t$, and $\int_0^T dW_t = W_T$ in integral form.
(ii) For $X_t = (W_t)^2$, take $f(t, x) = x^2$. Since $f_x = 2x$ and $f_{xx} = 2$, we have $dX_t = 2W_t dW_t + \frac{1}{2} \times 2dt = 2W_t dW_t + dt$, and

$$\int_0^T W_t dW_t = \frac{1}{2}(W_T)^2 - \frac{1}{2}T$$

in integral form. (See also Remark 10.1.)

- (iii) For $X_t = \frac{1}{3}(W_t)^3$, take $f(t, x) = \frac{1}{3}x^3$. Since $f_x = x^2$ and $f_{xx} = 2x$, we have $dX_t = (W_t)^2 dW_t + \frac{1}{2} \times 2W_t dt$ and

$$\int_0^T (W_t)^2 dW_t = \frac{1}{3}(W_T)^3 - \int_0^T W_t dt .$$

- (iv) For $X_t = tW_t$, take $f(t, x) = tx$. Since $f_t = x$, $f_x = t$ and $f_{xx} = 0$, we have $dX_t = W_t dt + t dW_t$ and

$$\int_0^T t dW_t = TW_T - \int_0^T W_t dt .$$

So far, we have considered examples when stochastic processes X_t are given and we derive the stochastic differential equations satisfied by X_t . We can evaluate a given Itô integral directly as in Exercises 10.5, 10.6, 10.7, 10.9.

10.3 The Itô Formula: General Form

Now we define a stochastic integral with respect to a general stochastic process other than a Brownian motion W_t . Let $\{\mathcal{F}_t\}_{t \geq 0}$ be the filtration associated with W_t .

Definition 10.1 (*Itô Process*) A stochastic process $\{X_t\}_{t \geq 0}$ of the form

$$X_t = X_0 + \int_0^t a_u \, du + \int_0^t b_u \, dW_u$$

is called an *Itô process* if a_t and b_t are stochastic processes adapted to $\{\mathcal{F}_t\}_{t \geq 0}$ such that

$$\int_0^T |a_t(\omega)| \, dt < \infty \quad \text{for almost every } \omega$$

and $\{b_t\}_{t \geq 0} \in \mathcal{H}_T^2$ for every T . We frequently write, for short,

$$dX_t = a_t \, dt + b_t \, dW_t.$$

Example 10.4 (i) The simplest Itô process is given by

$$X_t = X_0 + \int_0^t a_s \, ds$$

where $b_t = 0$.

(ii) For $X_t = W_t$ we have $dX_t = dW_t$, and hence $a_t = 0$, $b_t = 1$.

(iii) For $X_t = (W_t)^2$ we have $dX_t = dt + 2W_t dW_t$, and hence $a_t = 1$, $b_t = 2W_t$.

(iv) For $X_t = \frac{1}{3}(W_t)^3$ we have $dX_t = W_t dt + (W_t)^2 dW_t$, and hence $a_t = W_t$, $b_t = (W_t)^2$.

Example 10.5 For $X_t = t W_t$ we have $dX_t = W_t dt + t dW_t$, and hence $a_t = W_t$, $b_t = t$. This formula looks as if it were the usual product rule for differentiation in calculus.

Definition 10.2 (*Itô Integral with Respect to Itô Processes*) Let $\{X_t\}_{t \geq 0}$ be an Itô process given by

$$dX_t = a_t \, dt + b_t \, dW_t.$$

For a process $\{Y_t\}_{t \geq 0}$ adapted to $\{\mathcal{F}_t\}_{t \geq 0}$ define the stochastic integral of Y_t with respect to X_t by

$$\int_0^t Y_u \, dX_u = \int_0^t Y_u a_u \, du + \int_0^t Y_u b_u \, dW_u ,$$

which can be rewritten as

$$Y_t \, dX_t = Y_t a_t \, dt + Y_t b_t \, dW_t .$$

Remark 10.2 (Itô Integral with Respect to a Martingale) If X_t is a martingale then it is represented as an Itô process by Theorem 9.6, and we can define a stochastic integral $\int_0^t Y_u \, dX_u$ using X_t as an integrator.

Let M_t be a continuous square-integrable martingale. Define the mesh size of the partition $0 = t_0 < t_1 < \dots < t_n = t$ by $\delta_n = \max_{0 \leq j \leq n-1} |t_{j+1} - t_j|$. Recall that the *quadratic variation* of M_t , denoted by $[M, M]_t$ or simply $[M]_t$, is defined by

$$[M, M]_t = \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} |M_{t_{j+1}} - M_{t_j}|^2 ,$$

where the limit is in probability as δ_n decreases to 0 as $n \rightarrow \infty$. If N_t is another martingale, then the *quadratic covariation* of M_t and N_t is given by

$$[M, N]_t = \lim_{n \rightarrow \infty} \sum_{j=0}^{n-1} (M_{t_{j+1}} - M_{t_j})(N_{t_{j+1}} - N_{t_j})$$

and satisfies

$$[M, N]_t = \frac{1}{2}([M + N, M + N]_t - [M, M]_t - [N, N]_t) .$$

(See Definition 6.7.)

Now we have the formula for stochastic integration by parts.

Theorem 10.2 (Product Rule) *Let X_t and Y_t be Itô processes. Then*

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X, Y]_t ,$$

or equivalently, in differential notation,

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + d[X, Y]_t .$$

Proof Since

$$\begin{aligned}
 & \sum_{j=0}^{n-1} (X_{t_{j+1}} - X_{t_j})(Y_{t_{j+1}} - Y_{t_j}) \\
 &= \sum_{j=0}^{n-1} (X_{t_{j+1}} Y_{t_{j+1}} - X_{t_j} Y_{t_j}) - \sum_{j=0}^{n-1} X_{t_j} (Y_{t_{j+1}} - Y_{t_j}) - \sum_{j=0}^{n-1} Y_{t_j} (X_{t_{j+1}} - X_{t_j}) \\
 &= X_t Y_t - X_0 Y_0 - \sum_{j=0}^{n-1} X_{t_j} (Y_{t_{j+1}} - Y_{t_j}) - \sum_{j=0}^{n-1} Y_{t_j} (X_{t_{j+1}} - X_{t_j}) .
 \end{aligned}$$

The last two terms in the above converge in probability to Itô integrals $\int_0^t X_s dY_s$ and $\int_0^t Y_s dX_s$. $\square$

Example 10.6 For a Brownian motion W_t we have $[W]_t = t$.

Remark 10.3 For an Itô process X_t given by $dX_t = a(t, X_t) dt + b(t, X_t) dW_t$, consider the martingale

$$M_t = \int_0^t b(s, X_s) dW_s .$$

Then

$$[M, M]_t = \int_0^t b(s, X_s)^2 ds .$$

Now we present the Itô formula for an Itô process $\{X_t\}_{t \geq 0}$. Its proof is similar to the one for Theorem 10.1, and is omitted.

Theorem 10.3 (General Itô Formula) *Consider an Itô process $\{X_t\}_{t \geq 0}$ given by*

$$dX_t = a_t dt + b_t dW_t .$$

Let $f(t, x) : [0, \infty) \times \mathbb{R} \rightarrow \mathbb{R}^1$ be a function whose partial derivatives f_t , f_x and f_{xx} exist and are continuous. Suppose that

$$f_x(t, X_t) b_t \in \mathcal{H}_T^2$$

for every T . Then $f(t, X_t)$ is an Itô process, and with probability 1 we have

Table 10.2 Multiplication rule for general Itô calculus

$\times$	dt	$a dt + b dW$
dt	0	0
$a dt + b dW$	0	$b^2 dt$

$$\begin{aligned}
& f(t, X_t) - f(0, X_0) \\
&= \int_0^t f_t(u, X_u) du + \int_0^t f_x(u, X_u) dX_u + \frac{1}{2} \int_0^t f_{xx}(u, X_u) d[X, X]_u \\
&= \int_0^t \left(f_t(u, X_u) + f_x(u, X_u) a_u + \frac{1}{2} f_{xx}(u, X_u) b_u^2 \right) du \\
&\quad + \int_0^t f_x(u, X_u) b_u dW_u
\end{aligned}$$

for $t \geq 0$.

The above formula can be rewritten in a differential form as

$$df(t, X_t) = \left(f_t + f_x a_t + \frac{1}{2} f_{xx} b_t^2 \right) dt + f_x b_t dW_t .$$

See Table 10.2.

10.4 Examples of the Itô Formula

10.4.1 Geometric Brownian Motion

Consider the geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dW_t ,$$

which is a fundamental model for asset price movement. First, we look for a solution of the form $S_t = f(t, W_t)$. By the Itô formula we have

$$dS_t = \left(f_t(t, W_t) + \frac{1}{2} f_{xx}(t, W_t) \right) dt + f_x(t, W_t) dW_t$$

and

$$\begin{cases} f_t(t, W_t) + \frac{1}{2} f_{xx}(t, W_t) = \mu f(t, W_t) \\ f_x(t, W_t) = \sigma f(t, W_t) . \end{cases}$$

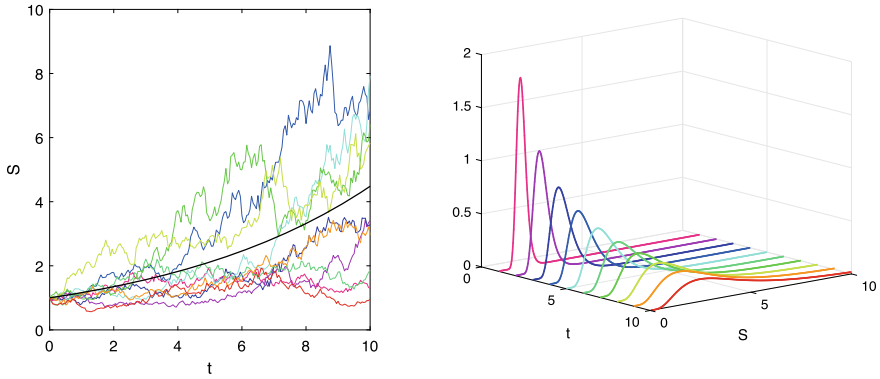


Fig. 10.4 Sample paths of geometric Brownian motion together with the average given by the curve $S = S_0 e^{\mu t}$ (left) and the probability density functions for S_t (right)

Since these equations must hold for an arbitrary real number $W_t(\omega)$ we have

$$\begin{cases} f_t(t, x) + \frac{1}{2} f_{xx}(t, x) = \mu f(t, x) \\ f_x(t, x) = \sigma f(t, x) . \end{cases}$$

From the second equation we see that $f(t, x) = g(t) e^{\sigma x}$ for some function $g(t)$. A necessary condition for g is

$$g'(t) + \frac{1}{2} \sigma^2 g(t) = \mu g(t) .$$

Hence $g(t) = C e^{(\mu - \frac{1}{2} \sigma^2)t}$ for some constant C . Thus $f(t, x) = C e^{(\mu - \frac{1}{2} \sigma^2)t + \sigma x}$ and

$$S_t = S_0 e^{(\mu - \frac{1}{2} \sigma^2)t + \sigma W_t} .$$

From Example 7.2 we have $\mathbb{E}[S_t] = S_0 e^{\mu t}$.

In Fig. 10.4 are given sample paths of S_t , $0 \leq t \leq 10$, with $\mu = 0.15$, $\sigma = 0.2$, $S_0 = 1$ together with the curve $S = S_0 e^{\mu t}$ (left) and the probability density functions for S_t , $t = 1, \dots, 10$ (right).

Remark 10.4 Note that $\{e^{-\mu t} S_t\}_{t \geq 0}$ is a martingale. For, if we put $M_t = e^{-\mu t} S_t = S_0 e^{-\frac{1}{2} \sigma^2 t + \sigma W_t}$, then $M_t = M_0 + \int_0^t \sigma M_u dW_u$, and hence M_t is a martingale. Thus $\mathbb{E}[M_t] = \mathbb{E}[M_0] = S_0$ for every $t \geq 0$, which implies that $\mathbb{E}[S_0 e^{-\frac{1}{2} \sigma^2 t + \sigma W_t}] = S_0$, and $\mathbb{E}[S_t] = S_0 e^{\mu t}$. (See also Example 10.6.)

Remark 10.5 Let Ω be a set of Brownian sample paths ω , and let $\mathbb{P}$ be a Wiener measure on Ω . For each t we may regard S_t as a function in L^p . For $p = 1$ we have

$$\|S_t - 0\|_1 = \int_{\Omega} |S_t(\omega) - 0| d\mathbb{P} = \mathbb{E}[S_t] = S_0 e^{\mu t} \rightarrow \infty ,$$

and for $p = 2$

$$\|S_t - 0\|_2^2 = \int_{\Omega} |S_t(\omega) - 0|^2 d\mathbb{P} = \mathbb{E}[S_t^2] = S_0^2 e^{(2\mu + \sigma^2)t} \rightarrow \infty$$

as $t \rightarrow \infty$. However, with probability 1,

$$\lim_{t \rightarrow \infty} S_t(\omega) = \begin{cases} \infty, & \mu - \frac{1}{2}\sigma^2 > 0 , \\ 0, & \mu - \frac{1}{2}\sigma^2 < 0 . \end{cases}$$

Remark 10.6 Since

$$S_t - S_0 = \int_0^t \mu S_u du + \int_0^t \sigma S_u dW_u , \quad (10.3)$$

Corollary 9.2 implies that

$$\mathbb{E} \left[\int_0^t S_u dW_u \right] = 0 .$$

(See Fig. 10.5 where we take $\mu = 0.1$, $\sigma = 0.3$, $T = 3$, $N = 300$ and $\delta t = \frac{T}{N} = 0.01$. Consult Simulation 9.5.) Hence

$$\mathbb{E}[S_t] - S_0 = \int_0^t \mu \mathbb{E}[S_u] du$$

from (10.3). By differentiating both sides and putting $g(t) = \mathbb{E}[S_t]$ we obtain $g'(t) = \mu g(t)$, whose solution is $g(t) = S_0 e^{\mu t}$. This result coincides with the result in Example 7.2. For $S_0 = 1$ and $\mu = \frac{1}{2}\sigma^2$ we have

$$\mathbb{E} [e^{\sigma W_t}] = e^{\frac{1}{2}\sigma^2 t} .$$

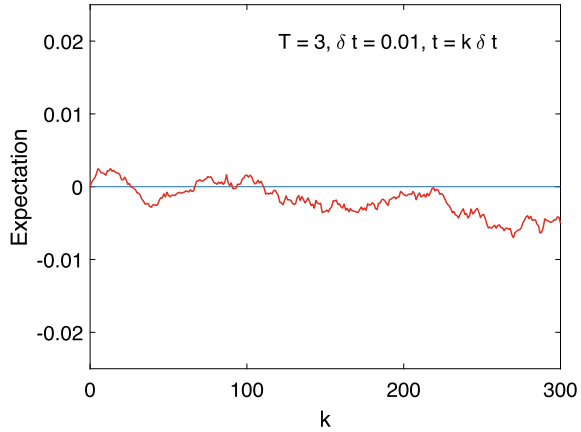
10.4.2 Existence of Singularity

Solve $dX_t = X_t^3 dt - X_t^2 dW_t$, $X_0 = 1$. Assuming $X_t = f(t, W_t)$, we have

$$dX_t = \left(f_t + \frac{1}{2} f_{xx} \right) dt + f_x dW_t ,$$

and hence

Fig. 10.5 For $T = 3$,
 $\mathbb{E}[\int_0^t S_u dW_u] = 0$,
 $0 \leq t \leq T$



$$\begin{cases} f_t + \frac{1}{2} f_{xx} = f^3, \\ f_x = -f^2. \end{cases}$$

From the second equation we have $-\frac{1}{f^2} f_x = 1$, and hence $\frac{\partial}{\partial x} \left(\frac{1}{f} \right) = 1$, and $\frac{1}{f(t,x)} = x + C(t)$, and finally $f(t, x) = \frac{1}{x+C(t)}$. From the first equation in the above, we have

$$-\frac{C'(t)}{(x+C(t))^2} + \frac{1}{2} \frac{(-1)(-2)}{(x+C(t))^3} = \frac{1}{(x+C(t))^3}.$$

Since $-C'(t)(x+C(t)) + 1 = 1$ holds for every (t, x) , we conclude that C is constant. Thus $f = \frac{1}{x+C}$. Since $f(0, 0) = \frac{1}{C} = 1$, we have $C = 1$. Therefore $f(t, x) = \frac{1}{x+1}$, and finally we obtain

$$f(t, W_t) = \frac{1}{W_t + 1}.$$

10.4.3 Brownian Bridge

A stochastic process X_t satisfying the SDE

$$dX_t = \frac{b - X_t}{T - t} dt + dW_t, \quad 0 \leq t \leq T, \quad X(0) = a,$$

is called a *Brownian bridge*. It can be written as

$$X_t = a\left(1 - \frac{t}{T}\right) + b\frac{t}{T} + (T - t) \int_0^t \frac{1}{T - s} dW_s .$$

Since $\int_0^t 1/(T - s)^2 ds < \infty$, the integral $\int_0^t 1/(T - s) dW_s$ is a martingale.

10.4.4 Differentiability Condition

We consider an example due to Tanaka. Let $f(t, x) = |x|$. Then f is continuous, and

$$f_x(t, x) = \begin{cases} +1, & x > 0, \\ -1, & x < 0. \end{cases}$$

The Itô formula cannot be directly applied in this case since f_x does not exist at $x = 0$. However, we may regard $2 \times \delta_0(x)$ as the second order derivative of f at $x = 0$ where $\delta_0(x)$ is the Dirac delta measure at 0. If the Itô formula were to hold for $f(t, x) = |x|$, then we would have

$$df = (f_t + \frac{1}{2} f_{xx}) dt + f_x dW_t = \delta_0(x) dt + \text{sign}(x) dW_t$$

and

$$|W_t| = \int_0^t \delta_0(W_s) ds + \int_0^t \text{sign}(W_s) dW_s$$

in a suitable sense where

$$\text{sign}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0. \end{cases}$$

To proceed rigorously, we define the *local time* for Brownian motion at 0 by

$$L_t(\omega) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \lambda(\{s \in [0, t] : W_s(\omega) \in (-\epsilon, \epsilon)\}) ,$$

where the limit is taken in the L^2 -sense, and obtain *Tanaka's formula*

$$|W_t| = L_t + \int_0^t \text{sign}(W_s) dW_s .$$

For a sketch of the proof, we modify f near $x = 0$ as follows:

$$f_\varepsilon(t, x) = \begin{cases} |x|, & |x| \geq \varepsilon \\ \frac{1}{2}\left(\varepsilon + \frac{x^2}{\varepsilon}\right), & |x| < \varepsilon. \end{cases}$$

Then f_ε is continuous, and

$$\frac{\partial}{\partial x} f_\varepsilon(t, x) = \begin{cases} +1, & x \geq \varepsilon \\ \frac{x}{\varepsilon}, & |x| < \varepsilon \\ -1, & x \leq -\varepsilon. \end{cases}$$

It can be shown that

$$f_\varepsilon(W_t) = f_\varepsilon(W_0) + \int_0^t f'_\varepsilon(W_s) dW_s + \frac{1}{2\varepsilon} \lambda(\{s \in [0, t] : W_s \in (-\varepsilon, \varepsilon)\}),$$

where λ denotes the Lebesgue measure on the real line. Applying the Itô isometry to

$$\mathbb{E} \left[\left(\int_0^t \frac{W_s}{\varepsilon} \mathbb{1}_{\{W_s \in (-\varepsilon, \varepsilon)\}} dW_s \right)^2 \right],$$

we obtain

$$\int_0^t f'_\varepsilon(W_s) \mathbb{1}_{\{W_s \in (-\varepsilon, \varepsilon)\}} dW_s = \int_0^t \frac{W_s}{\varepsilon} \mathbb{1}_{\{W_s \in (-\varepsilon, \varepsilon)\}} dW_s \rightarrow 0$$

in L^2 as $\varepsilon \rightarrow 0$. Finally, we let $\varepsilon \rightarrow 0$. For more details, consult Rogers and Williams (2000b). See also Example 12.3.

10.5 Computer Experiments

Simulation 10.1 (Itô Formula) We simulate the result in Example 10.2. See the right plot in Fig. 10.3.

```
T = 3;
K = 15;
M = 1000; % number of Brownian paths
L2 = zeros(1,K);
for k = 1:K
    dt = T*2^(-k);
    dW = sqrt(dt)*randn(M,2^k);
    W = zeros(M,2^k+1);
    error = zeros(M,1);
    epsilon = zeros(M,2^k);
```

```

for m = 1:M
    for i = 1:2^k
        W(m,i+1) = W(m,i) + dW(m,i);
        epsilon(m,i)=0.5*sin(W(m,i))*(dt-dW(m,i)^2);
    end
    error(m,1) = sum( epsilon(m,1:2^k) );
end
L2(k) = sqrt(mean(error(1:M,1).^2)); % L2-norm
end
x = 0:0.01:K;
plot(x,T*2.^(-(x+1)/2),'k-');
hold on;
plot(L2,'or-')

```

Simulation 10.2 (Geometric Brownian Motion) We plot sample paths of geometric Brownian motion. For an output see Fig. 10.4.

```

N = 200; % number of time steps
T = 10;
dt = T/N;
mu = 0.15;
sigma = 0.25;
time = 0:dt:T;
M = 30; % number of sample paths
S = zeros(M,N + 1); % dW and S are matrices.
S0 = 1;
S(:,1)=S0;

dW = sqrt(dt)*randn(M,N);

for i=1:N
    S(:,i+1) = S(:,i) + mu*S(:,i)*dt + sigma*S(:,i).*dW(:,i);
end

for j = 1:M
    plot(time,S(j,1:N + 1),'b');
hold on;
end

```

Exercises

10.1 (*Continuous Dividend*) If a stock S_t pays dividend continuously, what would be a modification of the geometric Brownian motion model for S_t ?

10.2 Let W_t denote a Brownian motion.

- (i) Suppose that X_t is a stochastic process given by $dX_t = a_t dt + b_t dW_t$ for some processes a_t and b_t . Compute $d(e^{X_t})$.
- (ii) For a stochastic process r_t , $t \geq 0$, define $Z_t = \exp(-\int_0^t r_s ds)$. Compute dZ_t .
- (iii) Suppose that a process $B_t > 0$ satisfies $dB_t = r_t B_t dt$ for some process r_t . Find B_t .

10.3 (*A proof of Remark 7.5*) Define $f(t, x) = e^{\frac{1}{2}t} \cos x$ and compute $df(t, W_t)$.

10.4 Solve the stochastic differential equation $dX_t = dt + 2\sqrt{X_t} dW_t$, $X_0 = 1$.

10.5 Compute $\int_0^t W_s dW_s$ by assuming that $\int_0^t W_s dW_s = f(t, W_t)$ for some f . (If this assumption is wrong, then we may try other candidates.)

10.6 Let W_t denote a Brownian motion.

- (i) Is there a sufficiently smooth function $f(t, x)$ such that $\int_0^t W_s^2 dW_s = f(t, W_t)$?
- (ii) Assume that $\int_0^t W_s^2 dW_s = f(t, W_t) + \int_0^t g(s, W_s) ds$ for some $f(t, x)$ and $g(t, x)$. Compute $\int_0^t W_s^2 dW_s$.

10.7 Under the assumption that $\int_0^t s dW_s = f(t, W_t) + \int_0^t g(s, W_s) ds$ for some $f(t, x)$ and $g(t, x)$, find $\int_0^t s dW_s$.

10.8 For a deterministic function $f(t)$, show that

$$\int_0^t f(s) dW_s = f(t)W_t - \int_0^t W_s f'(s) ds.$$

10.9 Find $\int_0^t e^{W_s} dW_s$ by assuming that $\int_0^t e^{W_s} dW_s = f(t, W_t) + \int_0^t g(s, W_s) ds$ for some f and g .

10.10 (i) Let $a_k(t) = \mathbb{E}[W_t^k]$ for $k \geq 0$ and $t \geq 0$. Apply the Itô formula to show that

$$a_k(t) = \frac{1}{2}k(k-1) \int_0^t a_{k-2}(s) ds, \quad k \geq 2.$$

(ii) Prove that $\mathbb{E}[W_t^{2k+1}] = 0$ and

$$\mathbb{E}[W_t^{2k}] = \frac{(2k)!t^k}{2^k k!}.$$

10.11 Let $X_t = \int_0^t a_s \, dW_s$ for a bounded process a_t . Prove that $X_t^2 - \int_0^t a_s^2 \, ds$ is a martingale. For example, $W_t^2 - t$ is a martingale.

10.12 Let $a(t)$, $b(t)$, $c(t)$, $d(t)$ be continuous deterministic functions, and consider a stochastic differential equation

$$X_t = \int_0^t (a(s)X_s + b(s)) \, ds + \int_0^t (c(s)X_s + d(s)) \, dW_s .$$

Let $\alpha(t) = \mathbb{E}[X_t]$ and $\beta(t) = \mathbb{E}[X_t^2]$. Show that

$$\begin{cases} \alpha' = a\alpha + b \\ \beta' = (2a + c^2)\beta + 2(b + cd)\alpha + d^2 . \end{cases}$$

Chapter 11

Girsanov's Theorem



We introduce Girsanov's theorem for the existence of an equivalent probability measure for which a Brownian motion with drift is also a Brownian motion. Let $\{W_t\}_{t \geq 0}$ be a Brownian motion with respect to a probability measure $\mathbb{P}$. Take a real constant θ , and consider $X_t = W_t + \theta t$, $0 \leq t < \infty$, which is called a Brownian motion with drift. Our goal is to find a probability measure $\mathbb{Q}$ for which X_t , $0 \leq t \leq T$, is a Brownian motion for a fixed $T < +\infty$. We require an additional condition that $\mathbb{Q}$ is equivalent to $\mathbb{P}$, i.e., $\mathbb{P}(A) = 0$ if and only if $\mathbb{Q}(A) = 0$. In other words, an event occurs with positive $\mathbb{P}$ -probability if and only if it happens with positive $\mathbb{Q}$ -probability. Such a condition is important in financial applications since we have to deal with the same set of asset price movements even when we switch to a new probability measure. Igor V. Girsanov (1934–1967) proved the existence of such a measure $\mathbb{Q}$. We will find first a necessary condition for the existence of an equivalent probability measure $\mathbb{Q}$ for which a Brownian motion with drift is a Brownian motion.

11.1 Motivation

Let $\mathbb{E}$ or $\mathbb{E}^{\mathbb{P}}$ denote expectation with respect to a probability measure $\mathbb{P}$. Let $\{W_t\}_{t \geq 0}$ denote a $\mathbb{P}$ -Brownian motion. For a given constant θ consider a new stochastic process

$$X_t = W_t + \theta t .$$

(See Fig. 11.1.) We want to find an equivalent probability measure $\mathbb{Q}$ such that $\{X_t\}_{0 \leq t \leq T}$ is a $\mathbb{Q}$ -Brownian motion for a fixed $0 < T < +\infty$. (See Fig. 11.2.)

First, we will see what a possible definition of $\mathbb{Q}$ would look like. Given a $\mathbb{P}$ -Brownian motion $\{W_t\}_{0 \leq t \leq T}$, let Ω be the set of all sample paths $\omega : [0, T] \rightarrow \mathbb{R}$, and let $\{\mathcal{F}_t\}$ be the filtration generated by $\{W_t\}$. For t , $0 \leq t \leq T$ and an interval $I = [a, b] \subset \mathbb{R}$, consider a cylinder subset of Ω defined by

Fig. 11.1 Sample paths of a Brownian motion with drift $X_t = W_t + \theta t$ with drift coefficient $\theta = 0.5$ where the straight lines are $x = \theta t$ (solid in black) and $x = 0$ (dotted in blue)

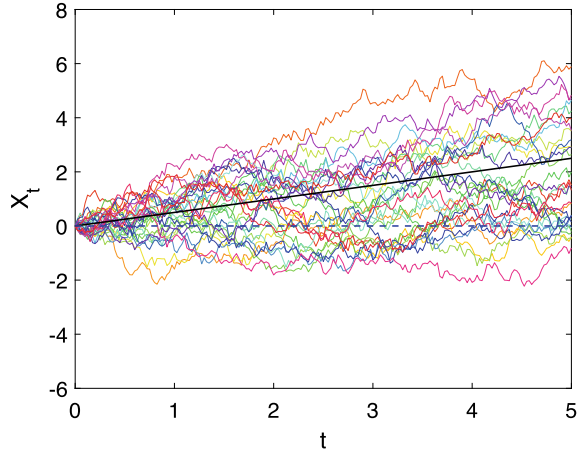
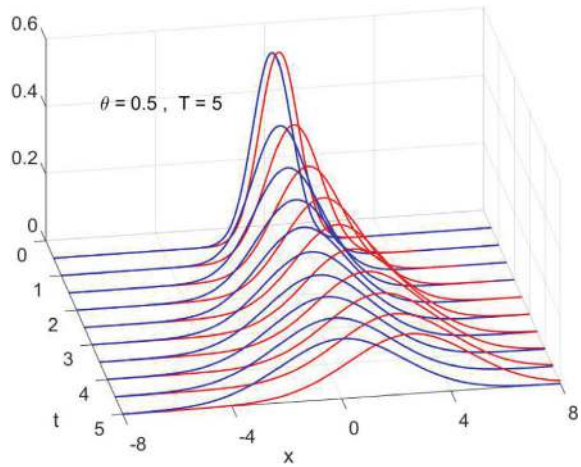


Fig. 11.2 Probability density functions of $X_t = W_t + \theta t$ with respect to $\mathbb{P}$ (red) and $\mathbb{Q}$ (blue)



$$C(t; I) = \{\omega \in \Omega : W_t(\omega) \in I\}.$$

Note that $C(t; I)$ is $\mathcal{F}_t$ -measurable since $C(t; I) = W_t^{-1}(I)$.

Suppose that there exists an equivalent probability measure $\mathbb{Q}$ such that

$$\begin{aligned} \int_{\Omega} f(W_t(\omega)) d\mathbb{P}(\omega) &= \int_{\Omega} f(X_t(\omega)) d\mathbb{Q}(\omega) \\ &= \int_{\Omega} f(W_t(\omega) + \theta t) d\mathbb{Q}(\omega) \end{aligned}$$

for every bounded measurable function $f : \mathbb{R} \rightarrow \mathbb{R}$. If we take an indicator function $f(x) = \mathbb{1}_I(x)$ for an interval $I = [a, b]$, then we have

$$\int_{\Omega} f(W_t) d\mathbb{P} = \int_{C(t;I)} d\mathbb{P} = \mathbb{P}(C(t; I))$$

and

$$\int_{\Omega} f(W_t + \theta t) d\mathbb{Q} = \int_{C(t; I - \theta t)} d\mathbb{Q} = \mathbb{Q}(C(t; I - \theta t)) .$$

Hence

$$\mathbb{P}(C(t; I)) = \mathbb{Q}(C(t; I - \theta t)) ,$$

or

$$\mathbb{P}(C(t; I + \theta t)) = \mathbb{Q}(C(t; I))$$

for every $0 \leq t \leq T$ and every I . By the definition of $\mathbb{P}$,

$$\mathbb{P}(C(t; I + \theta t)) = \int_{a+\theta t}^{b+\theta t} \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx = \int_a^b \frac{1}{\sqrt{2\pi t}} e^{-(y+\theta t)^2/2t} dy . \quad (11.1)$$

Since $\mathbb{Q}$ is absolutely continuous, the Radon–Nikodym theorem implies that there exists the Radon–Nikodym derivative $L = \frac{d\mathbb{Q}}{d\mathbb{P}}$ of $\mathbb{Q}$ with respect to $\mathbb{P}$. Let $L_t = \mathbb{E}[L|\mathcal{F}_t]$. Note that $L = L_T$.¹ Now we find a possible formula for L_t assuming that there exists a function $\rho_t : \mathbb{R} \rightarrow \mathbb{R}$ such that $L_t(\omega) = \rho_t(W_t(\omega))$, i.e., $L_t(\omega)$ depends on the present time t and does not depend on the past history of ω over the time interval $[0, t)$. Then

$$\begin{aligned} \mathbb{Q}(C(t; I)) &= \mathbb{E}^{\mathbb{Q}}[\mathbf{1}_{C(t;I)}] \\ &= \mathbb{E}^{\mathbb{P}}[\mathbf{1}_{C(t;I)} L] \\ &= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\mathbf{1}_{C(t;I)} L | \mathcal{F}_t]] \quad (\text{by the tower property}) \\ &= \mathbb{E}^{\mathbb{P}}[\mathbf{1}_{C(t;I)} \mathbb{E}^{\mathbb{P}}[L | \mathcal{F}_t]] \quad (\text{take out what is known}) \\ &= \mathbb{E}^{\mathbb{P}}[\mathbf{1}_{C(t;I)} L_t] \\ &= \int_{C(t;I)} \rho_t(W_t) d\mathbb{P} \\ &= \int_a^b \rho_t(x) \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx . \end{aligned}$$

Using (11.1), we obtain

$$\rho_t(x) e^{-x^2/2t} = e^{-(x+\theta t)^2/2t} . \quad (11.2)$$

Hence

$$\rho_t(x) = e^{-\theta x - \frac{1}{2}\theta^2 t}$$

¹ Later it will be shown that $L_t = \exp(-\frac{1}{2}\theta^2 t - \theta W_t)$, which is a $\mathbb{P}$ -martingale.

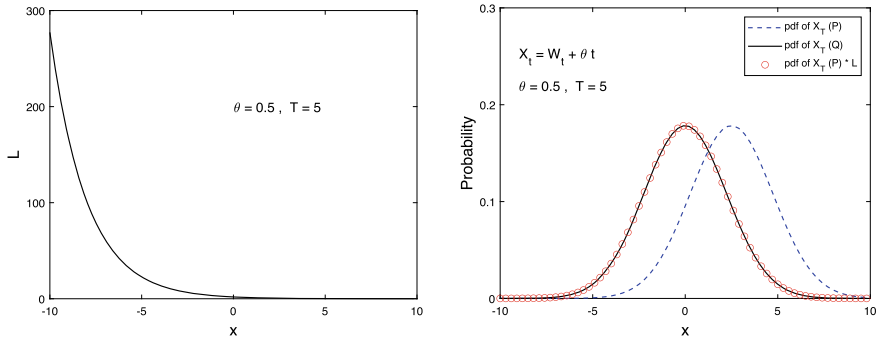


Fig. 11.3 The graph of $L(x) = \rho_T(x)$ (left), and the probability density functions of $X_T = W_T + \theta T$ with respect to $\mathbb{P}$ and $\mathbb{Q}$ (right), where $\theta = 0.5$ and $T = 5$

and

$$L_t = \rho_t(W_t) = e^{-\theta W_t - \frac{1}{2}\theta^2 t} . \quad (11.3)$$

Note that

$$L_t = \rho_t(X_t - \theta t) = e^{-\theta X_t + \frac{1}{2}\theta^2 t} .$$

See Fig. 11.3 for the graph of $L(x) = e^{-\theta x + \frac{1}{2}\theta^2 T}$. Note that the multiplication by L moves the normal pdf with respect to $\mathbb{P}$, which has mean equal to θT , to the normal pdf with respect to $\mathbb{Q}$, which has mean equal to 0. In general, for every $0 < t \leq T$, we have

$$\frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} = \rho_t(x) \frac{1}{\sqrt{2\pi t}} e^{-(x-\theta t)^2/2t} .$$

Remark 11.1 The same conclusion can be obtained using the Fourier transformation as follows: Since the probability density functions of X_t and W_t with respect to $\mathbb{Q}$ and $\mathbb{P}$, respectively, are identical, the corresponding Fourier transforms are equal. In fact, $\mathbb{E}^{\mathbb{Q}}[e^{i\zeta X_t}] = \mathbb{E}^{\mathbb{P}}[e^{i\zeta W_t}] = e^{-\frac{1}{2}\zeta^2 t}$ for real ζ . Since

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[e^{i\zeta X_t}] &= \mathbb{E}^{\mathbb{P}}[e^{i\zeta(W_t + \theta t)} \rho_t(W_t)] \\ &= \int_{-\infty}^{\infty} e^{i\zeta(x + \theta t)} \rho_t(x) \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx \\ &= \int_{-\infty}^{\infty} e^{i\zeta y} \rho_t(y - \theta t) \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y - \theta t)^2}{2t}} dy \end{aligned}$$

and since

$$\mathbb{E}^{\mathbb{P}}[e^{i\zeta W_t}] = \int_{-\infty}^{\infty} e^{i\zeta y} \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}} dy ,$$

by the uniqueness of the inverse Fourier transform, we have

$$\rho_t(y - \theta t) \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-\theta t)^2}{2t}} = \frac{1}{\sqrt{2\pi t}} e^{-\frac{y^2}{2t}},$$

which is equivalent to (11.2), and we obtain (11.3).

11.2 Equivalent Probability Measure

Now we consider a sequence $\{L_t\}_{0 \leq t \leq T}$ given by $L_t = \mathbb{E}[L | \mathcal{F}_t]$.

Definition 11.1 Recall that $L_t = e^{-\frac{1}{2}\theta^2 t - \theta W_t}$ is a $\mathbb{P}$ -martingale and $\mathbb{E}[L_t] = 1$ for $t \geq 0$. Define a probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_T)$ by $d\mathbb{Q} = L_T d\mathbb{P}$, i.e.,

$$\mathbb{Q}(A) = \mathbb{E}^{\mathbb{P}}[\mathbf{1}_A L_T] = \int_A L_T d\mathbb{P}$$

for $A \in \mathcal{F}_T$. (Note that $\mathbb{Q}$ is equivalent to $\mathbb{P}$ since $L_T > 0$.)

Lemma 11.1 For $0 \leq s \leq t$, and an $\mathcal{F}_t$ -measurable random variable ϕ_t , we have

$$\mathbb{E}^{\mathbb{Q}}[\phi_t | \mathcal{F}_s] = \mathbb{E}^{\mathbb{P}}\left[\phi_t \frac{L_t}{L_s} \middle| \mathcal{F}_s\right].$$

In particular, for $s = 0$, $\mathbb{E}^{\mathbb{Q}}[\phi_t] = \mathbb{E}^{\mathbb{P}}[\phi_t L_t]$.

Proof For $A \in \mathcal{F}_s$, we have

$$\begin{aligned} \int_A \phi_t d\mathbb{Q} &= \int_A \mathbb{E}^{\mathbb{Q}}[\phi_t | \mathcal{F}_s] d\mathbb{Q} \quad (\text{definition of conditional expectation}) \\ &= \int_A \mathbb{E}^{\mathbb{Q}}[\phi_t | \mathcal{F}_s] L_T d\mathbb{P} \\ &= \int_A \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{Q}}[\phi_t | \mathcal{F}_s] L_T | \mathcal{F}_s] d\mathbb{P} \quad (\text{by the tower property}) \\ &= \int_A \mathbb{E}^{\mathbb{Q}}[\phi_t | \mathcal{F}_s] \mathbb{E}^{\mathbb{P}}[L_T | \mathcal{F}_s] d\mathbb{P} \quad (\text{take out what is known}) \\ &= \int_A \mathbb{E}^{\mathbb{Q}}[\phi_t | \mathcal{F}_s] L_s d\mathbb{P}. \end{aligned}$$

On the other hand,

$$\begin{aligned}
 \int_A \phi_t \, d\mathbb{Q} &= \int_A \phi_t L_T \, d\mathbb{P} \\
 &= \int_A \mathbb{E}^{\mathbb{P}}[\phi_t L_T | \mathcal{F}_s] \, d\mathbb{P} \quad (\text{definition of conditional expectation}) \\
 &= \int_A \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\phi_t L_T | \mathcal{F}_t] | \mathcal{F}_s] \, d\mathbb{P} \quad (\text{by the tower property}) \\
 &= \int_A \mathbb{E}^{\mathbb{P}}[\phi_t \mathbb{E}^{\mathbb{P}}[L_T | \mathcal{F}_t] | \mathcal{F}_s] \, d\mathbb{P} \quad (\text{take out what is known}) \\
 &= \int_A \mathbb{E}^{\mathbb{P}}[\phi_t L_t | \mathcal{F}_s] \, d\mathbb{P} .
 \end{aligned}$$

Thus we have

$$\int_A \mathbb{E}^{\mathbb{Q}}[\phi_t | \mathcal{F}_s] L_s \, d\mathbb{P} = \int_A \mathbb{E}^{\mathbb{P}}[\phi_t L_t | \mathcal{F}_s] \, d\mathbb{P}$$

for every $A \in \mathcal{F}_s$, and hence $\mathbb{E}^{\mathbb{Q}}[\phi_t | \mathcal{F}_s] L_s = \mathbb{E}^{\mathbb{P}}[\phi_t L_t | \mathcal{F}_s]$. □

11.3 Brownian Motion with Drift

Lemma 11.2 *Let $\mathbb{Q}$ be the equivalent measure given in Definition 11.1, and let $X_t = W_t + \theta t$. Then $\mathbb{E}^{\mathbb{Q}}[X_t] = 0$ and $\mathbb{E}^{\mathbb{Q}}[X_t^2] = t$.*

Proof Recall the results in Exercises 7.8, 7.9. Then we have

$$\begin{aligned}
 \mathbb{E}^{\mathbb{Q}}[X_t] &= \mathbb{E}[(W_t + \theta t) e^{-\frac{1}{2}\theta^2 t - \theta W_t}] \\
 &= e^{-\frac{1}{2}\theta^2 t} (\mathbb{E}[W_t e^{-\theta W_t}] + \theta t \mathbb{E}[e^{-\theta W_t}]) \\
 &= e^{-\frac{1}{2}\theta^2 t} (-\theta t e^{\frac{1}{2}\theta^2 t} + \theta t e^{\frac{1}{2}\theta^2 t}) \\
 &= 0
 \end{aligned}$$

and

$$\begin{aligned}
 \mathbb{E}^{\mathbb{Q}}[X_t^2] &= \mathbb{E}[(W_t + \theta t)^2 e^{-\frac{1}{2}\theta^2 t - \theta W_t}] \\
 &= e^{-\frac{1}{2}\theta^2 t} (\mathbb{E}[W_t^2 e^{-\theta W_t}] + 2\theta t \mathbb{E}[W_t e^{-\theta W_t}] + \theta^2 t^2 \mathbb{E}[e^{-\theta W_t}]) \\
 &= e^{-\frac{1}{2}\theta^2 t} \{(t + \theta^2 t^2) e^{\frac{1}{2}\theta^2 t} + 2\theta t (-\theta t) e^{\frac{1}{2}\theta^2 t} + \theta^2 t^2 e^{\frac{1}{2}\theta^2 t}\} \\
 &= t .
 \end{aligned}$$

□

For higher moments, we have $\mathbb{E}^{\mathbb{Q}}[X_t^3] = 0$ and $\mathbb{E}^{\mathbb{Q}}[X_t^4] = 3t^2$. For a numerical confirmation, consult Simulation 11.2.

Lemma 11.3 *Let $\mathbb{Q}$ be the equivalent measure given in Definition 11.1, and let $X_t = W_t + \theta t$. Then X_t is a $\mathbb{Q}$ -martingale.*

Proof Since, by the result in Exercise 11.4,

$$\mathbb{E}[W_t L_t | \mathcal{F}_s] = e^{-\frac{1}{2}\theta^2 t} \mathbb{E}[W_t e^{-\theta W_t} | \mathcal{F}_s] = (W_s - \theta(t - s)) L_s ,$$

we have

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[X_t | \mathcal{F}_s] &= \mathbb{E}[X_t L_t L_s^{-1} | \mathcal{F}_s] \quad (\text{by Lemma 11.1}) \\ &= L_s^{-1} \mathbb{E}[(W_t + \theta t) L_t | \mathcal{F}_s] \\ &= L_s^{-1} \mathbb{E}[W_t L_t | \mathcal{F}_s] + L_s^{-1} \theta t \mathbb{E}[L_t | \mathcal{F}_s] \\ &= (W_s - \theta(t - s)) + \theta t \\ &= W_s + \theta s = X_s . \end{aligned}$$

□

11.4 Girsanov's Theorem

Theorem 11.1 (Girsanov) *Let $\{W_t\}_{0 \leq t \leq T}$ be a $\mathbb{P}$ -Brownian motion with respect to a filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let θ be a real constant. Let*

$$X_t = W_t + \theta t$$

and define an equivalent probability measure $\mathbb{Q}$ by $d\mathbb{Q} = L_T d\mathbb{P}$, where

$$L_T = e^{-\frac{1}{2}\theta^2 T - \theta W_T} .$$

Then $\{X_t\}_{0 \leq t \leq T}$ is a $\mathbb{Q}$ -Brownian motion.

Proof Two measures are equivalent since the Radon–Nikodym derivative is positive. Clearly, $X_0 = 0$ and $dX dX = dt$. To prove that $\{X_t\}$ is a $\mathbb{Q}$ -Brownian motion, use Theorem 7.10 and Lemma 11.3. □

Note that our proof of Theorem 11.1 does not rely on Itô's lemma explicitly even though Itô's lemma is used via Theorem 7.10. A generalized version of Girsanov's theorem can be proved using Itô's lemma.

Theorem 11.2 (Generalized Girsanov's Theorem) *Fix a constant $0 < T < \infty$. Let $\{W_t\}_{0 \leq t \leq T}$ be a $\mathbb{P}$ -Brownian motion with respect to a filtration $\{\mathcal{F}_t\}_{0 \leq t \leq T}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{\theta_t\}_{0 \leq t \leq T}$ be an adapted process to $\{\mathcal{F}_t\}$. Let*

$$X_t = W_t + \int_0^t \theta_s ds$$

and

$$L_t = \exp \left(-\frac{1}{2} \int_0^t \theta_s^2 ds - \int_0^t \theta_s dW_s \right).$$

Assume that the Novikov condition holds, i.e.,

$$\mathbb{E}^{\mathbb{P}} \left[\exp \left(\frac{1}{2} \int_0^T \theta_t^2 dt \right) \right] < \infty.$$

Then $\{L_t\}_{0 \leq t \leq T}$ is a $\mathbb{P}$ -martingale, and $\mathbb{E}[L_t] = \mathbb{E}[L_0] = 1$ for $0 \leq t \leq T$. Using L_T as the Radon–Nikodym derivative, define an equivalent probability measure $\mathbb{Q}$ by $d\mathbb{Q} = L_T d\mathbb{P}$. Then $\{X_t\}_{0 \leq t \leq T}$ is a $\mathbb{Q}$ -Brownian motion.

Proof Clearly, $dX_t dX_t = (dW_t + \theta_t dt)(dW_t + \theta_t dt) = dt$. By the generalized Itô formula,

$$dL_t = L_t \left(-\frac{1}{2} \theta_t^2 dt - \theta_t dW_t \right) + \frac{1}{2} L_t \theta_t^2 dt = -L_t \theta_t dW_t.$$

Hence $\{L_t\}$ is a $\mathbb{P}$ -martingale. Since

$$\begin{aligned} d(X_t L_t) &= L_t dX_t + X_t dL_t + (dX_t)(dL_t) \\ &= L_t (dW_t + \theta_t dt) + X_t (-L_t \theta_t dW_t) + (dW_t + \theta_t dt)(-L_t \theta_t dW_t) \\ &= (-X_t \theta_t + 1) L_t dW_t, \end{aligned}$$

$X_t L_t$ is also a $\mathbb{P}$ -martingale. Hence, for $s \leq t$,

$$\mathbb{E}^{\mathbb{Q}}[X_t | \mathcal{F}_s] = \mathbb{E}^{\mathbb{P}}[X_t \frac{L_t}{L_s} | \mathcal{F}_s] = \frac{1}{L_s} \mathbb{E}^{\mathbb{P}}[X_t L_t | \mathcal{F}_s] = \frac{1}{L_s} X_s L_s = X_s.$$

Now we apply Theorem 7.10. □

11.5 The Martingale Property of the Radon–Nikodym Derivative

As shown in the proof of Theorem 11.2, $\{L_t\}$ is a $\mathbb{P}$ -martingale and

$$L_t = \mathbb{E}^{\mathbb{P}}[L_T | \mathcal{F}_t] = e^{-\frac{1}{2} \theta^2 t - \theta W_t}, \quad 0 \leq t \leq T.$$

Fix s , $0 \leq s \leq T$. For an $\mathcal{F}_s$ -measurable random variable ϕ_s , and for t such that $s \leq t \leq T$, we have

$$\begin{aligned}
 \mathbb{E}^{\mathbb{Q}}[\phi_s] &= \mathbb{E}^{\mathbb{P}}[\phi_s L_T] \\
 &= \mathbb{E}^{\mathbb{P}}[\phi_s L_T | \mathcal{F}_0] \\
 &= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\phi_s L_T | \mathcal{F}_t] | \mathcal{F}_0] \quad (\text{by the tower property}) \\
 &= \mathbb{E}^{\mathbb{P}}[\phi_s \mathbb{E}^{\mathbb{P}}[L_T | \mathcal{F}_t] | \mathcal{F}_0] \quad (\text{take out what is known}) \\
 &= \mathbb{E}^{\mathbb{P}}[\phi_s L_t | \mathcal{F}_0] \\
 &= \mathbb{E}^{\mathbb{P}}[\phi_s L_t] .
 \end{aligned}$$

Example 11.1 Take $\phi_s = W_s$ for some $0 < s < T$. Then for $0 \leq u \leq s$ we have

$$\begin{aligned}
 \mathbb{E}^{\mathbb{P}}[W_s L_u] &= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[W_s L_u | \mathcal{F}_u]] \quad (\text{by the tower property}) \\
 &= \mathbb{E}^{\mathbb{P}}[L_u \mathbb{E}^{\mathbb{P}}[W_s | \mathcal{F}_u]] \quad (\text{take out what is known}) \\
 &= \mathbb{E}^{\mathbb{P}}[L_u W_u] \\
 &= \mathbb{E}^{\mathbb{Q}}[W_u] \\
 &= \mathbb{E}^{\mathbb{Q}}[X_u - \theta u] \\
 &= \mathbb{E}^{\mathbb{Q}}[X_u] - \theta u \\
 &= -\theta u .
 \end{aligned}$$

See Fig. 11.4 and Simulation 11.3 where we choose $\theta = 0.5$, $T = 5$ and $s = 3$. The numerical estimation is marked by red circles, which agrees with the theoretical conclusion. The dotted horizontal line indicates the value $\mathbb{E}^{\mathbb{P}}[W_s L_t] = \mathbb{E}^{\mathbb{Q}}[W_s] = -\theta s$, $s \leq t \leq T$.

Fig. 11.4 The martingale property of $L_t = \exp(-\frac{1}{2}\theta^2 t - \theta W_t)$ is used in computing $\mathbb{E}^{\mathbb{P}}[W_s L_t]$, $0 \leq t \leq T$, where $\theta = 0.5$, $T = 5$ and $s = 3$

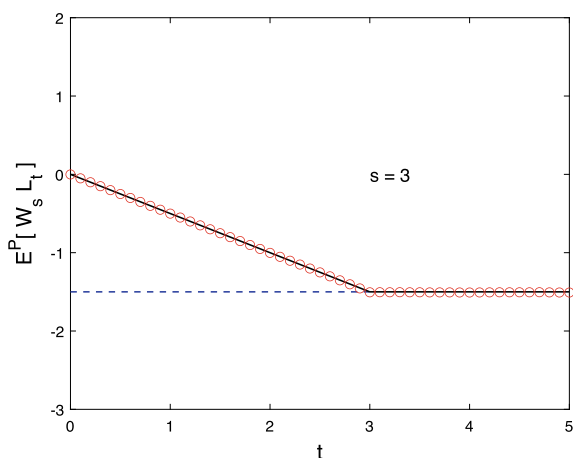
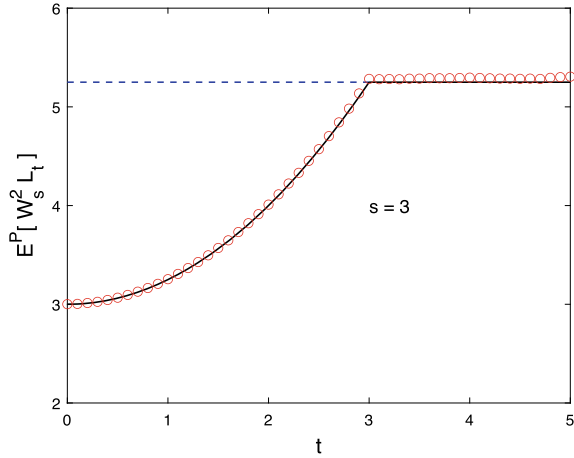


Fig. 11.5 The martingale property of $L_t = \exp(-\frac{1}{2}\theta^2 t - \theta W_t)$ is used in computing $\mathbb{E}^{\mathbb{P}}[W_s^2 L_t]$, $0 \leq t \leq T$, where $\theta = 0.5$, $T = 5$ and $s = 3$



Example 11.2 Take $\phi_s = W_s^2$ for some $0 < s < T$. Then for $0 \leq u \leq s$ we have

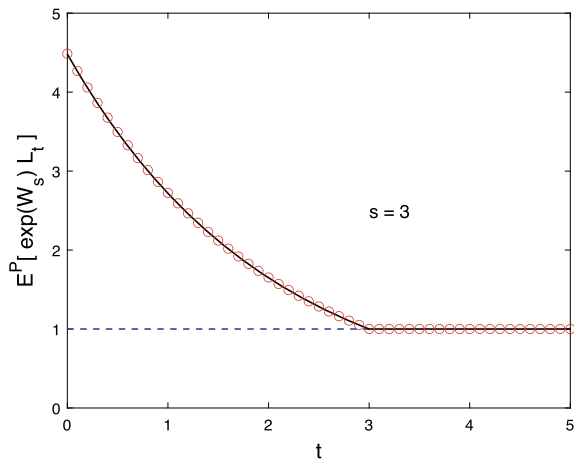
$$\begin{aligned}
 \mathbb{E}^{\mathbb{P}}[W_s^2 L_u] &= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[W_s^2 L_u | \mathcal{F}_u]] \quad (\text{by the tower property}) \\
 &= \mathbb{E}^{\mathbb{P}}[L_u \mathbb{E}^{\mathbb{P}}[W_s^2 | \mathcal{F}_u]] \quad (\text{take out what is known}) \\
 &= \mathbb{E}^{\mathbb{P}}[L_u \mathbb{E}^{\mathbb{P}}[(W_s^2 - s) + s | \mathcal{F}_u]] \\
 &= \mathbb{E}^{\mathbb{P}}[L_u \{(W_u^2 - u) + s\}] \\
 &= \mathbb{E}^{\mathbb{P}}[L_u (W_u^2 - u + s)] \\
 &= \mathbb{E}^{\mathbb{Q}}[W_u^2 - u + s] \\
 &= \mathbb{E}^{\mathbb{Q}}[(X_u - \theta u)^2 - u + s] \\
 &= \mathbb{E}^{\mathbb{Q}}[X_u^2 - 2\theta u X_u + \theta^2 u^2 - u + s] \\
 &= u - 0 + \theta^2 u^2 - u + s \\
 &= \theta^2 u^2 + s.
 \end{aligned}$$

See Fig. 11.5 and Simulation 11.3 where we choose $\theta = 0.5$, $T = 5$ and $s = 3$. The numerical estimation is marked by red circles, which agrees with the theoretical conclusion. The dotted line indicates the value $\mathbb{E}^{\mathbb{P}}[W_s^2 L_t] = \mathbb{E}^{\mathbb{Q}}[W_s^2] = \theta^2 s^2 + s$, $s \leq t \leq T$.

Example 11.3 Take $\phi_s = \exp(W_s)$ for some $0 < s < T$. Then for $0 \leq u \leq s$ we have

$$\begin{aligned}
 \mathbb{E}^{\mathbb{P}}[\exp(W_s) L_u] &= \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[\exp(W_s) L_u | \mathcal{F}_u]] \quad (\text{by the tower property}) \\
 &= \mathbb{E}^{\mathbb{P}}[L_u \mathbb{E}^{\mathbb{P}}[\exp(W_s) | \mathcal{F}_u]] \quad (\text{take out what is known}) \\
 &= \mathbb{E}^{\mathbb{P}}[L_u \mathbb{E}^{\mathbb{P}}[\exp(W_s - \frac{1}{2}s) e^{\frac{1}{2}s} | \mathcal{F}_u]] \\
 &= e^{\frac{1}{2}s} \mathbb{E}^{\mathbb{P}}[L_u \mathbb{E}^{\mathbb{P}}[\exp(W_s - \frac{1}{2}s) | \mathcal{F}_u]]
 \end{aligned}$$

Fig. 11.6 The martingale property of $L_t = \exp(-\frac{1}{2}\theta^2 t - \theta W_t)$ is used in computing $\mathbb{E}^{\mathbb{P}}[\exp(W_s)L_t], 0 \leq t \leq T$, where $\theta = 0.5, T = 5$ and $s = 3$



$$\begin{aligned}
 &= e^{\frac{1}{2}s} \mathbb{E}^{\mathbb{P}}[L_u \exp(W_u - \frac{1}{2}u)] \quad (\text{by Theorem 7.8}) \\
 &= e^{\frac{1}{2}s} \mathbb{E}^{\mathbb{Q}}[\exp(W_u - \frac{1}{2}u)] \\
 &= e^{\frac{1}{2}s} \mathbb{E}^{\mathbb{Q}}[\exp(X_u - \theta u - \frac{1}{2}u)] \\
 &= e^{\frac{1}{2}s - \theta u - \frac{1}{2}u} \mathbb{E}^{\mathbb{Q}}[\exp(X_u)] \\
 &= e^{\frac{1}{2}s - \theta u - \frac{1}{2}u} e^{\frac{1}{2}u} \\
 &= e^{\frac{1}{2}s - \theta u}.
 \end{aligned}$$

See Fig. 11.6 and Simulation 11.3 where we choose $\theta = 0.5, T = 5$ and $s = 3$. The numerical estimation is marked by red circles, which agrees with the theoretical conclusion. The dotted line indicates the value $\mathbb{E}^{\mathbb{P}}[W_s^2 L_t] = \mathbb{E}^{\mathbb{Q}}[\exp(W_s)] = \exp(\frac{1}{2}s - \theta s), s \leq t \leq T$.

11.6 Computer Experiments

Simulation 11.1 (Brownian Motion with Drift) We plot 30 sample paths of Brownian motion with drift. For the output see Fig. 11.1.

```

N = 200; % number of times steps
T = 5;
dt = T/N;
theta = 0.5;

```

```

M = 30; % number of sample paths
X = zeros(M,N+1); % dW and X are matrices.

dW = sqrt(dt)*randn(M,N);
for i = 1:N
    X(:,i+1) = X(:,i) + dW(:,i) + theta*dt;
end

for j = 1:M
    plot(0:dt:T,X(j,:), 'color',hsv2rgb([j/M 1 1]));
hold on;
end

x = 0:dt:T;
plot(x,theta*x, 'k')
hold on
plot(x,zeros(length(x),1), 'b--')
xlabel('t');
ylabel('X_t');

```

Simulation 11.2 (Moments of the Brownian Motion with Drift) For $1 \leq k \leq 4$ we compute the k th moments $m_k = \mathbb{E}^{\mathbb{Q}}[X_t^k]$ of $X_t = W_t + \theta t$, $0 \leq t \leq T$, with respect to the martingale measure $\mathbb{Q}$. Note that $m_1 = 0$, $m_2 = t$, $m_3 = 0$ and $m_4 = 3t^2$, which are respectively equal to mean, variance, skewness and kurtosis of the normal distribution $N(0, t)$. In the simulation we take $t = T = 5$.

```

T = 5
theta = 0.35
M = 10^8; % number of sample paths
W = sqrt(T)*randn(M,1);
X = W + theta*T;
L = exp(-0.5*theta^2*T)*exp(-theta*W); % Radon-Nikodym derivative
aveQ_X = mean(X.*L)
varQ_X = mean(X.^2.*L)
skewQ_X = mean(X.^3.*L)
kurtoQ_X = mean(X.^4.*L)

```

The output is given in the following:

```

aveQ_X =
    -3.6653e-04

varQ_X =
    5.0020

```

```
skewQ_X =
    -0.0115

kurtoQ_X =
    75.0654
```

Simulation 11.3 (The Martingale Property of the Radon–Nikodym Derivative) We check the fact that for an $\mathcal{F}_s$ -measurable random variable ϕ_s , and for t such that $s \leq t \leq T$,

$$\mathbb{E}^{\mathbb{Q}}[\phi_s] \mathbb{E}^{\mathbb{P}}[\phi_s L_t] .$$

(Consult Sect. 11.5.) For the output of numerical experiments see Figs. 11.4, 11.5 and 11.6.

```
theta = 0.5
T = 5;
N = T*10;
dt = T/N;
M = 10^6; % number of sample paths
W = zeros(M,N+1); % Brownian motion
L = zeros(M,N+1); % Radon-Nikodym derivative
dW = sqrt(dt)*randn(M,N);
for i = 1:N
    W(:,i+1) = W(:,i) + dW(:,i);
end
for i = 1:N+1
    L(:,i) = exp(-0.5*theta^2*(i-1)*dt)*exp(-theta*W(:,i));
end
ExpQ1 = zeros(1,N+1); % expectation of W with respect to Q
ExpQ2 = zeros(1,N+1); % expectation of W^2 with respect to Q
ExpQ3 = zeros(1,N+1); % expectation of exp(W_s) with respect to Q

s = 30*dt % s = 3

for i = 1:N+1
    ExpQ1(i) = mean(W(:,31).*L(:,i));
    ExpQ2(i) = mean(W(:,31).^2.*L(:,i));
    ExpQ3(i) = mean(exp(W(:,31)).*L(:,i));
end

figure(1)
plot((0:s*10)*dt, (-1)*theta*(0:s*10)*dt, 'k-');
hold on
plot((s*10:N)*dt, (-1)*theta*s*ones(N+1-s*10,1), 'k-');
hold on
```

```

plot((0:N)*dt,ExpQ1,'ro')
ylabel('E^P[ W_s L_t ]');

figure(2)
plot((0:s*10)*dt,s+theta^2*((0:s*10)*dt).^2,'k-');
hold on
plot((s*10:N)*dt,(s+theta^2*s^2)*ones(N+1-s*10,1),'k-');
hold on
plot((0:N)*dt,ExpQ2,'ro')
ylabel('E^P[ W_s^2 L_t ]');

figure(3)
plot((0:s*10)*dt,exp(0.5*s-theta*(0:s*10)*dt),'k-');
hold on
plot((s*10:N)*dt,exp(0.5*s-theta*s)*ones(N+1-s*10,1),'k-');
hold on
plot((0:N)*dt,ExpQ3,'ro')
ylabel('E^P[ exp(W_s) L_t ]');

```

Exercises

11.1 Let Z denote a standard normal variable.

- (i) Show that $\mathbb{E}[e^{W_t}] = \mathbb{E}[e^{\sqrt{t}Z}] = e^{\frac{1}{2}t}$.
- (ii) For a real constant α , show that $e^{\alpha Z} \in L^p$ for $1 \leq p < \infty$.

11.2 Show that $\mathbb{E}^{\mathbb{Q}}[W_t] = -\theta t$ and $\mathbb{E}^{\mathbb{Q}}[W_t^2] = \theta^2 t^2 + t$ by direct computation.

11.3 Show directly that $\mathbb{E}^{\mathbb{P}}[e^{\alpha W_t}] = \mathbb{E}^{\mathbb{Q}}[e^{\alpha X_t}]$ for a real constant α using the Radon–Nikodym derivative.

11.4 For a real constant θ and for $s < t$, show that the following holds:

$$\mathbb{E}[W_t e^{\theta W_t} | \mathcal{F}_s] = (W_s + \theta(t-s)) e^{\frac{1}{2}\theta^2(t-s)} e^{\theta W_s}.$$

11.5 For a real constant θ and for $s < t$, show that the following holds:

$$\mathbb{E}[W_t^2 e^{\theta W_t} | \mathcal{F}_s] = \{(t-s) + (W_s + \theta(t-s))^2\} e^{\frac{1}{2}\theta^2(t-s)} e^{\theta W_s}.$$

11.6 Let $\mathbb{Q}$ be the equivalent measure given in Definition 11.1, and let $X_t = W_t + \theta t$. Show that $X_t^2 - t$ is a $\mathbb{Q}$ -martingale without using the fact that X_t is a $\mathbb{Q}$ -Brownian motion.

11.7 (Infinite time horizon) Let W_t , $0 \leq t < \infty$, be a Brownian motion with respect to a probability measure $\mathbb{P}$. Take a constant $\theta \neq 0$. Let $\mathbb{Q}$ be a probability measure for which $X_t = W_t + \theta t$, $0 \leq t < \infty$, is a Brownian motion. Show that $\mathbb{Q}$ is not equivalent to $\mathbb{P}$.

11.8 A possible discrete version of Girsanov's theorem can be regarded as a coin tossing problem using a biased coin. Assume that the probability of showing heads when we flip a given coin is equal to p , $\frac{1}{2} < p < 1$. If the outcome of a toss is heads, then the game under consideration pays a player the amount $\$A$, and if the outcome is tails it pays $-\$B$. What is a condition for A and B to have a fair game?

Chapter 12

Stochastic Differential Equations



Let $\mathbf{W} = (W^1, \dots, W^m)$ be an m -dimensional Brownian motion, and let

$$\boldsymbol{\sigma} = (\sigma_{ij})_{1 \leq i \leq d, 1 \leq j \leq m} : [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^m$$

and

$$\boldsymbol{\mu} = (\mu^1, \dots, \mu^d) : [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$$

be continuous functions where $\boldsymbol{\sigma}$ is regarded as a $d \times m$ matrix.

Consider a stochastic differential equation (SDE) given by

$$X_t^i = x_0^i + \int_0^t \mu^i(s, \mathbf{X}_s) ds + \sum_{j=1}^m \int_0^t \sigma_{ij}(s, \mathbf{X}_s) dW_s^j,$$

or equivalently,

$$\mathbf{X}_t = \mathbf{x}_0 + \int_0^t \boldsymbol{\mu}(s, \mathbf{X}_s) ds + \int_0^t \boldsymbol{\sigma}(s, \mathbf{X}_s) d\mathbf{W}_s,$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are called coefficients. By a solution of the preceding SDE we mean a stochastic process $\mathbf{X}_t$ satisfying the SDE. We will show that there exists a unique solution. In the differential notation the given SDE is written as

$$dX_t^i = \mu^i(t, \mathbf{X}_t) dt + \sum_{j=1}^m \sigma_{ij}(t, \mathbf{X}_t) dW_t^j,$$

or

$$d\mathbf{X}_t = \boldsymbol{\mu}(t, \mathbf{X}_t) dt + \boldsymbol{\sigma}(t, \mathbf{X}_t) d\mathbf{W}_t. \quad (12.1)$$

Throughout the chapter we consider only the case that $d = 1$ and $m = 1$ for the sake of notational simplicity. The statements of facts and the corresponding proofs for multidimensional case are almost identical to those for the one-dimensional case.

Also presented in this chapter is how to construct sample paths of a Brownian bridge with given conditions specified at the beginning and the end of the time interval.

12.1 Strong Solutions

Definition 12.1 A function $g : \mathbb{R}^k \rightarrow \mathbb{R}^\ell$ is *Lipschitz continuous* if there exists a constant $0 < L < \infty$ such that

$$\|g(\mathbf{x}) - g(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|.$$

Definition 12.2 We are given a Brownian motion $\{W_t\}$ and a stochastic process $\{X_t\}$. If the integrals $\int_0^t \mu(s, X_s) ds$ and $\int_0^t \sigma(s, X_s) dW_s$ exist for every $t \geq 0$, and if

$$X_t = X_0 + \int_0^t \mu(s, X_s) ds + \int_0^t \sigma(s, X_s) dW_s,$$

then X_t is called a *strong solution* of the SDE given by (12.1).

Theorem 12.1 (Existence and uniqueness) *We are given a stochastic differential equation*

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t, \quad X_0 = x_0.$$

Suppose that μ and σ are Lipschitz continuous. Then there exists a unique solution X_t that is continuous and adapted.

Proof Recall that for a locally bounded predictable process Y_t , we have by Doob's L^2 -inequality

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s Y_u dW_u \right|^2 \right] \leq 4 \mathbb{E} \left[\int_0^t Y_u^2 du \right]$$

and by the Cauchy–Schwarz inequality

$$\sup_{0 \leq s \leq t} \left| \int_0^s Y_u du \right|^2 \leq t \int_0^t Y_u^2 du.$$

Start with $X_t^0 = x_0$, $t \geq 0$, and define X^n inductively for $n \geq 0$ by

$$X_t^{n+1} = x_0 + \int_0^t \mu(X_s^n) ds + \int_0^t \sigma(X_s^n) dW_s. \quad (12.2)$$

For $n \geq 1$ let

$$\mu^n(t) = \mathbb{E} \left[\sup_{0 \leq s \leq t} |X_s^n - X_s^{n-1}|^2 \right].$$

Fix $T \geq 1$. Then for $0 \leq t \leq T$ we have

$$\begin{aligned} \mu^1(t) &= \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s \mu(x_0) dr + \int_0^s \sigma(x_0) dW_r \right|^2 \right] \\ &\leq 2 \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s \mu(x_0) dr \right|^2 \right] + 2 \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s \sigma(x_0) dW_r \right|^2 \right] \\ &\leq 2t \mathbb{E} \left[\int_0^t |\mu(x_0)|^2 dr \right] + 8 \mathbb{E} \left[\int_0^t |\sigma(x_0)|^2 ds \right] \\ &\leq 2t^2 |\mu(x_0)|^2 + 8t |\sigma(x_0)|^2 \\ &\leq 10Tt (|\mu(x_0)|^2 + |\sigma(x_0)|^2), \end{aligned}$$

where the first inequality holds since $(a + b)^2 \leq 2a^2 + 2b^2$ and the second inequality is due to the Cauchy–Schwarz inequality and Doob's L^2 -inequality. For $n \geq 1$ we have

$$\begin{aligned} \mu^{n+1}(t) &\leq 2 \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s (\mu(X_r^n) - \mu(X_r^{n-1})) dr \right|^2 \right] \\ &\quad + 2 \mathbb{E} \left[\sup_{0 \leq s \leq t} \left| \int_0^s (\sigma(X_r^n) - \sigma(X_r^{n-1})) dr \right|^2 \right] \\ &\leq 10K^2T \int_0^t \mu^n(r) dr. \end{aligned}$$

By mathematical induction, we have

$$\mu^n(t) \leq C \frac{(10TKt)^n}{n!}$$

for $0 \leq t \leq T$, where

$$C = \frac{\mu(x_0)^2 + \sigma(x_0)^2}{K^2}.$$

Hence

$$\left\| \sum_{n=1}^{\infty} \sup_{0 \leq s \leq T} |X_s^n - X_s^{n-1}| \right\|_2 \leq \sum_{n=1}^{\infty} \sqrt{\mu^n(t)} < \infty,$$

where $\|\cdot\|_2$ denotes the L^2 -norm. Therefore there exists a continuous adapted process X_t such that

$$\lim_{n \rightarrow \infty} \sup_{0 \leq s \leq T} |X_s^n - X_s| = 0$$

with probability 1 and in L^2 . Thus we have

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\sup_{0 \leq s \leq T} \left| \int_0^s \mu(X_r^n) dr - \int_0^s \mu(X_r) dr \right|^2 \right] = 0 ,$$

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\sup_{0 \leq s \leq T} \left| \int_0^s \sigma(X_r^n) dW_r - \int_0^s \sigma(X_r) dW_r \right|^2 \right] = 0 .$$

Letting $n \rightarrow \infty$ in (12.2), we observe that X must satisfy the given SDE.

As for the uniqueness, suppose that X and Y are solutions of the given SDE. Define

$$g(t) = \mathbb{E} \left[\sup_{0 \leq s \leq t} |X_s - Y_s|^2 \right] .$$

Proceeding as before we deduce

$$g(t) \leq 10K^2T \int_0^t g(s) ds , \quad 0 \leq t \leq T$$

and conclude that $g = 0$ by Gronwall's inequality. (See Lemma D.1.) $\square$

Example 12.1 (*Langevin Equation*) Let α and σ be constants. The solution of the Langevin equation

$$dX_t = -\alpha X_t dt + \sigma dW_t , \quad X_0 = x_0 ,$$

is called the Ornstein–Uhlenbeck process. If $\sigma = 0$, then the SDE becomes an ordinary differential equation, and the solution is deterministic and given by $X_t = x_0 e^{-\alpha t}$. Hence we see that $X_t e^{\alpha t} = x_0$ is constant. For $\sigma \neq 0$ the process $Y_t = X_t e^{\alpha t}$ will move up and down about the constant x_0 . Since $dY_t = \sigma e^{\alpha t} dW_t$, we have a solution

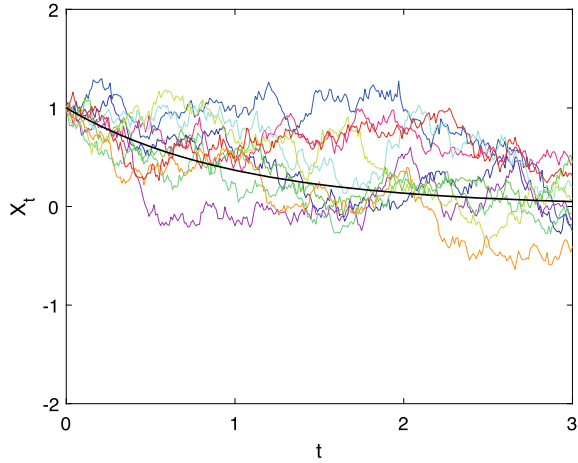
$$Y_t = Y_0 + \sigma \int_0^t e^{\alpha s} dW_s ,$$

and hence

$$X_t = e^{-\alpha t} x_0 + e^{-\alpha t} \sigma \int_0^t e^{\alpha s} dW_s . \quad (12.3)$$

For $\alpha > 0$, the mean and the variance are given by $\mathbb{E}[X_t] = e^{-\alpha t} x_0$ and

Fig. 12.1 Sample paths of the Ornstein–Uhlenbeck process



$$\begin{aligned}
 \text{Var}(X_t) &= \mathbb{E}[(X_t - e^{-at}x_0)^2] \\
 &= \mathbb{E}\left[\left(e^{-at}\sigma \int_0^t e^{as}dW_s\right)^2\right] \\
 &= e^{-2at}\sigma^2 \mathbb{E}\left[\int_0^t e^{2as}ds\right] \quad (\text{Itô isometry}) \\
 &= \sigma^2 \frac{1 - e^{-2at}}{2\alpha} .
 \end{aligned}$$

In Fig. 12.1 are plotted sample paths of the Ornstein–Uhlenbeck process with $\alpha = 1$, $\sigma = 0.5$ and $x_0 = 1$ together with its mean $e^{-at}x_0$ for $0 \leq t \leq 3$. See Simulation 12.2.

Theorem 12.2 *For a continuous stochastic process X_t of finite variation, suppose that $\mathcal{E}_t$ satisfies*

$$d\mathcal{E}_t = \mathcal{E}_t dX_t, \quad \mathcal{E}_0 = 1 .$$

Then the solution, called a stochastic exponential, is given by

$$\mathcal{E}_t = \exp\left(X_t - X_0 - \frac{1}{2}[X, X]_t\right) .$$

Proof Let $Y_t = X_t - X_0 - \frac{1}{2}[X, X]_t$, then

$$d(e^{Y_t}) = e^{Y_t}dY_t + \frac{1}{2}e^{Y_t}d[Y, Y]_t .$$

Since $[X, [X, X]]_t = 0$, we have $[Y, Y]_t = [X, X]_t$, and hence

$$d\mathcal{E}_t = e^{Y_t}dX_t - \frac{1}{2}e^{Y_t}d[X, X]_t + \frac{1}{2}e^{Y_t}d[X, X]_t = e^{Y_t}dX_t .$$

To prove the uniqueness, suppose that V_t is another solution, and show that $d(V_t/\mathcal{E}_t) = 0$. $\square$

Example 12.2 (i) If $X_t = W_t$ is a Brownian motion in Theorem 12.2, then

$$\mathcal{E}_t = e^{-\frac{1}{2}t + W_t}.$$

(ii) If $\mathcal{E}_t = S_t$ is the price of a stock and $X_t = R_t$ is the return on the investment in the stock defined by $dR_t = \frac{dS_t}{S_t}$, then $dS_t = S_t dR_t$, and hence

$$S_t = S_0 e^{R_t - R_0 - \frac{1}{2}[R, R]_t}.$$

If S_t is given by a geometric Brownian motion $dS_t = \mu S_t dt + \sigma S_t dW_t$, then $R_t = \mu t + \sigma W_t$ and

$$S_t = S_0^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t}.$$

(iii) If $\mathcal{E}_t = B_t$ is the price of a risk-free bond and $X_t = R_t$ the return on the investment in the bond defined by $dR_t = \frac{dB_t}{B_t}$, then $dB_t = B_t dR_t$ and R_t satisfies $dR_t = g(t) dt$ for some deterministic function $g(t)$. Hence $R(t) - R(0) = \int_0^t g(t) dt$ and $[R, R]_t = 0$. Thus

$$B_t = B_0 e^{R_t - R_0 - \frac{1}{2}[R, R]_t} = B_0 e^{\int_0^t g(t) dt}.$$

Remark 12.1 The Markov property means that given a present state of a process, the future is independent of the past. A solution X_t of the SDE in Theorem 12.1 has the Markov property. Consult Oksendal (2003).

12.2 Weak Solutions

Definition 12.3 Suppose that there exists a probability space with a filtration $\tilde{\mathcal{F}}_t$, a Brownian motion $\tilde{W}_t$ and a process $\tilde{X}_t$ adapted to $\tilde{\mathcal{F}}_t$ such that $\tilde{X}_0$ has the given distribution, and $\tilde{X}_t$ satisfies

$$\tilde{X}_t = \tilde{X}_0 + \int_0^t \mu(s, \tilde{X}_s) ds + \int_0^t \sigma(s, \tilde{X}_s) d\tilde{W}_s.$$

Then $\tilde{X}_t$ is called a *weak solution* of the SDE

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t.$$

By definition, a strong solution is a weak solution. In the following we consider an SDE with a discontinuous coefficient for which Theorem 12.1 is not applicable since the coefficient is not Lipschitz.

Example 12.3 Consider the stochastic differential equation $dX_t = \text{sign}(X_t) dW_t$ where

$$\text{sign}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}.$$

The function $\text{sign}(x)$ does not satisfy the Lipschitz condition, and it is known that the equation does not have a strong solution. However, the Brownian motion is the unique weak solution. For, if we take any Brownian motion $\tilde{W}_t$ for X_t , and define

$$Y_t = \int_0^t \text{sign}(\tilde{W}_s) d\tilde{W}_s = \int_0^t \text{sign}(X_s) dX_s ,$$

then $\text{sign}(\tilde{W}_s)$ is adapted,

$$\int_0^T \text{sign}(\tilde{W}_s)^2 ds = \int_0^T 1 ds = T < \infty ,$$

and Y_t is a continuous martingale such that

$$[Y, Y]_t = \int_0^t \text{sign}(\tilde{W}_s)^2 d[\tilde{W}, \tilde{W}]_s = \int_0^t 1 ds = t .$$

Then Y_t is a Brownian motion by Lévy's theorem. (See Theorem 7.9.) Finally, since

$$dY_t = \text{sign}(X_t) dX_t ,$$

multiplying both sides by $\text{sign}(X_t)$, we have

$$dX_t = \text{sign}(X_t) dY_t .$$

See also Example 10.4.4. For the proof of the uniqueness of the weak solution, consult Oksendal (2003).

12.3 Brownian Bridges

In Theorem 7.6 we showed how to construct sample paths of a Brownian motion constrained at $t = 1$ using successive approximations. In this section we present a constrained Brownian motion as a stochastic process.

Definition 12.4 A stochastic process $\{X_t\}_{t \geq 0}$ is called a *Gaussian process* if for $0 < t_1 < \dots < t_n$ any linear combination of $X_{t_1}, \dots, X_{t_n}$ is jointly normally distributed.

Example 12.4 A Brownian motion W_t is a Gaussian process. To see why, note that for $0 < t_1 < t_2 < \dots < t_n$, the increments $\delta W_1 = W_{t_1} - W_0, \dots, \delta W_n = W_{t_n} - W_{t_{n-1}}$ are independent and normal. Since $W_{t_1} = \delta W_1, \dots, W_{t_n} = \delta W_1 + \dots + \delta W_n$, the random variables $W_{t_1}, \dots, W_{t_n}$ are jointly normally distributed. (See Remark 4.12.) Recall that the covariance function is given by $c(s, t) = \min\{s, t\}$, $0 \leq s \leq t$.

Example 12.5 Let $f(t)$ be a deterministic function of time t , and define $X_t = \int_0^t f(s) dW_s$, where W_t is a Brownian motion. Then X_t is a Gaussian process. For the proof, recall that for a real constant θ the process

$$M_t = \exp \left(\theta X_t - \frac{1}{2} \theta^2 \int_0^t f(s)^2 ds \right)$$

is a martingale. Hence

$$1 = M_0 = \mathbb{E}[M_t] = \exp \left(-\frac{1}{2} \theta^2 \int_0^t f(s)^2 ds \right) \mathbb{E}[e^{\theta X_t}],$$

and hence the moment generating function is given by

$$\mathbb{E}[e^{\theta X_t}] = \exp \left(\frac{1}{2} \theta^2 \int_0^t f(s)^2 ds \right),$$

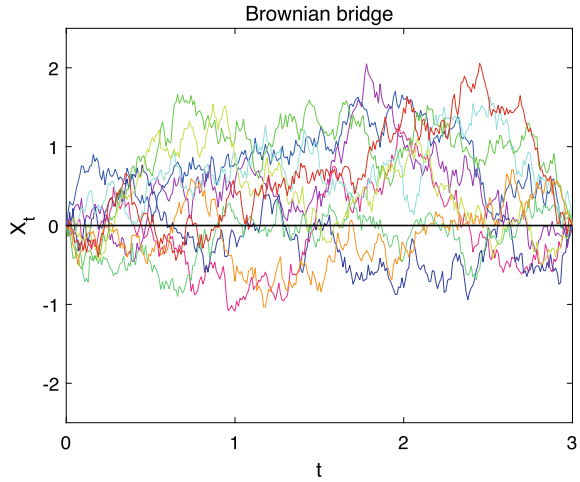
which is a moment generating function for a normal distribution with mean 0 and variance $\int_0^t f(s)^2 ds$. Thus X_t is normally distributed with the same mean and variance. Now it remains to prove that for $0 < t_1 < \dots < t_n$ the random variables $X_{t_1}, \dots, X_{t_n}$ are jointly normal. Note that the increments $\delta X_1 = X_{t_1} - X_0 = X_{t_1}, \dots, \delta X_n = X_{t_n} - X_{t_{n-1}}$ are normally distributed and independent. Since $X_{t_1} = \delta X_1, \dots, X_{t_n} = \delta X_1 + \dots + \delta X_n$, the random variables $X_{t_1}, \dots, X_{t_n}$ are jointly normally distributed. (See Remark 4.12.) For the details, consult Shreve (2004b).

Definition 12.5 (*Brownian Bridge*) Let W_t be a Brownian motion. For $T > 0$, the *Brownian bridge from 0 to 0* is defined by

$$X_t = W_t - \frac{t}{T} W_T, \quad 0 \leq t \leq T.$$

More generally, the *Brownian bridge from a to b* defined by

Fig. 12.2 Sample paths of the Brownian bridge $X_t = X_t^{0 \rightarrow 0}$ from $a = 0$ to $b = 0$ for $T = 3$



$$X_t^{a \rightarrow b} = a + \frac{t}{T}(b - a) + W_t - \frac{t}{T}W_T, \quad 0 \leq t \leq T.$$

See Simulation 12.3 and Fig. 12.2 for ten sample paths of the Brownian bridge from $a = 0$ to $b = 2$ for $T = 3$.

- Remark 12.2** (i) Note that $X_t = X_t^{0 \rightarrow 0}$ and that $X_0^{a \rightarrow b} = a$, $X_T^{a \rightarrow b} = b$.
(ii) Since we have to know the values of W_T to define X_t , $0 \leq t \leq T$, the Brownian bridge X_t is not adapted to the filtration generated by W_t .
(iii) For $0 < t_1 < \dots < t_n < T$, the random variables

$$X_{t_1} = W_{t_1} - \frac{t_1}{T}W_T, \dots, X_{t_n} = W_{t_n} - \frac{t_n}{T}W_T$$

are jointly normal since $W_{t_1}, \dots, W_{t_n}, W_T$ are jointly normal. Hence the Brownian bridge from 0 to 0 is a Gaussian process, and so is $X_t^{a \rightarrow b}$ since addition of a deterministic function to a Gaussian process X_t gives a new Gaussian process.

(iv) $\mathbb{E}[X_t] = \mathbb{E}[W_t - \frac{t}{T}W_T] = 0$ and $\mathbb{E}[X_t^{a \rightarrow b}] = a + \frac{t}{T}(b - a)$.

(v) For $0 < s \leq t < T$, the covariance function $c(s, t) = c(t, s)$ is given by

$$\begin{aligned} c(s, t) &= \mathbb{E} \left[\left(W_s - \frac{s}{T}W_T \right) \left(W_t - \frac{t}{T}W_T \right) \right] \\ &= \mathbb{E} \left[W_s W_t - \frac{s}{T}W_t W_T - \frac{t}{T}W_s W_T + \frac{st}{T^2}W_T^2 \right] \\ &= s - \frac{s}{T}t - \frac{t}{T}s + \frac{st}{T^2}T = \frac{s(T-t)}{T}. \end{aligned}$$

The covariance function of $X_t^{a \rightarrow b}$ is the same.

(vi) The Brownian bridge cannot be expressed as an Itô integral $\int_0^t f(s) dW_s$ of a

deterministic function $f(t)$ since

$$\text{Var}(X_t) = \mathbb{E}[X_t^2] = \frac{t(T-t)}{T},$$

which increases as t increases for $0 \leq t \leq \frac{T}{2}$ and then decreases to zero for $\frac{T}{2} \leq t \leq T$. However, the Itô isometry implies that the variance of the Itô integral is given by

$$\text{Var}\left(\int_0^t f(s) dW_s\right) = \mathbb{E}\left[\left(\int_0^t f(s) dW_s\right)^2\right] = \int_0^t f(s)^2 ds,$$

which increases monotonically for $0 \leq t \leq T$.

(vii) For an application of the Brownian bridge in simulating an at-the-money digital option, see Simulation 17.3.

Theorem 12.3 (Brownian Bridge) *Define a process Y_t , $0 \leq t \leq T$, by*

$$Y_t = (T-t) \int_0^t \frac{1}{T-s} dW_s, \quad 0 \leq t < T,$$

and $Y_T = 0$. Then the following facts hold:

- (i) Y_t is adapted to the filtration generated by the Brownian motion W_t .
- (ii) Y_t satisfies the stochastic differential equation

$$dY_t = -\frac{1}{T-t} Y_t dt + dW_t.$$

(iii) Y_t is a continuous Gaussian process on $[0, T]$ and has mean $m(t) = 0$ for every t and covariance

$$c(s, t) = \frac{s(T-t)}{T}, \quad 0 \leq s \leq t \leq T.$$

(iv) Y_t has the same distribution as the Brownian bridge X_t from 0 to 0.

Proof (i) Since Y_t is defined by W_s , $0 \leq s \leq t$, it is adapted to the filtration generated by W_s , $0 \leq s \leq t$.

(ii) Use the fact

$$dY_t = \left(\int_0^t \frac{1}{T-s} dW_s\right)(-dt) + (T-t) \frac{1}{T-t} dW_t.$$

(iii) For $0 \leq t < T$, the process

$$I_t = \int_0^t \frac{1}{T-s} dW_s$$

is a Gaussian process by the argument in Example 12.5. Since $I_{t_1}, \dots, I_{t_n}$ are jointly normal, the random variables $Y_{t_1} = (T - t_1)I_{t_1}, \dots, Y_{t_n} = (T - t_n)I_{t_n}$ are jointly normal. Hence Y_t is a Gaussian process for $0 \leq t < T$. The mean of Y_t is equal to 0 since the expectation of an Itô integral is 0. For $0 \leq s \leq t < T$, the Itô isometry implies that the covariance of Y is given by

$$\begin{aligned}
 c(s, t) &= \mathbb{E} \left[(T - s) \int_0^s \frac{1}{T - u} dW_u \times (T - t) \int_0^t \frac{1}{T - u} dW_u \right] \\
 &= (T - s)(T - t) \mathbb{E} \left[\int_0^t \mathbb{1}_{[0, s]}(u) \frac{1}{T - u} dW_u \times \int_0^t \frac{1}{T - u} dW_u \right] \\
 &= (T - s)(T - t) \mathbb{E} \left[\int_0^t \mathbb{1}_{[0, s]}(u) \frac{1}{(T - u)^2} du \right] \\
 &= (T - s)(T - t) \int_0^t \mathbb{1}_{[0, s]}(u) \frac{1}{(T - u)^2} du \\
 &= (T - s)(T - t) \int_0^s \frac{1}{(T - u)^2} du \\
 &= (T - s)(T - t) \left(\frac{1}{T - s} - \frac{1}{T} \right) = \frac{s(T - t)}{T}.
 \end{aligned}$$

The variance of Y_t , which is equal to $\frac{t(T-t)}{T}$, $0 \leq t < T$, converges to 0 as $t \uparrow T$. In combination with the fact that the mean of Y_t is always equal to 0, the definition $Y_T = 0$ enables Y_t to be continuous at $t = T$.

(iv) Since Y_t has the same mean and covariance functions $m(t)$ and $c(s, t)$ as the Brownian bridge X_t from 0 to 0 and since a Gaussian process is completely determined by $m(t)$ and $c(s, t)$, the process Y_t has the same distribution as X_t . $\square$

See Simulation 12.4 and Fig. 12.4 for ten sample paths of the process

$$Y_t = (T - t) \int_0^t \frac{1}{T - s} dW_s$$

with $a = 0$, $b = 2$ and $T = 3$, and see Simulation 12.5 and Fig. 12.5 for the corresponding SDE

$$dY_t = -\frac{1}{T - t} Y_t dt + dW_t.$$

Theorem 12.4 (Probability Density of Brownian Bridge) *Let X_t be the Brownian bridge from 0 to 0 for $0 \leq t \leq T$ in Definition 12.5. Take $0 = t_0 < t_1 < \dots < t_n < T$ and put $\delta t_i = t_i - t_{i-1}$. The joint density $f_{X_{t_1}, \dots, X_{t_n}}$ for $X_{t_1}, \dots, X_{t_n}$ is given by*

$$f_{X_{t_1}, \dots, X_{t_n}}(x_1, \dots, x_n) = \frac{p(T - t_n, x_n, 0)}{p(T, 0, 0)} \prod_{i=1}^n p(\delta t_i, x_{i-1}, x_i)$$

where $x_0 = 0$ and

$$p(\delta t, x, x') = \frac{1}{\sqrt{2\pi\delta t}} \exp\left(-\frac{(x' - x)^2}{2\delta t}\right)$$

is the transition density for Brownian motion. Note that $f_{X_{t_1}, \dots, X_{t_n}}$ is equal to the joint density of a Brownian motion W_t at $t_1, \dots, t_n$ conditional on $W_T = 0$, and hence X_t has the same probability distribution as W_t conditional on $W_T = 0$.

Proof Put $\tau_0 = T$ and $\tau_i = T - t_i$. Define

$$Y_i = \frac{1}{\tau_i} X_{t_i} - \frac{1}{\tau_{i-1}} X_{t_{i-1}}.$$

Since each of $Y_{t_1}, \dots, Y_{t_n}$ is a linear combination of $X_{t_1}, \dots, X_{t_n}$, which are jointly normal, we see that $Y_{t_1}, \dots, Y_{t_n}$ are also jointly normal. Then

$$\mathbb{E}[Y_i] = \frac{1}{\tau_i} \mathbb{E}[X_{t_i}] - \frac{1}{\tau_{i-1}} \mathbb{E}[X_{t_{i-1}}] = 0,$$

$$\begin{aligned} \text{Var}(Y_i) &= \mathbb{E}[Y_i^2] \\ &= \frac{1}{\tau_i^2} \text{Var}(X_{t_i}) - \frac{2}{\tau_i \tau_{i-1}} \text{Cov}(X_{t_i}, X_{t_{i-1}}) + \frac{1}{\tau_{i-1}^2} \text{Var}(X_{t_{i-1}}) \\ &= \frac{1}{\tau_i^2} \frac{t_i(T - t_i)}{T} - \frac{2}{\tau_i \tau_{i-1}} \frac{t_{i-1}(T - t_i)}{T} + \frac{1}{\tau_{i-1}^2} \frac{t_{i-1}(T - t_{i-1})}{T} \\ &= \frac{1}{\tau_i} \frac{t_i}{T} - \frac{2}{\tau_{i-1}} \frac{t_{i-1}}{T} + \frac{1}{\tau_{i-1}} \frac{t_{i-1}}{T} \\ &= \frac{(T - t_{i-1})t_i - t_{i-1}(T - t_i)}{\tau_{i-1} \tau_i T} \quad (\text{since } \tau_i = T - t_i) \\ &= \frac{\delta t_i}{\tau_{i-1} \tau_i}. \end{aligned}$$

For $i < j$, we have $t_{i-1} < t_i \leq t_{j-1} < t_j$ and

$$\begin{aligned} &\text{Cov}(Y_i, Y_j) \\ &= \text{Cov}\left(\frac{1}{\tau_i} X_{t_i} - \frac{1}{\tau_{i-1}} X_{t_{i-1}}, \frac{1}{\tau_j} X_{t_j} - \frac{1}{\tau_{j-1}} X_{t_{j-1}}\right) \\ &= \frac{1}{\tau_i \tau_j} c(t_i, t_j) - \frac{1}{\tau_i \tau_{j-1}} c(t_i, t_{j-1}) - \frac{1}{\tau_{i-1} \tau_j} c(t_{i-1}, t_j) + \frac{1}{\tau_{i-1} \tau_{j-1}} c(t_{i-1}, t_{j-1}) \\ &= \frac{t_i}{\tau_i T} - \frac{t_i}{\tau_i T} - \frac{t_{i-1}}{\tau_{i-1} T} + \frac{t_{i-1}}{\tau_{i-1} T} = 0. \end{aligned}$$

Hence the jointly normal variables $Y_{t_1}, \dots, Y_{t_n}$ are independent and their joint density is given by

$$f_{Y_{t_1}, \dots, Y_{t_n}}(y_1, \dots, y_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi \frac{\delta t_i}{\tau_{i-1}\tau_i}}} \exp\left(-\frac{1}{2 \frac{\delta t_i}{\tau_{i-1}\tau_i}} y_i^2\right).$$

Now we change the variables from $(y_1, \dots, y_n)$ to $(x_1, \dots, x_n)$ by the rule

$$y_i = \frac{x_i}{\tau_i} - \frac{x_{i-1}}{\tau_{i-1}}, \quad 1 \leq i \leq n,$$

where x_i represents the value taken by X_{t_i} , the Brownian bridge from 0 to 0 at time t_i . Since

$$\begin{aligned} \frac{1}{\frac{\delta t_i}{\tau_{i-1}\tau_i}} y_i^2 &= \frac{\tau_{i-1}\tau_i}{\delta t_i} \left(\frac{x_i^2}{\tau_i^2} - 2 \frac{x_i x_{i-1}}{\tau_{i-1}\tau_i} + \frac{x_{i-1}^2}{\tau_{i-1}^2} \right) \\ &= \frac{1}{\delta t_i} \left(\frac{\tau_{i-1}}{\tau_i} x_i^2 - 2x_i x_{i-1} + \frac{\tau_i}{\tau_{i-1}} x_{i-1}^2 \right) \\ &= \frac{1}{\delta t_i} \left(\left(1 + \frac{\delta t_i}{\tau_i}\right) x_i^2 - 2x_i x_{i-1} + \left(1 - \frac{\delta t_i}{\tau_{i-1}}\right) x_{i-1}^2 \right) \\ &= \frac{1}{\delta t_i} \left((x_i - x_{i-1})^2 + \frac{\delta t_i}{\tau_i} x_i^2 - \frac{\delta t_i}{\tau_{i-1}} x_{i-1}^2 \right) \\ &= \frac{1}{\delta t_i} (x_i - x_{i-1})^2 + \frac{1}{\tau_i} x_i^2 - \frac{1}{\tau_{i-1}} x_{i-1}^2, \end{aligned}$$

we have

$$\begin{aligned} \prod_{i=1}^n \exp\left(-\frac{1}{2 \frac{\delta t_i}{\tau_{i-1}\tau_i}} y_i^2\right) &= \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{1}{\frac{\delta t_i}{\tau_{i-1}\tau_i}} y_i^2\right) \\ &= \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{1}{\delta t_i} (x_i - x_{i-1})^2 - \frac{1}{2} \frac{1}{T - t_n} x_n^2\right). \end{aligned}$$

Also note that

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi \frac{\delta t_i}{\tau_{i-1}\tau_i}}} = \frac{\sqrt{T}}{\sqrt{T - t_n}} \prod_{i=1}^n \frac{1}{\sqrt{2\pi \delta t_i}} \prod_{i=1}^n \tau_i.$$

Since $\det \left[\frac{\partial y_i}{\partial x_j} \right] = \prod_{i=1}^n \frac{1}{\tau_i}$ and $p(T, 0, 0) = \sqrt{2\pi T}$, we have

$$\begin{aligned}
& f_{X_{t_1}, \dots, X_{t_n}}(x_1, \dots, x_n) \\
&= f_{Y_{t_1}, \dots, Y_{t_n}}(y_1, \dots, y_n) \prod_{i=1}^n \frac{1}{\tau_i} \\
&= \frac{\sqrt{T}}{\sqrt{T-t_n}} \exp\left(-\frac{1}{2} \frac{x_n^2}{T-t_n}\right) \prod_{i=1}^n \frac{1}{\sqrt{2\pi\delta t_i}} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{\delta t_i}\right) \\
&= \frac{\sqrt{2\pi T}}{\sqrt{2\pi(T-t_n)}} \exp\left(-\frac{1}{2} \frac{x_n^2}{T-t_n}\right) \prod_{i=1}^n \frac{1}{\sqrt{2\pi\delta t_i}} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{\delta t_i}\right) \\
&= \frac{p(T-t_n, x_n, 0)}{p(T, 0, 0)} \prod_{i=1}^n p(\delta t_i, x_{i-1}, x_i). \quad \square
\end{aligned}$$

Corollary 12.1 *The Brownian bridge $X_t^{a \rightarrow b}$ from a to b in Definition 12.5 has the same probability distribution as the Brownian motion conditional on $W_0 = a$ and $W_T = b$.*

Proof Let

$$X_t^{a \rightarrow b} = a + \frac{t}{T}(b-a) + W_t - \frac{t}{T}W_T$$

be the Brownian bridge from a to b in Definition 12.5 for some probability measure $\mathbb{P}$ and a $\mathbb{P}$ -Brownian motion W_t . Then

$$X_t^{0 \rightarrow b-a} = \frac{t}{T}(b-a) + W_t - \frac{t}{T}W_T$$

is the Brownian bridge from 0 to $b-a$. Recall that

$$X_t = X_t^{0 \rightarrow 0} = W_t - \frac{t}{T}W_T$$

is the Brownian bridge from 0 to 0, and has the same probability distribution as W_t for $0 \leq t \leq T$ conditional on $W_T = 0$. Let $\theta = \frac{b-a}{T}$. Define a new probability measure $\mathbb{Q}$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\frac{1}{2}\theta^2 T - \theta W_T}.$$

By Girsanov's theorem, $\tilde{W}_t = W_t + \theta t$, $0 \leq t \leq T$, is a $\mathbb{Q}$ -Brownian motion. Note that

$$X_t^{0 \rightarrow b-a} = X_t + \theta t$$

has the same probability distribution as $W_t + \theta t$ for $0 \leq t \leq T$ conditional on $W_T = 0$. Since $W_T = 0$ if and only if $\tilde{W}_T = b-a$, we note that $X_t^{0 \rightarrow b-a}$ has the

same probability distribution as a Brownian motion $\tilde{W}_t$ conditional on $\tilde{W}_T = b - a$. Therefore

$$X_t^{a \rightarrow b} = a + X_t^{0 \rightarrow b-a}$$

has the same probability distribution as a generalized Brownian motion $\widehat{W}_t = a + \tilde{W}_t$ conditional on $\widehat{W}_T = b$. $\square$

Remark 12.3 (i) In the proof of Corollary 12.1 we may directly derive the probability distribution of the Brownian bridge from a to b and show that it is equal to the probability distribution of the Brownian motion conditional on $W_0 = a$ and $W_T = b$ as done in Shreve (2004b). Our approach is to simplify the notational burden in doing so and to prove Theorem 12.4 first, because its proof is relatively simple.

(ii) Corollary 12.1 enables us to use Definition 12.5 to simulate the Brownian motion conditional on $W_T = b$ instead of constructing a Brownian motion using the successive approximations given in Theorem 7.6. See Fig. 15.5 for a simulation of Delta-hedging for an at-the-money European call option.

12.4 Computer Experiments

Simulation 12.1 (Numerical Solution of an SDE) We compare a numerical solution of the SDE given by

$$dX_t = 2W_t dW_t + dt$$

with the exact solution $X_t = W_t^2 + X_0$. See Fig. 12.3.

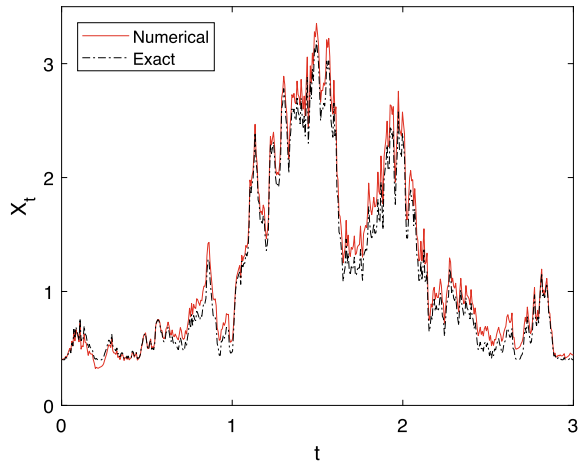
```
T= 3.0;
N = 500;
dt = T/N;
t = 0:dt:T;
dW = sqrt(dt)*randn(1,N);
W = zeros(1,N+1);

X = zeros(1,N+1);
X0 = 0.4;
X(1) = X0;

Exact = zeros(1,N+1);
Exact(1) = X0;

for i = 1:N
    W(i+1) = W(i) + dW(i);
    X(i+1) = X(i) + 2*W(i)*dW(i) + dt;
    Exact(i+1)=W(i+1)^2 + X0;
```

Fig. 12.3 A numerical solution of $dX_t = 2W_t dW_t + dt$ and the exact solution $X_t = W_t^2 + X_0$



```
end
plot(t,X,'r-',t,Exact,'k-.');
hlegend=legend('approx','exact');
```

Simulation 12.2 (Ornstein–Uhlenbeck Process) We generate M sample paths of the Ornstein–Uhlenbeck process

$$dX_t = -\alpha X_t dt + \sigma dW_t$$

together with the curve $\mathbb{E}[X_t] = e^{-\alpha t} x_0$, $0 \leq t \leq T$. See Fig. 12.1.

```
T= 3;
N = 200;
dt = T/N;
t = 0:dt:T;
M = 10;
X = zeros(M,N);
alpha = 1;
sigma = 0.5;
x0 = 1;
X(:,1) = x0 ;
dW = sqrt(dt)*randn(M,N);
for i = 1:N
    X(:,i+1) = X(:,i) - alpha*X(:,i)*dt + sigma*dW(:,i);
end
for j = 1:M
    plot(t,X(j,:));
```

```

        hold on
    end
    t = 0:0.01:T;
    plot(t,x0*exp(-alpha*t))

```

Simulation 12.3 (Brownian Bridge) We generate numerical realizations of the Brownian bridge

$$X_t = X_t^{0 \rightarrow 0} = W_t - \frac{t}{T} W_T$$

from $a = 0$ to $b = 0$ for $T = 3$ in Definition 12.5. We take $N = 100$ for the number of subintervals, $\delta t = \frac{T}{N}$, and $M = 10$ for the number of sample paths. See Fig. 12.2.

```

T= 3.0;
N = 100;
dt = T/N;
t = 0:dt:T;
M = 10;
W = zeros(M,N);
X = zeros(M,N);
a = 0;
b = 0;
X(:,1) = a;
dW = sqrt(dt)*randn(M,N);

for i = 1:N
    W(:,i+1) = W(:,i) + dW(:,i);
end

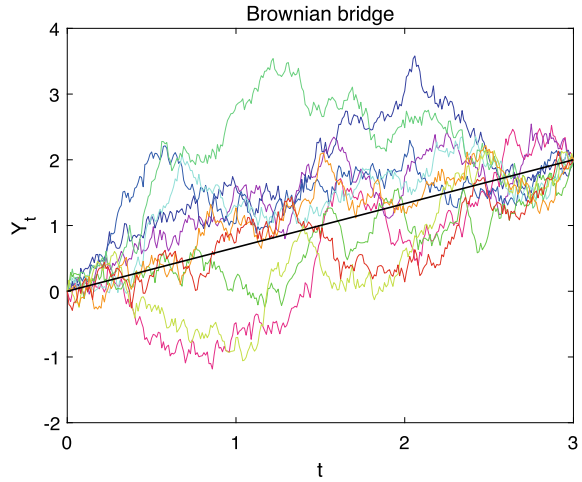
for i = 1:N+1
    X(:,i) = a*(1-(i-1)*dt)/T + b*(i-1)*dt/T + W(:,i) - (i-1)*dt/T*W(:,N+1);
end

for j = 1:M
    plot(t,X(j,:));
    hold on
end

```

Simulation 12.4 (Brownian Bridge Given by an Integral) We generate the sample paths of the process Y_t defined in Theorem 12.3. As in Simulation 12.3 we take $a = 0$, $b = 2$, $T = 3$, $N = 100$, $\delta t = \frac{T}{N}$ and $M = 10$. See Fig. 12.4.

Fig. 12.4 Sample paths of $Y_t = (T - t) \int_0^t \frac{1}{T-s} dW_s$ for $0 \leq t \leq 3$



```

Y = zeros(M,N+1);
Y(:,1) = a;
Integral = zeros(M,N+1);
Integral(:,1) = 0;
dW = sqrt(dt)*randn(M,N);

for i = 1:N
    Integral(:,i+1) = Integral(:,i) + (1/(T-(i-1)*dt))*dW(:,i);
end

for i = 1:N+1
    Y(:,i) = a*(1-(i-1)*dt/T) + b*(i-1)*dt/T + (T-(i-1)*dt)*Integral(:,i);
end

for j = 1:M
    plot(t,Y(j,:));
    hold on
end

```

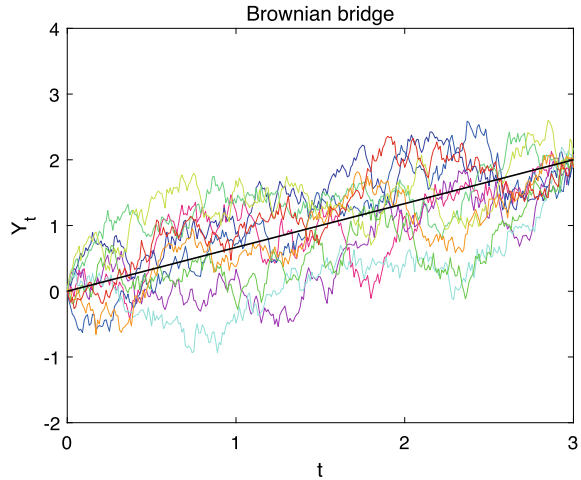
Simulation 12.5 (Brownian Bridge Given by an SDE) We generate the sample paths of the SDE given in Theorem 12.3. As in Simulations 12.3, 12.4 we take $a = 0, b = 2, T = 3, N = 100, \delta t = \frac{T}{N}$ and $M = 10$.

```

Y = zeros(M,N+1);
Y(:,1) = a;
dW = sqrt(dt)*randn(M,N);

```

Fig. 12.5 Brownian bridge:
sample paths of $dY_t =$
 $\frac{1}{T-t}(b - Y_t)dt + dW_t$,
 $Y_0 = a$ and $Y_T = b$, for
 $0 \leq t \leq T = 3$



```

for i = 1:N
    Y(:,i+1) = Y(:,i) + (b-Y(:,i))/(T-(i-1)*dt)*dt + dW(:,i);
    hold on
end

for j = 1:M
    plot(t,Y(j,:));
    hold on
end

```

See Fig. 12.5 for ten sample paths of the SDE $dY_t = \frac{1}{T-t}(b - Y_t)dt + dW_t$, $Y_0 = a$ with $a = 0$, $b = 2$ and $T = 3$. In the numerical realizations of the SDE we take $\delta t = \frac{T}{N} = \frac{3}{100}$, and due to the discretization error in the final time step, Y_T cannot be uniquely determined, but is distributed around $b = 2$. As N becomes larger, Y_T becomes more concentrated around b .

Exercises

12.1 What is a stochastic differential equation satisfied by $Y_t = W_t^3$?

12.2 Show that $dX_t = 3X_t^{1/3}dt + 3X_t^{2/3}dW_t$, $X_0 = 1$, has infinitely many solutions.

12.3 Show that the solution of $dX_t = dt + 2\sqrt{X_t}dW_t$ satisfies $X_t = (W_t + \sqrt{X_0})^2$.

12.4 Show that the solution of

$$dX_t = -X_t(2 \log X_t + 1) dt - 2X_t \sqrt{-\log X_t} dW_t$$

is given by $X_t = \exp \left(- (W_t + \sqrt{-\log X_0})^2 \right)$.

12.5 Show that the solution of the generalized Langevin equation or the mean-reverting Ornstein–Uhlenbeck process $dX_t = \alpha(m - X_t)dt + \sigma dW_t$ is given by

$$X_t = m + e^{-\alpha t}(X_0 - m) + \sigma \int_0^t e^{-\alpha(t-s)} dW_s .$$

(Hint: Take $Y_t = X_t - m$ and apply the result in Example 12.1.)

12.6 Find the solution of the SDE

$$dX_t = (\alpha_t + \beta_t X_t)dt + (\gamma_t + \delta_t X_t)dW_t ,$$

where $\alpha_t, \beta_t, \gamma_t, \delta_t$ are adapted continuous processes. (Hint: Consider first the case that $\alpha_t = \gamma_t = 0$.)

12.7 Solve the SDE $dX_t = X_t^3 dt - X_t^2 dW_t$, where $X_0 = 1$.

12.8 Suppose that X_t is the Ornstein–Uhlenbeck process $dX_t = -\alpha X_t dt + \sigma dW_t$ given in (12.3). Does $Y_t = X_t^2$ satisfy $dY_t = (\sigma^2 - 2\alpha Y_t) dt + 2\sigma \sqrt{Y_t} dW_t$?

12.9 Solve the SDE

$$dX_t = \frac{3}{4} X_t^2 dt - X_t^{3/2} dW_t , \quad X_0 > 0 .$$

Chapter 13

The Feynman–Kac Theorem



In this chapter we present the Feynman–Kac Theorem, which plays the role of a bridge between partial differential equations and stochastic processes. Historically, as an alternative method for the derivation of the Schrödinger partial differential equation in quantum mechanics, the path integral approach was introduced in the 1960s by the American physicist Richard Feynman (1918–1988). Along a similar line a certain type of partial differential equation can be solved using the expectation over the sample paths of a stochastic process. Its mathematical formulation, the Feynman–Kac Theorem, provides a link between two option pricing methods, one based on the Black–Scholes–Merton partial differential equation and the other on the martingale method.

13.1 The Feynman–Kac Theorem

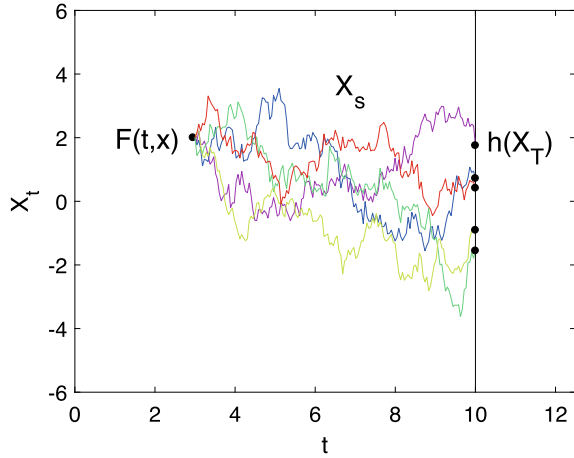
Let $F(t, x)$ be a function of t and x representing time and space, respectively. Given a stochastic process X_t , when is $F(t, X_t)$ a martingale? The Martingale Representation Theorem implies that $dF(t, X_t)$ should have no drift term. If that is the case, we have $F(t, X_t) = \mathbb{E}[F(T, X_T)|\mathcal{F}_t]$ for $0 \leq t \leq T$ by the definition of a martingale, and hence $F(t, x) = \mathbb{E}[F(T, X_T)|X_t = x]$. The following theorem shows how to find a solution of a differential equation using a probabilistic idea.

Theorem 13.1 (Feynman–Kac) *Let $F, \mu, \sigma : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ be functions of t and x , and let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a function of x . Consider a partial differential equation*

$$\left\{ \begin{array}{l} \frac{\partial F}{\partial t}(t, x) + \mu(t, x) \frac{\partial F}{\partial x}(t, x) + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2 F}{\partial x^2}(t, x) = 0, \quad 0 < t < T, \\ F(T, x) = h(x). \end{array} \right.$$

If a stochastic process X_t satisfies a stochastic differential equation

Fig. 13.1 The Feynman–Kac Theorem: Expectation of the final value $h(X_T)$ is the present value $F(t, x)$



$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t,$$

then we have

$$F(t, x) = \mathbb{E}[h(X_T)|\mathcal{F}_t] \Big|_{X_t=x} = \mathbb{E}[h(X_T)|X_t = x].$$

See Fig. 13.1 where we take $(t, x) = (3, 2)$ and $T = 10$.

Proof By the Itô formula we have

$$\begin{aligned} & F(T, X_T) - F(t, X_t) \\ &= \int_t^T \left\{ \frac{\partial F}{\partial s}(s, X_s) + \mu(s, X_s) \frac{\partial F}{\partial x}(s, X_s) + \frac{1}{2} \sigma^2(s, X_s) \frac{\partial^2 F}{\partial x^2}(s, X_s) \right\} ds \\ & \quad + \int_t^T \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) dW_s \\ &= \int_t^T \sigma(s, X_s) \frac{\partial F}{\partial x}(s, X_s) dW_s. \end{aligned}$$

Taking the conditional expectation $\mathbb{E}[\cdot | \mathcal{F}_t]$ on both sides, we obtain

$$\mathbb{E}[F(T, X_T)|\mathcal{F}_t] - F(t, X_t) = 0.$$

Thus $F(t, X_t) = \mathbb{E}[F(T, X_T)|\mathcal{F}_t] = \mathbb{E}[h(X_T)|\mathcal{F}_t]$. □

Example 13.1 If $\mu = \sigma = 0$, then X_t is constant. In this case the conclusion of the theorem implies that F depends only on x and $F(t, x) = h(x)$.

Example 13.2 If $\sigma = 0$, then X_t is a solution of an ordinary differential equation $\frac{dX}{dt} = \mu(t, X)$, and may be regarded as a deterministic function $\phi(t)$. Then

$$\begin{aligned} \frac{d}{dt} F(t, \phi(t)) &= \frac{\partial F}{\partial t}(t, \phi(t)) + \frac{\partial F}{\partial x}(t, \phi(t)) \phi'(t) \\ &= \frac{\partial F}{\partial t}(t, \phi(t)) + \frac{\partial F}{\partial x}(t, \phi(t)) \mu(t, \phi(t)) = 0 \end{aligned}$$

by the chain rule. Hence F is constant along the curve $t \mapsto (t, \phi(t))$, and

$$F(t, X_t) = F(t, \phi(t)) = F(T, \phi(T)) = F(T, X_T) = h(X_T)$$

for all t , i.e., $F(t, \phi(t)) = F(T, \phi(T))$.

Example 13.3 (Brownian Motion) For the case when $\mu = 0$ and $\sigma = 1$, we have the simplest stochastic differential equation $dX_t = dW_t$. Then

$$X_t = X_0 + \int_0^t dW_u = X_0 + W_t .$$

Thus, if $F(t, x)$ satisfies

$$\frac{\partial F}{\partial t} + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} = 0 , \quad F(T, x) = h(x) ,$$

then

$$\begin{aligned} F(t, x) &= \mathbb{E}[h(X_T) | \mathcal{F}_t] \Big|_{X_t=x} \\ &= \mathbb{E}[h(W_T + X_0) | \mathcal{F}_t] \Big|_{W_t+X_0=x} \\ &= \mathbb{E}[h(W_T - W_t + x) | \mathcal{F}_t] \Big|_{W_t=x-X_0} \\ &= \mathbb{E}[h(W_T - W_t + x)] \quad (\text{by the independence of } W_T - W_t \text{ and } \mathcal{F}_t) \\ &= \mathbb{E}[h(W_{T-t} + x)] \quad (W_T - W_t \text{ and } W_{T-t} \text{ have the same distribution}) \\ &= \int_{-\infty}^{\infty} h(z+x) \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{z^2}{2(T-t)}\right) dz \\ &= \int_{-\infty}^{\infty} h(y) \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{(y-x)^2}{2(T-t)}\right) dy . \end{aligned}$$

The last integral is the average of $h(y)$, $-\infty < y < \infty$, with respect to the normal distribution with mean x and variance $T - t$. Since the problem has the final condition at T , the formula contains the time to expiry $T - t$ rather than t which appears in the initial condition problem.

Example 13.4 (*Brownian Motion with Drift*) Consider $dX_t = \theta dt + dW_t$, where θ is a constant. Thus, if $F(t, x)$ satisfies

$$\frac{\partial F}{\partial t} + \theta \frac{\partial F}{\partial x} + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} = 0, \quad F(T, x) = h(x).$$

Assume that $F(t, x)$ is of the form $F(t, x) = e^{\alpha t + \beta x} G(t, x)$ for some constants α, β and a function $G(t, x)$. Then $F_t = \alpha e^{\alpha t + \beta x} G + e^{\alpha t + \beta x} G_t$, $F_x = \beta e^{\alpha t + \beta x} G + e^{\alpha t + \beta x} G_x$, and $F_{xx} = \beta^2 e^{\alpha t + \beta x} G + \beta e^{\alpha t + \beta x} G_x + \beta e^{\alpha t + \beta x} G_x + e^{\alpha t + \beta x} G_{xx}$. Hence

$$\begin{aligned} 0 &= F_t + \theta F_x + \frac{1}{2} F_{xx} \\ &= e^{\alpha t + \beta x} \left[\alpha G + G_t + \theta(\beta G + G_x) + \frac{1}{2} \{ \beta^2 G + 2\beta G_x + G_{xx} \} \right]. \end{aligned}$$

After dividing by $e^{\alpha t + \beta x}$, we have

$$0 = (\alpha + \beta(\theta + \frac{1}{2}\beta))G + G_t + (\theta + \beta)G_x + \frac{1}{2}G_{xx}.$$

Now we choose $\beta = -\theta$ and $\alpha = -\beta(\theta + \frac{1}{2}\beta) = \frac{1}{2}\theta^2$. Then

$$0 = G_t + \frac{1}{2}G_{xx}.$$

Note that $G(T, x) = e^{-\alpha T - \beta x} h(x) = e^{-\frac{1}{2}\theta^2 T + \theta x} h(x)$. By the result in Example 13.3, we have

$$G(t, x) = \int_{-\infty}^{\infty} e^{-\frac{1}{2}\theta^2 T + \theta y} h(y) \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{(x-y)^2}{2(T-t)}\right) dy.$$

Thus

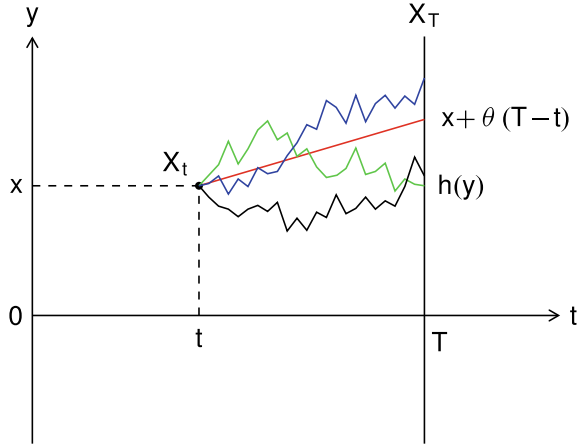
$$\begin{aligned} F(t, x) &= e^{\frac{1}{2}\theta^2 t - \theta x} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\theta^2 T + \theta y} h(y) \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{(x-y)^2}{2(T-t)}\right) dy \\ &= \int_{-\infty}^{\infty} h(y) \frac{1}{\sqrt{2\pi(T-t)}} \exp\left(-\frac{(y-x-\theta(T-t))^2}{2(T-t)}\right) dy. \end{aligned}$$

The last integral is the average of h with respect to the normal distribution with mean $x + \theta(T-t)$ and variance $T-t$. See Fig. 13.2.

Example 13.5 (*Geometric Brownian Motion*) Consider

$$dX_t = \mu_0 X_t dt + \sigma_0 X_t dW_t$$

Fig. 13.2 The Feynman–Kac theorem for a Brownian motion with drift given $X_t = x$



for which $\mu(t, x) = \mu_0 x$ and $\sigma(t, x) = \sigma_0 x$. Then

$$X_T = X_t \exp \left(\left(\mu_0 - \frac{1}{2} \sigma_0^2 \right) (T - t) + \sigma_0 W_{T-t} \right).$$

Consider

$$F_t + \mu_0 x F_x + \frac{1}{2} \sigma_0^2 x^2 F_{xx} = 0.$$

Then the Feynman–Kac Theorem implies

$$F(t, x) = \mathbb{E}[F(T, X_T) | X_t = x].$$

Hence

$$F(t, x) = \mathbb{E} \left[F(T, x \exp \left(\left(\mu_0 - \frac{1}{2} \sigma_0^2 \right) (T - t) + \sigma_0 W_{T-t} \right)) \right].$$

Definition 13.1 (*Infinitesimal Generator*) For a stochastic differential equation

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t \quad (13.1)$$

we define an *infinitesimal generator* by

$$\mathcal{A} = \mu(s, x) \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2(s, x) \frac{\partial^2}{\partial x^2}.$$

13.2 Application to the Black–Scholes–Merton Equation

Corollary 13.1 (Discounted Feynman–Kac Theorem) Let $V, \mu, \sigma : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ be functions of t and x , and let $h : \mathbb{R} \rightarrow \mathbb{R}$ be a function of x . Consider a partial differential equation

$$\left\{ \begin{array}{l} \frac{\partial V}{\partial t}(t, x) + \mu(t, x) \frac{\partial V}{\partial x}(t, x) + \frac{1}{2} \sigma^2(t, x) \frac{\partial^2 V}{\partial x^2}(t, x) = rV(t, x), \quad 0 < t < T, \\ V(T, x) = h(x). \end{array} \right.$$

If a stochastic process X_t satisfies a stochastic differential equation

$$dX_t = \mu(t, X_t) dt + \sigma(t, X_t) dW_t,$$

then we have

$$V(t, x) = \mathbb{E}[e^{-r(T-t)} h(X_T) | \mathcal{F}_t] \Big|_{X_t=x} = \mathbb{E}[e^{-r(T-t)} h(X_T) | X_t = x].$$

Proof Let

$$F(t, x) = e^{-rt} V(t, x).$$

Then $F_t = -re^{-rt} V + e^{-rt} V_t = -rF + e^{-rt} V_t$, $F_x = e^{-rt} V_x$, $F_{xx} = e^{-rt} V_{xx}$ and $F(T, x) = e^{-rT} h(X_T)$. Hence F satisfies the partial differential equation given in Theorem 13.1, and

$$e^{-rt} V(t, x) = \mathbb{E}[h(X_T) | \mathcal{F}_t] \Big|_{X_t=x},$$

which implies that

$$F(t, x) = \mathbb{E}[e^{-rT} h(X_T) | \mathcal{F}_t] \Big|_{X_t=x},$$

and the proof is complete. $\square$

Example 13.6 (Option Pricing) For a stochastic process X_t in Corollary 13.1 take a geometric Brownian motion S_t given by

$$dS_t = rS_t dt + \sigma_0 S_t d\tilde{W}_t$$

for some Brownian motion $\tilde{W}_t$, where r and σ_0 are constants, i.e., $\mu(t, x) = rx$ and $\sigma(t, x) = \sigma_0 x$. If V denotes the price of a European option, then it will be shown that the Black–Scholes–Merton partial differential equation holds:

$$\frac{\partial V}{\partial t}(t, x) + rx \frac{\partial V}{\partial x}(t, x) + \frac{1}{2} \sigma_0^2 x^2 \frac{\partial^2 V}{\partial x^2}(t, x) = rV(t, x).$$

(Consult Sect. 15.1 for the derivation.) Then Corollary 13.1 implies that

$$V(t, x) = e^{-r(T-t)} \mathbb{E}[h(S_T) | S_t = x].$$

Consult Sect. 16.1.3. See also Sect. 16.4.

13.3 The Kolmogorov Equations

In this section we compute the density function of the transition probability of Brownian motion.

13.3.1 The Kolmogorov Backward Equation

Theorem 13.2 (Kolmogorov Backward Equation) *If X_t satisfies (13.1), then for a Borel subset $B \subset \mathbb{R}^1$ the transition probability*

$$P(s, x; t, B) = \mathbb{P}(X_t \in B | X_s = x)$$

satisfies the Kolmogorov backward equation given by

$$\left(\frac{\partial}{\partial s} + \mathcal{A} \right) P(s, x; t, B) = 0, \quad 0 < s < t,$$

$$P(t, x; t, B) = \mathbb{1}_B(x).$$

In other words,

$$\frac{\partial}{\partial s} P = -\mathcal{A}P$$

and

$$P = e^{-t\mathcal{A}} P_0.$$

Proof In Theorem 13.1 use

$$\mathbb{P}(X_t \in B | X_s = x) = \mathbb{E}[\mathbb{1}_B(X_t) | X_s = x].$$

Consult Fig. 13.3, where B is represented by a vertical segment. □

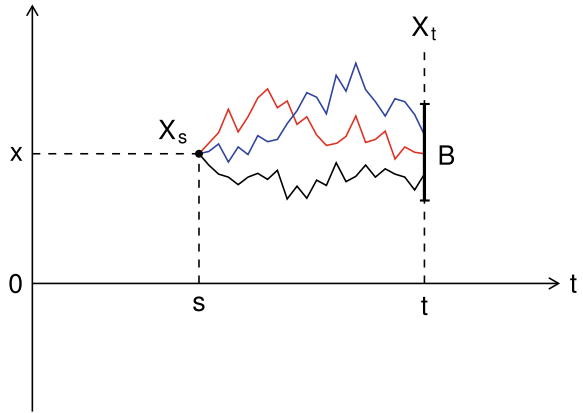
Corollary 13.2 *For an infinitesimal case, we choose an infinitesimally short interval $B = [y, y + dy]$. Suppose that the probability density is given by*

$$P(s, x; t, [y, y + dy]) = p(s, x; t, y) dy.$$

Then p is a solution of the differential equation

$$\left(\frac{\partial}{\partial s} + \mathcal{A} \right) p(s, x; t, y) = 0, \quad 0 < s < t,$$

Fig. 13.3 Derivation of the Kolmogorov backward equation



and $p(s, x; t, y)$ converges to the Dirac delta measure $\delta_x(y)$ as $s \rightarrow t$ in a suitable sense.

Proof Since

$$p(s, x; t, y) dy = P(s, x; t, [y, y + dy]) \approx \mathbb{1}_{[y, y+dy]}(x)$$

as $s \rightarrow t$, we use the approximation

$$p(s, x; t, y) \approx \frac{1}{dy} \mathbb{1}_{[y, y+dy]}(x) \approx \delta_y(x) .$$

□

Example 13.7 Consider the Ornstein–Uhlenbeck process

$$dX_t = -\alpha X_t dt + \sigma dW_t .$$

Take $B = (-\infty, y]$ and let

$$F(s, x; t, y) = P(s, x; t, B) = \mathbb{P}(X_t \in B | X_s = x) .$$

By the Kolmogorov backward equation, we have

$$\frac{\partial F}{\partial s} - \alpha x \frac{\partial F}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 F}{\partial x^2} = 0 , \quad 0 < s < t , \quad (13.2)$$

with the final condition

$$F(t, x; t, y) = \mathbb{1}_{(-\infty, y]}(x) .$$

Define

$$p(s, x; t, y) = \frac{\partial}{\partial y} F(s, x; t, y) .$$

By differentiating the left-hand side of (13.2) with respect to y , we obtain

$$\frac{\partial p}{\partial s} - \alpha x \frac{\partial p}{\partial x} + \frac{1}{2} \sigma^2 \frac{\partial^2 p}{\partial x^2} = 0 , \quad 0 < s < t , \quad (13.3)$$

with the final condition

$$p(t, x; t, y) = \delta_y(x) ,$$

where δ_y denotes the Dirac delta measure at y . From (13.3), we obtain

$$p(s, y; t, x) = \frac{1}{\sqrt{\pi(1 - e^{-2\alpha(t-s)}) \frac{\sigma^2}{\alpha}}} \exp\left(-\frac{(x - e^{-\alpha(t-s)}y)^2}{(1 - e^{-2\alpha(t-s)}) \frac{\sigma^2}{\alpha}}\right) .$$

13.3.2 The Kolmogorov Forward Equation

Now we derive the Kolmogorov forward equation, which is also called the Fokker–Planck equation.

Definition 13.2 (*Adjoint Operator*) For an infinitesimal generator

$$\mathcal{A} = \mu(s, x) \frac{\partial}{\partial x} + \frac{1}{2} \sigma^2(s, x) \frac{\partial^2}{\partial x^2}$$

its *adjoint operator* $\mathcal{A}^*$ is defined by

$$(\mathcal{A}^* f)(t, x) = -\frac{\partial}{\partial x} (\mu(t, x) f(t, x)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (\sigma^2(t, x) f(t, x)) .$$

Theorem 13.3 (Kolmogorov Forward Equation) *Let $p(s, x; t, y)$ be the transition probability density of the Itô process X_t introduced previously. Then p satisfies the Kolmogorov forward equation given by*

$$\frac{\partial}{\partial t} p(s, x; t, y) = \mathcal{A}^* p(s, x; t, y) , \quad 0 < t < T$$

$$p(s, x; t, y) \rightarrow \delta_x(y) , \quad s \rightarrow t .$$

Proof For any function $\phi(t, x)$ that is sufficiently smooth and compactly supported in the domain $(s, T) \times \mathbb{R}^1$ we have

$$\phi(T, X_T) = \phi(s, X_s) + \int_s^T \left(\frac{\partial \phi}{\partial t} + \mathcal{A} \phi \right) (t, X_t) dt + \sigma \int_s^T \frac{\partial \phi}{\partial x} (t, X_t) dW_t$$

by the Itô formula. Here, since $t = T$ and $t = s$ are a part of the boundary of the domain $(s, T) \times \mathbb{R}^1$, we have $\phi(T, X_T) = \phi(s, X_s) = 0$. Furthermore, if we take the conditional expectation $\mathbb{E}[\cdot | X_s = x]$, then the last integral on the right-hand side is zero. Hence

$$\mathbb{E} \left[\int_s^T \left(\frac{\partial \phi}{\partial t} + \mathcal{A}\phi \right) (t, X_t) dt \middle| X_s = x \right] = 0 ,$$

and by the definition of a transition probability density $p(s, x; t, y)$ we obtain

$$\int_{-\infty}^{\infty} p(s, x; t, y) \int_s^T \left(\frac{\partial \phi}{\partial t} + \mathcal{A}\phi \right) (t, y) dt dy = 0 .$$

Now by integration by parts with respect to t and y we obtain

$$\int_{-\infty}^{\infty} \int_s^T \phi(t, y) \left(-\frac{\partial}{\partial t} + \mathcal{A}^* \right) p(s, x; t, y) dt dy = 0 .$$

Since the equation holds for arbitrary ϕ we conclude that

$$\left(-\frac{\partial}{\partial t} + \mathcal{A}^* \right) p(s, x; t, y) = 0 .$$

Since $s > 0$ is arbitrary, the equation holds for $0 < t < T$. □

Example 13.8 Consider the Ornstein–Uhlenbeck process

$$dX_t = -\alpha X_t dt + \sigma dW_t .$$

From the Kolmogorov forward equation

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(-\alpha x p) + \frac{\partial^2}{\partial x^2} \left(\frac{1}{2} \sigma^2 p \right) , \quad (13.4)$$

we obtain

$$p(s, y; t, x) = \frac{1}{\sqrt{\pi(1 - e^{-2\alpha(t-s)})\frac{\sigma^2}{\alpha}}} \exp \left(-\frac{(x - e^{-\alpha(t-s)}y)^2}{(1 - e^{-2\alpha(t-s)})\frac{\sigma^2}{\alpha}} \right) .$$

This can be obtained directly from the closed form solution of the given SDE. For $0 \leq s \leq t$, we have

$$\begin{aligned}
X_t &= e^{-\alpha t} x_0 + e^{-\alpha t} \sigma \int_0^t e^{au} dW_u \\
&= e^{-\alpha t} x_0 + e^{-\alpha t} \sigma \left(\int_0^s e^{au} dW_u + \int_s^t e^{au} dW_u \right) \\
&= e^{-\alpha(t-s)} \left(e^{-\alpha s} x_0 + e^{-\alpha s} \sigma \int_0^s e^{au} dW_u \right) + e^{-\alpha t} \sigma \int_s^t e^{au} dW_u \\
&= e^{-\alpha(t-s)} X_s + e^{-\alpha t} \sigma \int_s^t e^{au} dW_u .
\end{aligned}$$

Let $y = X_s$. Then

$$\mathbb{E}[X_t | X_s = y] = e^{-\alpha(t-s)} y$$

$$\text{Var}(X_t | X_s = y) = e^{-2\alpha t} \sigma^2 \int_s^t e^{2au} du = \sigma^2 \frac{1 - e^{-2\alpha(t-s)}}{2\alpha}$$

by the Itô isometry. Since X_t is Gaussian, we have

$$p(s, y; t, x) = \frac{1}{\sqrt{2\pi(1 - e^{-2\alpha(t-s)})\frac{\sigma^2}{2\alpha}}} \exp\left(-\frac{(x - e^{-\alpha(t-s)}y)^2}{2(1 - e^{-2\alpha(t-s)})\frac{\sigma^2}{2\alpha}}\right).$$

By letting $t \rightarrow \infty$, we can obtain the limiting probability density

$$f(x) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2\alpha}}} \exp\left(-\frac{x^2}{2\frac{\sigma^2}{2\alpha}}\right). \quad (13.5)$$

If we want to find only the limiting probability density $f(x)$ directly, then we take

$\frac{\partial p}{\partial t} = 0$ in (13.4), and solve

$$0 = -\frac{\partial}{\partial x}(-\alpha x f(x)) + \frac{\partial^2}{\partial x^2} \left(\frac{1}{2} \sigma^2 f(x) \right).$$

Hence

$$C = \alpha x f(x) + \frac{\partial}{\partial x} \left(\frac{1}{2} \sigma^2 f(x) \right)$$

for some constant C . If we look for a probability density $f(x)$ satisfying

$$\lim_{x \rightarrow \pm\infty} xf(x) = 0$$

then $C = 0$. Thus we obtain (13.5).

13.4 Computer Experiments

Simulation 13.1 (Feynman–Kac Theorem) We produce sample paths of the Ornstein–Uhlenbeck process $dX_t = -\alpha X_t dt + \sigma dW_t$ in Fig. 13.1.

```

t0 = 3;          % initial time
X0 = 2 ;         % initial space point
T = 10;          % terminal time
N = 150;         % number of time steps from t0 to T
dt = (T-t0)/N;   % length of time interval
t = t0:dt:T;
M = 5;           % number of sample paths
alpha = 0.5;
sigma = 1.5;

X = zeros(M,N);
X(:,1) = X0;
dW = sqrt(dt)*randn(M,N);
for i = 1:N
    X(:,i+1) = X(:,i) - alpha*X(:,i)*dt + sigma*dW(:,i);
end

for j = 1:M
    plot(t(:),X(j,:), 'color', hsv2rgb([1-j/M 1 1]));
    plot(T,X(j,N+1), '. ' ) ;
    hold on
end

```

Now we choose $h(x) = \mathbb{1}_B$ where $B = [a, b]$, and evaluate $\mathbb{E}[h(X_T)]$.

```

a = 1;
b = 3;
X = zeros(M,N);
X(:,1) = X0;
dW = sqrt(dt)*randn(M,N);
for i = 1:N

```

```

X(:,i+1) = X(:,i) - alpha*X(:,i)*dt + sigma*dW(:,i);
end
mean(heaviside(X(:,N)-a).*heaviside(b-X(:,N)))

```

The command `heaviside` is the Heaviside function $H(x) = \mathbb{1}_{[0,\infty)}(x)$, and we have $H(x-a)H(b-x) = \mathbb{1}_{[a,b]}(x)$. Now we obtain an estimate for $F(t, x)$.

```
ans = 0.2409
```


Part V

Option Pricing Methods

Chapter 14

The Binomial Tree Method for Option Pricing



In this chapter we introduce the binomial tree method for option pricing, which is much simpler than the partial differential equation approach. Later, it will be shown that the solution from the binomial tree method converges to the price obtained from the Black–Scholes–Merton formula under certain conditions. Historically, not long after the partial differential equation approach was developed for option pricing by Black and Scholes (1973) and Merton (1973) in 1973, the binomial tree method was introduced by Cox et al. (1979) in 1979, which is a discrete time model and much easier to understand and implement in practice. The method is quite flexible and allows us to use parameters which may depend on time and asset prices. In this chapter only the binomial tree method is introduced, and the Black–Scholes–Merton differential equation will be presented in Chap. 15. For rigorous presentations of the binomial tree method, see Bjork (2009), Gianin and Sgarra (2013), Shreve (2004a). For historical accounts of the development of the binomial tree method, see Davis and Etheridge (2006), Harrison and Kreps (1979), Harrison and Pliska (1981).

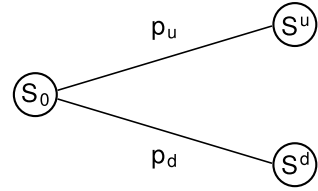
14.1 Motivation for the Binomial Tree Method

Consider a European call option on an underlying stock with its present price $S_0 = \$50$ per share. Suppose that at the expiry date T the stock has only two values $S^u = \$80$ and $S^d = \$40$ with probabilities $p_u = \frac{1}{2}$, $p_d = \frac{1}{2}$. See Fig. 14.1.

If the strike price is given by $K = \$60$ and the risk-free interest rate is $r = 0$, then we might conclude that the present value V_0 of the option is given by taking the average of payoffs $\max(S^u - K, 0) = \$20$ and $\max(S^d - K, 0) = \$0$ when the stock price rises and falls, respectively, and hence $V_0 = p_u \times \$20 + p_d \times \$0 = \$10$, which will turn out to be wrong.

For, if we construct a portfolio consisting of a borrowed cash \$20 at time $t = 0$ and $\frac{1}{2}$ shares of the stock, then its value is equal to

Fig. 14.1 A one period binomial tree and real probabilities



$$-\$20 + \frac{1}{2} \times \$S_0 = -\$20 + \frac{1}{2} \times \$50 = \$5 .$$

At the expiry date the value of the portfolio will be equal to either

$$-\$20 + \frac{1}{2} \times \$S^u = -\$20 + \frac{1}{2} \times \$80 = \$20$$

if the stock price goes up, or

$$-\$20 + \frac{1}{2} \times \$S^d = -\$20 + \frac{1}{2} \times \$40 = \$0$$

if the stock price goes down. In either case, the value of the portfolio will be equal to the value of the option. Hence the price of the option at $t = 0$ is equal to the value of the portfolio, which is \$5. We say that such a portfolio replicates the given option.

14.2 The Single-Period Binomial Tree Method

Now we give a more systematic analysis. Let r denote the risk-free interest rate, and let T be the expiry date of the option. Assume that the underlying asset S_t can have two values S^u and S^d at time T depending on the up and the down states, respectively. See Fig. 14.2, where the probability space is given by $\Omega = \{\text{up}, \text{down}\}$ and the bond price increases from 1 at time $t = 0$ to e^{rT} at time T regardless of the state. Similarly, the option price V_t has two values V^u and V^d given by the option payoff at T depending on the up and the down states, respectively. Under the no arbitrage principle, we have $S^d < e^{rT} S_0 < S^u$. For, if $e^{rT} S_0 \leq S^d$ then no investor would deposit his/her money in the bank, and if $S^u \leq e^{rT} S_0$ then no investor would invest in the stock market.

14.2.1 Pricing by Hedging

We construct a risk-free, or hedged, portfolio to compute the option price. Consider a portfolio Π consisting of three assets: an option sold, a risk-free bank deposit, and

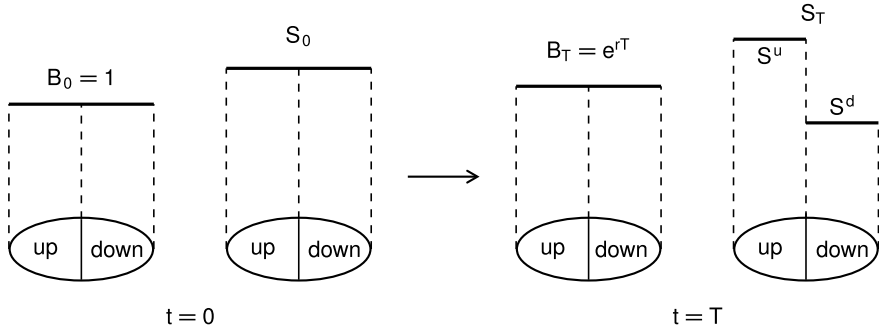


Fig. 14.2 Prices of risk-free and risky assets in the one period binomial tree model

Δ shares of the underlying risky asset or stock. More precisely, Π is given by

$$\Pi_t = -V_t + B_t + \Delta S_t$$

for $t = 0, T$, which is held by an option seller to hedge the risk. Note that Δ is decided at time $t = 0$, i.e., predictable, and is constant in the time interval $[0, T]$. Since short-selling is allowed, Δ can be negative. We find Δ for which Π becomes a risk-free portfolio by solving

$$\begin{cases} -V^u + e^{rT} B_0 + \Delta S^u = e^{rT} (-V_0 + B_0 + \Delta S_0) \\ -V^d + e^{rT} B_0 + \Delta S^d = e^{rT} (-V_0 + B_0 + \Delta S_0) \end{cases}$$

where V^u and V^d are values taken by the payoff V_T depending on the up and down states of S_T , respectively. Hence we have

$$\begin{bmatrix} e^{rT} & S^u - e^{rT} S_0 \\ e^{rT} & S^d - e^{rT} S_0 \end{bmatrix} \begin{bmatrix} V_0 \\ \Delta \end{bmatrix} = \begin{bmatrix} V^u \\ V^d \end{bmatrix},$$

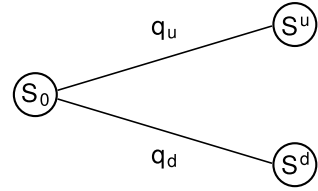
and

$$\begin{aligned} \begin{bmatrix} V_0 \\ \Delta \end{bmatrix} &= \frac{1}{e^{rT}(S^d - S^u)} \begin{bmatrix} S^d - e^{rT} S_0 & -(S^u - e^{rT} S_0) \\ -e^{rT} & e^{rT} \end{bmatrix} \begin{bmatrix} V^u \\ V^d \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{e^{rT}} \left(\frac{e^{rT} S_0 - S^d}{S^u - S^d} V^u + \frac{S^u - e^{rT} S_0}{S^u - S^d} V^d \right) \\ \frac{V^u - V^d}{S^u - S^d} \end{bmatrix}. \end{aligned}$$

Put

$$q_u = \frac{e^{rT} S_0 - S^d}{S^u - S^d}, \quad q_d = \frac{S^u - e^{rT} S_0}{S^u - S^d}. \quad (14.1)$$

Fig. 14.3 A single-period binomial tree and the risk-neutral probability



Then

$$V_0 = e^{-rT} (q_u V^u + q_d V^d) \quad (14.2)$$

and

$$\Delta = \frac{V^u - V^d}{S^u - S^d} . \quad (14.3)$$

Note that $0 < q_u < 1, 0 < q_d < 1$ under the no arbitrage assumption, and $q_u + q_d = 1$. Then the pair (q_u, q_d) represents a probability distribution $\mathbb{Q}$, which is called a *risk-neutral probability*. See Fig. 14.3.

The expectation with respect to $\mathbb{Q}$ is denoted by $\mathbb{E}^{\mathbb{Q}}$. The option price at $t = 0$ is given by

$$V_0 = e^{-rT} (q_u V^u + q_d V^d) = e^{-rT} \mathbb{E}^{\mathbb{Q}}[V_T] ,$$

where V_T is a random variable defined by the payoff at the expiry date. This method is called the *risk-neutral pricing*.

Remark 14.1 (*Martingale Property*) Consider the trivial case when $V_T = S_T$, i.e., the European call option with strike price $K = 0$. Then (14.2) implies that $S_0 = e^{-rT} (q_u S^u + q_d S^d)$.

14.2.2 Pricing by Replication

We replicate a given payoff at T and apply the no arbitrage principle to compute the European option price. Consider a portfolio Π consisting of a riskless asset and an underlying stock. More precisely,

$$\Pi_t = B_t + \Delta S_t ,$$

where Δ is the number of shares of the underlying stock. We find Δ for which $\Pi_T = V_T$ by solving

$$\begin{cases} e^{rT} B_0 + \Delta S^u = V^u \\ e^{rT} B_0 + \Delta S^d = V^d \end{cases}$$

or equivalently,

$$\begin{bmatrix} e^{rT} & S^u \\ e^{rT} & S^d \end{bmatrix} \begin{bmatrix} B_0 \\ \Delta \end{bmatrix} = \begin{bmatrix} V^u \\ V^d \end{bmatrix} .$$

Hence

$$\begin{bmatrix} B_0 \\ \Delta \end{bmatrix} = \frac{1}{e^{rT}(S^d - S^u)} \begin{bmatrix} S^d & -S^u \\ -e^{rT} & e^{rT} \end{bmatrix} \begin{bmatrix} V^u \\ V^d \end{bmatrix} .$$

Thus

$$\begin{aligned} B_0 &= \frac{1}{e^{rT}} \left(\frac{-S^d}{S^u - S^d} V^u + \frac{S^u}{S^u - S^d} V^d \right) , \\ \Delta &= \frac{V^u - V^d}{S^u - S^d} . \end{aligned}$$

Note that we have the same formula for Δ as in (14.3), and the option and the portfolio that replicates the option at $t = T$ have the same value at $t = 0$ by the no arbitrage principle. Hence

$$\begin{aligned} V_0 &= \Pi_0 \\ &= B_0 + \Delta S_0 \\ &= \frac{1}{e^{rT}} \left(\frac{-S^d}{S^u - S^d} V^u + \frac{S^u}{S^u - S^d} V^d \right) + \frac{V^u - V^d}{S^u - S^d} S_0 \\ &= \frac{1}{e^{rT}} \left(\frac{e^{rT} S_0 - S^d}{S^u - S^d} V^u + \frac{S^u - e^{rT} S_0}{S^u - S^d} V^d \right) , \end{aligned}$$

which is identical to the result obtained by the hedging method.

Example 14.1 (European Call Option) Note that $0 \leq \Delta \leq 1$ in the single-period binomial tree model for a European call option with payoff $\max\{S_T - K\}$ at expiry date T . To see why, consider the three possibilities:

$$S^d < S^u \leq K , \quad S^d < K < S^u \quad \text{and} \quad K \leq S^d < S^u .$$

In the first case,

$$\Delta = \frac{V^u - V^d}{S^u - S^d} = \frac{\max\{S^u - K\} - \max\{S^d - K\}}{S^u - S^d} = \frac{0 - 0}{S^u - S^d} = 0 .$$

In the second case,

$$0 < \Delta = \frac{(S^u - K) - 0}{S^u - S^d} < 1 .$$

In the third case,

$$\Delta = \frac{(S^u - K) - (S^d - K)}{S^u - S^d} = \frac{S^u - S^d}{S^u - S^d} = 1 .$$

Therefore, if the asset prices in the next step are all below or equal to the strike price, the Delta is zero, i.e., there is no need to hold any underlying asset. If the asset prices in the next step are all above or equal to the strike price, the Delta is equal to one, i.e., the issuer of the option should have one share of the underlying for one contract of the option.

14.2.3 A Special Case: The Multiplicative Model

Consider the time interval $[0, T]$ and assume that the asset price moves from S_0 to $S^u = uS_0$ or $S^d = dS_0$ in the one-period binomial tree from time 0 to T . (For example, take $u = e^{\sigma\sqrt{T}}$ and $d = e^{-\sigma\sqrt{T}}$.) Then the risk-neutral probability $\mathbb{Q} = (q_u, q_d)$ in (14.1) is given by

$$q_u = \frac{e^{rT} - d}{u - d} = q, \quad q_d = 1 - q .$$

Then

$$\Delta = \frac{V^u - V^d}{(u - d)S_0}$$

and

$$V_0 = e^{-rT} \{qV^u + (1 - q)V^d\} = e^{-rT} \mathbb{E}^{\mathbb{Q}}[V_T] .$$

14.3 The Multiperiod Binomial Tree Method

14.3.1 Construction of the Multiperiod Binomial Tree

Given an option with expiry date T , we consider a multiperiod binomial tree of length N , obtained by stringing together single-period binomial trees, where the length of the time interval for each single-period binomial tree is $\delta t = \frac{T}{N}$. At time $T = N \times \delta t$, the asset price can take one of $N + 1$ possible values. We define $\mathbb{Q}$ on the whole space Ω by

$$\mathbb{Q}([a_1, \dots, a_N]) = q_u^k q_d^{N-k} ,$$

where k is equal to the number of up movements.

A rigorous presentation of the binomial tree method may be given as follows: Let Ω be the set of all paths ω of length N . Consider a subset of Ω , called a cylinder set or a block, $[a_1, \dots, a_n]$, $1 \leq n \leq N$, defined by

$$[a_1, \dots, a_n] = \{\omega = a_1 \cdots a_n \omega_{n+1} \cdots \omega_N \mid \omega_{n+1}, \dots, \omega_N \in \{u, d\}\},$$

where we write ‘u’ and ‘d’ to represent up and down movements, respectively. Note that the number of paths in $[a_1, \dots, a_n]$ is equal to 2^{N-n} and that

$$\begin{aligned} \Omega &= [u] \cup [d] \\ &= [u,u] \cup [u,d] \cup [d,u] \cup [d,d] \\ &= [u,u,u] \cup [u,u,d] \cup [u,d,u] \cup [u,d,d] \cup [d,u,u] \cup [d,u,d] \cup [d,d,u] \cup [d,d,d] \\ &= \dots \end{aligned}$$

Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and let $\mathcal{F}_n$, $1 \leq n \leq N$, be the σ -algebra on Ω consisting of cylinder sets $[a_1, \dots, a_n]$. Then we have a filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}_N$. Note that

$$\begin{aligned} S_n &= e^{-r\delta t} (q_u S_{n+1}^u + q_d S_{n+1}^d) \\ &= e^{-r\delta t} \mathbb{E}^{\mathbb{Q}}[S_{n+1} | S_n] \\ &= e^{-r\delta t} \mathbb{E}^{\mathbb{Q}}[S_{n+1} | \mathcal{F}_n], \end{aligned}$$

where the first equality holds as in Remark 14.1 and the third equality by the Markov property of $\{S_n : 0 \leq n \leq N\}$. Hence the discounted asset price process $\{e^{-rn\delta t} S_n : 0 \leq n \leq N\}$ is a $\mathbb{Q}$ -martingale, i.e.,

$$e^{-rn\delta t} S_n = \mathbb{E}^{\mathbb{Q}}[e^{-r(n+1)\delta t} S_{n+1} | \mathcal{F}_n].$$

The payoff function of a European option is measurable with respect to $\mathcal{F}_N$, and hence is denoted by C_N to emphasize the fact. Define

$$C_{N-1} = e^{-r\delta t} \mathbb{E}^{\mathbb{Q}}[C_N | \mathcal{F}_{N-1}],$$

which is the option price at time $(N-1)\delta t$ and depends on the asset price at time $(N-1)\delta t$. This is measurable with respect to $\mathcal{F}_{N-1}$.

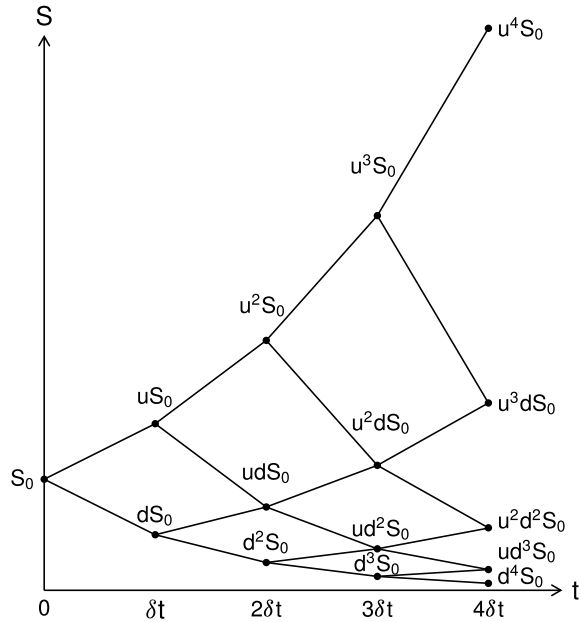
In the rest of the chapter we consider the multiperiod multiplicative binomial tree model over time $0 \leq t \leq T = N\delta t$, i.e., the asset price moves from S_n to $S_{n+1}^u = uS_n$ or $S_{n+1}^d = dS_n$ from time n to $n+1$. Such a binomial tree market model is complete, i.e., every contingent claim is attainable, in other words, there exists a strategy consisting of a suitable amount of riskless bank deposit and underlying asset exactly replicating the claim at any time and in any state.

In Table 14.1 a spreadsheet lists the asset prices at time $0, \delta t, 2\delta t, 3\delta t, 4\delta t$ as an upper triangular matrix, and in Fig. 14.4 a binomial tree of length $N = 4$ is plotted where $u = \frac{3}{2}, d = \frac{1}{2}$. Observe the exponential growth and decay of the top and bottom paths.

Table 14.1 A spreadsheet for a multiperiod binomial tree

32	48	72	108	162
	16	24	36	54
		8	12	18
			4	6
				2

Fig. 14.4 A multiperiod binomial tree



In Sect. 14.4 it will be shown that the European option price obtained by the binomial tree method converges to a formula from the continuous time model. (See Theorem 14.1.)

Example 14.2 (European Call) We now compute the price of a European call option with exercise price $K = 10$ using a three-step binomial tree. For the sake of computational convenience we assume $r = 0$. Other parameters are given by $S^u = uS_0$, $S^d = dS_0$, where $u = \frac{3}{2}$, $d = \frac{1}{2}$. Note that $q_u = q_d = \frac{1}{2}$.

In Figs. 14.5, 14.6, 14.7 the numbers in the circles represent stock prices while the numbers in the rectangles are call option prices for a given time and stock price. In Fig. 14.5 the numbers in the rectangles are payoffs at maturity date. As in Fig. 14.6, from the payoff specified at time $t = 3$ we obtain the option price at time $t = 2$ by the single-period binomial tree method, and from the option price at $t = 2$ we obtain the option price at $t = 1$ by the single-period binomial tree method, and finally from the price at $t = 1$ we obtain the price at $t = 0$. For example, at $t = 2$ and with $S_2 = 36$

Fig. 14.5 Payoff of a European call option in a binomial tree

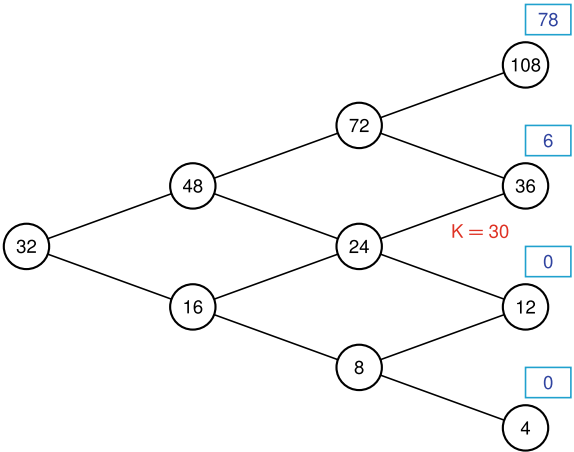
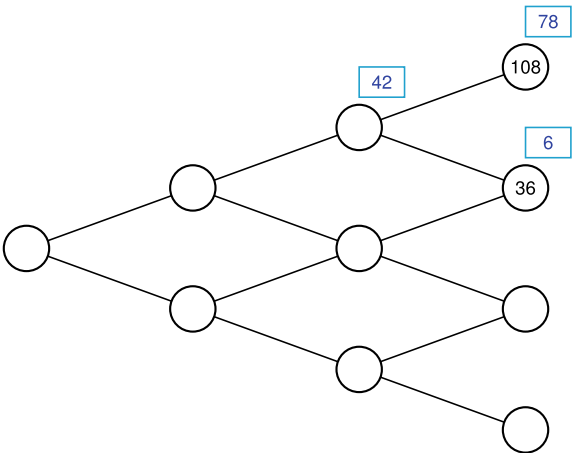


Fig. 14.6 The first step in pricing a European call option by a binomial tree



the option price is equal to 26. The option price at $t = 0$ is 8.5. For more examples of option pricing by the binomial tree method, see Simulations 14.1 and 14.2.

The binomial model, which is based on the risk-neutral valuation, is a popular approach to option pricing. It is based on the simplified assumption that over a single period (of very short duration δt), the underlying asset can only move from its current price to two possible levels. As will be shown in Sect. 14.4, the Black–Scholes–Merton formula is the limiting value resulting from taking $N \rightarrow \infty$ in the binomial period where $\delta t = \frac{T}{N}$. See Table 14.2 for the formal comparison of continuous time and discrete time cases.

Recall that to compute the values $\binom{N}{i}$, $1 \leq i \leq N$, we consider the binomial coefficients in the expansion of $(x + y)^N$ and *Pascal's triangle*, which is a recombining tree. That is why the tree employed in option pricing is called a *binomial tree*.

Fig. 14.7 Pricing of a European call option by a binomial tree

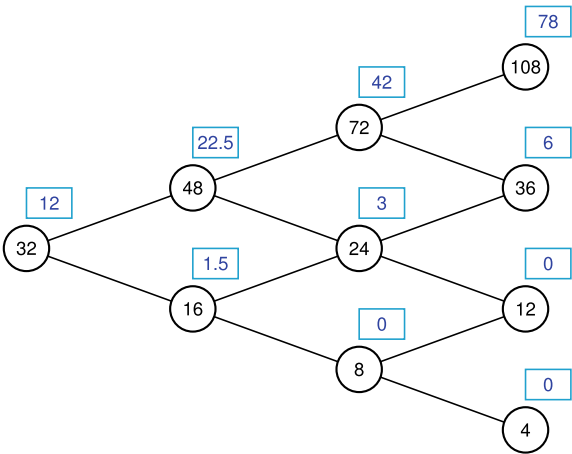


Table 14.2 Option pricing by risk-neutral method

	Continuous time	Discrete time
Observation time	$0 \leq t \leq T$	$n = 0, 1, \dots, N$
Expiry date	T	$N \times \delta t$
Random motion	Brownian motion W_t	Random walk
Range of random motion	$\mathbb{R}^1$	Integral multiples of $\hat{\alpha}$
Probability distribution	Normal distribution	Binomial distribution
Asset price movement	Geometric Brownian motion	Binomial tree
Size of randomness	Volatility	Standard deviation
Physical probability	$\mathbb{P}$	(p_u, p_d)
Risk-neutral probability	$\mathbb{Q}$	(q_u, q_d)
Payoff	C_T	C_N
Option price	$e^{-rT} \mathbb{E}[C_T]$	$e^{-rT} \mathbb{E}[C_N]$

It is impossible to use a *binary tree* as a model for an asset price movement. The fact that a binomial tree is recombinant reflects the usual assumption that future asset price movement is not dependent on the past history, while a binary tree tracks all the price history and the price movement depends not only on the present but also on the past information. Furthermore, the number of nodes of a binomial tree is equal to $N + 1$, where N is the length of the tree, while the number of nodes for a binary tree grows exponentially, i.e., the number of nodes is equal to 2^N . For example, for an option with expiry date $T = 1$ year and $\delta t = 1$ day, we have $N = 252$, in which case $2^{252} \approx 10^{75}$ is too large to deal with computationally.

Example 14.3 (*Delta*) We consider the multiperiod multiplicative model of length $N = 5$ where each time step is given by the single-period multiplicative model given in Sect. 14.2.3. Take $T = 1$, $S_0 = 110$, $K = 110$, $r = 0.05$ and $\sigma = 0.3$. For the

Fig. 14.8 The binomial tree for the asset price with $S_0 = 110$ and $K = 110$

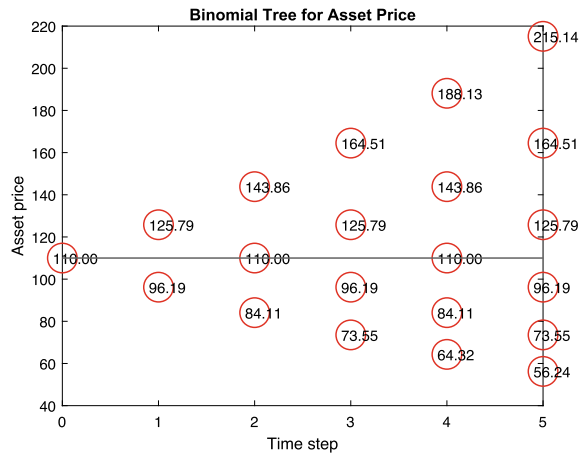
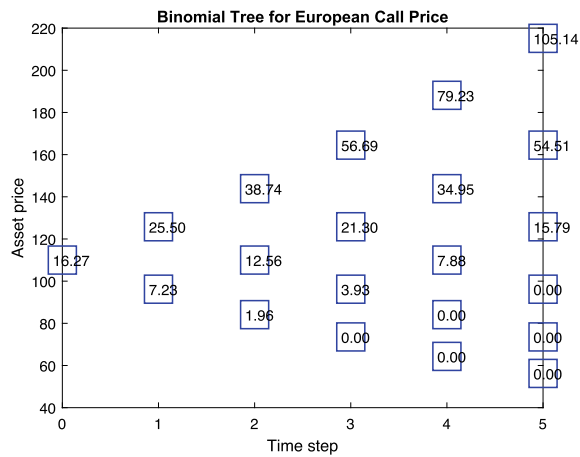


Fig. 14.9 The binomial tree for the European call option price with $S_0 = 110$ and $K = 110$

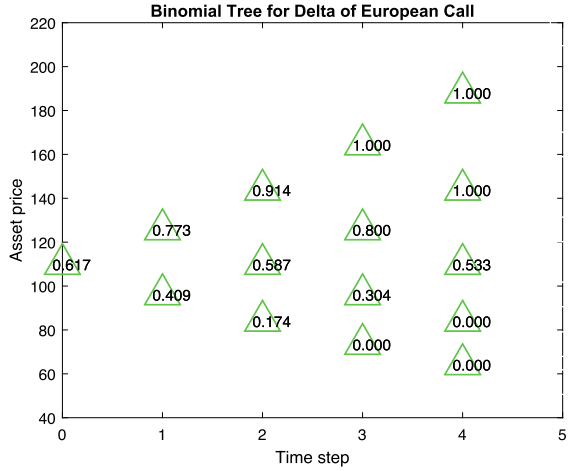


binomial trees for the asset price, the European call price and the Delta see Figs. 14.8, 14.9 and 14.10, respectively. In Fig. 14.8, a horizontal line marks the exercise price. In Fig. 14.9 the American put prices are printed at the nodes defined by the asset prices. Asset prices are marked by red circles, the option prices by blue squares, and the values of Delta by green triangles.¹

The option price by the binomial tree method is equal to 16.2682 and the Delta equal to 0.6171 while by the Black–Scholes–Merton formula the option price is equal to 15.6544 and the Delta equal to 0.6243. (See also Simulation 14.2 for a European put option.)

¹ The symbol Δ for Delta looks like a triangle Δ .

Fig. 14.10 The binomial tree for the Delta of the European call option with $S_0 = 110$ and $K = 110$



14.3.2 Delta-Hedging in the Multiperiod Binomial Tree Method

Using the hedging idea given in Sects. 14.2.1 and 14.3.1 for the multiperiod binomial tree method, we construct a self-financing multiperiod portfolio process

$$\Pi = -V + B + \Delta S$$

and check whether the risk-free portfolio grows exponentially, i.e.,

$$\Pi_n = \Pi_0 e^{rn \times \delta t}$$

for any asset price movement, where $\delta t = \frac{T}{N}$ and $1 \leq n \leq N$. The self-financing condition means that there is no net inflow of outflow of wealth. In Simulation 14.3, we consider a multiperiod binomial tree of length $N = 15$ for the pricing of a European call option with $T = 3$ and $K = 15$. We take $S_0 = 10$, risk-free interest rate $r = 0.15$ and volatility $\sigma = 0.3$. Then $dt = T/N$, $u = \exp(\sigma \sqrt{dt})$, $d = \exp(-\sigma \sqrt{dt})$. The simulation result is shown in Fig. 14.11. In the first panel we show the outcome of a biased coin tossing with probability q for upward movement and probability $1 - q$ for downward movement at each time step $i = 0, 1, \dots, N - 1$, where the risk-neutral probability $\mathbb{Q} = (q, 1 - q)$ is given by $q = (e^{r \delta t} - d)/(u - d)$. In our simulation, we have $q = 0.5797$, so we obtain considerably more upward movements than downward movements on average. In the second panel, the entire binomial tree for the asset is displayed and a single sample path defined by the coin tossing outcome is shown in red. In the third panel the option price movement defined by the formula given by Sect. 14.2.3 is shown. Note that $0 \leq \Delta \leq 1$ in the fourth panel. In the fifth panel, the self-financing condition implies that when we buy a stock we

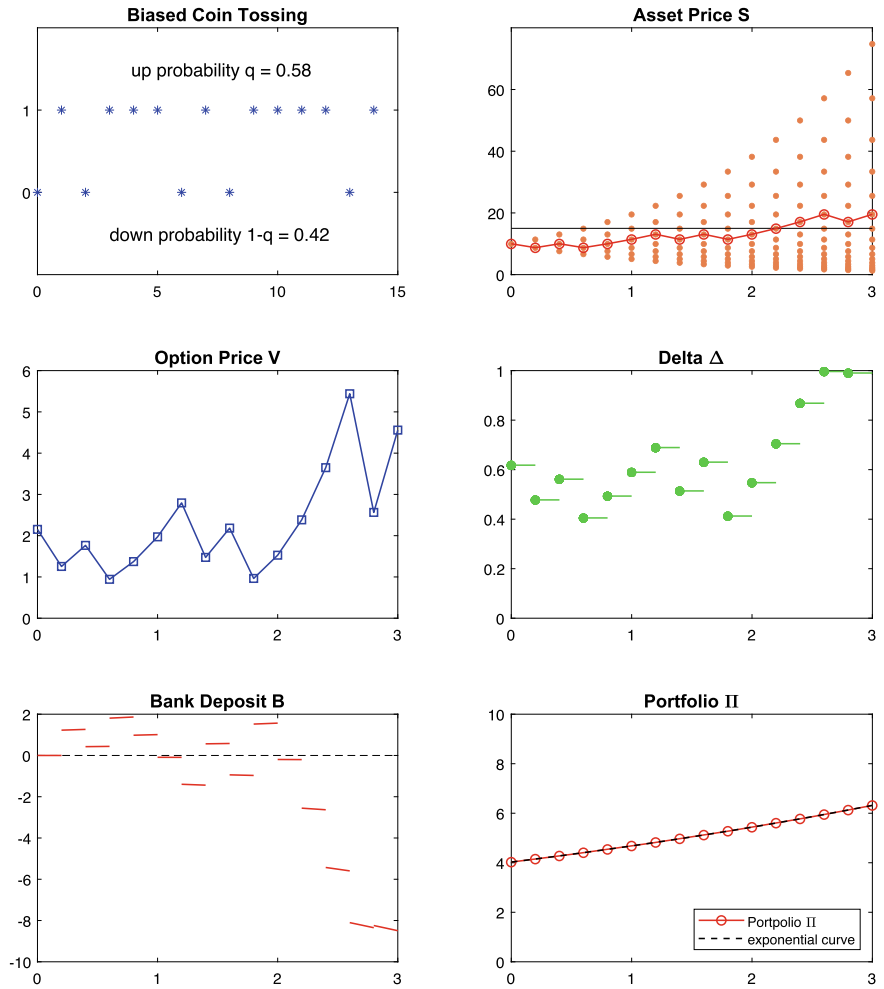


Fig. 14.11 Delta-hedging in the multiperiod binomial tree method for a European call option using a self-financing portfolio $\Pi = -V + B + \Delta S$ with $T = 3$, $S_0 = 10$, $K = 15$ and $N = 15$ (from top left to bottom right)

borrow money, i.e., B decreases, and sometimes we can have $B < 0$. We choose a large value for r in the simulation to make the exponential growth more visible in the sixth panel.

14.4 Convergence to the Black–Scholes–Merton Formula

In this section we show that the European call option price obtained from the N -period binomial tree model converges to the solution of the Black–Scholes–Merton partial differential equation in Chap. 15 as $N \rightarrow \infty$.

Let $\delta t = \frac{T}{N}$, and use $r\delta t$ as the interest rate in each single-period. If we take

$$S^u = uS_0, \quad S^d = dS_0$$

in the single period binomial model, then

$$\frac{S^u - S_0}{S_0} = u - 1, \quad \frac{S^d - S_0}{S_0} = d - 1.$$

Then, by the binomial tree method, the price of the European call option is given by

$$V_0 = e^{-rT} \sum_{i=0}^N \binom{N}{i} q^i (1-q)^{N-i} (S_0 u^i d^{N-i} - K)^+,$$

where N is the number of time steps, and

$$q = \frac{e^{r\delta t} - d}{u - d}.$$

Note that $e^{-r\delta t} qu + e^{-r\delta t} (1-q)d = 1$. Let m be the smallest integer such that

$$S_0 u^m d^{N-m} > K,$$

or equivalently, m is the unique integer such that

$$S_0 u^{m-1} d^{N-m+1} \leq K < S_0 u^m d^{N-m}.$$

Then

$$\frac{\log \frac{K}{S_0} - N \log d}{\log \frac{u}{d}} < m \leq \frac{\log \frac{K}{S_0} - N \log d}{\log \frac{u}{d}} + 1$$

and

$$V_0 = e^{-rT} \sum_{i=m}^N \binom{N}{i} q^i (1-q)^{N-i} (S_0 u^i d^{N-i} - K).$$

Put

$$A = e^{-rT} \sum_{i=m}^N \binom{N}{i} q^i (1-q)^{N-i} u^i d^{N-i}$$

and

$$B = \sum_{i=m}^N \binom{N}{i} q^i (1-q)^{N-i}.$$

Then $V_0 = AS_0 - Ke^{-rT}B$. To find the limit of B as $N \rightarrow \infty$, recall that the binomial distribution $B(N, q)$ is approximated by the normal distribution $N(Nq, Nq(1-q))$ with mean Nq and variance $Nq(1-q)$ for sufficiently large N , and note that

$$\begin{aligned} \sum_{i=m}^N \binom{N}{i} q^i (1-q)^{N-i} &= \Pr(B(N, q) \geq m) \\ &\approx \Pr(N(Nq, Nq(1-q)) \geq m) \\ &= \int_m^\infty \frac{1}{\sqrt{2\pi Nq(1-q)}} \exp\left(-\frac{1}{2} \frac{(x - Nq)^2}{Nq(1-q)}\right) dx \\ &= \int_{-d_2}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \\ &= N(d_2), \end{aligned}$$

where N denotes the standard normal distribution function,

$$d_2 = \lim_{N \rightarrow \infty} \frac{-m + Nq}{\sqrt{Nq(1-q)}} = \lim_{N \rightarrow \infty} \frac{\frac{\log \frac{S_0}{K} + N \log d}{\log \frac{u}{d}} + Nq}{\sqrt{Nq(1-q)}}$$

and we used the substitution

$$y = \frac{x - Nq}{\sqrt{Nq(1-q)}}.$$

Similarly,

$$\begin{aligned} \lim_{N \rightarrow \infty} A &= \lim_{N \rightarrow \infty} e^{-rT} \sum_{i=m}^N \binom{N}{i} q^i (1-q)^{N-i} u^i d^{N-i} \\ &= \lim_{N \rightarrow \infty} \sum_{i=m}^N \binom{N}{i} (e^{-r\delta t} qu)^i (e^{-r\delta t} (1-q)d)^{N-i} \\ &= \int_{d_1}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy \\ &= 1 - N(d_1), \end{aligned}$$

where

$$d_1 = \lim_{N \rightarrow \infty} \frac{\frac{\log \frac{S_0}{K} + N \log d}{\log \frac{u}{d}} + Nq^*}{\sqrt{Nq^*(1-q^*)}}$$

and $q^* = e^{-r\delta t}qu$. Hence we conclude that

$$\lim_{N \rightarrow \infty} V_0 = N(d_1)S_0 - Ke^{-rT}N(d_2).$$

Now we show that by choosing suitable values for u and d as in Sect. 14.2.3 the option price formula obtained from the binomial tree method converges to the Black–Scholes–Merton formula given in Theorem 15.2.

Theorem 14.1 *Take*

$$u = e^{\sigma\sqrt{\delta t}}, \quad d = e^{-\sigma\sqrt{\delta t}}.$$

Then the aforementioned d_1, d_2 converge to

$$\frac{\log \frac{S_0}{K} \pm (r + \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}$$

defined in the Black–Scholes–Merton formula.

Proof We will show that d_1 and d_2 given in the above converge to the usual d_1 and d_2 in the closed form solution of the Black–Scholes–Merton equation. Note that

$$\begin{aligned} q &= \frac{e^{r\delta t} - e^{-\sigma\sqrt{\delta t}}}{e^{\sigma\sqrt{\delta t}} - e^{-\sigma\sqrt{\delta t}}} \\ &= \frac{(1 + r\delta t) - (1 - \sigma\sqrt{\delta t} + \frac{1}{2}\sigma^2\delta t + \dots)}{(1 + \sigma\sqrt{\delta t} + \frac{1}{2}\sigma^2\delta t + \dots) - (1 - \sigma\sqrt{\delta t} + \frac{1}{2}\sigma^2\delta t + \dots)} \\ &\approx \frac{1}{2} + \frac{r - \frac{1}{2}\sigma^2}{2\sigma}\sqrt{\delta t}. \end{aligned}$$

Note that

$$\begin{aligned} d_2 &\approx \frac{\log \frac{S_0}{K} + N \log d + Nq \log \frac{u}{d}}{\sqrt{Nq(1-q)} \log \frac{u}{d}} \\ &= \frac{\log \frac{S_0}{K} - N\sigma\sqrt{\delta t} + N(\frac{1}{2} + \frac{r - \frac{1}{2}\sigma^2}{2\sigma}\sqrt{\delta t})2\sigma\sqrt{\delta t}}{\sqrt{N\left(\frac{1}{4} - \left(\frac{r - \frac{1}{2}\sigma^2}{2\sigma}\right)^2\delta t\right)}2\sigma\sqrt{\delta t}} \\ &\approx \frac{\log \frac{S_0}{K} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} \end{aligned}$$

and that

$$B \approx -K \left(e^{-r\delta t} \right)^N \int_{d_2}^{\infty} \frac{1}{2\pi} \exp\left(-\frac{y^2}{2}\right) dy = -K e^{-rT} N(d_2) .$$

As for A , note that

$$q^* = e^{-r\delta t} \left(\frac{1}{2} + \frac{r + \frac{1}{2}\sigma^2}{2\sigma} \sqrt{\delta t} \right) \approx \frac{1}{2} + \frac{r + \frac{1}{2}\sigma^2}{2\sigma} \sqrt{\delta t} .$$

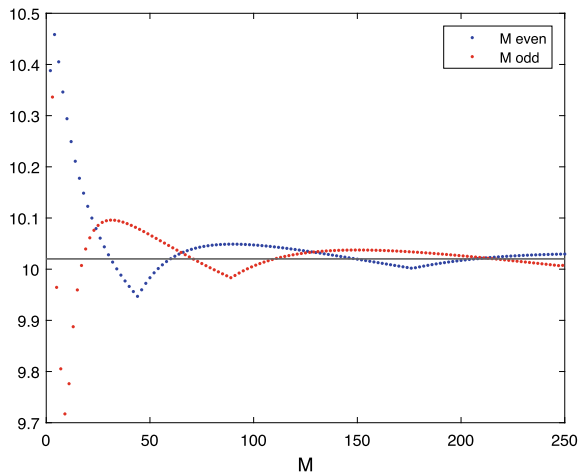
Hence

$$\begin{aligned} d_1 &\approx \frac{\log \frac{S_0}{K} + N \log d + N q^* \log \frac{u}{d}}{\sqrt{N q^* (1 - q^*)} \log \frac{u}{d}} \\ &\approx \frac{\log \frac{S_0}{K} - N \sigma \sqrt{\delta t} + N \left(\frac{1}{2} + \frac{r + \frac{1}{2}\sigma^2}{2\sigma} \sqrt{\delta t} \right) 2\sigma \sqrt{\delta t}}{\sqrt{N \left(\frac{1}{4} - \left(\frac{r + \frac{1}{2}\sigma^2}{2\sigma} \right)^2 \delta t \right)} 2\sigma \sqrt{\delta t}} \\ &\approx \frac{\log \frac{S_0}{K} + \left(r + \frac{1}{2}\sigma^2 \right) T}{\sigma \sqrt{T}} \end{aligned}$$

as $N \rightarrow \infty$. □

For a numerical experiment illustrating Theorem 14.1, see Fig. 14.12 and Simulation 14.4.

Fig. 14.12 Convergence of the price obtained by the binomial tree method to the Black–Scholes–Merton price



Remark 14.2 For $u, d = e^{\pm\sigma\sqrt{\delta t}} \approx 1 \pm \sigma\sqrt{\delta t} + \frac{1}{2}\sigma^2\delta t$, $\delta t > 0$, in the one-period binomial tree from time t to time $t + \delta t$, we have

$$\frac{\delta S_t}{S_t} = \frac{S_{t+\delta t} - S_t}{S_t} \approx \pm\sigma\sqrt{\delta t} + \frac{1}{2}\sigma^2\delta t.$$

This result is consistent with the geometric Brownian motion used in the continuous time model for the derivation of the Black–Scholes–Merton differential equation in Chap. 15 if we regard the first term $\pm\sigma\sqrt{\delta t}$ as $\sigma\delta W_t$ and the second term $\frac{1}{2}\sigma^2\delta t$ as $\mu\delta t$. Since the drift coefficient μ does not appear in the solution of the Black–Scholes–Merton equation, it does not matter that the above choice of u and d produces μ as a function of σ instead of an unrelated independent parameter. Note that the continuous limit of $\Delta = \frac{V^u - V^d}{S^u - S^d}$ is $\frac{\partial V}{\partial S}$.

Theorem 14.2 (Derivation of the Black–Scholes–Merton Equation) *Consider the binomial tree model. Over the time interval $[t, t + \delta t]$, we take*

$$u, d = e^{\pm\sigma\sqrt{\delta t}}.$$

The asset price moves from S_t to $S_{t+\delta t}^u = uS_t$ or $S_{t+\delta t}^d = dS_t$ with risk-neutral probability

$$q_u = \frac{e^{r\delta t} - d}{u - d} = q, \quad q_d = 1 - q.$$

Suppose that the limit $V_t = V(t, S_t)$ of the option price in the binomial tree model exists as $\delta t \rightarrow 0$. Then $V(t, S)$ satisfies the Black–Scholes–Merton equation given in Theorem 15.1, Chap. 15.

Proof The option price at time $t + \delta t$ is denoted by V^u or V^d . Then, as shown in Sect. 14.2.3, the option price V_t at time t is given by

$$V_t = e^{-r\delta t} \{qV^u + (1 - q)V^d\} = e^{-r\delta t} \{q(V^u - V^d) + V^d\}.$$

Hence

$$(1 + r\delta t)V_t - V^d \approx q(V^u - V^d). \quad (14.4)$$

Then Itô's formula implies that

$$\begin{aligned} V^u &= V(t + \delta t, uS_t) \\ &\approx V_t + \frac{\partial V}{\partial S}(uS_t - S_t) + \frac{1}{2} \frac{\partial^2 V}{\partial S^2}(uS_t - S_t)^2 + \frac{\partial V}{\partial t}\delta t, \\ &= V_t + \frac{\partial V}{\partial S}S_t(u - 1) + \frac{1}{2} \frac{\partial^2 V}{\partial S^2}S_t^2(u - 1)^2 + \frac{\partial V}{\partial t}\delta t. \end{aligned}$$

Similarly, we have

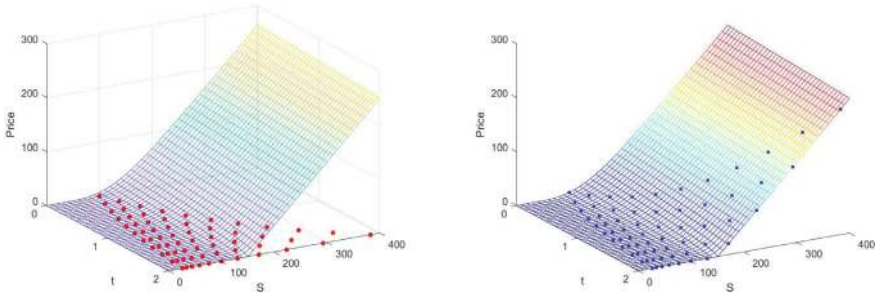


Fig. 14.13 The binomial trees for the asset price (*left, red dots*) and the European call option price (*right, blue dots*) together with the graph of the Black–Scholes–Merton formula

$$V^d \approx V_t + \frac{\partial V}{\partial S} S_t (d - 1) + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} S_t^2 (d - 1)^2 + \frac{\partial V}{\partial t} \delta t .$$

Hence

$$V^u - V^d \approx \frac{\partial V}{\partial S} S_t (u - d)$$

since $u - d = 2\sigma\sqrt{\delta t}$ and $(u - 1)^2 = (d - 1)^2 \approx \sigma^2 \delta t$. Thus (14.4) implies that

$$\begin{aligned} (1 + r\delta t)V_t - \left(V_t + \frac{\partial V}{\partial S} S_t (-\sigma\sqrt{\delta t} + \frac{1}{2}\sigma^2 \delta t) + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} S_t^2 \sigma^2 \delta t + \frac{\partial V}{\partial t} \delta t \right) \\ \approx \left(\sigma\sqrt{\delta t} + \left(r - \frac{1}{2}\sigma^2 \right) \delta t \right) \frac{\partial V}{\partial S} S_t , \end{aligned}$$

which is reduced to

$$r\delta t V_t - \left(\frac{1}{2} \frac{\partial^2 V}{\partial S^2} S_t^2 \sigma^2 \delta t + \frac{\partial V}{\partial t} \delta t \right) \approx r\delta t \frac{\partial V}{\partial S} S_t$$

after cancelling common terms. Now we divide both sides by δt and take the limit as $\delta t \rightarrow 0$. $\square$

The solution of the Black–Scholes–Merton partial differential equation is called the Black–Scholes–Merton formula, and is given in Theorem 15.2. For a numerical justification of Theorem 14.2 see Fig. 14.13. In the left panel the binomial tree for the asset price (*red dots*) is plotted on the (t, S) -plane, and in the right panel the European call option price (*blue dots*) is plotted together with the graph of the Black–Scholes–Merton option price as a function of (t, S) . We take $S_0 = 100$, $K = 150$, $T = 2$, $r = 0.05$ and $\sigma = 0.3$. Observe that the option price represented by blue dots on the right are located close to the option price surface. Consult Simulation 14.5.

Table 14.3 The spreadsheet corresponding to the binomial tree for the underlying asset

100	118.9110	141.3982	168.1381
	84.0965	100.0000	118.9110
		70.7222	84.0965
			59.4749

Table 14.4 The spreadsheet corresponding to the binomial tree for the European call option price

10.3363	18.6541	33.2164	58.1381
	2.1987	4.4264	8.9110
		0	0
			0

14.5 Computer Experiments

Simulation 14.1 (The Binomial Tree Method for a European Call Price) For a European call option we apply the binomial tree method and obtain the value 10.3363, which is a rough approximation to the value 10.0201 that is obtained by the Black–Scholes–Merton formula. (In Tables 14.3 and 14.4 we use upper triangular matrices to keep records of the asset price and the option price. Entries below the diagonals are ignored. The first entries in the upper triangles represent S_0 and the option price at $t = 0$, respectively. The $(i + 1)$ st column corresponds to the i th time step for $1 \leq i \leq 3$.) For a better estimation we need to increase the number of time steps. The following MATLAB code produces two binomial trees for the asset price and a European call option price. Parameters have to satisfy the condition that $0 < q = \frac{e^{rdt} - d}{u - d} < 1$, or equivalently, $rdt < \sigma \sqrt{dt}$. (For the simulations for American options, see Chap. 19.)

```

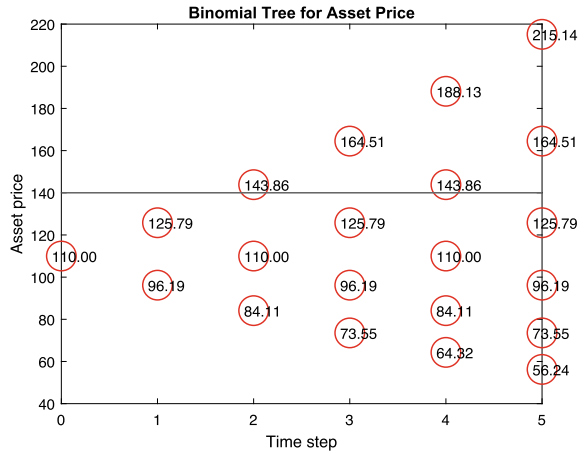
S0 = 100; % asset price at t=0
K = 110; % strike price
T = 1; % expiry date
r = 0.05; % interest rate
sigma = 0.3; % volatility
M = 3; % number of time steps
dt = T/M;

u = exp(sigma*sqrt(dt));
d = exp(-sigma*sqrt(dt));
q = (exp(r*dt)-d)/(u-d);

% We compute asset prices.

```

Fig. 14.14 The binomial tree for the asset price with $S_0 = 110$



```

for i = 0:1:M
    fprintf('time = %i\n', i)
    S = S0*u.^([i:-1:0]').*d.^([0:1:i]')
end

fprintf('Payoff at expiry\n')
Call = max(S0*u.^([M:-1:0]')*d.^([0:1:M]') - K, 0)

% We proceed backward to compute option value at time 0.
for i = M:-1:1
    fprintf('time = %i\n', i-1)
    Call = exp(-r*dt)*(q*Call(1:i) + (1-q)*Call(2:i+1))
end

```

Simulation 14.2 (The Binomial Trees for the Asset, the European Put Price and the Delta) We consider a European put option with $S_0 = 110$, exercise price $K = 140$, expiry date $T = 1$, $r = 0.05$, $\sigma = 0.3$ and $N = 5$ time periods, and plot the binomial trees for the asset and the put option price. Parameters have to satisfy the condition that $0 < q < 1$. In Fig. 14.14 for the asset, the exercise price is plotted as a horizontal line. In Fig. 14.15 for the European put price at the corresponding nodes the price at time 0 is given by 29.22. (Compare the result with Simulation 19.1.) In Fig. 14.16 the Delta is shown at the nodes. (See also Example 14.3 for a European call option.)

```

T = 1; % expiry date
N = 5; % number of time periods
S0 = 110; % asset price at time 0
S = ones(N+1,N+1); % asset price
S(1,:) = S0;
P = zeros(N+1,N+1); % European put price

```

Fig. 14.15 The binomial tree for the European put option price

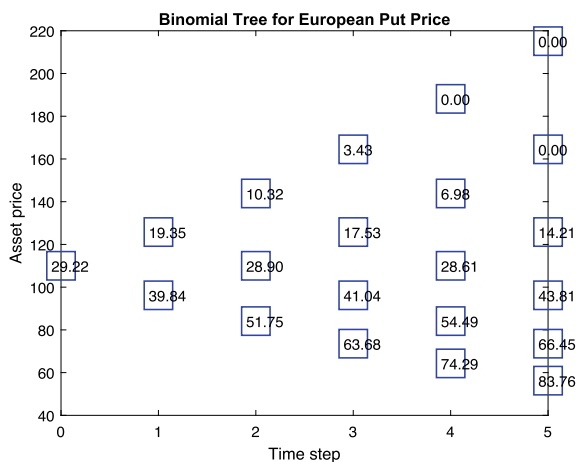
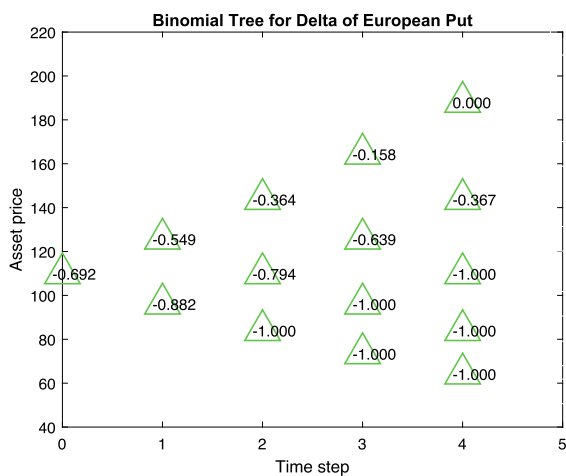


Fig. 14.16 The binomial tree for the Delta of the European put option



```
K = 140; % strike Price
r = 0.05; % interest rate
sigma = 0.3; % volatility
dt = T/N;
u = exp(sigma*sqrt(dt));
d = exp(-sigma*sqrt(dt));
q = (exp(r*dt)-d)/(u-d);
```

```
% i for time, j for asset price level
for i = 2:N+1
    for j = 1:i
        S(i,j) = S0*u^(j-1)*d^(i-j); % asset price
```

```

        end
    end

    for j = 1:N+1
        P(N+1,j) = max(K-S(N+1,j),0); % European put payoff at T
    end

    for i = N:-1:1 % backward in time
        for j = 1:i
            P(i,j) = exp(-r*dt)*(q*P(i+1,j+1)+(1-q)*P(i+1,j));
        end
    end

    for i = N:-1:1 % backward in time
        for j = 1:i
            Delta(i,j) = (P(i+1,j+1)-P(i+1,j))/(S(i+1,j+1)-S(i+1,j));
        end
    end

    figure(1)
    for i = 1:N+1
        plot(i-1,S(i,1:i),'or')
        for j = 1:i
            text(i-1-0.1,S(i,j),num2str(S(i,j),'%.2f'))
        end
        hold on
    end

    line([0 N],[K K])
    xlabel('Time')
    ylabel('Asset')
    title('Binomial Tree for Asset Price');

    figure(2)
    for i = 1:N+1
        plot(i-1,S(i,1:i),'sb')
        for j = 1:i
            text(i-1-0.1,S(i,j),num2str(P(i,j),'%.2f'))
        end
        hold on
    end

    xlabel('Time')
    ylabel('European Put')
    title('Binomial Tree for European Put Price');

    figure(3)

```



```

for i = 1:N
    plot(i-1,S(i,1:i),'^g')
    for j = 1:i
        text(i-1-0.1,S(i,j),num2str(Delta(i,j),'%.2f'),'Color','k')
    end
    hold on
end
xlabel('Time')
ylabel('Delta')
xticks([0:1:N])
title('Binomial Tree for Delta of European Put');

```

In the following we compare the results from the binomial tree method and the values obtained from the Black–Scholes–Merton formula given in Theorem 15.2 and the formula in Theorem 15.3 and the identity (15.6).

```

d1 = (log(S0/K) + (r+0.5*sigma^2)*T)/sigma/sqrt(T) ;
d2 = d1 - sigma*sqrt(T);
BSM_price_put = S0*normcdf(d1) - K*exp(-r*T)*normcdf(d2) - S0 + K*exp(-r*T)
BSM_Delta_put = normcdf(d1)-1

binomial_tree_European_put_price =
                                29.2207

binomial_tree_European_put_Delta =
                                -0.6922

BSM_European_put_price =
                                28.8990

BSM_European_put_Delta =
                                -0.6869

```

Simulation 14.3 (Delta-hedging in the Binomial Tree Method) To check the idea given in Sect. 14.3.2, we consider a European call option and simulate Delta-hedging in the binomial tree method. We take $N = 15$, $T = 3$, $K = 15$, $S_0 = 10$, risk-free interest rate $r = 0.15$ and volatility $\sigma = 0.3$. Then $u = \exp(\sigma\sqrt{dt}) = 1.1436$, $d = \exp(-\sigma\sqrt{dt}) = 0.8744$ and $q = 0.5797$. Parameters have to satisfy the condition that $0 < q < 1$. See Fig. 14.11.

```

T = 3;
r = 0.15; % interest rate
S0 = 10; % asset price at time t=0
K = 15; % strike price
sigma = 0.3; % volatility

```

```

N = 15; % number of time steps
dt = T/N;
u = exp(sigma*sqrt(dt));
d = exp(-sigma*sqrt(dt));
q = (exp(r*dt)-d)/(u-d);
t_value = 0:dt:T;
S = ones(N+1,N+1); % asset price
S(1,:) = S0;
S_path = zeros(1,N+1); % asset price
S_path(1,1)= S0;
V = zeros(1,N+1); % option price
Delta = zeros(N+1,N); % Delta on the binomial tree
Delta_path = zeros(1,N+1); % Delta along the path
B = zeros(1,N+1); % bank deposit
B(1,1) = 0.0; % Choose any number for the initial cash amount.
Pi = zeros(1,N+1); % portfolio

% i for time, j for asset price level
for i = 2:N+1
    for j = 1:i
        S(i,j) = S0*u^(j-1)*d^(i-j); % asset price
    end
end

for j = 1:N+1
    C(N+1,j) = max(S(N+1,j)-K,0); % European call payoff at T
end

for i = N:-1:1 % backward in time
    for j = 1:i
        C(i,j) = exp(-r*dt)*(q*C(i+1,j+1)+(1-q)*C(i+1,j));
    end
end

for i = N:-1:1 % backward in time
    for j = 1:i
        Delta(i,j) = (C(i+1,j+1)-C(i+1,j))/(S(i+1,j+1)-S(i+1,j));
    end
end

toss = zeros(1,N); % biased coin tossing with outcome 0 and 1
for i = 1:N
    if rand < 1-q
        toss(1,i) = 0; % down movement
    else

```

```

        toss(1,i) = 1; % up movement
    end
end

node_level = ones(1,N+1);
node_level(1,1) = 1;
for i = 1:N
    node_level(1,i+1) = node_level(1,i) + toss(1,i);
end

for i = 1:N+1
    S_path(i) = S(i,node_level(i));
end

for i = 1:N+1
    V(i) = C(i,node_level(1,i));
end

for i = 1:N
    Delta_path(i) = Delta(i,node_level(i));
end

% the self-financing condition
for i = 1:N-1
    B(i+1) = exp(r*dt)*B(i)+(Delta_path(i)-Delta_path(i+1))*S_path(i+1);
end

for i = 1:N
    Pi(i) = -V(i) + B(i) + Delta_path(i)*S_path(i);
end
Pi(N+1) = -V(N+1) + exp(r*dt)*B(N) + Delta_path(N)*S_path(N+1);

subplot(3,2,1);
plot([0:1:N-1],toss,'b*')
text(3,-0.5,num2str(1-q,'down probability 1-q = %.3f'))
text(3.9,1.5,num2str(q,'up probability q = %.3f'))
title('Biased Coin Tossing')
set(gca,'xlim',[0,N],'ylim',[-1,2],'xtick',[0:5:N],'ytick',[0:1:1]);

subplot(3,2,2);
% the binomial tree
for i = 1:N+1 % backward in time
    plot((i-1)*dt,S(i,1:i),'Marker','.', 'Color',[0.9 0.5 0.3]);
    hold on
end

```

```

plot([0:dt:T],K*ones(1,N+1),'k-')
hold on
% a sample asset path
plot([0:dt:T],S_path,'ro-');
title('Asset Price S')

subplot(3,2,3);
plot(0:dt:T,V(:,),'bs-');
title('Option Price V')

subplot(3,2,4);
for k = 1:N
    x = (k-1)*dt:dt/10:k*dt;
    y = Delta_path(k)*ones(11,1);
    plot(x,y,'Color','g'); % Plot the graph on each subinterval.
    hold on
    scatter(0:dt:T-dt,Delta_path(1:N),'go','filled')
end
title('Delta \Delta')
set(gca,'xlim',[0,T],'xtick',[0:1:T],'ylim',[0,1]);

subplot(3,2,5);
plot([0:dt:T],zeros(1,N+1),'k--');
hold on
for i=1:N
    x = (i-1)*dt:dt/10:i*dt;
    y = B(i);
    plot(x,y*exp(r*(x-(i-1)*dt)),'r-'); % Plot the graph on each subinterval.
    hold on
end
title('Bank Deposit B')

subplot(3,2,6);
plot([0:dt:T],Pi,'ro-');
hold on
plot([0:dt:T],Pi(1)*exp(r*[0:dt:T]),'k--');
legend('Portfolio \Pi','exponential curve');
title('Portfolio \Pi')

```

Simulation 14.4 (Convergence of the Binomial Tree Method) For the simulation of Theorem 14.1 we take the same set of parameter values for S_0 , K , T , r and σ as in Simulation 14.1. Parameters have to satisfy the condition that $0 < q < 1$, or equivalently, $r\delta t < \sigma\sqrt{\delta t}$. The option price obtained by the binomial tree method converges to the price from the Black–Scholes–Merton formula, represented by the horizontal line in Fig. 14.12, as $M \rightarrow \infty$. The oscillating behavior of the graph is

inherent in the binomial tree method due to the imperfect stability of the numerical algorithm as explained in Chap. 6, Vol. II.

```

S0 = 100;
K = 110;
T = 1;
r = 0.05;
sigma = 0.3;
M_values = [3:1:250];
L = length(M_values);
Call_prices = zeros(L,1);

for j = 1:L
    M = M_values(j);
    dt = T/M;
    u = exp(sigma*sqrt(dt));
    d = exp(-sigma*sqrt(dt));
    q = (exp(r*dt)-d)/(u-d);
    Call = max(S0 * u.^([M:-1:0]') .* d.^([0:1:M]') - K,0);
    for i = M:-1:1
        Call = exp(-r*dt)*(q*Call(1:i)+(1-q)*Call(2:i+1));
    end
    Call_prices(j) = Call;
end
plot(M_values,Call_prices,'.');
hold on;

% the Black-Scholes-Merton formula.
d1 = (log(S0/K)+(r+0.5*sigma^2)*T)/(sigma*sqrt(T));
d2 = d1-sigma*sqrt(T);
Call_BSM = S0*normcdf(d1) - K*exp(-r*T)*normcdf(d2)
yline(Call_BSM,'k');
```

Simulation 14.5 (The Binomial Tree Method in Agreement with the Black–Scholes–Merton Partial Differential Equation) We plot the price of a European call option over the binomial tree for the asset price, and observe that the price obtained by the binomial tree method is close to the solution of the Black–Scholes–Merton partial differential equation. Parameters have to satisfy the condition that $0 < q < 1$. See Fig. 14.13.

```

S0 = 100;
K = 150;
T = 2;
r = 0.05;
sigma = 0.3;
M = 40;
```

```

N = 10; % number of time steps in the binomial tree
dt = T/N;
Smax = 400;
S = linspace(0,Smax,M+1); % Divide the interval into M subintervals.
t = linspace(0,T,M+1);
C = zeros(M+1,M+1);
tau = T-t;
[t,S] = meshgrid(t,S);
d1 = (log(S/K)+(r+0.5*sigma^2)*tau)./(sigma*sqrt(tau));
d2 = d1-sigma*sqrt(tau);
N1 = normcdf(d1);
N2 = normcdf(d2);
C = S.*N1-K*exp(-r*tau).*N2;
C(:,M+1) = max(S(:,M+1)-K,0);

bi_price = zeros(N+1,N+1); % option price by the binomial tree method
u = exp(sigma*sqrt(dt));
d = exp(-sigma*sqrt(dt));
q = (exp(r*dt)-d)/(u-d);

% i for time, j for asset price level
for i = 1:N+1
    for j = 1:i
        S_binom(i,j) = S0*u^(j-1)*d^(i-j); % binomial tree asset price
    end
end

for j = 1:N+1
    bi_price(N+1,j) = max(S_binom(N+1,j)-K,0); % European call payoff at T
end

for i = N:-1:1 % backward in time
    for j = 1:i
        bi_price(i,j) = exp(-r*dt)*(q*bi_price(i+1,j+1)+(1-q)*bi_price(i+1,j));
    end
end

A = zeros(3,(N+1)*(N+2)/2);

for i = 1:N+1
    A(1,(i-1)*i/2+1:(i-1)*i/2+i) = (i-1)*dt;
    A(2,(i-1)*i/2+1:(i-1)*i/2+i) = S_binom(i,1:i);
    A(3,(i-1)*i/2+1:(i-1)*i/2+i) = bi_price(i,1:i);
end

```

```

figure(1);
mesh(t,S,C,'FaceAlpha',0.0);
hold on
scatter3(A(1,:),A(2,:),zeros(1,(N+1)*(N+2)/2),20,'ro','filled');
%title('The binomial tree for the asset');
hold off

figure(2);
hold on
mesh(t,S,C,'FaceAlpha',0.0);
hold on
scatter3(A(1,:),A(2,:),A(3,:),20,'bs','filled');
%title('The binomial tree for the European call price');

```

Simulation 14.6 (American Call Option: the Binomial Tree Method) We use the binomial tree method to price an American option, and observe that the prices of a European call and an American call are equal with the same parameter values. Parameters have to satisfy the condition that $0 < q < 1$. See Exercise 14.2 for a justification. (The same statement is true in general. Consult Theorem 19.1.)

```

S0 = 100;
K = 110;
T = 1;
r = 0.05;
sigma = 0.3;
M = 100; % number of time steps
dt = T/M;

u = exp(sigma*sqrt(dt) + (r-0.5*sigma^2)*dt);
d = exp(-sigma*sqrt(dt) + (r-0.5*sigma^2)*dt);
q = (exp(r*dt)-d)/(u-d);

Call_Am = max(S0*u.^([M:-1:0]')*.d.^([0:1:M]')- K,0); % payoff at T
Call_Eu = max(S0*u.^([M:-1:0]')*.d.^([0:1:M]')- K,0); % payoff at T

% We proceed backward to compute option value at time 0.
for i = M:-1:1
    S = S0*u.^([i-1:-1:0]')*.d.^([0:1:i-1]');
    Call_Am = max(max(S-K,0),exp(-r*dt)*(q*Call_Am(1:i)+(1-q)*Call_Am(2:i+1)));
    Call_Eu = max(max(S-K,0),exp(-r*dt)*(q*Call_Eu(1:i)+(1-q)*Call_Eu(2:i+1)));
end
Call_Am
Call_Eu
d1 = (log(S0/K)+(r+0.5*sigma^2)*T)/(sigma*sqrt(T));
d2 = d1-sigma*sqrt(T);
Call_BSM = S0*normcdf(d1) - K*exp(-r*T)*normcdf(d2)

```

The output is

```
Call_Am = 10.0473
Call_Eu = 10.0473
Call_BSM = 10.0201
```

Exercises

14.1 In the single-period binomial tree model given in Sect. 14.2, find the hedge ratio Δ of a European put option in the following cases: (i) $K > S^u$, (ii) $S^d < K < S^u$, and (iii) $K < S^d$.

14.2 An American option is an option that can be exercised before expiry date. Its early exercise possibility makes it at least as valuable as the European option with the same underlying asset and expiry date. Show that the American call price is equal to the European call price in the binomial tree model. (Hint: Show that the American call price is equal to the European call price in the single period model.)

14.3 Suppose that we want to use the simple interest rate r in Sect. 14.2, i.e., $B_T = (1 + r)^T B_0$. Show that under the assumption that $S^u = uS_0$ and $S^d = dS_0$ we have

$$V_0 = \frac{1}{1+r} \left(\frac{1+r-d}{u-d} V^u + \frac{u-(1+r)}{u-d} V^d \right).$$

In this case, the no arbitrage principle implies that $0 < d < 1 + r < u$.

14.4 (Greeks) We compute the Greeks of an option by the binomial tree model. For example, the binomial trees in Figs. 14.17 and 14.18 represent asset price and option price respectively, where the number of time periods equals $N = 4$ and the time step size equals $\delta t = \frac{T}{N}$. In the notations S_i^j and V_i^j the subscript i denotes the time $0 \leq i\delta t \leq N\delta t = T$ and the superscript j , $0 \leq j \leq N$, denotes the level of the asset price. To approximate mathematical derivatives we use various finite differences.

- (i) Show that Delta at the node (i, j) , $0 \leq i \leq N - 1$, can be obtained by the formula

$$\Delta_i^j = \frac{V_{i+1}^{j+1} - V_{i+1}^j}{S_{i+1}^{j+1} - S_{i+1}^j}.$$

- (ii) Show that Gamma at the node (i, j) , $0 \leq i \leq N - 2$, can be obtained by the formula

$$\Gamma_i^j = \frac{\Delta_{i+1}^{j+1} - \Delta_{i+1}^j}{S_{i+1}^{j+1} - S_{i+1}^j},$$

where Δ_{i+1}^{j+1} and Δ_{i+1}^j can be obtained by (i). Or, we may use the second order central difference.

Fig. 14.17 Asset price S in the binomial tree model

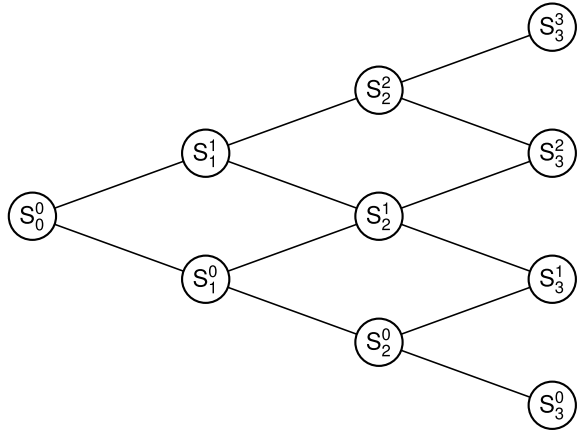
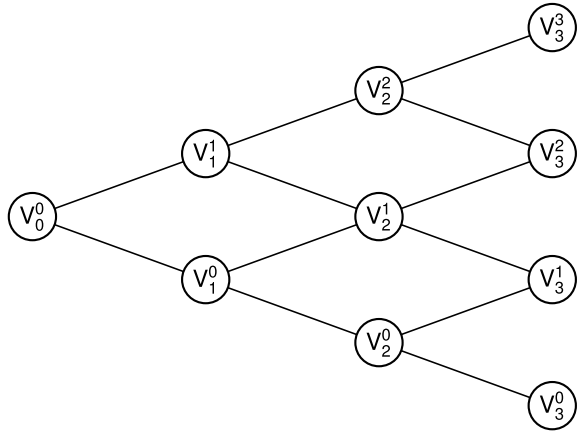


Fig. 14.18 Option price V in the binomial tree model



- (iii) Suppose that the binomial tree for asset is multiplicative, i.e., $S_{i+1}^j = dS_i^j$ and $S_{i+1}^{j+1} = uS_i^j$ for some $0 < d < u$ and satisfies the condition that $ud = 1$. Show that Theta at the node (i, j) , $0 \leq i \leq N - 2$, can be obtained by

$$\Theta_i^j = \frac{V_{i+2}^{j+1} - V_i^j}{2\delta t}.$$

Chapter 15

The Black–Scholes–Merton Differential Equation



In this chapter we derive the Black–Scholes–Merton partial differential equation, solve it, and investigate the properties of its solution. Historically, the simultaneous publications of Black and Scholes (Black and Scholes, 1973) and Merton (Merton, 1973) in 1973 mark the beginning of the theory of option pricing. Using stochastic calculus, they derived the so-called Black–Scholes–Merton differential equation. It is essentially a heat equation with the direction of time reversed. To help the readers to understand the derivation of the Black–Scholes–Merton equation, we include computer simulations of self-financing portfolios, combining the concepts of geometric Brownian motion, delta hedging, the Black–Scholes–Merton formula and exponential growth of risk-free portfolios. A stress test for the Black–Scholes–Merton formula is presented to check whether the formula makes sense in extreme cases of financial market. Also included is the Laplace Transform method to solve the Black–Scholes–Merton partial differential equation. For a more detailed introduction to the Black–Scholes–Merton equation consult (Bjork, 2009; Capinski and Kopp, 2012; Shreve, 2004b; Steele, 2001).

15.1 Derivation of the Black–Scholes–Merton Differential Equation

Here is a list of the assumptions for the Black–Scholes–Merton model.

Assumptions on the Underlying Asset

- (i) The asset follows a geometric Brownian motion with constant volatility.
- (ii) There are no dividends or stock splits.

Assumptions on the Financial Market

- (i) (Liquidity) It is possible to buy and sell any amount of the asset at any time.
- (ii) The bid and the ask prices are equal, i.e., the bid–ask spread is zero.
- (iii) There are no transaction costs or taxes.
- (iv) Short selling is allowed without any cost. Borrowing money is possible at any time.
- (v) The risk-free interest rate is known and constant.

These assumptions constitute a simplified model of the financial market, and we need take into account the real market behavior. For example, the distribution of return of the asset has fatter tails than the normal distribution, and the volatility can change. Furthermore, in time of market turmoil liquidity is not guaranteed. The bid and ask prices differ, and there are costs and taxes. In a financial crisis short selling might not be allowed, especially for the stocks of financial companies. Finally, the interest rate can be random.

Recall the general version of Itô's lemma. (See Theorem 10.3). Its idea can be summarized in the following heuristic argument. Since the option price V depends on t and S we will write $V(S, t)$. As time changes from t to $t + \delta t$, V also changes. Using the second order Taylor series expansion we obtain

$$\delta V = \frac{\partial V}{\partial t} \delta t + \frac{\partial V}{\partial S} \delta S + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} (\delta S)^2 + \frac{\partial^2 V}{\partial S \partial t} \delta S \delta t + \frac{1}{2} \frac{\partial^2 V}{\partial t^2} (\delta t)^2 .$$

The increment δt is close to 0, and we ignore any term whose order is greater than 1 by Itô's lemma. Consult Tables 10.1 and 10.2. Since

$$(\delta S)^2 = \sigma^2 S^2 (\delta W)^2 + \text{higher order terms} \approx \sigma^2 S^2 \delta t ,$$

we can classify the increments as follows:

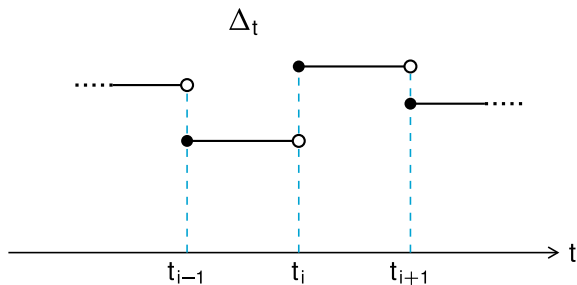
$$\delta V = \underbrace{\left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) \delta t}_{\text{risk-free}} + \underbrace{\frac{\partial V}{\partial S} \delta S}_{\text{risky}} . \quad (15.1)$$

Now we are ready to derive a partial differential equation called the Black–Scholes–Merton equation for a European option price V . Throughout the book, an uppercase delta (Δ) represents the number of shares needed in hedging, not the increment as in calculus and classical analysis. To avoid the possibility of confusion, we use a lowercase delta (δ) to denote the increment in stochastic calculus.

Theorem 15.1 (Black–Scholes–Merton Equation) *The price $V(S, t)$ of a European option at time t with expiry date T and strike price K satisfies*

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} = rV .$$

Fig. 15.1 Adjustment of Δ_t in discrete time hedging



The final condition $V(S, T)$ is given by the payoff function of the option.

Proof To derive a partial differential equation for V we take the viewpoint of an option seller who manages a portfolio based on the idea of hedging to eliminate asset price risk. More precisely, we construct a portfolio Π that is self-financing and risk-free as follows:

$$\Pi(S, t) = -V(S, t) + D(S, t) + \Delta(S, t)S.$$

In other words, Π consists of an option that has been sold, a bank deposit or risk-free asset D , and Δ shares of risky asset S . Here Δ is a function of t and S , and is called the hedge ratio. If $\Delta < 0$, it represents short selling. (The reason why we have $-V$ in the portfolio V is that we have one sold contract of an option, it is regarded as a debt).

As a fund manager we maintain the same number of shares of a stock from time t to $t + \delta t$, which fits common sense. More precisely, since we do not know how much S would change, we wait until we obtain the information on the stock value at time $t + \delta t$, and make an investment decision upon that information. Suppose that while the hedge ratio Δ_t is fixed, S_t changes to $S_{t+\delta t}$, and Π_t changes to $\Pi_t + \delta\Pi$. (See Fig. 15.1).

The risk-free asset D_t gains interest equal to the amount

$$\delta D_t = r D_t \delta t.$$

Hence (15.1) implies that

$$\begin{aligned} \delta\Pi &= -\delta V + \delta D + \Delta \delta S \\ &= -\delta V + r D \delta t + \Delta \delta S \\ &= \left(-\frac{\partial V}{\partial t} - \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + r D \right) \delta t + \left(\Delta - \frac{\partial V}{\partial S} \right) \delta S. \end{aligned}$$

If we take

$$\Delta_t = \frac{\partial V}{\partial S}(t, S_t)$$

The diagram shows the equation $\delta\Pi = -\delta V + rD\delta t + \Delta\delta S = r\Pi\delta t$. Annotations include:

- An upward arrow from $\delta\Pi$ labeled "increment".
- A bracket under $-\delta V$ with a downward arrow labeled "Itô lemma".
- A bracket under $rD\delta t$ with a downward arrow labeled "risk-free interest".
- A bracket under $\Delta\delta S$ with a downward arrow labeled "geometric Brownian motion".
- An upward arrow from Δ labeled "hedge ratio".
- An upward arrow from $r\Pi\delta t$ labeled "risk-free portfolio".

Fig. 15.2 Components of $\delta\Pi$ in Delta hedging where $\Pi = -V + D + \Delta S$

for every t , then we have

$$\delta\Pi = \left(-\frac{\partial V}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rD \right) \delta t. \quad (15.2)$$

Now the δS term has disappeared and $\delta\Pi$ is risk-free, and hence it is equivalent to a bank deposit for a time duration δt . Thus we obtain

$$\delta\Pi = r \Pi \delta t \quad (15.3)$$

by the no arbitrage principle. Since the right-hand sides of (15.2) and (15.3) are equal, we have

$$\left(-\frac{\partial V}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rD \right) \delta t = r \Pi \delta t.$$

(See Fig. 15.2). Therefore

$$\left(-\frac{\partial V}{\partial t} - \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rD \right) = r \left(-V + D + \frac{\partial V}{\partial S} S \right)$$

and we obtain the Black–Scholes–Merton equation after canceling the rD terms. $\square$

The hedging method introduced in the proof of Theorem 15.1 is called Delta hedging. For a different derivation of the Black–Scholes–Merton equation based on the martingale method see Sect. 16.4. For the logical implications among the facts related to option pricing, see the diagram in Fig. 2.7.

Definition 15.1 Options can be classified according to their current market price S_t in comparison with the strike price K .

- (i) An *in-the-money* (ITM) option is an option that would have positive value if exercised now. For example, if the asset price is above the strike price then a call option is in-the-money, i.e., $S_t > K$.

- (ii) An *out-of-the-money* (OTM) option is an option that would have zero value if it is exercised now. For example, if the asset price is below the strike price then a call option is out-of-the-money, i.e., $S_t \leq K$.
- (iii) An *at-the-money* (ATM) option is an option for which the current market price of the underlying asset is equal to the strike price, $S_t = K$.

For a simulation of Delta hedging for a European call option consult Simulation 15.1 where we take $T = 5$, $S_0 = 10$, $K = 15$ for the option contract. For the simulation of the geometric Brownian motion S_t we choose $\mu = 0.14$, $\sigma = 0.3$ and $r = 0.1$. To make exponential growth of Π more visible in the sixth panel, we choose a sufficiently large number for r .

In Figs. 15.3 and 15.4 for the in-the-money case and out-of-the-money case, respectively, the six panels in each case show sample paths of the Brownian motion W_t , the geometric Brownian motion S_t , the European call option price V_t given in Theorem 15.2, the hedge ratio Δ_t , the amount of bank deposit D_t , and finally the portfolio $\Pi_t = -V_t + D_t + \Delta_t S_t$, which is risk-free after being Delta-hedged and satisfies $\Pi_t = \Pi_0 e^{rt}$, $0 \leq t \leq T$. Consult Theorem 15.3 for the formula of $\Delta = N(d_1)$, where $N(x)$ is the cumulative density function of the standard normal distribution and

$$d_1 = \frac{\log \frac{S}{K} + (r + \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}}.$$

Since $0 \leq \Delta \leq 1$, the hedge ratio is between 0 and 1, as shown in the fourth panel in each of Figs. 15.3 and 15.4. Depending on whether $S_t > K$ as t nears the expiry date T , the option price V_t stays positive (ITM) or falls to zero (OTM), and Δ converges to either 1 or 0. At $t \approx T$, if the option is in-the-money, then keeping $\Delta \approx 1$ shares will hedge the risk, and if the option is out-of-the-money, then there is no need to hedge and we have $\Delta \approx 0$.

In Fig. 15.5 we simulate the at-the-money option using the Brownian bridge X_t such that $S_0 e^{(\mu - \frac{1}{2}\sigma^2)T + \sigma X_T} = K$. To generate a sample asset path such that

$$S_T = S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma W_T\right) = K,$$

we use a Brownian bridge path $X_t^{0 \rightarrow b}$ from 0 to b , where b satisfies

$$S_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma b\right) = K,$$

i.e.,

$$b = \frac{1}{\sigma} \left\{ \log \frac{K}{S_0} - \left(\mu - \frac{1}{2}\sigma^2\right)T \right\}.$$

See Sect. 12.3.

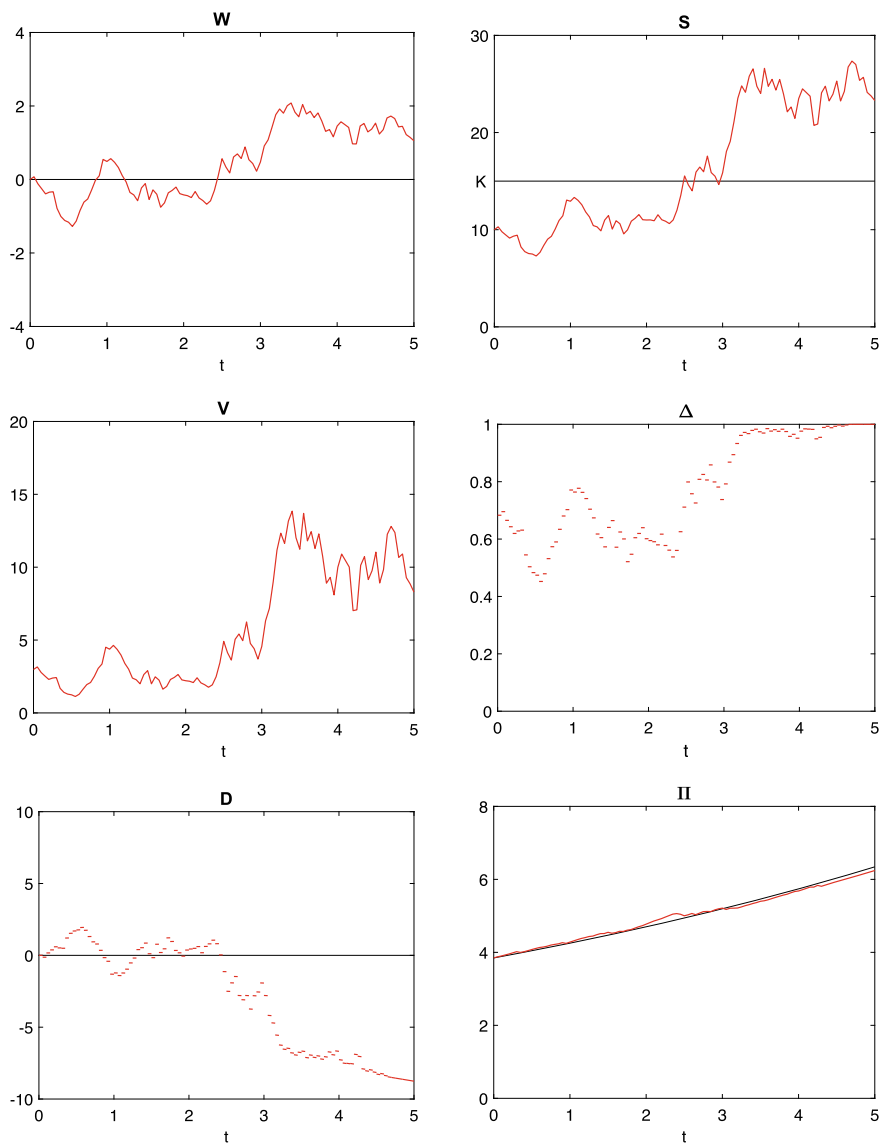


Fig. 15.3 In-the-money European call option: Delta hedging using a self-financing portfolio $\Pi = -V + D + \Delta S$ (from top left to bottom right)

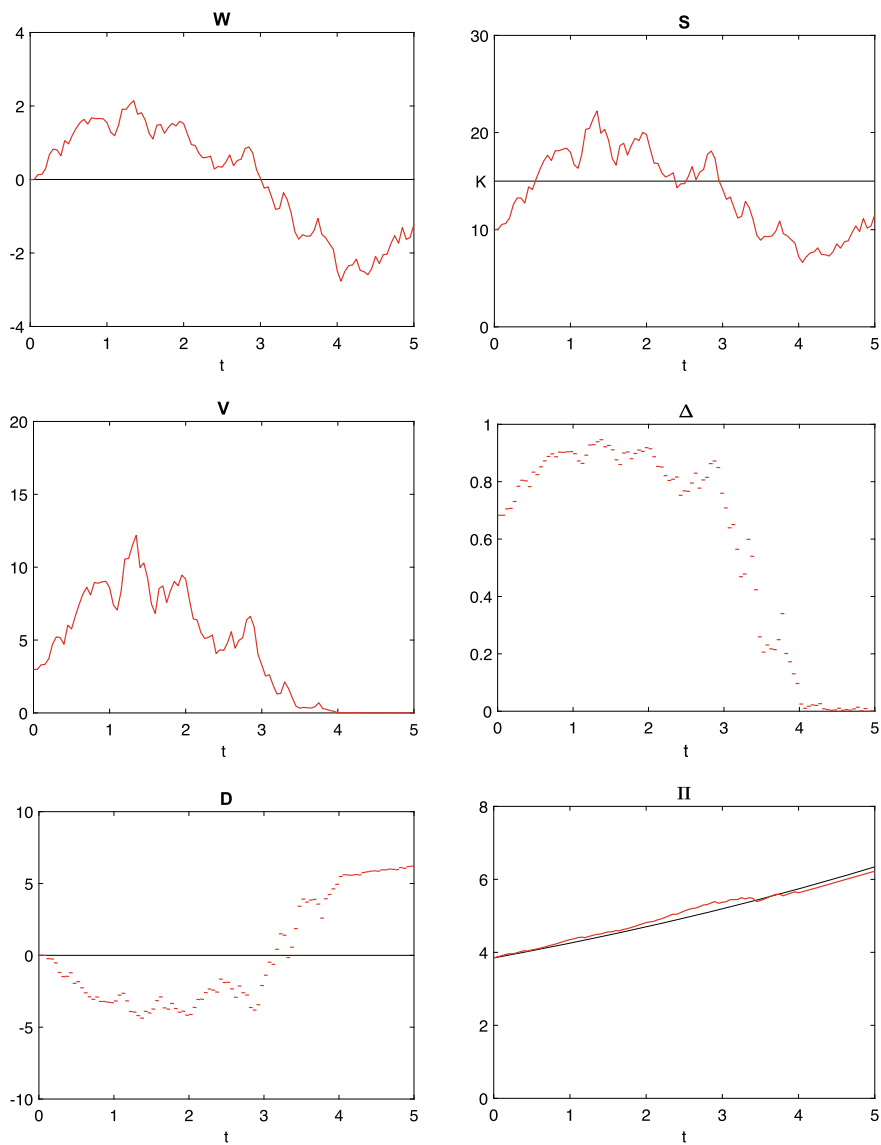


Fig. 15.4 Out-of-the-money European call option: Delta hedging using a self-financing portfolio $\Pi = -V + D + \Delta S$ (from top left to bottom right)

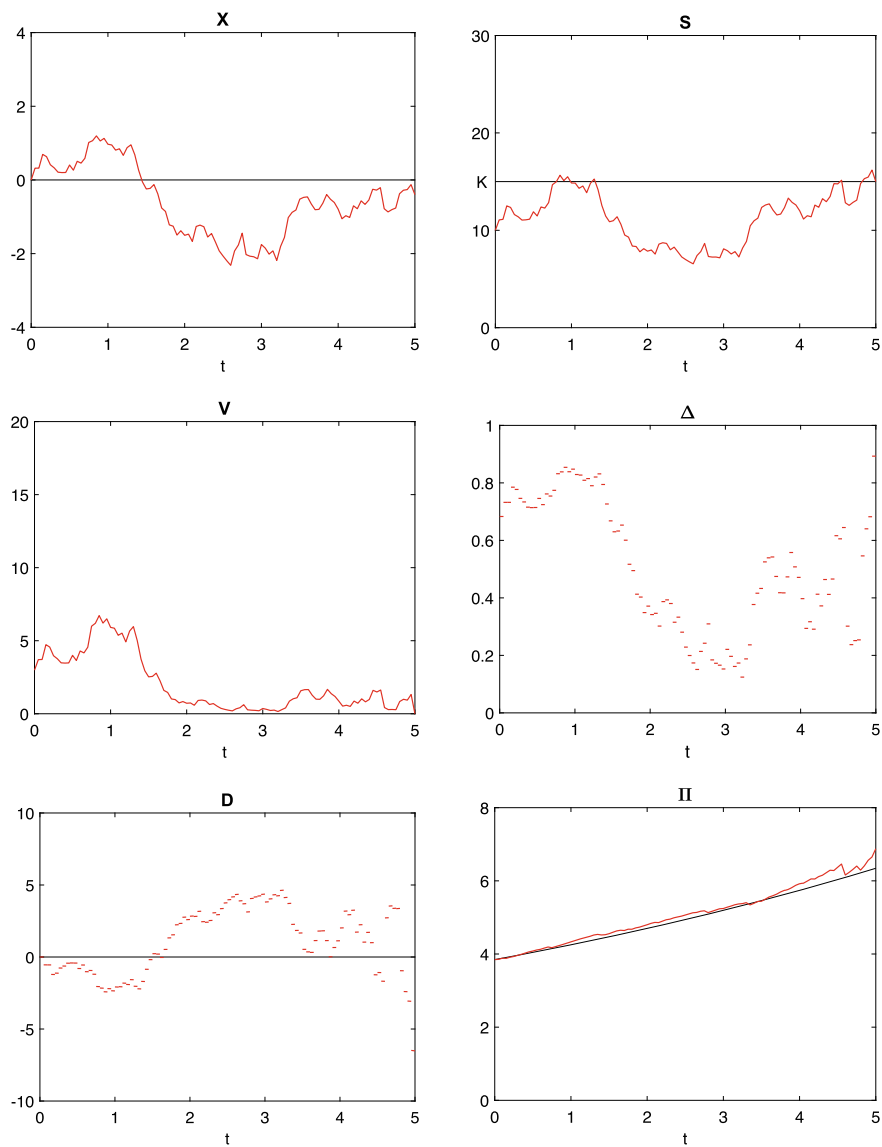


Fig. 15.5 At-the-money European call option: Delta hedging using a self-financing portfolio $\Pi = -V + D + \Delta S$ (from top left to bottom right)

15.2 Price of a European Call Option

Now we solve the Black–Scholes–Merton equation and find the price $V(S, t)$ of a European call option with expiry T . The domain of V is given by $0 \leq t \leq T$, $0 \leq S < \infty$. The boundary conditions are given by

$$\begin{aligned} V(0, t) &= 0, \\ V(S, t) &\approx S - K \end{aligned}$$

for sufficiently large S , and the final condition is given by

$$V(S, T) = \max\{S - K, 0\}.$$

(See Fig. 15.6). The Black–Scholes–Merton equation resembles the heat equation

$$\frac{\partial V}{\partial t} = \frac{1}{2} \frac{\partial^2 V}{\partial S^2}$$

except for the sign of the time derivative, and instead of an initial condition at time $t = 0$ a final condition at $t = T$ is given. The reason is that the payoff at the expiry is fixed from the beginning of the life of an option.

Theorem 15.2 (Black–Scholes–Merton Formula) *The price V_t at time $0 \leq t \leq T$ of a European call option with expiry T and strike price K is given by*

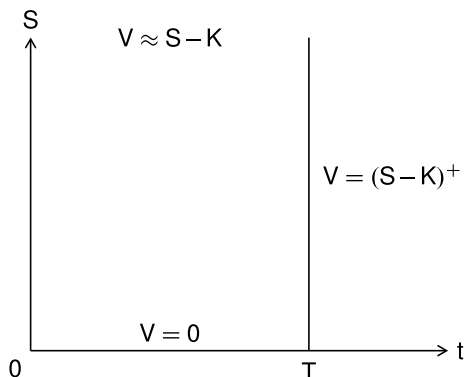
$$V_t = S_t N(d_1) - K e^{-r(T-t)} N(d_2)$$

where

$$d_{1,2} = \frac{\log \frac{S_t}{K} + (r \pm \frac{1}{2} \sigma^2)(T - t)}{\sigma \sqrt{T - t}}$$

and $N(\cdot)$ denotes the standard normal cumulative distribution function.

Fig. 15.6 The boundary condition of the Black–Scholes–Merton equation for a European call option



Proof We convert the Black–Scholes–Merton equation into the heat equation with an initial condition going through several steps of change of variables. Put

$$\begin{cases} S &= K e^x \\ t &= T - \frac{1}{\sigma^2} \tau \\ V(S, t) &= K v(x, \tau) . \end{cases}$$

In other words,

$$v(x, \tau) = \frac{1}{K} V(S, t) = \frac{1}{K} V(K e^x, T - \frac{1}{\sigma^2} \tau) .$$

Then $-\infty < x < \infty$, $0 \leq \tau \leq \sigma^2 T$ and

$$\sigma^2 \frac{\partial v}{\partial \tau} = \frac{1}{2} \sigma^2 \frac{\partial^2 v}{\partial x^2} + \left(r - \frac{1}{2} \sigma^2 \right) \frac{\partial v}{\partial x} - r v .$$

If we let $C = \frac{r}{\sigma^2}$, then

$$\frac{\partial v}{\partial \tau} = \frac{1}{2} \frac{\partial^2 v}{\partial x^2} + \left(C - \frac{1}{2} \right) \frac{\partial v}{\partial x} - C v .$$

Since $V(S, T) = \max\{S - K, 0\}$, the initial condition for v is given by

$$v(x, 0) = \frac{1}{K} V(K e^x, T) = \frac{1}{K} \max\{K e^x - K, 0\} = \max\{e^x - 1, 0\} .$$

Observe that the parameters K , T , σ^2 , r have disappeared, and only one parameter C remains in the equation. When several parameters appear in one equation simultaneously and make the equation look complicated, we can sometimes combine some of the parameters into one new parameter so that the new equation becomes simpler and easier to solve. If such a new parameter is dimensionless, then we may say that it is an essential parameter in describing the equation. For example, the dimensions of r and σ^2 are both time^{-1} , and hence $C = r/\sigma^2$ is dimensionless.

Now we look for v of the form

$$v(x, \tau) = e^{\alpha x + \beta \tau} u(x, \tau)$$

for some α , β , $u(x, \tau)$. If such a solution exists, then we have

$$\beta u + \frac{\partial u}{\partial \tau} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} + \left(C - \frac{1}{2} + \alpha \right) \frac{\partial u}{\partial x} + \left(\frac{1}{2} \alpha^2 + (C - \frac{1}{2}) \alpha - C \right) u .$$

If we take

$$\alpha = -C + \frac{1}{2}$$

and

$$\beta = \frac{1}{2}\alpha^2 + (C - \frac{1}{2})\alpha - C = -\frac{1}{2}(C + \frac{1}{2})^2,$$

then a necessary condition for u becomes the heat equation or diffusion equation

$$\frac{\partial u}{\partial \tau} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}.$$

Note that the initial condition for u is given by

$$u(x, 0) = u_0(x) = e^{-\alpha x} v(x, 0) = e^{(C - \frac{1}{2})x} \max\{e^x - 1, 0\}.$$

Hence

$$\begin{aligned} u(x, \tau) &= \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^{\infty} u_0(\xi) e^{-(x-\xi)^2/(2\tau)} d\xi \\ &= \frac{1}{\sqrt{2\pi\tau}} \int_0^{\infty} e^{(C - \frac{1}{2})\xi} (e^{\xi} - 1) e^{-(x-\xi)^2/(2\tau)} d\xi. \end{aligned}$$

(Consult Sect. E.2 in Appendix E). Put

$$I_1, I_2 = \frac{1}{\sqrt{2\pi\tau}} \int_0^{\infty} e^{(C \pm \frac{1}{2})\xi} e^{-(x-\xi)^2/(2\tau)} d\xi,$$

then $u(x, \tau) = I_1 - I_2$. Note that

$$\begin{aligned} & e^{\alpha x + \beta \tau} I_1 \\ &= e^{(-C + \frac{1}{2})x - \frac{1}{2}(C + \frac{1}{2})^2 \tau} \frac{1}{\sqrt{2\pi\tau}} \int_0^{\infty} e^{-(x-\xi)^2/(2\tau) + (C + \frac{1}{2})\xi} d\xi \quad (\text{take } \eta = -\xi) \\ &= e^{(-C + \frac{1}{2})x - \frac{1}{2}(C + \frac{1}{2})^2 \tau} \frac{1}{\sqrt{2\pi\tau}} \int_0^{-\infty} e^{-(x+\eta)^2/(2\tau) - (C + \frac{1}{2})\eta} (-1) d\eta \\ &= e^{(-C + \frac{1}{2})x - \frac{1}{2}(C + \frac{1}{2})^2 \tau} \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^0 e^{-(x+\eta)^2/(2\tau) - (C + \frac{1}{2})\eta} d\eta \\ &= e^{(-C + \frac{1}{2})x - \frac{1}{2}(C + \frac{1}{2})^2 \tau} \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^0 e^{-(\eta+x+(C+\frac{1}{2})\tau)^2/(2\tau) + (C+\frac{1}{2})x + \frac{1}{2}(C+\frac{1}{2})^2 \tau} d\eta \\ &= e^x \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^0 e^{-(\eta+x+(C+\frac{1}{2})\tau)^2/(2\tau)} d\eta \end{aligned}$$

$$\begin{aligned}
&= e^x \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^{x+(C+\frac{1}{2})\tau} e^{-\zeta^2/(2\tau)} d\zeta \quad (\text{take } \zeta = \frac{\zeta}{\sqrt{\tau}}) \\
&= e^x \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x+(C+\frac{1}{2})\tau)/\sqrt{\tau}} e^{-\zeta^2/2} d\zeta = e^x N(d_1)
\end{aligned}$$

where

$$d_1 = \frac{x + (C + \frac{1}{2})\tau}{\sqrt{\tau}} = \frac{\log \frac{S}{K} + (\frac{r}{\sigma^2} + \frac{1}{2})\sigma^2(T-t)}{\sigma\sqrt{T-t}}.$$

Similarly, for the second term we obtain

$$\begin{aligned}
&e^{ax+\beta\tau} I_2 \\
&= e^{(-C+\frac{1}{2})x - \frac{1}{2}(C+\frac{1}{2})^2\tau} \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^0 e^{-(\eta+x+(C-\frac{1}{2})\tau)^2/2\tau + (C-\frac{1}{2})x + \frac{1}{2}(C-\frac{1}{2})^2\tau} d\eta \\
&= e^{-C\tau} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x+(C-\frac{1}{2})\tau)/\sqrt{\tau}} e^{-\eta^2/2} d\zeta \\
&= e^{-C\tau} N(d_2),
\end{aligned}$$

where

$$d_2 = \frac{x + (C - \frac{1}{2})\tau}{\sqrt{\tau}} = \frac{\log \frac{S}{K} + (\frac{r}{\sigma^2} - \frac{1}{2})\sigma^2(T-t)}{\sigma\sqrt{T-t}}.$$

Since $\tau = \sigma^2(T-t)$, we have

$$C\tau = \frac{r}{\sigma^2}\sigma^2(T-t) = r(T-t).$$

Hence

$$\begin{aligned}
V(S, t) &= Kv(x, \tau) \\
&= Ke^x N(d_1) - Ke^{-C\tau} N(d_2) \\
&= SN(d_1) - Ke^{-r(T-t)} N(d_2).
\end{aligned}$$

Corollary 15.1 *The price of a European put option is given by*

$$\begin{aligned}
P(S_t, t) &= S_t N(d_1) - Ke^{-r(T-t)} N(d_2) - S_t + Ke^{-r(T-t)} \\
&= Ke^{-r(T-t)} N(-d_2) - S_t N(-d_1).
\end{aligned}$$

Proof Apply Theorem 2.1. □

Remark 15.1 The Black–Scholes–Merton formula given in Theorem 15.2 is still meaningful in some trivial cases. We check the European call option case.

- (i) If $T = 0$, then the payoff $(S_0 - K)^+$ should agree with the formula given by $S_0 N(d_1) - K N(d_2)$ since

$$\begin{aligned}
\lim_{T \rightarrow +0} d_1, d_2 &= \lim_{T \rightarrow +0} \frac{\log \frac{S_T}{K} + (r \pm \frac{1}{2}\sigma^2)T}{\sigma \sqrt{T}} \\
&= \lim_{T \rightarrow +0} \frac{\log \frac{S_0}{K}}{\sigma \sqrt{T}} \\
&= \begin{cases} +\infty, & S_0 > K, \\ 0, & S_0 = K, \\ -\infty, & S_0 < K, \end{cases}
\end{aligned}$$

and since

$$\lim_{T \rightarrow +0} N(d_1) = \lim_{T \rightarrow +0} N(d_2) = \begin{cases} 1, & S_0 > K, \\ \frac{1}{2}, & S_0 = K, \\ 0, & S_0 < K. \end{cases}$$

- (ii) If $S_0 = 0$ then $S_t = 0$ for every $0 \leq t \leq T$, and hence the payoff is 0 for any K . Therefore $V_t = 0$, which can be seen from the formula since

$$\lim_{S_0 \rightarrow 0+} d_1 = \lim_{S_0 \rightarrow 0+} d_2 = -\infty$$

and

$$\lim_{S_0 \rightarrow 0+} N(d_1) = \lim_{S_0 \rightarrow 0+} N(d_2) = 0.$$

- (iii) If $K = 0$, then the payoff is equal to the stock price at expiry date, and hence $V_T = S_T$, thus $V_t = S_t$ for every $0 \leq t \leq T$. This is obvious from the formula since

$$\lim_{K \rightarrow 0+} d_1 = +\infty$$

and

$$\lim_{K \rightarrow 0+} N(d_1) = 1.$$

- (iv) If $\sigma = 0$, then there is no risk in stock investment and $\mu = r$. Note that $S_t = S_0 e^{rt}$ is deterministic. Hence

$$V_T = (S_T - K)^+ = (S_0 e^{rT} - K)^+$$

is also deterministic, and

$$V_t = e^{-r(T-t)} (S_0 e^{rT} - K)^+.$$

This is obvious from the formula since if $S_t > K e^{-r(T-t)}$ then

$$\lim_{\sigma \rightarrow 0+} d_1 = \lim_{\sigma \rightarrow 0+} d_2 = +\infty,$$

$$\lim_{\sigma \rightarrow 0+} N(d_1) = \lim_{\sigma \rightarrow 0+} N(d_2) = 1 ,$$

thus $V_t = S_t - Ke^{-r(T-t)}$ for every $0 \leq t \leq T$. If $S_t < Ke^{-r(T-t)}$ then

$$\lim_{\sigma \rightarrow 0+} d_1 = \lim_{\sigma \rightarrow 0+} d_2 = -\infty$$

$$\lim_{\sigma \rightarrow 0+} N(d_1) = \lim_{\sigma \rightarrow 0+} N(d_2) = 0 ,$$

and $V_t = 0$. See also Sect. 5.4, Vol. II.

For the European put options we proceed similarly.

Remark 15.2 If $T \approx +\infty$, then a little change in t would not affect the option price V_t at time $t < T$, and we conclude that $\frac{\partial V}{\partial t} \approx 0$. Thus V approximately satisfies the Cauchy–Euler equation

$$\cancel{\frac{\partial V}{\partial t}} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0.$$

We look for solutions of the form $V = S^\lambda$. (See Sect. D.3). Then we obtain the quadratic equation

$$\frac{1}{2}\sigma^2 \lambda(\lambda - 1) + r\lambda - r = 0$$

and obtain

$$\lambda = 1 \quad \text{and} \quad \lambda = -\frac{2r}{\sigma^2} .$$

Thus the general solution is of the form

$$c_1 S + c_2 S^{-2r/\sigma^2} ,$$

where the constants c_1 and c_2 are determined by boundary conditions. For example, as $S \rightarrow \infty$ we have $V \approx S$ for a call and $V \approx 0$ for a put. Hence $c_1 = 1$ for a European call and $c_1 = 0$ for a European put. Furthermore, for a call we have $V = 0$ for $S = 0$, and hence $c_2 = 0$. (See also Theorem 19.3 for a perpetual American put option).

In Figs. 15.7 and 15.8 the call price and put price surfaces are plotted for the values of $0 \leq t \leq 1$, $0 \leq S_0 \leq 20$, where we take strike price $K = 10$, interest rate $r = 0.05$ and volatility $\sigma = 0.3$.

In Fig. 15.9 we take present date $t_0 = 0$, expiry date $T = \frac{1}{12}$, 1, 2 and plot the graphs on the interval $40 \leq S \leq 160$, where $K = 100$, $r = 0.05$ and $\sigma = 0.3$.

In Fig. 15.10 we plot the option prices as functions of time to expiry $T - t$ for various strike prices.

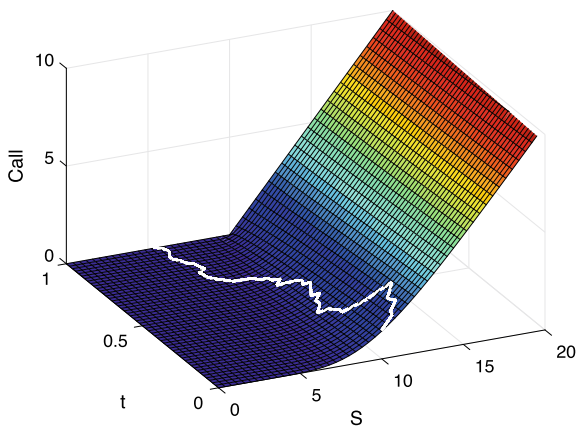


Fig. 15.7 Price of a European call option

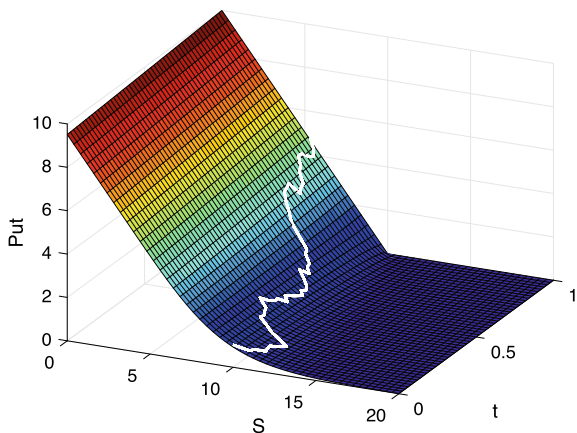


Fig. 15.8 Price of a European put option

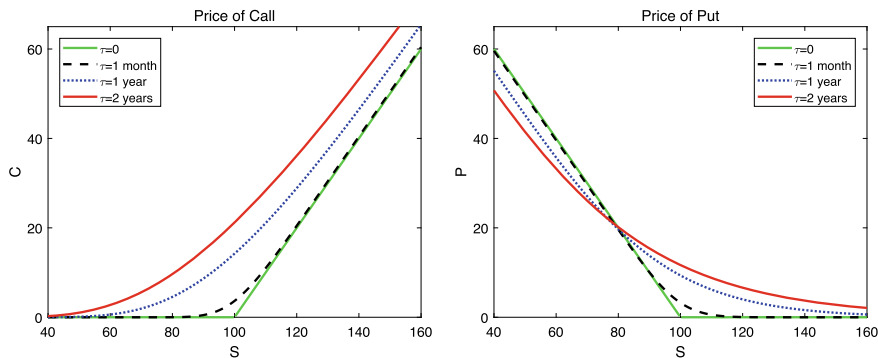


Fig. 15.9 Prices of European call and put options as functions of asset price with different times to expiration $\tau = T - t$

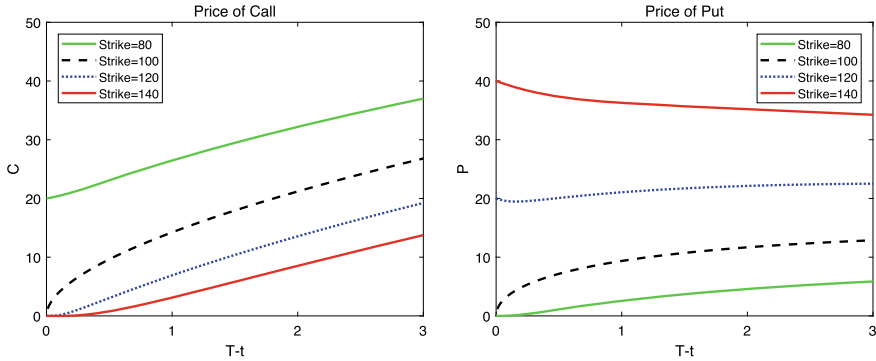


Fig. 15.10 Prices of European call and put options as functions of time to expiry $\tau = T - t$ with different strike prices

15.3 Stress Test for the Black–Scholes–Merton Formula

Stress test is used to test the resilience of financial institutions and investment portfolios against various possibilities in the future. In times of financial turmoil such as financial crunch or hyperinflation when previously unforeseen extreme events of small probability can happen, mathematical models might not agree with economic common sense. By choosing extreme values for the parameters in the Black–Scholes–Merton model we check whether the model can still endure the stress test. First, we check the behavior of the Black–Scholes–Merton formula for extremely large values of volatility and interest rate, and next, check the Black–Scholes–Merton partial differential equation itself.¹

Extreme Cases for the Black–Scholes–Merton Formula

- (i) If $\sigma \approx +\infty$, then $d_1 \approx +\infty$ and $d_2 \approx -\infty$. Hence $N(d_1) \approx 1$ and $N(d_2) \approx 0$. Thus the Black–Scholes–Merton formula for a European call option price is approximately equal to S_t . The conclusion can be interpreted as follows: Since the volatility is very high, the underlying asset may take very large values easily and the payoff at expiry is close to the asset price S_T , and its price at t is nothing but S_t .
- (ii) If $r \approx +\infty$, then $d_1 \approx +\infty$ and $d_2 \approx +\infty$. Hence $N(d_1) \approx 1$ and $N(d_2) \approx 1$. Thus the Black–Scholes–Merton formula for a European call option price is approximately equal to S_t . During hyperinflation the strike price at expiry T amounts to almost zero at time $t < T$, and the payoff approximately equals S_T , whose price at t is S_t .

¹ The presentation given here is for theoretical curiosity, not for real financial situations.

Extreme Cases for the Black–Scholes–Merton Equation

- (i) For $\sigma \approx +\infty$, we can ignore relatively small terms in the following:

$$\cancel{\frac{\partial V}{\partial t}}^0 + \cancel{\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}}^0 + rS \cancel{\frac{\partial V}{\partial S}}^0 = rV^0.$$

Then we obtain

$$\cancel{\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}}^0 = 0,$$

which has a general solution of the form

$$V(t, S) = c_1(t)S + c_2(t)$$

for some $c_1(t)$ and $c_2(t)$. Since $V(t, S) \approx S$ for $S \approx +\infty$ for a European call option, we have $c_1(t) = 1$ and $c_2(t) = 0$, and hence $V(t, S) = S$ in agreement with Theorem 15.2. For a European put option, we have $V(t, S) \approx 0$ for $S \approx +\infty$, and $c_1(t) = 0$. Hence $V(t, S) = c_2(t)$ for every $S \geq 0$. Since $V(t, S) = V(t, 0)$, we conclude that $V(t, S) = Ke^{r(T-t)}$ in agreement with Corollary 15.1.

- (ii) For $r \approx +\infty$, we can ignore relatively small terms in the following:

$$\cancel{\frac{\partial V}{\partial t}}^0 + \cancel{\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2}}^0 + rS \frac{\partial V}{\partial S} = rV.$$

Then we obtain

$$\cancel{S} \frac{\partial V}{\partial S} = \cancel{V}.$$

To solve

$$S \frac{\partial V}{\partial S} = V,$$

we use separation of variables, and obtain

$$\log V = \log S + c_1(t),$$

that is,

$$V(t, S) = C(t)S$$

for some $C(t)$. Since $V(t, S) \approx S$ for $S \approx +\infty$ for a European call option, we have $C(t) = 1$, and hence $V(t, S) = S$ in agreement with Theorem 15.2. For a European put option, we have $V(t, S) \approx 0$ for $S \approx +\infty$, and $C(t) = 0$. Hence $V(t, S) = 0$ in agreement with Corollary 15.1.

15.4 Greeks

The formula for the option price contains several parameters, and we can take partial derivatives with respect to these parameters. Those partial derivatives of the option price are denoted by various Greek letters, called *Greeks*, and we list some of them in the following:

$$\begin{aligned}\Delta &= \frac{\partial V}{\partial S} \\ \Gamma &= \frac{\partial^2 V}{\partial S^2} \\ \rho &= \frac{\partial V}{\partial r} \\ \Theta &= \frac{\partial V}{\partial t} \\ \text{Vega} &= \frac{\partial V}{\partial \sigma}\end{aligned}$$

Here is a convenient way to memorize which Greek letter represents which partial derivative: The Greek letter ρ corresponds to the Roman letter r for interest rate, and θ to t for time. However, Vega is an invented letter and not in the Greek alphabet. To represent the English word ‘volatility’, we need a letter corresponding to the Roman symbol ‘v’ but the Greek alphabet does not have such a symbol. Sometimes we write ‘V’ to denote Vega.² For the list of Greek letters consult Table 15.1, in which Greek letters are given with their sounds written in Roman letters.

Theorem 15.3 *Consider a European call option, and let*

$$d_1, d_2 = \frac{\log \frac{S}{K} + (r \pm \frac{1}{2}\sigma^2)(T - t)}{\sigma \sqrt{T - t}}.$$

Then the Greeks are given by

$$\begin{aligned}\Delta &= N(d_1) \\ \Gamma &= \frac{N'(d_1)}{S\sigma\sqrt{T - t}} \\ \rho &= (T - t)Ke^{-r(T-t)}N(d_2)\end{aligned}$$

² Vega is also a name of a star, located 25 light-years from Earth. It is the brightest star in the constellation Lyra and the fifth brightest in the night sky, with a visual magnitude of 0.03. Vega, not Polaris, was the North Star about fourteen thousand years ago, and it will be so again in about twelve thousand years due to the precession of Earth’s rotation axis.

Table 15.1 The Greek alphabet

Uppercase	Lowercase	Name	Pronunciation
A	α	alpha	a
B	β	beta	b (v in modern Greek)
Γ	γ	gamma	g
Δ	δ	delta	d
E	ϵ	epsilon	e (i in modern Greek)
Z	ζ	zeta	z
H	η	eta	h
Θ	θ	theta	th
I	ι	iota	i
K	κ	kappa	k
Λ	λ	lambda	l
M	μ	mu	m
N	ν	nu	n
Ξ	ξ	xi	ks
O	$\omicron$	omicron	o
Π	π	pi	p
P	ρ	rho	r
Σ	σ	sigma	s
T	τ	tau	t
Υ	υ	upsilon	y (i in modern Greek)
Φ	ϕ	phi	ph
X	χ	chi	kh
Ψ	ψ	psi	ps
Ω	ω	omega	o

$$\Theta = \frac{-S\sigma}{2\sqrt{T-t}} N'(d_1) - rK e^{-r(T-t)} N(d_2)$$

$$\text{Vega} = S\sqrt{T-t} N'(d_1) ,$$

where

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz .$$

Proof (Δ) By direct computation, we obtain

$$\frac{\partial V}{\partial S} = N(d_1) + SN'(d_1) \frac{1}{S} \frac{1}{\sigma\sqrt{T-t}} - K e^{-r(T-t)} N'(d_2) \frac{1}{S} \frac{1}{\sigma\sqrt{T-t}} .$$

Now it suffices to show

$$SN'(d_1) = Ke^{-r(T-t)}N'(d_2) .$$

Since

$$N'(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2} ,$$

it suffices to show

$$\log \frac{S}{K} + r(T-t) = \frac{1}{2}(d_1^2 - d_2^2) .$$

Now we use

$$\begin{aligned} \frac{1}{2}(d_1^2 - d_2^2) &= \frac{1}{2}(d_1 + d_2)(d_1 - d_2) \\ &= \frac{1}{2} \times 2 \frac{\log \frac{S}{K} + r(T-t)}{\sigma \sqrt{T-t}} \sigma \sqrt{T-t} \\ &= \log \frac{S}{K} + r(T-t) . \end{aligned} \tag{15.4}$$

(Γ) Since

$$\frac{\partial d_1}{\partial S} = \frac{\partial}{\partial S} \left(\frac{\log \frac{S}{K} + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma \sqrt{T-t}} \right) = \frac{1}{S\sigma \sqrt{T-t}} ,$$

we have

$$\begin{aligned} \frac{\partial^2 V}{\partial S^2} &= \frac{\partial \Delta}{\partial S} \\ &= N'(d_1) \frac{\partial d_1}{\partial S} \\ &= \frac{N'(d_1)}{S\sigma \sqrt{T-t}} . \end{aligned}$$

(ρ) Since

$$\frac{\partial d_1}{\partial r} = \frac{\partial d_2}{\partial r} = \frac{1}{\sigma} \sqrt{T-t}$$

and since (15.4) implies that

$$\begin{aligned}
& SN'(d_1) - Ke^{-r(T-t)}N'(d_2) \\
&= S \frac{1}{\sqrt{2\pi}} e^{-d_1^2/2} - Ke^{-r(T-t)} \frac{1}{\sqrt{2\pi}} e^{-d_2^2/2} \\
&= \frac{1}{\sqrt{2\pi}} Ke^{-r(T-t)} e^{-d_1^2/2} \left\{ \frac{S}{K} e^{r(T-t)} - e^{(d_1^2-d_2^2)/2} \right\} \\
&= \frac{1}{\sqrt{2\pi}} Ke^{-r(T-t)} e^{-d_1^2/2} \left\{ e^{\log \frac{S}{K}} e^{r(T-t)} - e^{(d_1^2-d_2^2)/2} \right\} \\
&= 0,
\end{aligned} \tag{15.5}$$

we have

$$\begin{aligned}
\frac{\partial V}{\partial r} &= \frac{\partial}{\partial r} (SN(d_1) - Ke^{-r(T-t)}N(d_2)) \\
&= SN'(d_1) \frac{\partial d_1}{\partial r} + K(T-t)e^{-r(T-t)}N(d_2) - Ke^{-r(T-t)}N'(d_2) \frac{\partial d_2}{\partial r} \\
&= \{SN'(d_1) - Ke^{-r(T-t)}N'(d_2)\} \frac{1}{\sigma} \sqrt{T-t} + K(T-t)e^{-r(T-t)}N(d_2) \\
&= K(T-t)e^{-r(T-t)}N(d_2).
\end{aligned}$$

(Θ) Since $d_1 - d_2 = \sigma \sqrt{T-t}$, we have

$$\frac{\partial d_1}{\partial t} - \frac{\partial d_2}{\partial t} = -\sigma \frac{1}{2\sqrt{T-t}}.$$

Hence

$$\begin{aligned}
\frac{\partial V}{\partial t} &= \frac{\partial}{\partial t} (SN(d_1) - Ke^{-r(T-t)}N(d_2)) \\
&= SN'(d_1) \frac{\partial d_1}{\partial t} - rKe^{-r(T-t)}N(d_2) - Ke^{-r(T-t)}N'(d_2) \frac{\partial d_2}{\partial t} \\
&= SN'(d_1) \frac{\partial d_1}{\partial t} - rKe^{-r(T-t)}N(d_2) - SN'(d_1) \frac{\partial d_2}{\partial t} \quad (\text{by (15.5)}) \\
&= SN'(d_1) \frac{\partial d_1}{\partial t} - rKe^{-r(T-t)}N(d_2) - SN'(d_1) \left(\frac{\partial d_1}{\partial t} + \sigma \frac{1}{2\sqrt{T-t}} \right) \\
&= -rKe^{-r(T-t)}N(d_2) - SN'(d_1) \sigma \frac{1}{2\sqrt{T-t}}.
\end{aligned}$$

(Vega) Since $d_1 - d_2 = \sigma \sqrt{T-t}$, we have

$$\frac{\partial d_1}{\partial \sigma} - \frac{\partial d_2}{\partial \sigma} = \sqrt{T-t}.$$

Hence

$$\begin{aligned}
 \frac{\partial V}{\partial \sigma} &= SN'(d_1) \frac{\partial d_1}{\partial \sigma} - Ke^{-r(T-t)} N'(d_2) \frac{\partial d_2}{\partial \sigma} \\
 &= SN'(d_1) \frac{\partial d_1}{\partial \sigma} - SN'(d_1) \frac{\partial d_2}{\partial \sigma} \quad (\text{by (15.5)}) \\
 &= SN'(d_1) \sqrt{T-t} .
 \end{aligned}$$

Now the proof is complete. $\square$

Remark 15.3 For a European call option the following statements are true:

- (i) When the underlying stock price rises, the price of the call option also rises, and hence $\Delta > 0$, which is observed in the formula for Δ .
- (ii) As the interest rate increases, there is more discount, and the discounted strike price decreases, and hence $\rho > 0$ as seen in the formula.
- (iii) As the time to expiry decreases, $\Theta < 0$ as seen in the formula, thus the option price decreases. For European put options, this need not be true. See Figs. 15.9 and 15.10, where for put options the price is not necessarily monotone with respect to time. See also Fig. 15.15 for the plot of Theta.
- (iv) Note that Vega is always positive since as volatility increases the possibility that the stock price goes up above the strike price, and therefore the chance that the option can be exercised, increases.
- (v) Since $\Gamma > 0$, V as a function of S , the graph of $V(t, S)$ is convex.
- (vi) (Kappa) Note that $\frac{\partial V}{\partial K} < 0$ since the call option price decreases as K increases. In fact, we have

$$\frac{\partial V}{\partial K} = -e^{-r(T-t)} N(d_2) .$$

Consult Exercise 15.5 for a proof. For an application see Theorem 17.2.

- (vii) Note that Vega is equal to $S^2 \sigma (T-t) \Gamma$.

We plot the Greeks using the Black–Scholes–Merton formula with strike price $K = 100$, interest rate $r = 0.05$ and volatility $\sigma = 0.3$. In Fig. 15.11 we take $T - t = \frac{1}{12}, 1, 2$ and plot the graphs on the interval $40 \leq S \leq 160$. Let Δ_C and Δ_P denote the Deltas of call and put options with the same K and T . Due to the put-call parity

$$C - P = S - Ke^{-rT} ,$$

they satisfy

$$\Delta_C - \Delta_P = 1 . \tag{15.6}$$

The slopes of the curves in Fig. 15.9 are plotted in Fig. 15.11.

For Gamma we take $T - t = \frac{1}{2}, 1, 2$. Since $\Delta_C - \Delta_P = 1$, we know that Γ of a call and a put is identical since $\Gamma = \partial \Delta / \partial S$. Recall that Γ is positive. See Fig. 15.12.

Rho is positive for a call option while it is negative for a put option. For, if the interest rate rises, the present value of the strike price decreases, and the call option

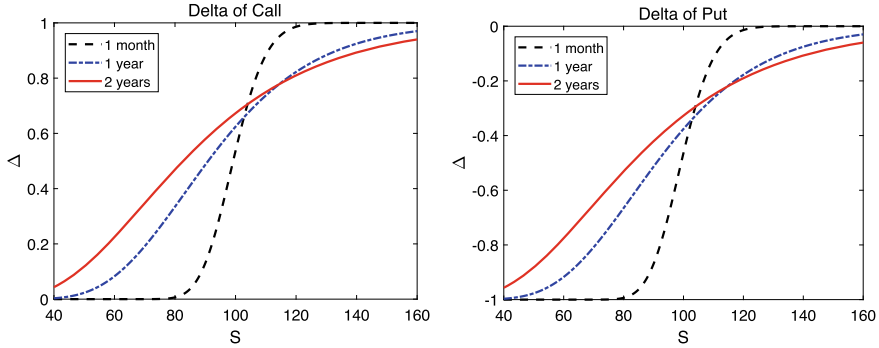


Fig. 15.11 Delta of European call and put options

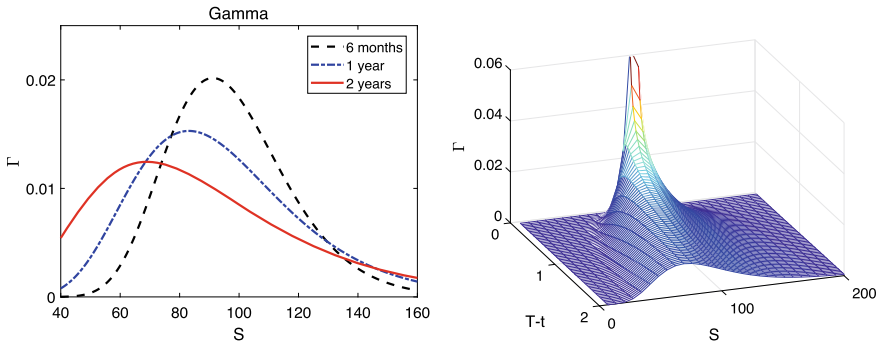


Fig. 15.12 Gamma of European call and put options

value increases while the put option value decreases. Rho has large absolute value, as seen in Fig. 15.13, which is compensated by the fact that δr is usually very small in the relation

$$V(r + \delta r) - V(r) \approx \rho \times \delta r .$$

Note that if $r = 5\%$ moves to 5.1% then $\delta r = 0.1\% = 0.001$.

Vega is large for asset values close to the strike price, and at-the-money options have large values of Vega. Out-of-the-money options have low values of Vega since their prices are low. Deep in-the-money options also have low values of Vega since their values are mostly intrinsic values. See Fig. 15.14.

For the rate of change of the option value as time changes, we consider Theta, which is negative for European call options. Let Θ_C and Θ_P denote the Thetas of call and put options. From the put-call parity

$$C(t) - P(t) = S - Ke^{-r(T-t)} ,$$

we have

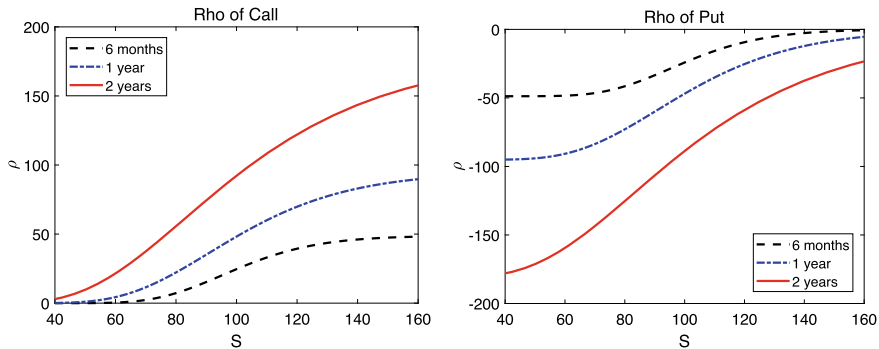


Fig. 15.13 Rho of European call and put options

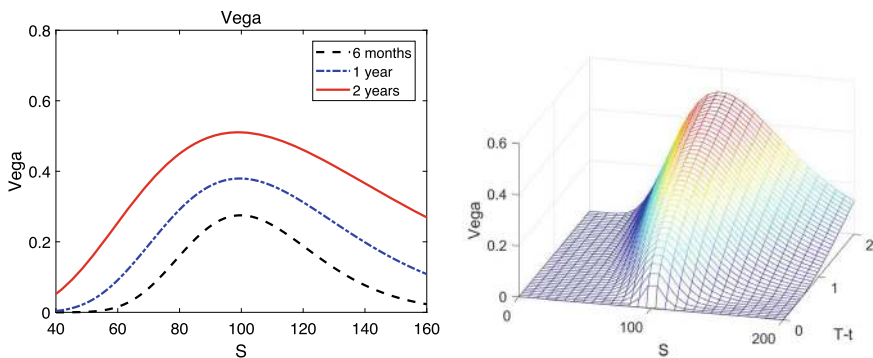


Fig. 15.14 Vega of European call and put options

$$\Theta_C - \Theta_P = -rKe^{-r(T-t)}.$$

In our example, $K = 100$, $r = 0.05$ and $T - t = 0.5, 1, 2$. Hence $\Theta_P = \Theta_C + rKe^{-r(T-t)}$ and $rKe^{-r(T-t)} = 4.8765, 4.7561, 4.5242$. See Fig. 15.15.

For more information on Greeks, consult (McDonald, 2013).

15.5 Solution by the Laplace Transform

In Sect. 15.2 we employed a more or less direct method to solve the Black–Scholes–Merton partial differential equation. In this section we use another method based on the Laplace transformation. The Black–Scholes–Merton equation is essentially a heat equation (or a diffusion equation), and we modify the Laplace transformation method for the heat equation, which is given in Appendix E.3.

To convert the given problem into an initial value problem, we let

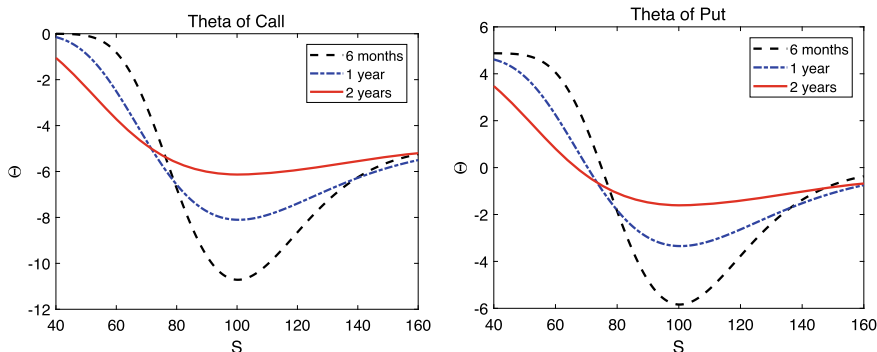


Fig. 15.15 Theta of European call and put options

$$\tau = T - t$$

and put

$$v(S, \tau) = V(S, T - \tau) .$$

By Theorem 15.1 we have

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 v}{\partial S^2} + rS \frac{\partial v}{\partial S} - \frac{\partial v}{\partial \tau} = rv . \quad (15.7)$$

The initial data is given by $v(S, 0) = V(S, T)$. Let $f(S, s)$ be the Laplace transformation $\mathcal{L}[v]$ of $v(S, \tau)$ with respect to τ . Since

$$\mathcal{L}\left[\frac{\partial v}{\partial \tau}\right](S, s) = s\mathcal{L}[v](S, s) - v(S, 0) ,$$

we take the Laplace transformations of both sides of (15.7) and obtain

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - (s + r)f = -v(S, 0) . \quad (15.8)$$

To solve it, find first a homogeneous solution of the homogeneous equation

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - (s + r)f = 0 . \quad (15.9)$$

In this case there exist two linearly independent solutions of the form

$$f(S, s) = S^\lambda . \quad (15.10)$$

(See Remark 15.2 on the Cauchy–Euler equation and also Sect. D.3. Or, to make (15.9) a little simpler, we put $y = \log S$, and obtain a constant coefficient differential equation

$$\frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial y^2} + \left(-\frac{1}{2}\sigma^2 + r\right) \frac{\partial f}{\partial y} - (s + r)f = 0$$

and find a solution of the form $c_1 e^{\lambda_1 y} + c_2 e^{\lambda_2 y}$ where λ_1 and λ_2 satisfy (15.11)). Substituting (15.10) into (15.9), we obtain

$$\frac{1}{2}\sigma^2 S^2 \lambda(\lambda - 1)S^{\lambda-2} + rS\lambda S^{\lambda-1} - (s + r)S^\lambda = 0 .$$

Eliminating S^λ from the both sides, we obtain the quadratic equation

$$\frac{1}{2}\sigma^2 \lambda^2 + \left(-\frac{1}{2}\sigma^2 + r\right)\lambda - (s + r) = 0 . \quad (15.11)$$

Hence there exist two zeros

$$\lambda_{1,2} = \frac{-\left(r - \frac{1}{2}\sigma^2\right) \pm \sqrt{\left(r - \frac{1}{2}\sigma^2\right)^2 + 2(s + r)\sigma^2}}{\sigma^2} .$$

For $s > 0$ we have

$$\lambda_1 > 1 , \quad \lambda_2 < -\frac{2r}{\sigma^2} < 0 .$$

In summary, a homogeneous solution f_H of (15.9) is of the form

$$f_H = c_1 S^{\lambda_1} + c_2 S^{\lambda_2} ,$$

where $c_1(s)$ and $c_2(s)$ are constants with respect to S .

Now we find the prices of financial derivatives.

Example 15.1 (Forward Contract) Consider a forward contract to buy a share of a stock at price K on a future date T . Let us find its price v . Note that the initial condition is given by

$$v(S, 0) = V(S, T) = S - K .$$

Hence (15.8) becomes

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - (s + r)f = K - S . \quad (15.12)$$

Let us find its particular solution f_P of the form

$$f_P(S, s) = c_1(s)S^{\lambda_1} + c_2(s)S^{\lambda_2} + \frac{S}{s} - \frac{K}{s + r} .$$

To determine the constants we check the behavior of v on the boundary of the domain. As $S \approx +\infty$, we have $v(S, \tau) \sim S$, and hence

$$f_P(S, s) \sim S \frac{1}{s}$$

and $c_1 = 0$. On the other hand, for $S \approx 0$, we have

$$v(S, \tau) \sim -K e^{-r\tau},$$

and hence

$$f_P(S, s) \sim -\frac{K}{s+r},$$

thus $c_2 = 0$. Hence we obtain a particular solution of (15.12) of the form

$$f_P(S, s) = \frac{S}{s} - \frac{K}{s+r}.$$

Now we take the inverse Laplace transformation and obtain

$$v(S, \tau) = S - K e^{-r\tau}$$

and

$$V(S, t) = S - K e^{-r(T-t)}.$$

Example 15.2 (European Call Option) The initial condition is given by

$$v(S, 0) = V(S, T) = \max\{S - K, 0\}.$$

Let $f(S, s)$ be the Laplace transformation $\mathcal{L}[v]$ of $v(S, \tau)$ with respect to τ . Then

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - (s+r)f = \min\{K - S, 0\}. \quad (15.13)$$

The homogeneous solution f_H of the above equation is of the form

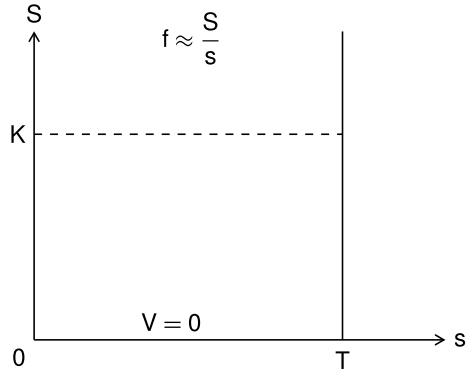
$$f_H = c_1 S^{\lambda_1} + c_2 S^{\lambda_2}.$$

Now we try to find a particular solution f_P of (15.13) of the following form

$$f_P(S, s) = \begin{cases} aS^{\lambda_1} + bS^{\lambda_2} + \frac{S}{s} - \frac{K}{s+r}, & K < S < \infty, \\ cS^{\lambda_1} + dS^{\lambda_2}, & 0 \leq S < K. \end{cases}$$

To determine the constants let us check the property of v on the boundary of the domain. As $S \approx +\infty$, we have $v(S, \tau) \sim S$, and hence $f_P(S, s) \sim S \frac{1}{s}$ and $a = 0$.

Fig. 15.16 Boundary condition for the Laplace transformation of a European call price



Similarly, as $S \approx 0$, we have $v(S, \tau) \sim 0$, and hence $f_P(S, s) \approx 0$ and $d = 0$. Thus

$$f_P(S, s) = \begin{cases} bS^{\lambda_2} + \frac{S}{s} - \frac{K}{s+r}, & K < S < \infty, \\ cS^{\lambda_1}, & 0 \leq S < K. \end{cases}$$

See Fig. 15.16.

Since $f_P(S, s)$ is continuous and differentiable along the boundary $S = K$ of two domains, it satisfies the following two conditions:

$$\begin{aligned} bK^{\lambda_2} + \frac{K}{s} - \frac{K}{s+r} &= cK^{\lambda_1}, \\ \lambda_2 bK^{\lambda_2-1} + \frac{1}{s} &= \lambda_1 cK^{\lambda_1-1}. \end{aligned}$$

Now we find b and c , and then take the inverse Laplace transformation of $f_P(S, s)$. For more algebraic details consult (Kumar, 2007; Shukla and Tomas, 2015).

15.6 Stocks Paying Dividends

In this section we consider the case when the underlying asset S_t pays a dividend continuously. For the sake of generality we assume that the risk-free interest rate $r_t = r(t)$ and the continuous dividend rate $q_t = q(t)$ are deterministic functions of time t . Then the increment of the portfolio

$$\Pi_t = -V_t + D_t + \Delta_t S_t$$

is given by

$$\Pi_{t+dt} - \Pi_t = \Pi_{t+dt} - \Pi_t = d\Pi_t = -dV_t + r_t D_t dt + \Delta_t dS_t + \Delta_t q_t S_t dt ,$$

where the last term on the right-hand side is due to the dividend. For dV_t we apply Itô's lemma and for dS_t we assume the geometric Brownian motion model

$$dS_t = \mu(t)S_t dt + \sigma(t)S_t dW_t ,$$

where $\mu(t)$ and the dividend rate $\sigma(t)$ are deterministic functions of t . To hedge the risk, take

$$\Delta_t = \frac{\partial V}{\partial S}(S_t, t) ,$$

then the random terms disappear, and we obtain the following result.

Theorem 15.4 (Black–Scholes–Merton Equation with Dividend) *Consider the European option with a dividend paying stock S as its underlying asset. The price $V(S, t)$ of the European option at time t with expiry date T satisfies*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma(t)^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r(t) - q(t))S \frac{\partial V}{\partial S} = r(t)V .$$

The final condition $V(S, T)$ is given by the payoff function of the option.

Remark 15.4 (i) For $K = 0$ we have $V(S_T, T) = S_T$ in Theorem 15.4. Suppose that one share of the stock pays a continuous dividend which is reinvested in the same asset. Then the number of shares $a(t)$ at time t satisfies

$$a(t + \delta t) = a(t) + q(t)a(t)\delta t , \quad a(0) = 1 .$$

Hence $a'(t) = q(t)a(t)$ and $a(t) = \log(\int_0^t q(u)du)$. Thus $\exp(-\int_0^T q(u)du)$ shares of the stock at $t = 0$ will grow to one share at T . In general, $\exp(-\int_t^T q(u)du)$ shares of the stock at time t will increase to one share at T . Hence a solution of the given partial differential equation for the European option price with payoff S_T , i.e., one share of the stock, is given by

$$V(S, t) = S \exp(-\int_t^T q(u)du) .$$

(ii) For $V_T = K$ in Theorem 15.4, we have a solution

$$V(S, t) = K \exp(-\int_t^T r(u)du) .$$

Theorem 15.5 (Black–Scholes–Merton Formula with Dividend) *Consider the European call option with a dividend paying stock S as its underlying asset. Then the price $V(S, t)$ of the European option at time t with expiry date T and strike price K is given by*

$$V(S, t) = S \exp \left(- \int_t^T q(u) du \right) N(\hat{d}_1) - K \exp \left(- \int_t^T r(u) du \right) N(\hat{d}_2) ,$$

where

$$\hat{d}_1 = \frac{\log \frac{S}{K} + \int_t^T (r(u) - q(u) + \frac{1}{2} \sigma(u)^2) du}{\sqrt{\int_t^T \sigma(u)^2 du}} ,$$

$$\hat{d}_2 = \hat{d}_1 - \sqrt{\int_t^T \sigma(u)^2 du} .$$

Proof For some functions $\alpha(t)$ and $\beta(t)$ to be determined, we define x and v by

$$x = S e^{\alpha(t)} ,$$

$$v(x, t) = V(S, t) e^{\beta(t)} .$$

Then we have

$$\frac{\partial V}{\partial t} = \left(\frac{\partial v}{\partial t} + x \frac{\partial v}{\partial x} \alpha'(t) - v \beta'(t) \right) e^{-\beta(t)} ,$$

$$S \frac{\partial V}{\partial S} = x \frac{\partial v}{\partial x} e^{-\beta(t)} ,$$

$$S^2 \frac{\partial^2 V}{\partial S^2} = x^2 \frac{\partial^2 v}{\partial x^2} e^{-\beta(t)} .$$

Hence the Black–Scholes–Merton equation in Theorem 15.4 is changed to

$$\frac{\partial v}{\partial t} + \frac{1}{2} \sigma(t)^2 x^2 \frac{\partial^2 v}{\partial x^2} + (r(t) - q(t) + \alpha'(t)) x \frac{\partial v}{\partial x} = (r(t) + \beta'(t)) v . \quad (15.14)$$

Now we choose $\alpha(t)$ and $\beta(t)$ satisfying

$$r(t) - q(t) + \alpha'(t) = 0 ,$$

$$r(t) + \beta'(t) = 0$$

with

$$\alpha(T) = 0 \text{ and } \beta(T) = 0 .$$

In other words,

$$\alpha(t) = \int_t^T (r(u) - q(u)) du , \quad (15.15)$$

$$\beta(t) = \int_t^T r(u) du . \quad (15.16)$$

Then (15.14) is simplified to

$$\frac{\partial v}{\partial t} + \frac{1}{2}\sigma(t)^2 x^2 \frac{\partial^2 v}{\partial x^2} = 0 \quad (15.17)$$

with the condition

$$v|_{t=T} = \max\{x - K, 0\} .$$

Define a monotonically increasing function τ in t by

$$\tau = \int_0^t \sigma(u)^2 du ,$$

and transform (15.17) into

$$\frac{\partial v}{\partial \tau} + \frac{1}{2}x^2 \frac{\partial^2 v}{\partial x^2} = 0 \quad (15.18)$$

with the condition

$$v(S, t)|_{\tau=T_1} = \max\{x - K, 0\} ,$$

where

$$T_1 = \int_0^T \sigma(u)^2 du .$$

Note that (15.18) is a special case of the Black–Scholes–Merton equation given in Theorem 15.1 with $\sigma = 1$ and $r = 0$. Hence by Theorem 15.2 we have

$$v(x, \tau) = xN(\hat{d}_1) - KN(\hat{d}_2) , \quad (15.19)$$

where

$$\hat{d}_1 = \frac{\log \frac{x}{K} + \frac{1}{2}(T_1 - \tau)}{\sqrt{T_1 - \tau}} , \quad (15.20)$$

$$\hat{d}_2 = \hat{d}_1 - \sqrt{T_1 - \tau} . \quad (15.21)$$

Finally, using (15.15) and (15.16), we express (15.19), (15.20) and (15.21) in terms of V and S . □

Corollary 15.2 (Put-Call Parity) *Suppose that the drift coefficient μ , the volatility σ , the risk-free interest rate r and the dividend rate σ are deterministic functions of t . Let $C(S, t)$ and $P(S, t)$ denote the prices of a European call and put options, respectively, with the same strike price K and expiry date T . Then*

$$C(S, t) + K \exp\left(-\int_t^T \tau(u) du\right) = P(S, t) + S \exp\left(-\int_t^T q(u) du\right) .$$

Proof Put $F(S, t) = C(S, t) - P(S, t)$. Then $F(S, t)$ satisfies the following partial differential equation

$$\frac{\partial F}{\partial t} + \frac{1}{2}\sigma(t)^2 S^2 \frac{\partial^2 F}{\partial S^2} + (r(t) - q(t))S \frac{\partial F}{\partial S} = r(t)F .$$

At $t = T$, we have

$$\begin{aligned} F(S_T, T) &= C(S_T, T) - P(S_T, T) \\ &= \max\{S_T - K, 0\} - \max\{K - S_T, 0\} \\ &= S_T - K . \end{aligned}$$

Hence F itself is an option with its payoff $S_T - K$, and it may be regarded as the difference of two options with payoffs S_T and K , respectively. Now Remark 15.4 implies that their prices are given by

$$S \exp\left(-\int_t^T q(u) du\right) \quad \text{and} \quad K \exp\left(-\int_t^T r(u) du\right) .$$

Thus

$$F(S, t) = S \exp\left(-\int_t^T q(u) du\right) - K \exp\left(-\int_t^T r(u) du\right) ,$$

and the proof is complete. $\square$

15.7 Computer Experiments

Simulation 15.1 (Delta Hedging) Let us review **discrete time** hedging in the derivation of the Black–Scholes–Merton equation. What is the meaning of $dD_t = rD_t dt$? In contrast to $d\Pi_t = r\Pi_t dt$ over the time interval $[0, T]$, which has a continuous global solution $\Pi_t = \Pi_0 e^{rt}$, the bank deposit D_t does not satisfy $D_t = D_0 e^{rt}$ for $0 \leq t \leq T$. The reason is that we readjust the investment into the risky asset upon receiving new stock market information by using a new value of Δ at t_i , and the investment into risk-free asset is also readjusted. More precisely, $D_t = D_{t_i} \times e^{r(t-t_i)}$ for $t_i \leq t < t_{i+1}$. See Fig. 15.17.

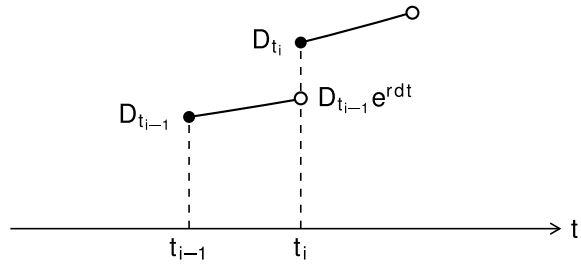
Since Π_t is self-financing, it is continuous at time t_i , and we have

$$\Pi_{t_i} = -V_{t_i} + D_{t_{i-1}} e^{r\delta t} + \Delta_{t_{i-1}} S_{t_i} = -V_{t_i} + D_{t_i} + \Delta_{t_i} S_{t_i} ,$$

which implies that

$$D_{t_i} = D_{t_{i-1}} e^{r\delta t} + (\Delta_{t_{i-1}} - \Delta_{t_i}) S_{t_i} .$$

Fig. 15.17 Readjustment of the bank deposit in discrete time hedging



We present in-the-money and out-of-the-money cases for a European call option in Figs. 15.3 and 15.4, respectively. The values of Δ approaches 1 or 0 depending on whether the option is in-the-money or out-of-the-money. In the first case, the option will be exercised with certainty, so the seller must hedge it fully, while in the second case there is not much chance to be exercised and the need to hedge is almost zero. In both cases the plots for Π_t are close to the exponential curve $\Pi_0 e^{rt}$ as expected. The graphs for Δ and D are step functions which will converge to continuous functions as the number of time steps increases to ∞ .

```

T = 5;
r = 0.10; % interest rate
mu = 0.15; % drift coefficient
sigma = 0.3; % volatility
S0 = 10; % asset price at time t=0
K = 15; % strike price

N = 100 ; % number of time steps
dt = T/N;
t_value = [0:dt:T];

W = zeros(1,N+1); % Brownian motion
S = zeros(1,N+1); % asset price
V = zeros(1,N+1); % option price
Delta = zeros(1,N+1); % Delta
D = zeros(1,N+1); % bank deposit
Pi = zeros(1,N+1); % portfolio

D(1) = 0.0; % Choose any number for the initial cash amount.
S(1)= S0;

for i=2:N+1
    dW = sqrt(dt)*randn;
    W(i) = W(i-1) + dW;
    S(i) = S(i-1) + mu*S(i-1)*dt + sigma*S(i-1)*dW;

```

```

end

for i=1:N+1
    tau = T-(i-1)*dt;
    d1 = (log(S(i)/K) + (r+0.5*sigma^2)*tau)/sigma/sqrt(tau);
    d2 = d1 - sigma*sqrt(tau);
    V(i)= S(i)*normcdf(d1) - K*exp(-r*tau)*normcdf(d2);
    Delta(i)=normcdf(d1);
end

for i = 1:N
    D(i+1) = exp(r*dt)*D(i) + (Delta(i)-Delta(i+1))*S(i+1); %self-financing
end

for i = 1:N+1
    Pi(i) = -V(i) + D(i) + Delta(i)*S(i);
end

figure(1);
plot([0:dt:T],W)
title('W')

figure(2);
plot([0:dt:T],S)
title('S')
hold on
plot([0:dt:T],K,'--')

figure(3);
plot([0:dt:T],V)
title('V')

figure(4);
for i=1:N
    x = (i-1)*dt:dt/(500/N):i*dt;
    y = Delta(i)*exp(0*x);
    plot(x,y) % Plot the graph on each subinterval.
    hold on
end
title('\Delta')

figure(5);
for i=1:N
    x = (i-1)*dt:dt/(500/N):i*dt;
    y = D(i);

```

```

plot(x,y*exp(r*(x-(i-1)*dt))) % Plot the graph on each subinterval.
hold on
end
title('D')

figure(6);
plot([0:dt:T],Pi(1)*exp(r*[0:dt:T]))
title('\Pi')
hold on
plot([0:dt:T],Pi)

```

Simulation 15.2 (Option Price Surface) We regard the European call and put option prices as functions of t and S , and plot their graphs. See Figs. 15.7 and 15.8.

```

K = 10;
S0 = 10;
r = 0.05;
sigma = 0.3;
mu = 0.1;
T = 1;
N = 51;
dt = T/(N-1);

S_path = zeros(N,1);
S_path(1)= S0; % asset price at time 0

for i=1:N-1
    dW = sqrt(dt)*randn;
    S_path(i+1) = S_path(i) + mu*S_path(i)*dt + sigma*S_path(i)*dW;
end

C_path = zeros(N,1);
P_path = zeros(N,1);

for i = 1:N
    S = S_path(i);
    tau = T-(i-1)*dt;
    d1 = (log(S/K) + (r+0.5*sigma^2)*tau)/(sigma*sqrt(tau));
    d2 = d1 - sigma*sqrt(tau);
    C_path(i) = S*normcdf(d1) - K*exp(-r*tau)*normcdf(d2);
    P_path(i) = C_path(i) - S + K*exp(-r*tau);
end

```

```

t_value = linspace(0,T,N);
S_value = linspace(0,20,N);
C = zeros(N,N);
P = zeros(N,N);

for j=1:N
    S = S_value(j);
    for i = 1:N-1
        tau = T-t_value(i);
        d1 = (log(S/K)+(r+0.5*sigma^2)*tau)/(sigma*sqrt(tau));
        d2 = d1-sigma*sqrt(tau);
        N1 = normcdf(d1);
        N2 = normcdf(d2);
        C(i,j) = S*N1 - K*exp(-r*tau)*N2;
        P(i,j) = C(i,j) + K*exp(-r*tau) - S;
    end
    C(N,j) = max(S-K,0); % payoff at T
    P(N,j) = max(K-S,0); % payoff at T
end

% Plot the superimposed image of asset price movement.
[S_grid,t_grid] = meshgrid(S_value,t_value);

figure(1);
surf(S_grid,t_grid,C)
hold on
plot3(S_path,t_grid,C_path);
xlabel('S'), ylabel('t'), zlabel('Call')
hold off

figure(2);
surf(S_grid,t_grid,P)
hold on
plot3(S_path,t_grid,P_path);
xlabel('S'), ylabel('t'), zlabel('Put')
hold off

```

Simulation 15.3 (The Greeks by the Monte Carlo Method) When there is no formula for the Greeks we may use the Monte Carlo method presented in Volume II. For example, we estimate

$$e^{-rT} \frac{\mathbb{E}[V(S_0 + \delta S) - \mathbb{E}[V(S_0)]]}{\delta S}$$

for Δ of a European put option at $t = 0$ by choosing sufficiently small $\delta S > 0$. Consult (Higham, 2004) for more information.

```
S0 = 110;
K = 140;
r = 0.05;
sigma = 0.3;
T = 1;
d1 = (log(S0/K) + (r + 0.5*sigma^2)*T)/(sigma*sqrt(T));
M = 10^6;
W = sqrt(T)*randn(M,1);
S = S0*exp((r-0.5*sigma^2)*T + sigma*W);
dS = 0.01;
S2 = (S0+dS)*exp((r-0.5*sigma^2)*T + sigma*W);
V = max(K-S,0);
V2 = max(K-S2,0);
Delta_MC = exp(-r*T)*mean(V2-V)/dS
Delta_BSM = normcdf(d1) - 1
```

Compare the following output with the result in Simulation 14.2.

```
Delta_MC =
    -0.6870

Delta_BSM =
    -0.6869
```

Exercises

15.1 What is the boundary condition of the Black–Scholes–Merton equation for a European put option? Consult Fig. 15.18.

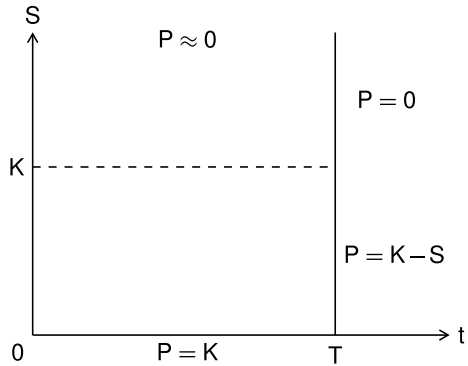
15.2 Define d_1 and d_2 as usual. Show that $d_1 - d_2 = \sigma\sqrt{T-t}$ and hence $d_2 = d_1 - \sigma\sqrt{T-t}$.

15.3 Recall that the price of a European put option is given by

$$P(S_t, t) = Ke^{-r(T-t)}N(-d_2) - S_tN(-d_1).$$

- (i) Find $\lim_{K \rightarrow 0+} P(S)$.
- (ii) Find $\lim_{\sigma \rightarrow 0+} P(S, t)$.

Fig. 15.18 The boundary condition of the Black–Scholes–Merton equation for a European put option



15.4 Let $C(S, t)$ denote the value of a European call option with strike price K . When $S_T > K$, compute

$$\lim_{t \rightarrow T^-} \frac{\partial C(S, t)}{\partial S}.$$

15.5 Prove that for a European call option price C we have

$$\frac{\partial C}{\partial K} = -e^{-r(T-t)} N(d_2).$$

15.6 Explain the behavior of the solution of the Black–Scholes–Merton equation for a European put option in the case when the volatility σ is zero. How about the case when σ is very large? Does it agree with common sense?

15.7 Derive the Black–Scholes–Merton equation under the assumption that there exists a replicating self-financing portfolio, i.e., $V_t = a_t S_t + b_t B_t$ where a_t and b_t determine the trading strategy, and $dV_t = a_t dS_t + b_t dB_t$. (Hint: Since

$$\left(\frac{\partial V}{\partial t} + \mu S \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \sigma S \frac{\partial V}{\partial S} dW = (a\mu S + brB)dt + a\sigma S dW.$$

By comparing coefficients, we have $a = \frac{\partial V}{\partial S}$, and hence

$$\frac{\partial V}{\partial t} + \mu S \frac{\partial V}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} = a\mu S + brB.$$

After cancelling two terms containing μ , we use $bB = V - aS = V - S \frac{\partial V}{\partial S}$.

15.8 The *elasticity* Ω of an option price is the ratio of the percentage change in the option price V with respect to the percentage change in the underlying asset price S . More precisely,

$$\Omega = \lim_{\delta S \rightarrow 0} \frac{\frac{\delta V}{V}}{\frac{\delta S}{S}}.$$

(It measures the leverage in the option investment).

- (i) Show that $\Omega = \Delta \frac{S}{V}$. (This means that for a deep out-of-the-money option, i.e., an option with $V \approx 0$, the elasticity can be large).
- (ii) Show that for a European call option, $\Omega \geq 1$. (This means that the investment in an option is riskier than that in the underlying asset).
- (iii) Show that for a European put option, $\Omega \leq 0$.

15.9 Let us consider some solutions of the Black–Scholes–Merton equation given in Theorem 15.1 without final time or boundary conditions. What is their financial meaning?

- (i) Show that $V(S, t) = S$ and $V(S, t) = e^{rt}$ are solutions of the Black–Scholes–Merton equation.
- (ii) Show that

$$V(S, t) = \frac{1}{S} e^{(\sigma^2 - 2r)(T-t)}$$

is a solution of the Black–Scholes–Merton equation.

15.10 Find the probability that a European put will expire in-the-money.

15.11 Compute the Gamma, Theta, Vega and Rho for a European put option.

Chapter 16

The Martingale Method



In this chapter we introduce two proofs of the risk-neutral option pricing formula given by the discounted expected payoff by applying martingale theory. The first method is based on hedging of a portfolio process, and the second on replication of the payoff at expiry date T . Also shown is the connection to the Black–Scholes–Merton partial differential equation.

16.1 Option Pricing by the Martingale Method

Let $\{W_t\}_{t \geq 0}$ be a Brownian motion with respect to a probability measure $\mathbb{P}$. Assume that the asset price movement $\{S_t\}_{t \geq 0}$ follows geometric Brownian motion $dS_t = \mu S_t dt + \sigma S_t dW_t$, where μ and σ are constant, which has a solution $S_t = S_0 e^{\sigma W_t + (\mu - \frac{1}{2}\sigma^2)t}$. Let $r > 0$ be risk-free interest rate, put $\theta = \frac{\mu - r}{\sigma}$ and define X_t by

$$X_t = W_t + \theta t .$$

Then $dX_t = dW_t + \theta dt$ and

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma S_t (dX_t - \theta dt) \\ &= (\mu - \sigma \theta) S_t dt + \sigma S_t dX_t \\ &= r S_t dt + \sigma S_t dX_t . \end{aligned}$$

Hence $dS^2 = \sigma^2 S^2 dt$.

Let $\mathbb{Q}$ be a probability measure such that $\mathbb{E} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \middle| \mathcal{F}_t \right] = e^{-\frac{1}{2}\theta^2 t - \theta W_t}$ for $0 \leq t \leq T$. Girsanov's theorem states that $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, and $\{X_t\}$ is a $\mathbb{Q}$ -Brownian motion. Let $\tilde{S}_t = e^{-rt} S_t$ be a discounted asset price. Since

$$\tilde{S}_t = S_0 e^{(\mu - r - \frac{1}{2}\sigma^2)t + \sigma W_t} = \tilde{S}_0 e^{-\frac{1}{2}\sigma^2 t + \sigma X_t},$$

we have $d\tilde{S}_t = \sigma \tilde{S}_t dX_t$ and $\{\tilde{S}_t\}$ is a $\mathbb{Q}$ -martingale.

Recall that if a payoff function at T depends on a process $\{S_t\}_{0 \leq t \leq T}$, then it is $\mathcal{F}_T$ -measurable.

16.1.1 Option Pricing by Hedging

Consider a portfolio process

$$\Pi_t = -V_t + D_t + \Delta_t S_t$$

as given in Sect. 15.1. Define the discounted processes

$$\tilde{\Pi}_t = e^{-rt} \Pi_t, \quad \tilde{V}_t = e^{-rt} V_t \quad \text{and} \quad \tilde{D}_t = e^{-rt} D_t.$$

Then

$$\tilde{\Pi}_t = -\tilde{V}_t + \tilde{D}_t + \Delta_t \tilde{S}_t.$$

Since, with the choice of

$$\Delta_t = \frac{\partial V}{\partial S},$$

the portfolio Π_t becomes risk-free, we have $\Pi_t = \Pi_0 e^{rt}$, $0 \leq t \leq T$. Hence $\tilde{\Pi}_t = \Pi_0$ and $d\tilde{\Pi}_t = 0$.

Let Y_{t-} denote the limit from the left of the process Y at time $t > 0$. Take a partition $0 = t_0 < t_1 < \dots < t_{N-1} < t_N = T$. Since the portfolio is self-financing, we have

$$\tilde{\Pi}_{t_{i+1}-} - \tilde{\Pi}_{t_i} = -(\tilde{V}_{t_{i+1}-} - \tilde{V}_{t_i}) + (\tilde{D}_{t_{i+1}-} - \tilde{D}_{t_i}) + \Delta_{t_i}(\tilde{S}_{t_{i+1}-} - \tilde{S}_{t_i}).$$

Since $\tilde{\Pi}_t$, $\tilde{V}_t$ and $\tilde{S}_t$ are continuous (with probability 1), we have

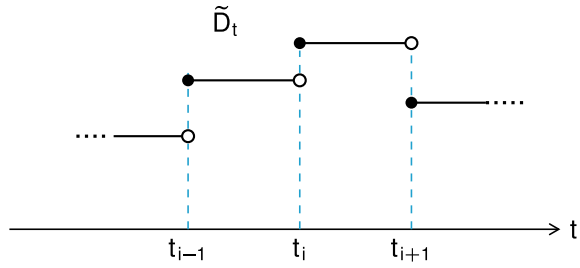
$$\tilde{\Pi}_{t_{i+1}} - \tilde{\Pi}_{t_i} = -(\tilde{V}_{t_{i+1}} - \tilde{V}_{t_i}) + (\tilde{D}_{t_{i+1}} - \tilde{D}_{t_i}) + \Delta_{t_i}(\tilde{S}_{t_{i+1}} - \tilde{S}_{t_i}).$$

Since $\tilde{\Pi}_t = \Pi_0$ for $0 \leq t \leq T$ and since $\tilde{D}_t = D_t e^{-rt_i}$ for $t_i \leq t < t_{i+1}$, we have

$$0 = -(\tilde{V}_{t_{i+1}} - \tilde{V}_{t_i}) + 0 + \Delta_{t_i}(\tilde{S}_{t_{i+1}} - \tilde{S}_{t_i}).$$

(See Fig. 16.1.)

Fig. 16.1 Time discretization for discounted bond price $\tilde{D}_t$



Hence

$$\tilde{V}_T - \tilde{V}_0 = \sum_{i=0}^{N-1} (\tilde{V}_{t_{i+1}} - \tilde{V}_{t_i}) = \sum_{i=0}^{N-1} \Delta_{t_i} (\tilde{S}_{t_{i+1}} - \tilde{S}_{t_i}) .$$

By taking the limit, we have

$$\tilde{V}_T - \tilde{V}_0 = \int_0^T \Delta_t d\tilde{S}_t . \quad (16.1)$$

The stochastic integral on the right-hand side of (16.1) is meaningfully defined since the integrator $\tilde{S}_t$ is a continuous martingale with respect to $\mathbb{Q}$. By the Martingale Representation Theorem there exists an α_t such that

$$d\tilde{S}_t = \alpha_t dX_t .$$

In fact, $\alpha_t = \sigma \tilde{S}_t$. Hence

$$\tilde{V}_T - \tilde{V}_0 = \int_0^T \Delta_t \sigma \tilde{S}_t dX_t .$$

By taking the expectations of both sides, we have

$$\mathbb{E}^{\mathbb{Q}}[\tilde{V}_T] - \tilde{V}_0 = 0 ,$$

since the expectation of the Itô integral is zero. Therefore,

$$V_0 = e^{-rT} \mathbb{E}^{\mathbb{Q}}[V_T] .$$

If we consider the time interval $[t, T]$ in the beginning of the derivation of the formula, we obtain

$$\tilde{V}_T - \tilde{V}_t = \int_t^T \Delta_u \sigma \tilde{S}_u dX_u$$

and

$$V_t = e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[V_T | \mathcal{F}_t] . \quad (16.2)$$

16.1.2 Option Pricing by Replication

Let C_T denote the payoff of a given European option, and $B_t = e^{rt}$ be the value of risk-free deposit of unit amount. Define V_t and $\tilde{V}_t$ by

$$V_t = \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)} C_T | \mathcal{F}_t]$$

and

$$\tilde{V}_t = e^{-rt} V_t = \mathbb{E}^{\mathbb{Q}}[e^{-rT} C_T | \mathcal{F}_t] .$$

Then we have a replication of the payoff since

$$V_T = e^{rT} \tilde{V}_T = \mathbb{E}^{\mathbb{Q}}[C_T | \mathcal{F}_T] = C_T .$$

It remains to show that V_t is self-financing. Since $\tilde{V}_t$ is a $\mathbb{Q}$ -martingale, by the Martingale Representation Theorem there exists a predictable process $\{\alpha_t\}_{t \geq 0}$ such that

$$\tilde{V}_t = \tilde{V}_0 + \int_0^t \alpha_u dX_u = \tilde{V}_0 + \int_0^t \alpha_u \frac{1}{\sigma \tilde{S}_u} d\tilde{S}_u ,$$

where the continuous martingale $\tilde{S}_t$ is the integrator for the given stochastic integral. Now define ϕ_t by

$$\phi_t = \alpha_t \frac{1}{\sigma \tilde{S}_t} .$$

Then we have

$$\tilde{V}_t = \tilde{V}_0 + \int_0^t \phi_u d\tilde{S}_u .$$

Since

$$d\tilde{V}_t = \phi_t d\tilde{S}_t , \quad (16.3)$$

we have

$$d(e^{-rt} V_t) = \phi_t d(e^{-rt} S_t) ,$$

and hence

$$-re^{-rt} V_t dt + e^{-rt} dV_t = \phi_t (-re^{-rt} S_t dt + e^{-rt} dS_t) ,$$

and after multiplying by e^{rt} we obtain

$$-r V_t dt + dV_t = \phi_t(-r S_t dt + dS_t) .$$

Note that since $e^{-rt} dB_t = r dt$,

$$dV_t = (V_t - \phi_t S_t) e^{-rt} dB_t + \phi_t dS_t .$$

Now define ψ_t by

$$\psi_t = (V_t - \phi_t S_t) e^{-rt} ,$$

or equivalently,

$$V_t = \psi_t B_t + \phi_t S_t . \quad (16.4)$$

Since

$$dV_t = \psi_t dB_t + \phi_t dS_t ,$$

ψ_t and ϕ_t produce a self-financing portfolio V_t . Thus V_t is the option price at time t .

The following results shows that the option price does not depend on the drift coefficient in the geometric Brownian motion, which is also observed in the Black–Scholes–Merton formula.

Theorem 16.1 (Risk Neutrality) *For a European option with its expiry date T and payoff C_T , its price V_0 at $t = 0$ is given by*

$$V_0 = e^{-rT} \mathbb{E}^{\mathbb{Q}}[C_T(S_0 e^{\sigma X_T + (r - \frac{1}{2}\sigma^2)T})] .$$

In other words, the option price is the discounted expected payoff taken over the sample paths of a geometric Brownian motion with the drift coefficient μ replaced by the risk-free interest rate r .

Proof Note that (16.2) implies

$$\begin{aligned} V_0 &= e^{-rT} \mathbb{E}^{\mathbb{Q}}[C_T(S_T)] \\ &= e^{-rT} \mathbb{E}^{\mathbb{Q}}[C_T(e^{rT} \tilde{S}_T)] \\ &= e^{-rT} \mathbb{E}^{\mathbb{Q}}[C_T(e^{rT} S_0 e^{\sigma X_T - \frac{1}{2}\sigma^2 T})] \\ &= e^{-rT} \mathbb{E}^{\mathbb{Q}}[C_T(S_0 e^{\sigma X_T + (r - \frac{1}{2}\sigma^2)T})] . \end{aligned}$$

Remark 16.1 In the above formula the only fact that is used is that S_t is a geometric Brownian motion with respect to some probability measure with drift coefficient given by r . Thus we may choose the pair $(\mathbb{P}, \{W_t\})$ instead of $(\mathbb{Q}, \{X_t\})$ as long as we replace μ by r . Thus $V_0 = e^{-rT} \mathbb{E}^{\mathbb{P}}[C_T(S_0 e^{\sigma W_T + (r - \frac{1}{2}\sigma^2)T})]$, which is a viewpoint employed in the Monte Carlo method.

16.1.3 The Black–Scholes–Merton Equation Connection

Two proofs of the option pricing formula (16.2) given in Sects. 16.1.1 and 16.1.2 are obtained independently of the Black–Scholes–Merton differential equation. Now we derive the same formula assuming the Black–Scholes–Merton differential equation. The proof is a restatement of the idea in Sect. 13.2. First, let V_t denote the option price at t and put $\tilde{V}_t = e^{-rt} V_t$. Then by Itô's lemma we have

$$\begin{aligned}
 d\tilde{V}_t &= d(e^{-rt} V_t) \\
 &= -re^{-rt} V dt + e^{-rt} dV_t \\
 &= -re^{-rt} V dt + e^{-rt} \left[\frac{\partial V(t, S)}{\partial t} dt + \frac{\partial V(t, S)}{\partial S} dS + \frac{1}{2} \frac{\partial^2 V(t, S)}{\partial S^2} dS^2 \right] \\
 &= -re^{-rt} V dt + e^{-rt} \left[\frac{\partial V}{\partial t} dt + \frac{\partial V}{\partial S} (rS dt + \sigma S dX) + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} dt \right] \\
 &= e^{-rt} \sigma S_t \frac{\partial V}{\partial S} dX_t,
 \end{aligned}$$

where the Black–Scholes–Merton differential equation was used for the last equality. Since there is no dt -term in $d\tilde{V}_t$, we conclude that $\{\tilde{V}_t\}$ is a $\mathbb{Q}$ -martingale, and that

$$e^{-rt} V_t = \tilde{V}_t = \mathbb{E}^{\mathbb{Q}}[\tilde{V}_T | \mathcal{F}_t] = \mathbb{E}^{\mathbb{Q}}[e^{-rT} C_T | \mathcal{F}_t].$$

Thus we obtain (16.2).

16.2 The Probability Distribution of Asset Price

Now we find the price of a European option with expiry date T and payoff $C_T(S_T)$, where S_T is the asset price at T , by using the probability density function of S_T .

If the asset price S_t follows the geometric Brownian motion, then at time $t = T$ we have $S_T = S_0 e^{\sigma W_T + (\mu - \frac{1}{2}\sigma^2)T}$. If x denotes a value taken by S_T , then we have

$$\frac{\log \frac{x}{S_0} - (\mu - \frac{1}{2}\sigma^2)T}{\sigma} = W_T.$$

If y denotes a value taken by W_T , the pdf of y is given by

$$f(y) = \frac{1}{\sqrt{2\pi T}} \exp \left(-\frac{1}{2T} \left(\frac{\log \frac{x}{S_0} - (\mu - \frac{1}{2}\sigma^2)T}{\sigma} \right)^2 \right).$$

Since $\frac{dy}{dx} = \frac{1}{\sigma x}$, the price of the option is given by

$$\begin{aligned}
e^{-rT} \mathbb{E}[C_T(S_T)] &= e^{-rT} \int_0^\infty C_T(x) f(y(x)) \frac{dy}{dx} dx \\
&= e^{-rT} \int_0^\infty \frac{C_T(x)}{x\sigma\sqrt{2\pi T}} \exp\left(-\frac{(\log \frac{x}{S_0} - (\mu - \frac{1}{2}\sigma^2)T)^2}{2T\sigma^2}\right) dx .
\end{aligned}$$

To apply the risk-neutral method, we put $\mu = r$. Then the option price equals

$$e^{-rT} \int_0^\infty \frac{C_T(x)}{x\sigma\sqrt{2\pi T}} \exp\left(-\frac{(\log \frac{x}{S_0} - (r - \frac{1}{2}\sigma^2)T)^2}{2T\sigma^2}\right) dx . \quad (16.5)$$

See Exercise 16.2.

16.3 The Black–Scholes–Merton Formula

In this section we compute the price of a European call option using the martingale method. Since the payoff is given by $C(S_T) = (S_T - K)^+$ at expiry date T , we have

$$\begin{aligned}
V_0 &= e^{-rT} \mathbb{E}^{\mathbb{Q}}[(S_T - K)^+] \\
&= e^{-rT} \mathbb{E}^{\mathbb{Q}}[(e^{rT} \tilde{S}_T - K)^+] \\
&= e^{-rT} \mathbb{E}^{\mathbb{Q}}[(e^{rT} S_0 e^{\sigma X_T - \frac{1}{2}\sigma^2 T} - K)^+] \\
&= e^{-rT} \mathbb{E}^{\mathbb{Q}}[(S_0 e^{\sigma\sqrt{T}Z + (r - \frac{1}{2}\sigma^2)T} - K)^+] \\
&= e^{-rT} \int_{-\infty}^\infty (S_0 e^{\sigma\sqrt{T}x + (r - \frac{1}{2}\sigma^2)T} - K)^+ \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx ,
\end{aligned}$$

where the symbol Z denotes the standard normal variable. If we consider a European call option with exercise price K , then the domain of integration is given by

$$\{x : S_0 e^{\sigma\sqrt{T}x + (r - \frac{1}{2}\sigma^2)T} \geq K\} .$$

Hence

$$x \geq x_0 = \frac{\log \frac{K}{S_0} - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} . \quad (16.6)$$

Therefore the given integral is equal to

$$e^{-rT} \int_{x_0}^\infty (S_0 e^{\sigma\sqrt{T}x + (r - \frac{1}{2}\sigma^2)T} - K) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx , \quad (16.7)$$

where x_0 is given by (16.6). Now we put

$$d_1, d_2 = \frac{\log \frac{S_0}{K} + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} .$$

If we let $y = -x$, then $dy = -dx$, and (16.7) is equal to

$$\begin{aligned} & e^{-rT} \int_{-\infty}^{d_2} (S_0 e^{-\sigma\sqrt{T}y + (r - \frac{1}{2}\sigma^2)T} - K) \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= S_0 \int_{-\infty}^{d_2} e^{-\sigma\sqrt{T}y - \frac{1}{2}\sigma^2 T} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy - K e^{-rT} \int_{-\infty}^{d_2} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy . \end{aligned}$$

Then we have

$$\begin{aligned} \int_{-\infty}^{d_2} \frac{1}{\sqrt{2\pi}} e^{-\sigma\sqrt{T}y - \frac{1}{2}\sigma^2 T - y^2/2} dy &= \int_{-\infty}^{d_2} \frac{1}{\sqrt{2\pi}} e^{-(y + \sigma\sqrt{T})^2/2} dy \\ &= N(d_2 + \sigma\sqrt{T}) \\ &= N(d_1) , \end{aligned}$$

and the original integral representing the option price is given by

$$S_0 N(d_1) - K e^{-rT} N(d_2) ,$$

which is identical with the solution given in Sect. 15.2.

The martingale method can derive a closed form solution in some cases, but in general the Monte Carlo method is used to find an approximate solution, and it is more widely used than the partial differential equation approach.

16.4 Derivation of the Black–Scholes–Merton Equation

In this section, using the martingale method combined with the Feynman–Kac Theorem, we derive the Black–Scholes–Merton equation for European options. Let V_t denote the option price at time t , and assume that the payoff is of the form $C_T(S_T)$. If a function $F : [0, T] \times [0, \infty) \rightarrow \mathbb{R}$ satisfies

$$V_t(\omega) = F(t, S_t(\omega)) ,$$

then

$$F(t, x) = \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)} C_T(S_T) | S_t = x] .$$

Now if we put

$$G(t, x) = \mathbb{E}^{\mathbb{Q}}[C_T(S_T) | S_t = x] ,$$

then

$$G(t, x) = e^{r(T-t)} F(t, x) .$$

Now we apply the Feynman–Kac Theorem to $G(t, x)$ and obtain

$$\begin{aligned} \frac{\partial}{\partial t} (e^{r(T-t)} F(t, x)) \Big|_{x=S_t} + r S_t \frac{\partial}{\partial x} (e^{r(T-t)} F(t, x)) \Big|_{x=S_t} \\ + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2}{\partial x^2} (e^{r(T-t)} F(t, x)) \Big|_{x=S_t} = 0 . \end{aligned}$$

After dividing by e^{rT} , then taking derivatives, and finally multiplying by e^{rt} , we obtain

$$\begin{aligned} -r F(t, S_t) + \frac{\partial F(t, x)}{\partial t} \Big|_{x=S_t} + r S_t \frac{\partial F(t, x)}{\partial x} \Big|_{x=S_t} \\ + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 F(t, x)}{\partial x^2} \Big|_{x=S_t} = 0 . \end{aligned}$$

Since S_t can take any arbitrary value, the following equation should hold:

$$-r F(t, x) + \frac{\partial F(t, x)}{\partial t} + r x \frac{\partial F(t, x)}{\partial x} + \frac{1}{2} \sigma^2 x^2 \frac{\partial^2 F(t, x)}{\partial x^2} = 0 , \quad (16.8)$$

which is nothing but the Black–Scholes–Merton equation. (See Example 13.6.)

16.5 Delta Hedging

For hedging, we will find the Delta (Δ) of a European call option. From (16.4) it is equal to ϕ_t , which is given by (16.3).

First, put

$$\tilde{F}(t, y) = e^{-rt} F(t, ye^{rt}) ,$$

then we have

$$\begin{aligned} & \frac{\partial}{\partial t} \tilde{F}(t, y) \\ &= -r e^{-rt} F(t, ye^{rt}) + e^{-rt} \left\{ \frac{\partial F}{\partial t}(t, ye^{rt}) + \frac{\partial F}{\partial x}(t, ye^{rt}) r y e^{rt} \right\} \\ &= e^{-rt} \left\{ -r F(t, ye^{rt}) + \frac{\partial F}{\partial t}(t, ye^{rt}) + r y e^{rt} \frac{\partial F}{\partial x}(t, ye^{rt}) \right\} \end{aligned}$$

$$\begin{aligned}
&= e^{-rt} \left\{ -\frac{1}{2} \sigma^2 y^2 e^{2rt} \frac{\partial^2 F}{\partial x^2}(t, ye^{rt}) \right\} \quad (\text{by (16.8)}) \\
&= -\frac{1}{2} \sigma^2 y^2 e^{rt} \frac{\partial^2 F}{\partial x^2}(t, ye^{rt}),
\end{aligned}$$

$$\frac{\partial \tilde{F}(t, x)}{\partial x} = \frac{\partial}{\partial x} (e^{-rt} F(t, xe^{rt})) = \frac{\partial F}{\partial x} \Big|_{(t, xe^{rt})},$$

and

$$\frac{\partial^2 \tilde{F}(t, x)}{\partial x^2} = \frac{\partial^2}{\partial x^2} (e^{-rt} F(t, xe^{rt})) = e^{rt} \frac{\partial^2 F}{\partial x^2} \Big|_{(t, xe^{rt})}.$$

Hence

$$\frac{\partial \tilde{F}}{\partial t}(t, x) = -\frac{1}{2} \sigma^2 x^2 \frac{\partial^2 \tilde{F}}{\partial x^2}(t, x). \quad (16.9)$$

On the other hand, from $d\tilde{S}_t = \sigma \tilde{S}_t dX_t$ we obtain

$$(d\tilde{S}_t)^2 = \sigma^2 \tilde{S}_t^2 dt,$$

and by applying Itô's lemma to $\tilde{F}(t, \tilde{S}_t)$, and using (16.9), we obtain

$$\begin{aligned}
d\tilde{F}(t, \tilde{S}_t) &= \frac{\partial \tilde{F}}{\partial t}(t, \tilde{S}_t)dt + \frac{\partial \tilde{F}}{\partial x}(t, \tilde{S}_t)d\tilde{S}_t + \frac{1}{2} \frac{\partial^2 \tilde{F}}{\partial x^2}(t, \tilde{S}_t)\sigma^2 \tilde{S}_t^2 dt \\
&= \frac{\partial \tilde{F}}{\partial x}(t, \tilde{S}_t)d\tilde{S}_t.
\end{aligned}$$

Since

$$\tilde{V}_t = e^{-rt} V_t = e^{-rt} F(t, S_t) = \tilde{F}(t, e^{-rt} S_t) = \tilde{F}(t, \tilde{S}_t),$$

we have

$$d\tilde{V}_t = d\tilde{F}(t, \tilde{S}_t) = \frac{\partial \tilde{F}}{\partial x}(t, \tilde{S}_t)d\tilde{S}_t.$$

Hence (16.3) implies that

$$\Delta = \phi_t = \frac{\partial \tilde{F}}{\partial x}(t, \tilde{S}_t) = \frac{\partial F}{\partial x}(t, S_t).$$

Remark 16.2 (i) In the case of a European call option we can show easily $\Delta = N(d_1)$ using the martingale method as done in Theorem 15.3. First, put

$$d_1, d_2 = \frac{\log \frac{S_t}{K} + (r \pm \frac{1}{2} \sigma^2)(T - t)}{\sigma \sqrt{T - t}}.$$

Now note that

$$V_t = F(t, S_t) = S_t N(d_1) - K e^{-rT} N(d_2) e^{rt}$$

and $B_t = e^{rt}$. Then by checking the coefficients of S_t and B_t , we find

$$\begin{cases} \phi_t = N(d_1) , \\ \psi_t = -K e^{-rT} N(d_2) . \end{cases}$$

- (ii) In the standard continuous hedging strategy (as in the derivation of the Black–Scholes–Merton equation) we consider a portfolio

$$\Pi_t = -V_t + D_t + \Delta S_t$$

from the viewpoint of the option seller, where D_t is the bank deposit. Then

$$K e^{-r(T-t)} N(d_2) = \Delta_t S_t - V_t = \Pi_t - D_t ,$$

which is risk-free. Suppose that we, as a seller of the option, start with no money, i.e., $\Pi_0 = 0$. This means that we sell an option and using the option premium received, we buy some shares of the underlying asset and invest the left-over cash in risk-free bond. Note that $0 = -V_0 + D_0 + \Delta_0 S_0$, and hence

$$D_0 = V_0 - \Delta_0 S_0 = -K e^{-rT} N(d_2) ,$$

which is the borrowed amount at the start.

16.6 Computer Experiments

Simulation 16.1 (Martingale Method) We compare the martingale method with the Black–Scholes–Merton formula. See Figs. 16.2, 16.3, where the call price and Delta are approximately equal to 5.5871 and 0.4108, respectively.

```
S0 = 100; K = 110; sigma = 0.3; r = 0.05; T = 0.5;

% Black-Scholes-Merton formula
d1 = (log(S0/K)+(r+0.5*sigma^2)*T)/(sigma*sqrt(T)); d2 = d1 - sigma*sqrt(T);
N1 = normcdf(d1); N2 = normcdf(d2);

Call_formula = S0*N1 - K*exp(-r*T)*N2;
Delta_formula = N1;
```

Fig. 16.2 Convergence to the Black–Scholes–Merton price

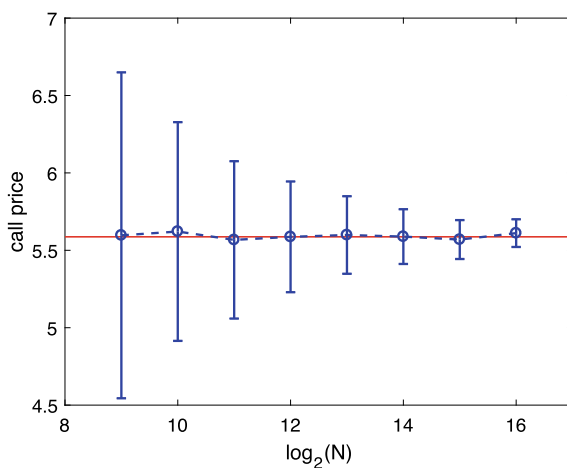
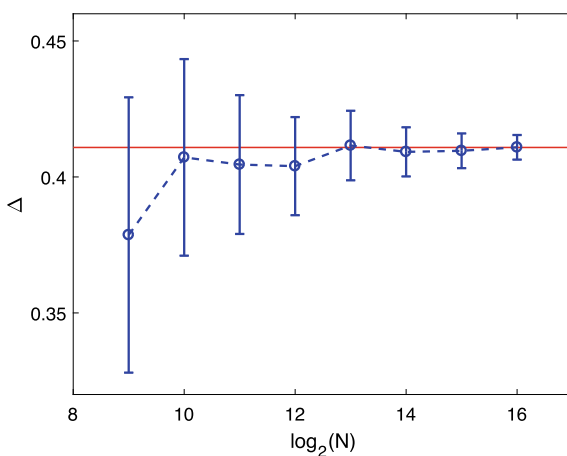


Fig. 16.3 Convergence to the Black–Scholes–Merton formula for Delta



```
% option price
J = 16; L = 2^J;
X = sqrt(T)*randn(L,1);
S = S0*exp((r-0.5*sigma^2)*T + sigma*X); % S_T
V = exp(-r*T)*max(S - K,0);
```

```
figure(1)
for j = 9:J
    M(j) = 2^j; % number of samples
    a(j) = mean(V(1:M(j)));
    b(j) = 1.96*std(V(1:M(j)))/sqrt(M(j));
end
x = 8:0.01:J+1;
```

```

plot(x,Call_formula,'r')
hold on
errorbar(9:J, a(9:J),b(9:J));
xlabel('log_2(N)')
ylabel('call price');
hold off

% Delta
dS = 0.001;
S1 = (S0+dS)*exp((r-0.5*sigma^2)*T + sigma*X);
dVdS = exp(-r*T)*(max(S1-K,0)-max(S-K,0))/dS;
hold off

figure(2)
for j = 9:J
M(j) = 2^j; % number of samples
a2(j) = mean(dVdS(1:M(j)));
b2(j) = 1.96*std(dVdS(1:M(j)))/sqrt(M(j));
end
x = 8:0.01:J+1;
plot(x,Delta_formula,'-r')
hold on
errorbar(9:J, a2(9:J),b2(9:J));
xlabel('log_2(N)')
ylabel('\Delta');

```

Exercises

16.1 (A trivial application of the martingale method) Consider a European option that pays \$1 at time T . Find its price V_0 at $t = 0$. Its price V_0 at $t = 0$ is given by the expectation with respect to a martingale measure $\mathbb{Q}$, and $V_0 = e^{-rT} \mathbb{E}^{\mathbb{Q}}[1] = e^{-rT}$, which is nothing but the risk-free bond price.

16.2 Plot the graph of the integrand in (16.5) as a function of x over an interval $[0, n]$ for some large n , and compute the integral numerically.

Part VI

Examples of Option Pricing

Chapter 17

Pricing of Vanilla Options



In this chapter we present more examples of pricing of vanilla options, including the cases when the underlying stock pays continuous or periodic dividends. Pricing formula for binary options are derived using five different approaches. For more information consult Bjork (2009) and Etheridge (2002).

17.1 Stocks with a Dividend

A dividend is a distribution of a company's profits to its shareholders. Dividends are usually issued as cash payments or shares of stock. The dividend rate may mean the cash amount each share receives, or a percent of the current market price.

When a stock pays a dividend, it is periodically paid. Although dividends are not paid continuously in the real financial world, for theoretical analysis we also consider the case when the dividend is paid continuously.

17.1.1 Continuous Dividends

Suppose that the asset price follows the geometric Brownian motion

$$S_t = S_0 \exp(\nu t + \sigma W_t) .$$

Over the time interval $[t, t + dt)$ for a sufficiently short time length $dt > 0$, we receive a dividend $\delta S_t dt$ per share for some constant $\delta > 0$, and buy δdt shares of stock and add them to the existing stocks. If we let $g(t)$ be the number of stocks at t , then $g(t + dt) - g(t) = g(t) \delta dt$ since we can use the dividend $g(t) \delta S_t dt$ to buy $g(t) \delta dt$ shares of stock. In other words, $g(t)$ satisfies an ordinary differential

equation $g'(t) = g(t) \delta$. Hence $g(t) = g(0) e^{\delta t}$. If we begin with one share at $t = 0$, i.e., $g(0) = 1$, then $g(t) = e^{\delta t}$. If we reinvest the dividend into stock continuously, then the total value of stock is equal to

$$Z_t = g(t) S_0 \exp(\nu t + \sigma W_t) = S_0 \exp((\nu + \delta)t + \sigma W_t) .$$

Regarding Z_t as a new underlying asset, we apply the martingale method to compute the option price.

First, let

$$\tilde{Z}_t = e^{-rt} Z_t = S_0 \exp((\nu + \delta - r)t + \sigma W_t) .$$

Then

$$d\tilde{Z}_t = (\nu + \delta + \frac{1}{2}\sigma^2 - r)\tilde{Z}_t dt + \sigma \tilde{Z}_t dW_t .$$

If we let

$$\theta = \frac{\nu + \delta + \frac{1}{2}\sigma^2 - r}{\sigma}$$

and define a probability measure $\mathbb{Q}$ by

$$\left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t} = \exp \left(-\theta W_t - \frac{1}{2}\theta^2 t \right) ,$$

then $X_t = W_t + \theta t$ is a $\mathbb{Q}$ -Brownian motion, and $\{\tilde{Z}_t\}_{t \geq 0}$ is a $\mathbb{Q}$ -martingale.

Thus the price of a call option is given by

$$\begin{aligned} V_t &= e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[(S_T - K)^+ | \mathcal{F}_t] \\ &= e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[(e^{-\delta T} Z_T - K)^+ | \mathcal{F}_t] \\ &= e^{-\delta T} e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[(Z_T - e^{\delta T} K)^+ | \mathcal{F}_t] . \end{aligned}$$

This value is equal to the price of $e^{-\delta T}$ contracts of call options on the underlying asset $\{Z_t\}_{t \geq 0}$ with maturity date T and exercise price $e^{\delta T} K$. If we substitute the result into the previous formula, then

$$\begin{aligned} V_t &= e^{-\delta T} (Z_t N(c_1) - e^{\delta T} K e^{-r(T-t)} N(c_2)) \\ &= S_t N(c_1) - K e^{-r(T-t)} N(c_2) , \end{aligned}$$

where

$$c_1, c_2 = \frac{\log \frac{Z_t}{e^{\delta T} K} + (r \pm \frac{1}{2}\sigma^2)(T-t)}{\sigma \sqrt{T-t}} = \frac{\log \frac{S_t}{K} + (r \pm \frac{1}{2}\sigma^2 - \delta)(T-t)}{\sigma \sqrt{T-t}} .$$

Now, if we let

$$d_1, d_2 = \frac{\log \frac{F_t}{K} \pm \frac{1}{2} \sigma^2 (T-t)}{\sigma \sqrt{T-t}} = \frac{\log \frac{S_t}{K} + (r - \delta \pm \frac{1}{2} \sigma^2) (T-t)}{\sigma \sqrt{T-t}} = c_1, c_2$$

and

$$F_t = e^{(r-\delta)(T-t)} S_t ,$$

i.e., if we let F_t be the forward price of the stock, then we can write

$$V_t = e^{-r(T-t)} (F_t N(d_1) - K N(d_2)) .$$

Remark 17.1 Let us find the replicating portfolio for $V_t = D_t + \Delta_t S_t$. If

$$\phi_t = N(c_1) = N(d_1) ,$$

then we have to keep $e^{-\delta T} \phi_t$ units of Z_t . In other words, we have to keep $e^{-\delta(T-t)} \phi_t$ units of S_t . On the other hand, since

$$V_t - Z_t e^{-\delta T} \phi_t = -K e^{-rT} N(c_2) e^{rt} ,$$

we have to keep $-K e^{-rT} N(c_2) = -K e^{-rT} N(d_2)$ units of bond.

17.1.2 Periodic Dividends

In the real world dividends are paid periodically, not continuously. If a dividend δS_{T_i} is paid at predetermined dates T_i , $i \geq 1$, then under the assumption that the price of the stock S_t is continuous from the left,

$$S_{T_i+} - S_{T_i-} = -\delta S_{T_i-}$$

since S_t decreases instantaneously by δS_{T_i-} as time passes by across T_i . (As for the notation, we write $\lim_{t \rightarrow T_i \pm} S_t(\omega) = S_{T_i \pm}(\omega)$ for every Brownian sample path ω .) Hence

$$S_{T_i+} = (1 - \delta) S_{T_i-} ,$$

and every time when dividend is paid the stock price is multiplied by $1 - \delta$. If we let $n[t]$ be the number of times when dividend is paid until time t , i.e.,

$$n[t] = \max\{i : T_i \leq t\} ,$$

then the stock price is given by

$$S_t = S_0 (1 - \delta)^{n[t]} \exp(\nu t + \sigma W_t) .$$

If at every time the dividend is paid we increase the number of shares of the stock by buying the stock using the dividend, the instantaneous increment in the number of shares across the time T_i is equal to

$$\frac{\delta S_{T_i-}}{S_{T_i+}} = \frac{\delta}{1 - \delta} .$$

Hence immediately after the time T_i the number of shares of the stock increases by the ratio of $1 + \frac{\delta}{1-\delta} = \frac{1}{1-\delta}$. If we denote such a portfolio by Z_t , then it satisfies

$$Z_t = \frac{1}{(1 - \delta)^{n[t]}} S_t = S_0 \exp(vt + \sigma W_t) .$$

Thus defined Z_t may be regarded as a stock without dividend. Put

$$\theta = \frac{v + \frac{1}{2}\sigma^2 - r}{\sigma}$$

and define a probability measure $\mathbb{Q}$ by

$$\left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t} = \exp \left(-\theta W_t - \frac{1}{2}\theta^2 t \right) .$$

Then $X_t = W_t + \theta t$ is a $\mathbb{Q}$ -Brownian motion, and the discounted process of Z_t , $\{\tilde{Z}_t\}_{t \geq 0}$, is a $\mathbb{Q}$ -martingale.

Theorem 17.1 *The fair price for a forward contract on a stock that periodically pays dividend is given by*

$$K = e^{rT} (1 - \delta)^{n[T]} S_0 .$$

Proof Since the value of the forward contract at expiry date T is equal to $C_T = S_T - K$, its value at time t is equal to

$$\begin{aligned} V_t &= \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)}(S_T - K)|\mathcal{F}_t] \\ &= \mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)}((1 - \delta)^{n[T]} Z_T - K)|\mathcal{F}_t] \\ &= (1 - \delta)^{n[T]} Z_t - e^{-r(T-t)} K \\ &= (1 - \delta)^{n[T]-n[t]} S_t - e^{-r(T-t)} K . \end{aligned}$$

Since $V_0 = 0$ for a fair contract, we have $(1 - \delta)^{n[T]} S_0 - e^{-rT} K = 0$. □

Remark 17.2 As seen in the proof, we have

$$V_t = (1 - \delta)^{n[T]} Z_t - e^{-r(T-t)} K ,$$

and hence we keep $(1 - \delta)^{n[T]}$ units of Z_t , i.e., $(1 - \delta)^{n[T]-n[t]}$ shares of the stock for hedging.

17.2 Bonds with Coupons

Consider a bond whose price changes stochastically in the financial market. Suppose that the bond pays coupons at predetermined dates $0 < T_1 < \dots < T_n < T$. Let C denote the coupon amount, and $r > 0$ the risk-free interest rate. If a bond investor saves the coupons in the form of a risk-free deposit at a bank, the sum of present values of the coupons until maturity is equal to $\sum_{i=1}^n C e^{-r T_i}$. Let

$$J(t) = \min \{i : t < T_i\} , \quad 0 \leq t \leq T .$$

That is, $J(t)$ is the first time a coupon payment is received after time t . For example, $J(T_k) = k + 1$ for $k = 1, \dots, n - 1$. Hence

$$\sum_{i=J(t)}^n C e^{-r(T_i-t)}$$

is the sum of values at the time t of coupons which will be received after t . Assume that the bond price S_t satisfies

$$S_t = \sum_{i=J(t)}^n C e^{-r(T_i-t)} + A e^{vt+\sigma W_t} \quad (17.1)$$

for some constants A , v and σ . Suppose that the bond investor deposits a coupon payment into a bank account immediately after it is paid. Then the value of the coupon payments until time t is equal to

$$\sum_{i=1}^{J(t)-1} C e^{-r(T_i-t)} .$$

Since the bond itself has value given by (17.1), the portfolio Z_t consisting of the bond itself and the coupons received until time t is given by

$$Z_t = \sum_{i=1}^n C e^{-r(T_i-t)} + A e^{vt+\sigma W_t} .$$

Now we look for a measure $\mathbb{Q}$ for which the discounted price $\tilde{Z}_t = e^{-rt} Z_t$ is a martingale. Since

$$\tilde{Z}_t = \sum_{i=1}^n C e^{-r T_i} + A e^{(v-r)t+\sigma W_t} ,$$

it suffices to find $\mathbb{Q}$ for which $A e^{(v-r)t+\sigma W_t}$ is a martingale. To find it, put

$$\theta = \frac{\nu + \frac{1}{2}\sigma^2 - r}{\sigma}$$

and let $\mathbb{Q}$ be the measure such that

$$\left. \frac{d\mathbb{Q}}{d\mathbb{P}} \right|_{\mathcal{F}_t} = e^{-\theta W_t - \frac{1}{2}\theta^2 t} ,$$

then $X_t = W_t + \theta t$ is a $\mathbb{Q}$ -Brownian motion, and $\{\tilde{Z}_t\}$ is a $\mathbb{Q}$ -martingale.

The value at time t of the option whose payoff is given by $C_T \in \mathcal{F}_T$ is given by

$$\mathbb{E}^{\mathbb{Q}}[e^{-r(T-t)} C_T | \mathcal{F}_t] .$$

At maturity T there remains no coupon to receive, and the bond price is equal to

$$S_T = Ae^{\nu T + \sigma W_T} = Ae^{(r - \frac{1}{2}\sigma^2)T + \sigma X_T} .$$

17.3 Binary Options

17.3.1 The Black–Scholes–Merton Formula for Binary Options

A European cash-or-nothing binary call option with strike price K pays an agreed amount A at expiry date T if the asset price S_T satisfies $S_T > K$, pays nothing if $S_T < K$, and $\frac{1}{2}A$ if $S_T = K$. If we ignore the probability zero case that $S_T = K$ in a continuous asset price model, then the payoff may be written as $A \times \mathbf{1}_{\{S_T \geq K\}}$. A European asset-or-nothing binary call option pays S_T if $S_T \geq K$, and pays nothing if $S_T < K$. Binary put options are defined similarly. A binary option is also called a *digital* option. The put-call parity relation for the cash-or-nothing option is given by

$$C(S_t, t) + P(S_t, t) = Ae^{-r(T-t)}$$

since

$$C(S_T, T) + P(S_T, T) = A ,$$

where $C(S, t)$ and $P(S, t)$ denote the prices of binary call and binary put with the same strike price Fig. 17.1.

If we combine a cash-or-nothing binary option and a European call option, we can construct an asset-or-nothing binary option since

$$K \times \mathbf{1}_{\{S_T \geq K\}} + (S_T - K)^+ = S_T \times \mathbf{1}_{\{S_T \geq K\}} .$$

Fig. 17.1 Payoffs of cash-or-nothing call and put (clo)

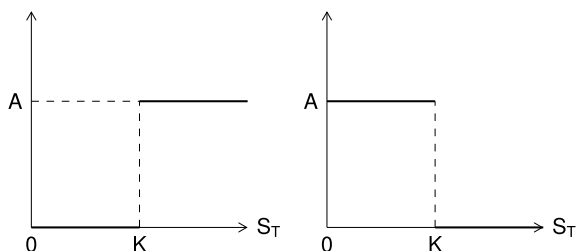
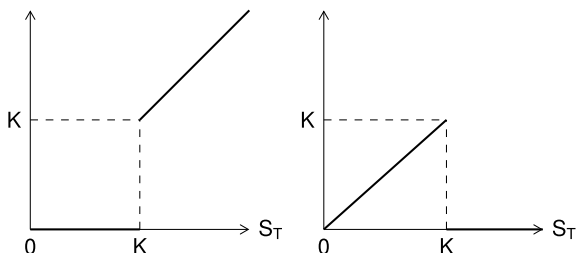


Fig. 17.2 Payoffs of asset-or-nothing call and put



The put-call parity for the asset-or-nothing option is given by

$$C(S_t, t) + P(S_t, t) = S_t .$$

(See Fig. 17.2.) In continuous asset price models such as the geometric Brownian motion model, it does not make any difference what value we assign to the payoff when $S_T = K$ since the probability of such an event is zero. However, in the real financial market stock price can have integral multiples of tick size, which is the minimum amount that the price of the stock can change. Thus we have to define the payoff for the event $S_T = K$. For example, we may take $\frac{1}{2}A$ as the payoff for $S_T = K$ in the case of a cash-or-nothing binary option.

A cash-or-nothing binary option can be dynamically hedged. However, if near the expiry date the asset price is close to the exercise price then hedging is difficult in practice.

Theorem 17.2 *The price C of a cash-or-nothing call option is given by*

$$C(S, t) = e^{-r(T-t)} N(d_2) ,$$

where d_2 is identical as in the formula for the European call.

Here we present five different proofs. For the sake of notational simplicity we find $C(S, t)$ for $t = 0$.

Proof (Method 1: Partial Differential Equation) To derive the option price formula, we solve the Black–Scholes–Merton equation with the boundary conditions

Fig. 17.3 Price of a cash-or-nothing call option

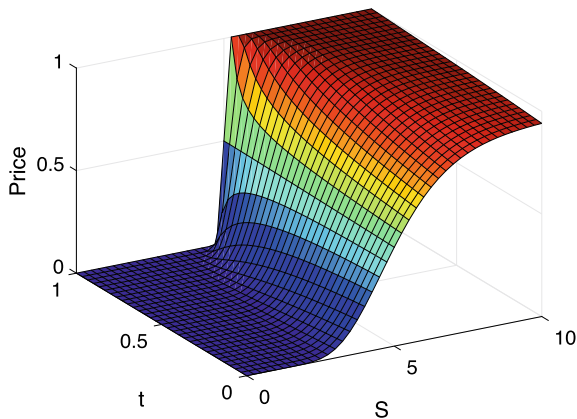
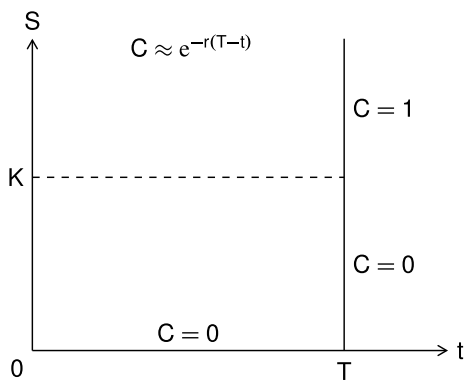


Fig. 17.4 Boundary conditions of a cash-or-nothing call option



$$\begin{cases} C(0, t) = 0, & 0 \leq t \leq T, \\ C(S, T) = 0, & S < K, \\ C(S, T) = 1, & K < S, \\ C(S, t) \approx e^{-r(T-t)}, & S \approx \infty. \end{cases}$$

See Fig. 17.4. The last condition means that if the stock price is very high then the probability of exercising the option at expiry date is close to 1 and its discounted value is the present option price.

As in the case for a European call option, put

$$\begin{cases} S &= K e^x, \\ t &= T - \frac{1}{\sigma^2} \tau, \\ V(S, t) &= K v(x, \tau). \end{cases}$$

Then

$$\sigma^2 \frac{\partial v}{\partial \tau} = \frac{1}{2} \sigma^2 \frac{\partial^2 v}{\partial x^2} + \left(r - \frac{1}{2} \sigma^2 \right) \frac{\partial v}{\partial x} - r v.$$

If we let $C = \frac{r}{\sigma^2}$, then

$$\frac{\partial v}{\partial \tau} = \frac{1}{2} \frac{\partial^2 v}{\partial x^2} + \left(C - \frac{1}{2}\right) \frac{\partial v}{\partial x} - C v .$$

Since $V(S, T) = \mathbb{1}_{\{S \geq K\}}$, the initial condition for v is given by

$$v(x, 0) = \frac{1}{K} V(K e^x, T) = \frac{1}{K} \mathbb{1}_{\{x \geq 0\}} .$$

Now we look for v of the form

$$v(x, \tau) = e^{ax + \beta \tau} u(x, \tau)$$

for some α, β, u . If such a solution exists, then we have

$$\beta u + \frac{\partial u}{\partial \tau} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2} + \left(C - \frac{1}{2} + \alpha\right) \frac{\partial u}{\partial x} + \left(\frac{1}{2} \alpha^2 + (C - \frac{1}{2}) \alpha - C\right) u .$$

If we take $\alpha = -C + \frac{1}{2}$ and

$$\beta = \frac{1}{2} \alpha^2 + \left(C - \frac{1}{2}\right) \alpha - C = -\frac{1}{2} \left(C + \frac{1}{2}\right)^2 ,$$

then

$$\frac{\partial u}{\partial \tau} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}$$

and

$$u(x, 0) = u_0(x) = e^{-\alpha x} v(x, 0) = \frac{1}{K} e^{(C - \frac{1}{2})x} \mathbb{1}_{\{x \geq 0\}} .$$

Hence

$$\begin{aligned} & e^{ax + \beta \tau} u(x, \tau) \\ &= e^{ax + \beta \tau} \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^{\infty} u_0(\xi) e^{-(x-\xi)^2/(2\tau)} d\xi \\ &= e^{ax + \beta \tau} \frac{1}{\sqrt{2\pi\tau}} \int_0^{\infty} \frac{1}{K} e^{(C - \frac{1}{2})\xi} e^{-(x-\xi)^2/(2\tau)} d\xi \\ &= \frac{1}{K} e^{(-C + \frac{1}{2})x - \frac{1}{2}(C + \frac{1}{2})^2 \tau} \frac{1}{\sqrt{2\pi\tau}} \int_{-\infty}^0 e^{-(\eta + x + (C - \frac{1}{2})\tau)^2/2\tau + (C - \frac{1}{2})x + \frac{1}{2}(C - \frac{1}{2})^2 \tau} d\eta \\ &= \frac{1}{K} e^{-C\tau} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x + (C - \frac{1}{2})\tau)/\sqrt{\tau}} e^{-\zeta^2/2} d\zeta \end{aligned}$$

$$= \frac{1}{K} e^{-C\tau} N(d_2) ,$$

where

$$d_2 = \frac{x + (C - \frac{1}{2})\tau}{\sqrt{\tau}} = \frac{\log \frac{S}{K} + (r - \frac{1}{2}\sigma^2)(T - t)}{\sigma \sqrt{T - t}} .$$

Since $C\tau = r(T - t)$, we conclude $V(S, t) = K v(x, \tau) = e^{-r(T-t)} N(d_2)$. $\square$

Proof (Method 2: Risk-neutral Method) Suppose that the underlying asset S_t follows a geometric Brownian motion $dS_t = rS_t dt + \sigma S_t dW_t$ under a risk-neutral measure $\mathbb{Q}$. The option price is given by

$$C(S, t) = \mathbb{E}[e^{-r(T-t)} \mathbb{1}_{\{S_T \geq K\}} | S_t] ,$$

where the conditional expectation is taken with respect to $\mathbb{Q}$. Under the condition S_t at time t , we have $S_T = S_t e^{(r-1/2\sigma^2)(T-t) + \sigma W_{T-t}}$ and

$$\begin{aligned} \mathbb{E}[\mathbb{1}_{\{S_T \geq K\}} | S_t] &= \mathbb{E}[\mathbb{1}_{\{S_t e^{(r-1/2\sigma^2)(T-t) + \sigma W_{T-t}} \geq K\}}] \\ &= \mathbb{Q}\{S_t e^{(r-1/2\sigma^2)(T-t) + \sigma W_{T-t}} \geq K\} \\ &= \mathbb{Q}\left\{\left(r - \frac{1}{2}\sigma^2\right)(T - t) + \sigma W_{T-t} \geq \log \frac{K}{S_t}\right\} \\ &= \mathbb{Q}\left\{W_{T-t} > \frac{\log \frac{K}{S_t} - (r - \frac{1}{2}\sigma^2)(T - t)}{\sigma}\right\} \\ &= \mathbb{Q}\left\{Z > \frac{\log \frac{K}{S_t} - (r - \frac{1}{2}\sigma^2)(T - t)}{\sigma \sqrt{T - t}}\right\} \quad (\text{where } Z \sim N(0, 1)) \\ &= \mathbb{Q}\left\{Z < \frac{\log \frac{S_t}{K} + (r - \frac{1}{2}\sigma^2)(T - t)}{\sigma \sqrt{T - t}}\right\} \quad (\text{since } -Z \sim N(0, 1)) \\ &= N(d_2) . \end{aligned}$$

Hence $C(S_t, t) = e^{-r(T-t)} N(d_2)$. $\square$

Proof (Method 3: Differentiation of a European Call) There is a simpler method of finding the price of a digital option if we already have the classical Black–Scholes–Merton formula for a European call option. Let $C(K)$ denote the price of a European call option with exercise price K . Assume $K < K_1$. Then the price of a European call option with exercise price K_1 is equal to $C(K_1)$. If we consider a new option defined by the difference of two European call options with exercise prices K and K_1 , respectively, then its payoff at expiry T is given by

$$\begin{cases} 0 & \text{if } 0 \leq S_T < K , \\ S_T - K & \text{if } K \leq S_T < K_1 , \\ K_1 - K & \text{if } K_1 \leq S_T , \end{cases}$$

Fig. 17.5 Difference of two European call options

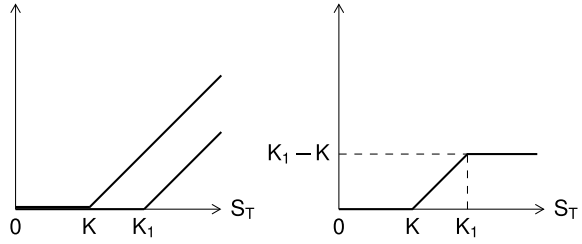
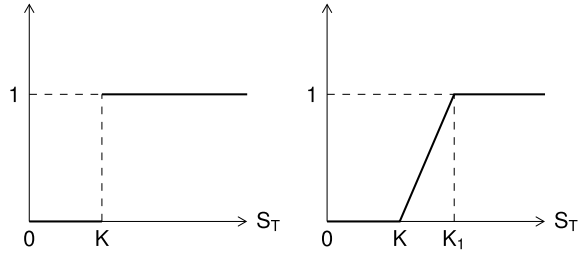


Fig. 17.6 Approximation of a cash-or-nothing call option



and its price is given by $C(K) - C(K_1)$. See Fig. 17.5. Now note that the cash-or-nothing call option with payoff $\mathbb{1}_{\{S_T \geq K\}}$ can be approximated by a European option whose payoff at expiry date T is given by the second graph in Fig. 17.6. Hence the price of the cash-or-nothing call option with exercise price K can be approximated by $\frac{1}{K_1 - K}(C(K) - C(K_1))$ as shown in Fig. 17.6. Therefore the price of the digital option is equal to the limit

$$\lim_{K_1 \rightarrow K} \frac{C(K) - C(K_1)}{K_1 - K} = -\frac{\partial C}{\partial K} = e^{-r(T-t)} N(d_2) .$$

See Remark 15.3 for the second equality. $\square$

Proof (Method 4: Differentiation of a Binary Call) Let B denote the price of a cash-or-nothing binary call option. It will be regarded as a function of strike price K while the other parameters are fixed. Then the price of the European option with the payoff given in Fig. 17.7 is equal to

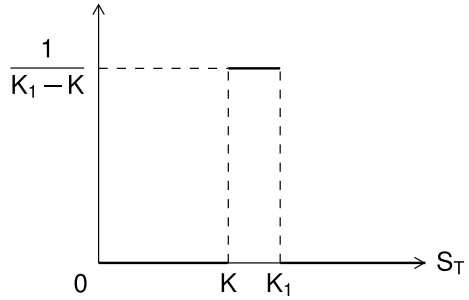
$$\frac{1}{K_1 - K}(B(K) - B(K_1)) ,$$

which converges to $-\frac{\partial B}{\partial K}$ as $K_1 \rightarrow K$.

On the other hand, according to the risk-neutral method the option price is equal to

$$e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}} \left[\frac{1}{K_1 - K} \mathbb{1}_{\{K \leq S_T \leq K_1\}}(S_T) \middle| S_t \right] . \quad (17.2)$$

Fig. 17.7 Difference of two binary call options



Since Fig. 17.7 is an approximation of a point mass $\delta_{\{S_T=K\}}$, called the Dirac delta functional at K , the conditional expectation in (17.2) converges to

$$e^{-r(T-t)} \mathbb{E}^{\mathbb{Q}}[\delta_{\{S_T=K\}}(S_T)|S_t] \quad (17.3)$$

as $K_1 \rightarrow K$.

By the formula in Example 4.17 the conditional pdf $f(y)$, $0 < y < \infty$, of S_T , under the condition defined by S_t , is given by

$$f(y) = \frac{1}{\sigma y \sqrt{2\pi(T-t)}} \exp\left(-\frac{(\log y - \log S_t - (r - \frac{1}{2}\sigma^2)(T-t))^2}{2\sigma^2(T-t)}\right).$$

Then the conditional expectation in (17.3) is equal to $e^{-r(T-t)} f(K)$. Since

$$\begin{aligned} f(K) &= \frac{1}{\sigma K \sqrt{2\pi(T-t)}} \exp\left(-\frac{(\log K - \log S_t - (r - \frac{1}{2}\sigma^2)(T-t))^2}{2\sigma^2(T-t)}\right) \\ &= \frac{1}{\sigma K \sqrt{2\pi(T-t)}} \exp\left(-\frac{1}{2}d_2^2\right) \end{aligned}$$

and since two limits should be equal, we have

$$-\frac{\partial B}{\partial K} = e^{-r(T-t)} \frac{1}{\sigma K \sqrt{2\pi(T-t)}} \exp\left(-\frac{1}{2}d_2^2\right).$$

Now, since

$$\frac{\partial N(d_2)}{\partial K} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}d_2^2\right) \left(-\frac{1}{K\sigma\sqrt{T-t}}\right),$$

we conclude that

$$\frac{\partial B}{\partial K} = e^{-r(T-t)} \frac{\partial N(d_2)}{\partial K}$$

and hence

$$B(K) = e^{-r(T-t)}N(d_2) + c$$

for some constant c . Since $B(0) = e^{-r(T-t)}$ and $N(d_2(0)) = 1$, we have $c = 0$. $\square$

Proof (Method 5: Laplace Transformation Approach) As in Sect. 15.5, we let $\tau = T - t$ and put $v(S, \tau) = V(S, T - \tau)$, and obtain

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 v}{\partial S^2} + rS \frac{\partial v}{\partial S} - \frac{\partial v}{\partial \tau} = rv \quad (17.4)$$

with the initial data

$$v(S, 0) = V(S, T) .$$

Let $f(S, s)$ be the Laplace transformation $\mathcal{L}[v]$ of $v(S, \tau)$. Since

$$\mathcal{L}\left[\frac{\partial v}{\partial \tau}\right](S, s) = s\mathcal{L}[v](S, s) - v(S, 0) ,$$

after taking the Laplace transformations of both sides of (17.4), we obtain

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - (s+r)f = -v(S, 0) . \quad (17.5)$$

To solve it, first find a solution of the homogeneous equation

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - (s+r)f = 0 . \quad (17.6)$$

In this case there exist two linearly independent solutions, and we look for solutions of the form

$$f(S, s) = S^\lambda . \quad (17.7)$$

Note that $\lambda \neq 1, 2$. Substituting (17.7) into (17.6), we obtain

$$\frac{1}{2}\sigma^2 S^2 \lambda(\lambda - 1)S^{\lambda-2} + rS\lambda S^{\lambda-1} - (s+r)S^\lambda = 0 .$$

Eliminating S^λ from the both sides, we obtain the quadratic equation

$$\frac{1}{2}\sigma^2 \lambda^2 + \left(-\frac{1}{2}\sigma^2 + r\right)\lambda - (s+r) = 0 . \quad (17.8)$$

Two zeros of (17.8) are given by

$$\lambda_1, \lambda_2 = \frac{-(r - \frac{1}{2}\sigma^2) \pm \sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s+r)\sigma^2}}{\sigma^2} .$$

For $s > 0$ we have

$$\lambda_1 > \frac{-(r - \frac{1}{2}\sigma^2) + \sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2r\sigma^2}}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1 ,$$

$$\lambda_2 < \frac{-(r - \frac{1}{2}\sigma^2) - \sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2r\sigma^2}}{\sigma^2} = -\frac{2r}{\sigma^2} < 0 .$$

In summary, a homogeneous solution f_H of (17.6) is of the form

$$f_H = c_1(s)S^{\lambda_1} + c_2(s)S^{\lambda_2} ,$$

where $c_1(s)$ and $c_2(s)$ depend only on s but not on S .

Now we recall that the initial condition is given by

$$v(S, 0) = V(S, T) = \mathbf{1}_{\{S \geq K\}} .$$

Hence (17.5) becomes

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - (s+r)f = -\mathbf{1}_{\{S \geq K\}} . \quad (17.9)$$

In the region $S \geq K$, we have

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + rS \frac{\partial f}{\partial S} - (s+r)f = -1 ,$$

and we have a particular solution

$$f_P(S, s) = \frac{1}{s+r} .$$

Therefore we try to find a general solution f of (17.9) of the form

$$f(S, s) = \begin{cases} a_1 S^{\lambda_1} + a_2 S^{\lambda_2} + \frac{1}{s+r} , & S \geq K , \\ a_3 S^{\lambda_1} + a_4 S^{\lambda_2} , & S < K , \end{cases}$$

where a_i may depend on s . To determine a_i in the region $S \geq K$ we note that $v(S, \tau) \rightarrow 1$ as $S \rightarrow +\infty$ and hence $f(S, s)$ must be bounded as $S \rightarrow +\infty$, which is possible only when $a_1 = 0$. On the other hand, if $S < K$, we have $v(S, \tau) \rightarrow 0$ as $S \downarrow 0$, and hence $f(S, s) \rightarrow 0$ as $S \downarrow 0$, thus we conclude that $a_4 = 0$. If we summarize what has been obtained so far, we have

$$f(S, s) = \begin{cases} c_2(s)S^{\lambda_2} + \frac{1}{s+r}, & S \geq K, \\ c_1(s)S^{\lambda_1}, & S < K, \end{cases}$$

for some c_1 and c_2 . Since $f(S, s)$ is continuous and differentiable at (K, s) , we have

$$\begin{cases} c_2(s)K^{\lambda_2} + \frac{1}{s+r} = c_1(s)K^{\lambda_1}, \\ c_2(s)\lambda_2 K^{\lambda_2-1} = c_1(s)\lambda_1 K^{\lambda_1-1}. \end{cases}$$

Hence

$$\begin{cases} c_1(s) = \frac{\lambda_2}{\lambda_1 - \lambda_2} K^{-\lambda_1} \frac{1}{s+r}, \\ c_2(s) = \frac{\lambda_1}{\lambda_1 - \lambda_2} K^{-\lambda_2} \frac{1}{s+r}. \end{cases}$$

Note that

$$\begin{aligned} & \frac{\lambda_1}{\lambda_1 - \lambda_2} \\ &= \frac{-(r - \frac{1}{2}\sigma^2) + \sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s+r)\sigma^2}}{2\sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s+r)\sigma^2}} \\ &= \frac{(s+r)\sigma^2}{\sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s+r)\sigma^2}} \frac{1}{r - \frac{1}{2}\sigma^2 + \sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s+r)\sigma^2}} \end{aligned}$$

and

$$\begin{aligned} & \frac{\lambda_2}{\lambda_1 - \lambda_2} \\ &= \frac{-(r - \frac{1}{2}\sigma^2) - \sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s+r)\sigma^2}}{2\sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s+r)\sigma^2}} \\ &= \frac{(s+r)\sigma^2}{\sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s+r)\sigma^2}} \frac{1}{r - \frac{1}{2}\sigma^2 - \sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s+r)\sigma^2}}. \end{aligned}$$

Since, for $S < K$, we have

$$\begin{aligned}
& f(S, s) \\
&= \frac{\lambda_2}{\lambda_1 - \lambda_2} \left(\frac{S}{K} \right)^{\lambda_1} \frac{1}{s + r} \\
&= \frac{\sigma^2}{\sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s + r)\sigma^2}} \frac{\exp\left(\lambda_1 \log \frac{S}{K}\right)}{r - \frac{1}{2}\sigma^2 - \sqrt{(r - \frac{1}{2}\sigma^2)^2 + 2(s + r)\sigma^2}}.
\end{aligned}$$

Now we take the inverse Laplace transformation and obtain $v(S, \tau) = e^{-r\tau} N(d_2)$ and finally we have $V(S, t) = e^{-r(T-t)} N(d_2)$. $\square$

17.3.2 Delta of the Binary Option

Theorem 17.3 Consider a cash-or-nothing call with payoff $\mathbf{1}_{\{S_T \geq K\}}$, i.e., it pays 1 at time T if $S_T \geq K$. The hedge ratio Δ is given by

$$\Delta = \frac{\partial C}{\partial S} = \frac{e^{-r(T-t)} N'(d_2)}{\sigma S \sqrt{T-t}}$$

and satisfies

$$\lim_{t \rightarrow T-} \Delta = \begin{cases} 0, & S_T > K \\ 0, & S_T < K \end{cases}$$

with probability 1 where the limit is taken as t converges to T from the left.

Proof In the following argument, we consider only continuous sample paths for S_t since they are continuous with probability 1.

(i) Suppose $S_T > K$. Since

$$d_2(t, S_t) \approx \frac{\log \frac{S_T}{K} + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma \sqrt{T-t}} \approx C_1 \frac{1}{\sqrt{T-t}}$$

for $T-t \approx 0$, we have

$$N'(d_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_2^2} \approx C_2 e^{-C_3/(T-t)}$$

and

$$\Delta \approx C_4 \frac{e^{-C_3/(T-t)}}{\sqrt{T-t}} \approx 0$$

for $T-t \approx 0$ where C_i , $1 \leq i \leq 4$, are positive constants.

(ii) Suppose $S_T < K$. Since there exists a constant $C > 0$ such that

$$\begin{aligned} d_2(t, S_t) &\approx d_2(t, S_T) \\ &= \frac{\log \frac{S_T}{K} + (r - \frac{1}{2}\sigma^2)(T - t)}{\sigma\sqrt{T - t}} \\ &\approx -C \frac{1}{\sqrt{T - t}} \end{aligned}$$

for $T - t \approx 0$, we have $\Delta \approx 0$ for $T - t \approx 0$ as in the case (i). $\square$

In Figs. 17.8 and 17.9 we take $K = 5$ and $T = 1$, and plot the graph of Delta as a function of t and S , and as a function of S for each fixed t as $t \rightarrow T$. For $t = 0.99998$

Fig. 17.8 Delta of a cash-or-nothing call option as a function of t and S with $K = 5$ and $T = 1$

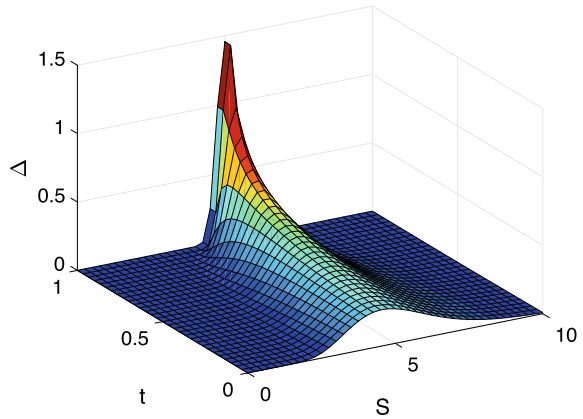
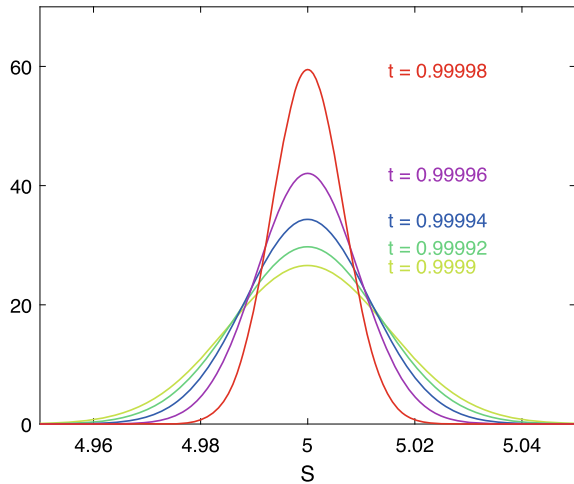


Fig. 17.9 Delta of a cash-or-nothing call option as a function of S with $K = 5$ and $T = 1$ as $t \rightarrow T$



the maximum of Delta at $S = K$ is about 60, and it increases to $+\infty$ as $t \rightarrow T$. (See also Fig. 17.3.)

- Remark 17.3** (i) The conclusion of the theorem implies that if $S > K$ or $S < K$ at t near expiry date T then with almost certainty the seller of the option has to pay to the buyer of the option either \$1 or nothing, and hence, close to expiry date, either there is no need for the option seller to hedge by keeping stocks or there is no way to hedge.
- (ii) Note that Delta is inversely proportional to the asset price in Theorem 17.3.
- (iii) Consider the case that $S_T = K$. For $t \approx T$, we have

$$d_2(t, S_t) \approx \frac{\log \frac{S_t}{K}}{\sigma \sqrt{T-t}},$$

whose limit cannot be determined. Hence the limits of $N'(d_2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_2^2}$ and Δ cannot be determined as $t \rightarrow T$.

For Delta-hedging of the at-the-money European digital call option we use a self-financing portfolio

$$\Pi = -V + D + \Delta S.$$

(See Fig. 17.10.) We investigate the limiting behavior of the cash-or-nothing call option along five individual sample paths when the asset price S_t converges to the strike price K as $t \rightarrow T$. The option pays 1 if $S_T > K$, and pays nothing if $S_T < K$. The value for D_0 can be chosen as we like, and in the simulation we take $D_0 = 0$. (See Simulation 17.3.)

To generate a sample asset path such that $S_T = K$, we use a Brownian bridge path $X_t^{0 \rightarrow b}$ from 0 to b where b satisfies

$$S_0 e^{(\mu - \frac{1}{2}\sigma^2)T + \sigma b} = K.$$

Observe that the option price V fluctuates between two extreme values 0 and 1, and that Δ may become very large as $t \rightarrow T$. As one can see easily, it is almost impossible to perform Delta-hedging and obtain a risk-free portfolio $\Pi_t = \Pi_0 e^{rt}$ as $t \rightarrow T$. In the last plot for Π the exponential growth of $\Pi_0 e^{rt}$ over the time interval $[0, T]$ is not visible due to the small value of $r > 0$.

Remark 17.4 (European Call Option Price) We can derive the formula of the price of a European call option from that of the price of a binary option. Let K be the strike price of a given European call option. Its payoff is approximated by a sum of payoffs of infinitely many binary options. More precisely, the payoff $(S_T - K)^+$ is approximated by

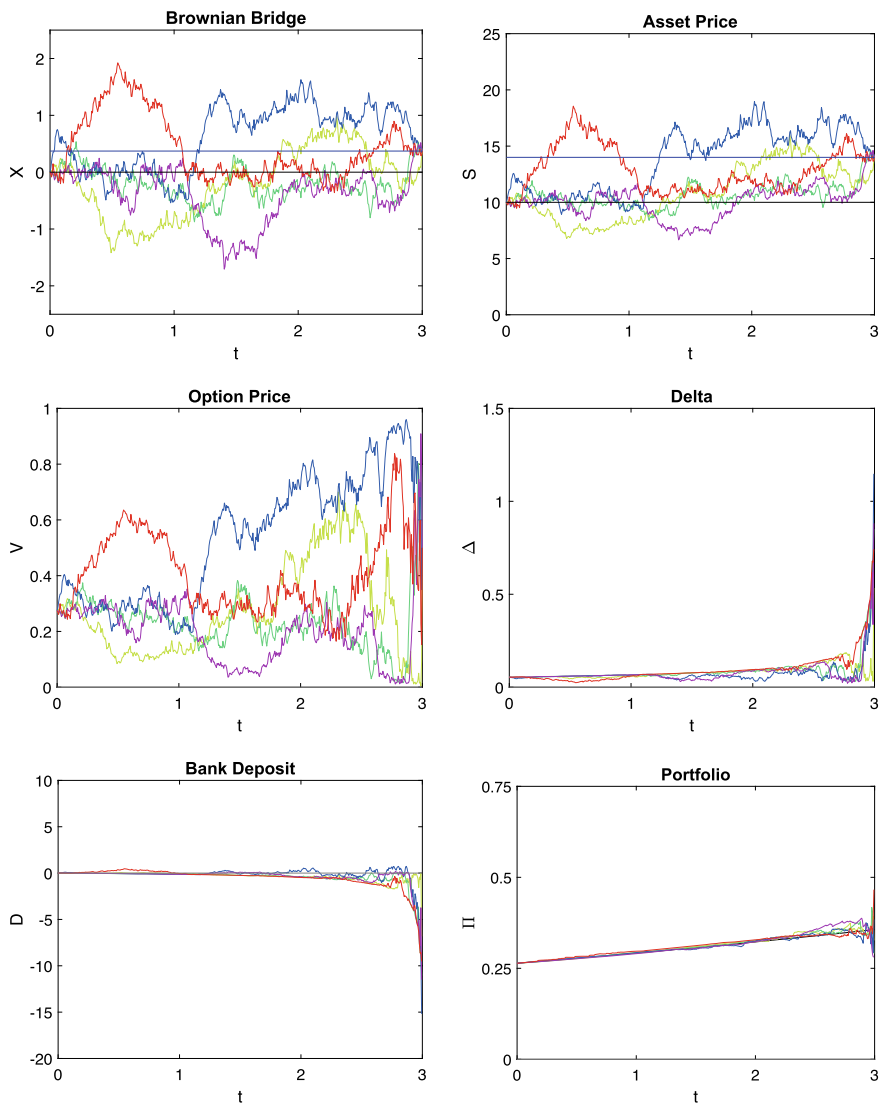
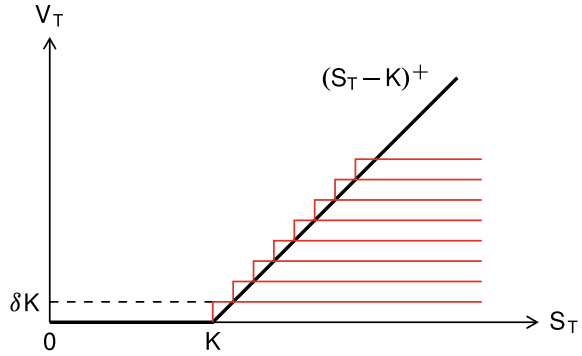


Fig. 17.10 Delta-hedging of the at-the-money digital call option using a self-financing portfolio $\Pi = -V + D + \Delta S$: The option price fluctuates between 0 and 1, and Delta can become large as $t \rightarrow T$ (from top left to bottom right)

$$\sum_{i=0}^{\infty} \delta K \times \mathbb{1}_{\{S_T \geq K + i \times \delta K\}}$$

if $\delta K > 0$ is sufficiently small. See Fig. 17.11.

Fig. 17.11 Difference of two binary options



Let $B(x)$ denote the time $t = 0$ price of a binary option with strike price x and payoff $\mathbb{1}_{\{S_T \geq x\}}$. Then the price of a European call option with strike price K , denoted by $C(K)$, is approximated by

$$\sum_{i=0}^{\infty} B(K + i \times \delta K) \delta K .$$

Hence

$$\begin{aligned} C(K) &= \lim_{\delta K \rightarrow 0} \sum_{i=0}^{\infty} B(K + i \times \delta K) \delta K \\ &= \int_K^{\infty} B(x) dx \\ &= \int_K^{\infty} e^{-rT} N(d_2) dx , \end{aligned}$$

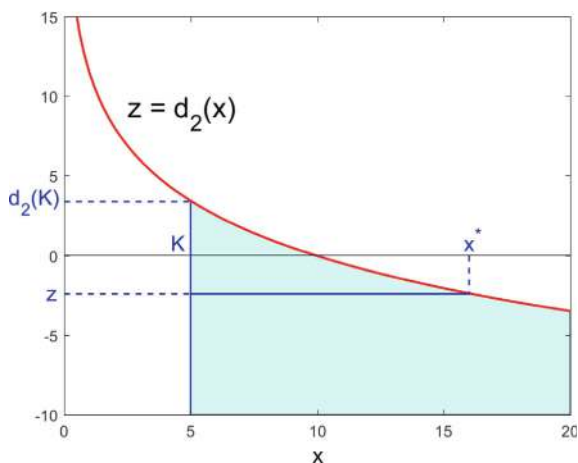
where

$$d_{1,2}(x) = \frac{-\log x + \log S_0 + (r \pm \frac{1}{2}\sigma^2)T}{\sigma \sqrt{T}} .$$

Note that

$$\begin{aligned} \int_K^{\infty} N(d_2) dx &= \int_K^{\infty} \int_{-\infty}^{d_2(x)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz dx \\ &= \int_{-\infty}^{d_2(K)} \int_K^{x^*} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dx dz \quad (x^* = S_0 e^{(r - \frac{1}{2}\sigma^2)T - \sigma \sqrt{T}z}) \end{aligned}$$

Fig. 17.12 $z = d_2(x)$ as a function of strike price x where $S = 10$, $r = 0.05$, $\sigma = 0.4$ and $T = 0.25$



$$\begin{aligned}
 &= \int_{-\infty}^{d_2(K)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} (x^* - K) dz \\
 &= \int_{-\infty}^{d_2(K)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \left(S_0 e^{(r - \frac{1}{2}\sigma^2)T - \sigma\sqrt{T}z} - K \right) dz \\
 &= S_0 e^{rT} \int_{-\infty}^{d_2(K)} \frac{1}{\sqrt{2\pi}} e^{-(z + \sigma\sqrt{T})^2/2} dz - K N(d_2(K)) \\
 &= S_0 e^{rT} N(d_2(K) + \sigma\sqrt{T}) - K N(d_2(K)) .
 \end{aligned}$$

For the domain of integration for the double integral see Fig. 17.12. Now we use the fact that $d_2(K) + \sigma\sqrt{T} = d_1(K)$.

17.3.3 The Binomial Tree Method for the Binary Option

We compare the prices obtained by the binomial tree method with the prices obtained by the Black–Scholes–Merton formula given by Theorem 17.2. In Fig. 17.13 the prices obtained by the binomial tree method appear to converge to the Black–Scholes–Merton prices plotted as horizontal lines. On the left we show the case $K = 90 < 100 = S_0$, and on the right we present the case $K = 110 > 100 = S_0$. In both cases the prices obtained by the binomial tree method oscillate as M increases. Consult Simulation 17.4.

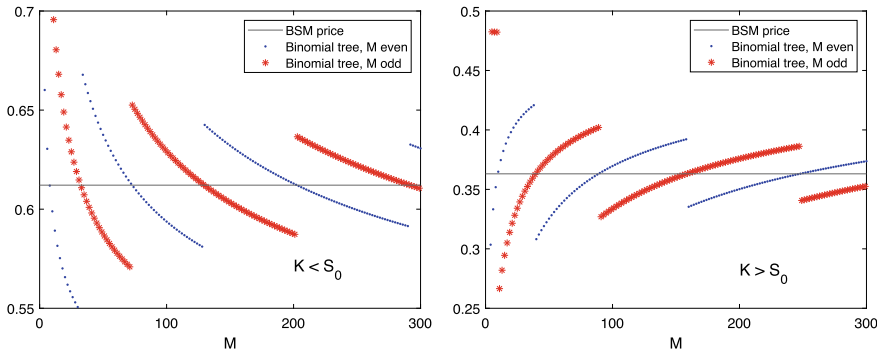


Fig. 17.13 Prices obtained by the binomial tree method and the Black–Scholes–Merton formula for binary call options where $K < S_0$ (left) and $K > S_0$ (right)

17.4 Computer Experiments

Simulation 17.1 (Cash-or-Nothing Call) The price surface defined by the Black–Scholes–Merton formula for a cash-or-nothing call option is given in Fig. 17.3. Also presented in Fig. 17.8 is the Delta for the option where the graph is unbounded at $t = T$, as proved in Theorem 17.3.

```

K = 5;
A = 1;
r = 0.05;
sigma = 0.3;

M= 70; % number of subintervals of the asset interval
N= 50; % number of subintervals of the time interval

mu = 0.1;

T = 1;
dt=T/N;

S0 = 4;
S_max = 10;
dS = S_max/M;

S_range = 0:dS:S_max;
t_range = 0:dt:T;

call = zeros(M+1,N+1);

```

```

Delta= zeros(M+1,N+1);

for i = 1:M+1
    S = S_range(i);
    for j = 1:N
        tau = T-t_range(j);
        d2 = (log(S/K)+(r-0.5*sigma^2)*tau)/(sigma*sqrt(tau));
        call(i,j) = A*exp(-r*tau)*normcdf(d2);
        Delta(i,j) = (A*exp(-r*tau)*normpdf(d2))/(sigma*S*sqrt(tau));
    end
    Delta(i,N+1) = (A*normpdf(d2))/(sigma*S*sqrt(tau));
    call(i,N+1) = A*heaviside(S-K); % payoff at T
end

Delta(1,:) = 0;
[t_grid,S_grid] = meshgrid(t_range,S_range);

figure(1)
surf(S_grid,t_grid,call);
hold off;

figure(2);
surf(S_grid,t_grid,delta);

```

Simulation 17.2 (Graph of the Delta of the At-the-Money Option) We plot the graph of the Delta of a cash-or-nothing call as a function of asset price S for t near the expiry date T . For the output see Fig. 17.9.

```

S0 = 4;
K = 5;
r = 0.05;
sigma = 0.3;
N = 5;
mu = 0.1;
T = 1;
T0 = T - T/10^4;
dt = (T-T0)/N;

t = T0:dt:T;
for i = 1:N
    S = 4.95:0.0001:5.05;
    tau = T - t(i);
    d2 = (log(S/K)+(r-0.5*sigma^2)*tau)/(sigma*sqrt(tau));

```

```

Delta = exp(-r*tau)*normpdf(d2)./(sigma*S*sqrt(tau));
plot(S,Delta,'Color',hsv2rgb([i/N 1 1]));
text(5.015,Delta(500),['t=',num2str(t(i))],'Color',hsv2rgb([i/N 1 1]));
hold on
end

figure(1);
xlabel('S');
set(gca,'xlim',[4.95,5.05]);

```

Simulation 17.3 (Brownian Bridge and At-the-Money Option) The Delta of an at-the-money cash-or-nothing call is computed, and the output is given in Fig. 17.10, where $\Delta \rightarrow \infty$ as $t \rightarrow T$. To generate a sample asset path S_t such that $S_0 \exp((\mu - \frac{1}{2}\sigma^2)T + \sigma W_T) = K$, we use a Brownian bridge path $X_t^{0 \rightarrow b}$ from 0 to b , where $b = \frac{1}{\sigma} \left\{ \log \frac{K}{S_0} - (\mu - \frac{1}{2}\sigma^2)T \right\}$. We take $\frac{1}{2}A$ as the payoff when $S_T = K$.

```

T = 3;
r = 0.1; % interest rate
mu = 0.12; % drift coefficient
sigma = 0.3; % volatility
S0 = 10; % asset price at time t=0
K = 14; % strike price

N = 500; % number of time steps
dt = T/N;
t = 0:dt:T;
M = 5; % number of sample paths
W = zeros(M,N + 1); % Brownian Motion
X = zeros(M,N+1); % Brownian Bridge
S = zeros(M,N+1); % Asset Price
S(:,1) = S0;
V = zeros(M,N+1); % option price
Delta = zeros(M,N+1); % Delta
D = zeros(M,N+1); % bank deposit
Pi = zeros(M,N+1); % portfolio
d2 = zeros(M,N+1);

D(:,1) = 0.0;
A = 1;

for i = 1:N
    dW = sqrt(dt)*randn(M,1);
    W(:,i+1) = W(:,i) + dW;
end

```

```

b = (log(K/S0) - (mu-sigma^2/2)*T)/sigma
for i = 1:N+1
    X(:,i) = b*(i-1)*dt/T + W(:,i) - W(:,N+1)*(i-1)*dt/T;
end

for i = 1:N+1
    S(:,i) = S0*exp((mu-(sigma^2)/2)*(i-1)*dt + sigma*X(:,i));
end

for i = 1:N
    tau = T-(i-1)*dt;
    d2(:,i) = (log(S(:,i)/K) + (r-0.5*sigma^2)*tau)/sigma/sqrt(tau);
    V(:,i) = A*exp(-r*tau)*normcdf(d2(:,i));
end
V(:,N+1) = A/2;

for i = 1:N+1
    tau = T-(i-1)*dt;
    d2(:,i) = (log(S(:,i)/K) + (r-0.5*sigma^2)*tau)/sigma/sqrt(tau);
    Delta(:,i) = A*exp(-r*tau)*normpdf(d2(:,i))./(S(:,i)*sigma*sqrt(tau));
end

for i = 1:N
    D(:,i+1) = exp(r*dt)*D(:,i) + (Delta(:,i)-Delta(:,i+1)).*S(:,i+1);
end

for i = 1:N+1
    Pi(:,i) = -V(:,i) + D(:,i) + Delta(:,i).*S(:,i);
end

figure(1)
for j = 1:M
    plot(t,X(j,1:N+1),'Color',hsv2rgb([j/M 1 1]));
hold on;
end
plot(t,zeros(1,N+1),'k-');
hold on
plot(t,b*ones(1,N+1),'b-');
hold on
xlabel('t');
ylabel('X');
title('Brownian Bridge');
hold off

figure(2)

```

```

for j = 1:M
    plot(t,S(j,:), 'Color',hsv2rgb([j/M 1 1]));
    hold on;
end
x = 0:dt:T;
plot(x,S0*ones(1,N+1), 'k-');
plot(x,K*ones(1,N+1), 'b-');
xlabel('t');
ylabel('S');
title('Asset Price');
hold off

figure(3)
for j = 1:M
    plot(t,V(j,:), 'Color',hsv2rgb([j/M 1 1]));
    hold on;
end

xlabel('t');
ylabel('V');
title('Option Price');
hold off;

figure(4)
for j = 1:M
    plot(t, Delta(j,:), 'Color',hsv2rgb([j/M 1 1]));
    hold on;
end

xlabel('t');
ylabel('\Delta');
title('Delta');
hold off;

figure(5);
for i = 2:N+1
    D(:,i) = exp(r*dt)*D(:,i-1) + (Delta(:,i-1)-Delta(:,i)).*S(:,i);
end
hold on

for i = 1:M
    plot(t, D(i,:), 'Color', hsv2rgb([i/M 1 1]));
    hold on;
end

```



```

title('Bank Deposit')
xlabel('t');
ylabel('D');
hold off

figure(6);
plot(t,Pi(1,1)*exp(r*t),'k-')
hold on
for j = 1:M
    plot(t, Pi(j,1:N+1), 'Color', hsv2rgb([j/M 1 1]));
    hold on;
end
xlabel('t','fontsize',14);
ylabel('\Pi');
title('Portfolio')

```

See Fig. 17.10.

Simulation 17.4 (Convergence of the Price by the Binomial Tree Method to the Black–Scholes–Merton Formula) . We check whether the price obtained by the binomial tree method converges to the price obtained by the Black–Scholes–Merton formula. The command `erf` is defined in Definition 4.2.

```

A = 1;
S0 = 100;
K = 90;
T = 1;
r = 0.05;
sigma = 0.3;
M1 = 4; % minimum for M, an even number
M2 = 300; % maximum for M
for M = M1:M2 % M = number of time steps
    dt = T/M;
    u = exp(sigma*sqrt(dt));
    d = exp(-sigma*sqrt(dt));
    q = (exp(r*dt)-d)/(u-d); % risk-neutral probability
    S = S0*u.^((M:-1:0)')*.d.^((0:1:M)');
    C = A*0.5*(1 + sign(S-K)); % option price at expiry
    % price at time 0.
    for i = M:-1:1
        C = exp(-r*dt)*(q*C(1:i) + (1-q)*C(2:i+1));
    end
    Digital_call(M) = C;
end
Digital_call(M2)

```

```
% European cash-or-nothing binary call option by BSM formula
d1 = (log(S0/K) + (r+0.5*sigma^2)*T)/sigma/sqrt(T);
d2 = d1 - sigma*sqrt(T);
BSM_price = A*exp(-r*T)*((1 + erf(d2/sqrt(2)))/2)
yline(BSM_price,'k');
hold on
plot(M1:2:M2,Digital_call(M1:2:M2),'b. ');
hold on
plot(M1+1:2:M2,Digital_call(M1+1:2:M2),'r* ');
legend('BSM price','Binomial tree, M even','Binomial tree, M odd');
xlabel('M');
text(200,0.57,'K < S_0','fontsize',14)
set(gca,'xlim',[0,M2]);
```

See Fig. 17.13.

Exercises

17.1 Show that the fair price for a forward contract on S_T at $t = 0$ is equal to $F = Ae^{rT}$.

17.2 What is the price of an asset-or-nothing binary call option?

Chapter 18

Pricing of Exotic Options



Options with nonstandard features are called exotic options. In this chapter we introduce exotic options such as Asian options and barrier options. For an encyclopedic collection of option pricing formulas consult (Haug, 2007).

18.1 Asian Options

The payoff of an Asian option with expiry date T is determined by a suitably defined average of underlying asset prices $S_1, \dots, S_n$ measured on predetermined dates $0 < t_1 < t_2 < \dots < t_n = T$. For example, we can choose any of the following definitions of average to define a payoff of an Asian option

$$A_1 = \text{arithmetic average} = \frac{S_1 + \dots + S_n}{n},$$

$$A_2 = \text{geometric average} = (S_1 \times \dots \times S_n)^{1/n}.$$

Or using weights w_i we may define weighted averages by

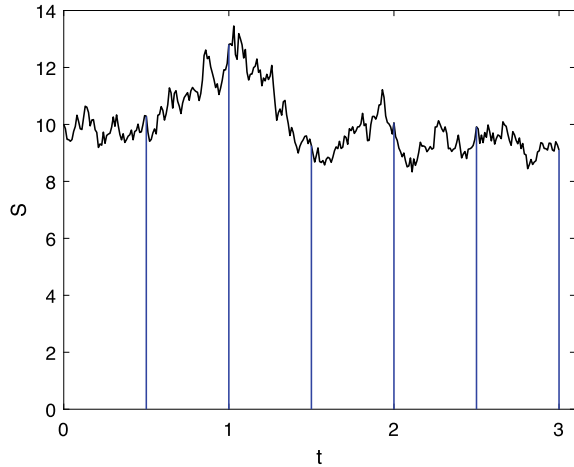
$$\text{weighted arithmetic average} = \frac{w_1 S_1 + \dots + w_n S_n}{w_1 + \dots + w_n},$$

$$\text{weighted geometric average} = (S_1^{w_1} \times \dots \times S_n^{w_n})^{1/(w_1 + \dots + w_n)}.$$

We can also define the corresponding continuous versions of averages. If we consider only equal weights in computing the average, there are four types of averages, as given in Table 18.1.

Table 18.1 Definitions of average for Asian options

Average	Discrete time	Continuous time
Arithmetic	$\frac{1}{n} \sum_{i=1}^n S_{t_i}$	$\frac{1}{T} \int_0^T S_t dt$
Geometric	$(S_{t_1} \cdots S_{t_n})^{1/n}$	$\exp\left(\frac{1}{T} \int_0^T \log S_t dt\right)$

Fig. 18.1 An asset price path for Asian option with six monitoring times

In Fig. 18.1 a sample asset price path is presented for an Asian option with $n = 6$ monitoring times for the period $0 \leq t \leq T = 3$.

For a call option with exercise price K the payoff is given by $(A - K)^+$ for some average A . Asian options are used to hedge the risk from fluctuations of metal prices, crude oil price or foreign currency rates. Due to lower volatility of an average compared to the original underlying asset price, Asian option prices are lower than those of vanilla options, in general.

Theorem 18.1 Suppose that the underlying asset price S_t follows a geometric Brownian motion, i.e., $S_t = S_0 \exp((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t)$ for some μ and σ . Then the price of an Asian call option with geometric average, with strike price K and expiry date T , is given by

$$S_0 e^{(\bar{\mu} - r)T} N(d_1) - K e^{-rT} N(d_2),$$

where

$$d_1, d_2 = \frac{\log \frac{S_0}{K} + (\bar{\mu} \pm \frac{1}{2}\bar{\sigma}^2)T}{\bar{\sigma}\sqrt{T}},$$

$$\bar{\sigma}^2 = \frac{1}{6}\sigma^2(1 + \frac{1}{n})(2 + \frac{1}{n}),$$

$$\bar{\mu} = \frac{1}{2}\bar{\sigma}^2 + \frac{1}{2}(r - \frac{1}{2}\sigma^2)(1 + \frac{1}{n}).$$

Proof We apply the risk-neutral method in Sect. 16.1 and take $\mu = r$. Put $\delta t = \frac{T}{n}$, then

$$A_2 = S_0 \exp \left(\frac{1}{n} \left\{ \left(r - \frac{1}{2} \sigma^2 \right) \sum_{i=1}^n i \delta t + \sigma \sum_{i=1}^n W_{i \delta t} \right\} \right).$$

Since

$$\begin{aligned} \sum_{i=1}^n W_{i \delta t} &= n (W_{\delta t} - W_0) \\ &\quad + (n-1) (W_{2\delta t} - W_{\delta t}) \\ &\quad + \quad \vdots \\ &\quad + 3 (W_{(n-2)\delta t} - W_{(n-3)\delta t}) \\ &\quad + 2 (W_{(n-1)\delta t} - W_{(n-2)\delta t}) \\ &\quad + (W_{n\delta t} - W_{(n-1)\delta t}) \end{aligned}$$

and since the increments $W_{i\delta t} - W_{(i-1)\delta t}$ are independent on the right-hand side, $\sum_{i=1}^n W_{i\delta t}$ is normally distributed with mean 0 and variance

$$\sum_{i=1}^n i^2 \delta t = \frac{1}{6} n(n+1)(2n+1) \delta t = \frac{1}{6} (n+1)(2n+1) T.$$

If we put

$$X = \frac{1}{n} \left\{ \left(r - \frac{1}{2} \sigma^2 \right) \sum_{i=1}^n i \delta t + \sigma \sum_{i=1}^n W_{i \delta t} \right\},$$

then $A_2 = S_0 e^X$. Note that X has normal distribution with mean $(\bar{\mu} - \frac{1}{2} \sigma^2)T$ and variance $\bar{\sigma}^2 T$, and hence $Y = X - (\bar{\mu} - r)T$ has normal distribution with mean $(r - \frac{1}{2} \sigma^2)T$ and variance $\bar{\sigma}^2 T$. Furthermore, we have

$$A_2 = S_0 e^{(\bar{\mu} - r)T} e^Y.$$

Now we find the price of an Asian call option where A_2 is used as the definition of average and the payoff is given by $\max\{0, A_2 - K\}$. Taking $S_0 e^{(\bar{\mu} - r)T}$ as the initial price in the Black–Scholes–Merton formula, we obtain the Asian call option price. $\square$

See Simulation 18.1 for a computational experiment.

Remark 18.1 (i) Note that for sufficiently large n , we have $\bar{\sigma}^2 \approx \frac{1}{3} \sigma^2$ and

$$\bar{\mu} - r \approx \frac{1}{2} \times \frac{1}{3} \sigma^2 + \frac{1}{2} \left(r - \frac{1}{2} \sigma^2 \right) - r = -\frac{1}{12} \sigma^2 - \frac{1}{2} r < 0,$$

which implies that the Asian call option price is cheaper since $\bar{\sigma} < \sigma$ and $S_0 e^{(\bar{\mu}-r)T} < S_0$.

(ii) In the above proof, we used the identity

$$\sum_{i=1}^n W_{i\delta t} = \sum_{k=0}^{n-1} (n-k) (W_{(k+1)\delta t} - W_{k\delta t}),$$

which is nothing but a discretized version of the Itô integral: After multiplying both sides by $\delta t = \frac{T}{n}$, we obtain

$$\sum_{i=1}^n W_{i\delta t} \times \frac{T}{n} = \sum_{k=0}^{n-1} (T - k\frac{T}{n}) (W_{(k+1)\delta t} - W_{k\delta t}),$$

whose continuous limit is given by

$$\int_0^T W_t dt = \int_0^T (T-t) dW_t.$$

The right-hand side is equal to

$$TW_T - \int_0^T t dW_t.$$

Hence we obtain the formula in Example 10.3 (iv).

18.2 Barrier Options

A barrier option has a barrier or barriers, and becomes effective only when the asset price stays within the barriers or goes outside the boundary depending on the type of the option during the lifetime of the option. Figure 18.2 illustrates a barrier option with an upper barrier. For a knock-out option, if the asset price crosses the barrier or barriers, the option loses its value immediately. For a knock-in option the barrier option has value only when the asset price reaches a certain level before or at expiry date. Barriers may be defined only on a part of the time to expiry date, and barriers of different heights can be set up.

For example, consider a call option with its exercise price equal to \$10 and knock-in barrier equal to \$11. The asset price at $t = 0$ was \$9, and the price at expiry date is \$10.5. If the asset price has never been equal to \$11, then the option at expiry date will be useless. However, the value of the option at expiry date is worth \$0.5 if the asset price has been at \$11 at least once. Or, we may consider a knock-out barrier option which becomes worthless if the asset price hits a given barrier at any time

Fig. 18.2 A knock-out barrier option with five sample asset paths

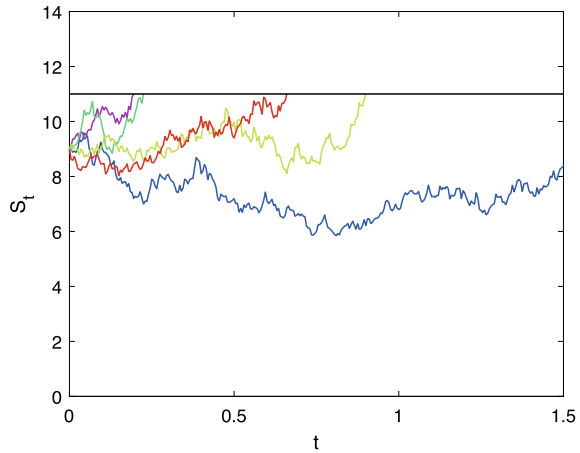
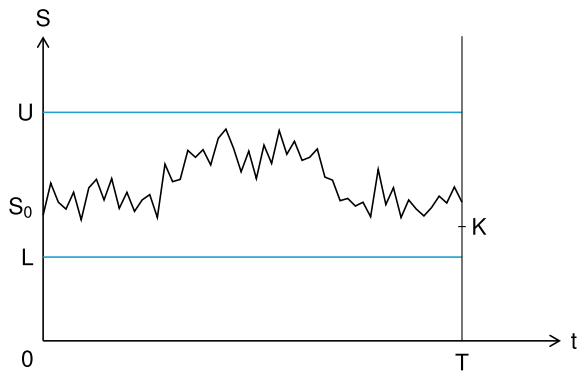


Fig. 18.3 A double barrier option



t , $0 \leq t \leq T$. See Fig. 18.2 for several scenarios of asset price movement. Consult Simulation 18.2.

A barrier option can have two barriers $L < U$. In Fig. 18.3 a double barrier option is shown with its lower and upper barriers L and U . In this case, knock-out barrier options can have many different kinds of knock-outs and knock-ins: up-and-out, down-and-out, up-and-in, down-and-in, and their various combinations.

Let $\tilde{W}_t$ be a $\tilde{\mathbb{P}}$ -Brownian motion. Let $\hat{W}_t = \tilde{W}_t + \theta t$, where

$$\theta = \frac{r - \frac{1}{2}\sigma^2}{\sigma} = \frac{r}{\sigma} - \frac{1}{2}\sigma.$$

Then

$$S_t = S_0 e^{\sigma \hat{W}_t + (r - \frac{1}{2}\sigma^2)t} = S_0 e^{\sigma \tilde{W}_t}.$$

Let us compute the price of an up-and-out call option where there is one barrier $B = B_1 > S_0$. Define

$$\widehat{M}_T = \max_{0 \leq t \leq T} \widehat{W}_t .$$

Then

$$\max_{0 \leq t \leq T} S_t = S_0 e^{\sigma \widehat{M}_T} .$$

The option knocks out if and only if $S_0 e^{\sigma \widehat{M}_T} > B$. Otherwise, the option pays off

$$(S_T - K)^+ = (S_0 e^{\sigma \widehat{W}_T} - K)^+ .$$

Hence the payoff at expiry date T is given by

$$\begin{aligned} V_T &= (S_0 e^{\sigma \widehat{W}_T} - K)^+ \mathbb{1}_{\{S_0 e^{\sigma \widehat{M}_T} \leq B\}} \\ &= (S_0 e^{\sigma \widehat{W}_T} - K) \mathbb{1}_{\{S_0 e^{\sigma \widehat{W}_T} \geq K, S_0 e^{\sigma \widehat{M}_T} \leq B\}} \\ &= (S_0 e^{\sigma \widehat{W}_T} - K) \mathbb{1}_{\{\widehat{W}_T \geq k, \widehat{M}_T \leq b\}} , \end{aligned}$$

where

$$k = \frac{1}{\sigma} \log \frac{K}{S_0} \quad \text{and} \quad b = \frac{1}{\sigma} \log \frac{B}{S_0} .$$

The option value at $t \leq T$ is given by

$$V_t = \mathbb{E}^{\widetilde{\mathbb{P}}} [e^{-r(T-t)} V_T | \mathcal{F}_t] ,$$

and hence

$$e^{-rt} V_t = \mathbb{E}^{\widetilde{\mathbb{P}}} [e^{-rT} V_T | \mathcal{F}_t]$$

is a martingale.

Let τ denote the first time at which the asset price reaches the barrier B , i.e.,

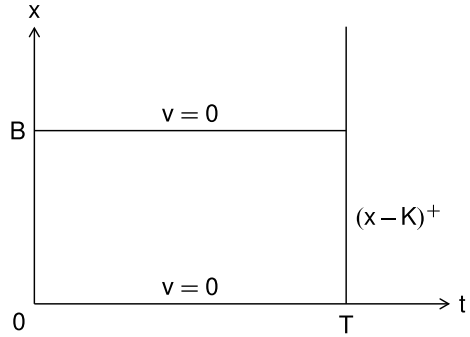
$$\tau = \begin{cases} \inf\{t \geq 0 : S_t = B\} & \text{if } S_t = B \text{ for some } 0 \leq t < \infty , \\ \infty , & \text{otherwise .} \end{cases}$$

(Recall that a sample path of S_t is continuous with probability 1.) Then τ is a stopping time. Define a stochastic process by

$$e^{-r(t \wedge \tau)} V_{t \wedge \tau} = \begin{cases} e^{-rt} V_t , & 0 \leq t \leq \tau , \\ e^{-r\tau} V_\tau , & \tau < t \leq T . \end{cases}$$

Since a stopped martingale is also a martingale, we see that $e^{-r(t \wedge \tau)} V_{t \wedge \tau}$ is a $\widetilde{\mathbb{P}}$ -martingale. (See Theorem 6.3.) More precisely, if we let $V_t = v(t, S_t)$, $0 \leq t \leq \tau$, then $e^{-r(t \wedge \tau)} v(t \wedge \tau, S_{t \wedge \tau})$ is a $\widetilde{\mathbb{P}}$ -martingale.

Fig. 18.4 Boundary and final conditions for an up-and-out call barrier option



Now we derive the Black–Scholes–Merton partial differential equation for an up-and-out call barrier option.

Theorem 18.2 (Up-and-Out Call) *Let $v(t, x)$ denote the option price at time t , $0 \leq t \leq T$, of an up-and-out call under the assumption that the call has not knocked out prior to time t and $S_t = x$. Then we have the Black–Scholes–Merton equation*

$$v_t + rxv_x + \frac{1}{2}\sigma^2 x^2 v_{xx} = rv, \quad (t, x) \in [0, T] \times [0, B)$$

with boundary and final conditions given by

$$\begin{cases} v(t, 0) = 0, & 0 \leq t \leq T, \\ v(t, B) = 0, & 0 \leq t \leq T, \\ v(T, x) = (x - K)^+, & 0 \leq x \leq B. \end{cases}$$

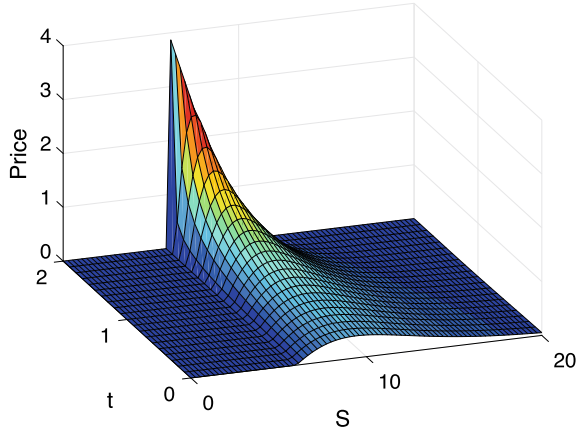
(See Fig. 18.4.) Its solution is given by

$$\begin{aligned} v(t, x) = & x \left\{ N\left(\delta_+\left(\tau, \frac{x}{K}\right)\right) - N\left(\delta_+\left(\tau, \frac{x}{B}\right)\right) \right\} \\ & - e^{-r\tau} K \left\{ N\left(\delta_-\left(\tau, \frac{x}{K}\right)\right) - N\left(\delta_-\left(\tau, \frac{x}{B}\right)\right) \right\} \\ & - B \left(\frac{x}{B}\right)^{-2r/\sigma^2} \left\{ N\left(\delta_+\left(\tau, \frac{B^2}{Kx}\right)\right) - N\left(\delta_+\left(\tau, \frac{B}{x}\right)\right) \right\} \\ & + e^{-r\tau} K \left(\frac{x}{B}\right)^{-2r/\sigma^2+1} \left\{ N\left(\delta_-\left(\tau, \frac{B^2}{Kx}\right)\right) - N\left(\delta_-\left(\tau, \frac{B}{x}\right)\right) \right\} \end{aligned}$$

where $\tau = T - t$,

$$\delta_{\pm}(\tau, u) = \frac{1}{\sigma\sqrt{\tau}} \left\{ \log u + \left(r \pm \frac{1}{2}\sigma^2\right)\tau \right\},$$

Fig. 18.5 Price of a down-and-out put option



and $N(\cdot)$ denotes the cumulative distribution function of the standard normal distribution.

Proof Note that

$$\begin{aligned} & d(e^{-rt}v(t, S_t)) \\ &= e^{-rt} \left\{ -rvdt + v_t dt + v_x dS + \frac{1}{2} v_{xx} (dS)^2 \right\} \\ &= e^{-rt} \left\{ -rv + v_t + rSv_x + \frac{1}{2} \sigma^2 S^2 v_{xx} \right\} dt + e^{-rt} \sigma S v_x d\tilde{W}. \end{aligned}$$

The drift term should vanish for $0 \leq t \leq \tau$. Since (t, S_t) is an arbitrary point in the domain $\{(t, x) : 0 \leq t \leq T, 0 \leq x \leq B\}$ before the up-and-out option strikes out, we obtain the Black–Scholes–Merton equation in the rectangle. $\square$

Now we find the formula for the price of a down-and-out put barrier option. See Fig. 18.5, where we observe that the option price is zero if the asset price falls below the given barrier.

Theorem 18.3 (Down-and-Out Put) *The price of a down-and-out put option P_{d-o} with strike price K , a barrier L , $L < K$ and expiry date T is given by*

$$\begin{aligned} P_{d-o} &= K e^{-rT} N(-d_2^K) - S_0 N(-d_1^K) \\ &\quad - K e^{-rT} N(-d_2^L) + S_0 N(-d_1^L) \\ &\quad - K e^{-rT} \left(\frac{L}{S_0} \right)^{\frac{2r}{\sigma^2}-1} (N(d_4) - N(d_6)) \\ &\quad + S_0 \left(\frac{L}{S_0} \right)^{\frac{2r}{\sigma^2}+1} (N(d_3) - N(d_5)), \end{aligned}$$

where

$$\begin{aligned} d_{1,2}^K &= \frac{\log \frac{S_0}{K} + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \\ d_{1,2}^L &= \frac{\log \frac{S_0}{L} + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \\ d_{3,4} &= \frac{\log \frac{L}{S_0} + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}, \\ d_{5,6} &= \frac{\log \frac{L^2}{S_0 K} + (r \pm \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}. \end{aligned}$$

Proof Note that

$$\begin{aligned} P_{d-o} &= e^{-rT} \mathbb{E} \left[(K - S_T) \mathbb{1}_{\{L < S_T < K\}} \mathbb{1}_{\{\min_{0 \leq t \leq T} S_t > L\}} \right] \\ &= e^{-rT} \mathbb{E} \left[(K - S_T) \mathbb{1}_{\{L < S_T < K\}} \right] \\ &\quad - e^{-rT} \mathbb{E} \left[(K - S_T) \mathbb{1}_{\{L < S_T < K\}} \mathbb{1}_{\{\min_{0 \leq t \leq T} S_t \leq L\}} \right] \\ &= e^{-rT} \mathbb{E} \left[(K - S_T) \mathbb{1}_{\{S_T < K\}} \right] \\ &\quad - e^{-rT} \mathbb{E} \left[(K - S_T) \mathbb{1}_{\{S_T < L\}} \right] \\ &\quad - e^{-rT} \mathbb{E} \left[K \mathbb{1}_{\{L < S_T < K\}} \mathbb{1}_{\{\min_{0 \leq t \leq T} S_t \leq L\}} \right] \\ &\quad + e^{-rT} \mathbb{E} \left[S_T \mathbb{1}_{\{L < S_T < K\}} \mathbb{1}_{\{\min_{0 \leq t \leq T} S_t \leq L\}} \right] \\ &= I_1 - I_2 - I_3 + I_4. \end{aligned}$$

Observe that I_1 is the price of a European put option, and hence

$$I_1 = K e^{-rT} N(-d_2^K) - S_0 N(-d_1^K).$$

Similarly, we see that

$$\begin{aligned} I_2 &= e^{-rT} \mathbb{E}[(K - L + L - S_T) \mathbb{1}_{\{S_T < L\}}] \\ &= e^{-rT} (K - L) \mathbb{P}(\{S_T < L\}) \\ &\quad + L e^{-rT} N(-d_2^L) \\ &\quad - S_0 N(-d_1^L). \end{aligned}$$

Let us compute $\mathbb{P}(\{S_T < L\})$. If W_t denotes the Brownian motion driving the price process, then we have

$$\begin{aligned}
\mathbb{P}(\{S_T < L\}) &= \mathbb{P}\left(\left\{W_1 < \frac{\log \frac{L}{S_0} - (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right\}\right) \\
&= \mathbb{P}\left(\left\{-W_1 > \frac{\log \frac{S_0}{L} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}}\right\}\right) \\
&= 1 - N(d_2^L) \\
&= N(-d_2^L)
\end{aligned}$$

since $-W_1$ is a standard normal variable. Hence

$$\begin{aligned}
I_2 &= e^{-rT}(K - L)N(-d_2^L) + Le^{-rT}N(-d_2^L) - S_0N(-d_1^L) \\
&= Ke^{-rT}N(-d_2^L) - S_0N(-d_1^L) .
\end{aligned}$$

Now put

$$\begin{aligned}
\lambda &= \frac{1}{\sigma} \log \frac{L}{S_0} , \\
\kappa &= \frac{1}{\sigma} \log \frac{K}{S_0} , \\
\theta &= \frac{r}{\sigma} - \frac{1}{2}\sigma .
\end{aligned}$$

Since

$$\overline{W}_t = -W_t$$

is also a Brownian motion, we define a Brownian motion with drift by

$$\widehat{W}_t = \overline{W}_t - \theta t$$

and define

$$\widehat{M}_T = \max_{0 \leq t \leq T} \widehat{W}_t .$$

Then $L < S_T < K$ if and only if $-\kappa < \widehat{W}_T < -\lambda$. Also note that the following three statements are equivalent:

$$\min_{0 \leq t \leq T} S_t \leq L ,$$

$$\min_{0 \leq t \leq T} (W_t + \theta t) \leq \lambda ,$$

and

$$\max_{0 \leq t \leq T} \widehat{W}_t \geq -\lambda .$$

Hence

$$\begin{aligned}
& \mathbb{E}[\mathbb{1}_{\{L < S_T < K\}} \mathbb{1}_{\{\min_{0 \leq t \leq T} S_t \leq L\}}] \\
&= \mathbb{P}\{-\kappa < \widehat{W}_T < -\lambda, \widehat{M}_T \geq -\lambda\} \\
&= \int_{-\kappa}^{-\lambda} \int_{-\lambda}^{\infty} \frac{2(2m-w)}{T\sqrt{2\pi T}} e^{-\theta w - \frac{1}{2}\theta^2 T - \frac{1}{2T}(2m-w)^2} dm dw \\
&= \int_{-\kappa}^{-\lambda} \frac{1}{\sqrt{2\pi T}} e^{-\theta w - \frac{1}{2}\theta^2 T - \frac{1}{2T}(-2\lambda-w)^2} dw \\
&= e^{2\theta\lambda} \int_{-\kappa}^{-\lambda} \frac{1}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(w+2\lambda+\theta T)^2} dw \\
&= e^{2\theta\lambda} \int_{z(-\kappa)}^{z(-\lambda)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\
&= e^{2\theta\lambda} \left\{ N\left(\frac{\lambda + \theta T}{\sqrt{T}}\right) - N\left(\frac{-\kappa + 2\lambda + \theta T}{\sqrt{T}}\right) \right\},
\end{aligned}$$

where we applied Theorem 8.1 in the second equality, and we used the identity

$$-\theta w - \frac{1}{2}\theta^2 T - \frac{1}{2T}(w+2\lambda)^2 = -\frac{1}{2T}(w+2\lambda+\theta T)^2 + 2\theta\lambda$$

in the fourth equality, and we took the substitution

$$z(w) = \frac{w + 2\lambda + \theta T}{\sqrt{T}}$$

in the fifth. Note that

$$e^{2\theta\lambda} = \exp\left(2\left(\frac{r}{\sigma} - \frac{1}{2}\sigma\right)\frac{1}{\sigma}\log\frac{L}{S_0}\right) = \left(\frac{L}{S_0}\right)^{\frac{2r}{\sigma^2}-1},$$

$$\frac{\lambda + \theta T}{\sqrt{T}} = \frac{\log\frac{L}{S_0} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = d_4,$$

and

$$\frac{-\kappa + 2\lambda + \theta T}{\sqrt{T}} = \frac{-\log\frac{K}{S_0} + 2\log\frac{L}{S_0} + (r - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = d_6.$$

Hence

$$I_3 = K e^{-rT} \left(\frac{L}{S_0}\right)^{\frac{2r}{\sigma^2}-1} \{N(d_4) - N(d_6)\}.$$

Finally, since

$$S_T = S_0 e^{\sigma \theta T - \sigma \bar{W}_T} = S_0 e^{-\sigma(-\theta T + \bar{W}_T)} = S_0 e^{-\sigma \hat{W}_T},$$

we have

$$\begin{aligned} & \mathbb{E}[S_T \mathbb{1}_{\{L < S_T < K\}} \mathbb{1}_{\{\min_{0 \leq t \leq T} S_t \leq L\}}] \\ &= \int_{-\kappa}^{-\lambda} \int_{-\lambda}^{\infty} S_0 e^{-\sigma w} \frac{2(2m - w)}{T \sqrt{2\pi T}} e^{-\theta w - \frac{1}{2}\theta^2 T - \frac{1}{2T}(2m - w)^2} dm dw \\ &= S_0 \int_{-\kappa}^{-\lambda} \frac{1}{\sqrt{2\pi T}} e^{-(\sigma + \theta)w - \frac{1}{2}\theta^2 T - \frac{1}{2T}(w + 2\lambda)^2} dw \\ &= S_0 \int_{-\kappa}^{-\lambda} \frac{1}{\sqrt{2\pi T}} e^{-\frac{1}{2T}(w + 2\lambda + (\sigma + \theta)T)^2 + 2\lambda(\sigma + \theta) + \frac{1}{2}\sigma^2 T + \sigma \theta T} dw \\ &= S_0 e^{2\lambda(\sigma + \theta) + \frac{1}{2}\sigma^2 T + \sigma \theta T} \int_{z(-\kappa)}^{z(-\lambda)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= S_0 e^{2\lambda(\sigma + \theta) + \frac{1}{2}\sigma^2 T + \sigma \theta T} \\ &\quad \times \left\{ N\left(\frac{\lambda + (\sigma + \theta)T}{\sqrt{T}}\right) + N\left(\frac{-\kappa + 2\lambda + (\sigma + \theta)T}{\sqrt{T}}\right) \right\}, \end{aligned}$$

where we used the substitution

$$z(w) = \frac{w + 2\lambda + (\sigma + \theta)T}{\sqrt{T}}$$

in the fourth equality. Note that

$$\begin{aligned} & 2\lambda(\sigma + \theta) + \frac{1}{2}\sigma^2 T + \sigma \theta T \\ &= 2\left(\frac{r}{\sigma^2} + \frac{1}{2}\right) \log \frac{L}{S_0} + rT \\ &= \log \left(\left(\frac{L}{S_0}\right)^{2r/\sigma^2 + 1} e^{rT} \right) \end{aligned}$$

and that

$$\begin{aligned} \frac{\lambda + (\sigma + \theta)T}{\sqrt{T}} &= \frac{\log \frac{L}{S_0} + (r + \frac{1}{2}\sigma^2)T}{\sigma \sqrt{T}} = d_3, \\ \frac{-\kappa + 2\lambda + (\sigma + \theta)T}{\sqrt{T}} &= \frac{\log \frac{L^2}{S_0 K} + (r + \frac{1}{2}\sigma^2)T}{\sigma \sqrt{T}} = d_5, \end{aligned}$$

$$I_4 = S_0 \left(\frac{L}{S_0} \right)^{2r/\sigma^2 + 1} (N(d_3) - N(d_5)) ,$$

which completes the proof. $\square$

18.3 Computer Experiments

Simulation 18.1 (Geometric Average Asian Option) Let A be the geometric average of asset prices at $t_j = j \times \frac{T}{L}$, $1 \leq j \leq L$. We estimate

$$e^{-rT} \mathbb{E}^{\mathbb{Q}}[\max\{A - K, 0\}]$$

and use it to check the geometric Asian option pricing formula given in Theorem 18.1.

```

S0 = 100;
K = 110;
r = 0.05;
sigma = 0.3;
T = 1;
L = 12; % number of observations
dt = T/L;
sigma_bar = sqrt(sigma^2*(L+1)*(2*L+1)/(6*L^2));
mu_bar = 1/2*sigma_bar^2 + (r-1/2*sigma^2)*(L+1) / (2*L);
d1 = (log(S0/K) + (mu_bar+1/2*sigma_bar^2)*T)/(sigma_bar*sqrt(T));
d2 = (log(S0/K) + (mu_bar-1/2*sigma_bar^2)*T)/(sigma_bar*sqrt(T));
V_geo_formula = S0*exp((mu_bar-r)*T)*normcdf(d1) - K*exp(-r*T)*normcdf(d2)

M = 10^5;
S = zeros(L,M);
S(1,:) = S0*exp((r-1/2*sigma^2)*dt + sigma*sqrt(dt)*randn(1,M));
for j = 2:L
    S(j,:) = S(j-1,:).*exp((r-1/2*sigma^2)*dt+sigma*sqrt(dt)*randn(1,M));
end

geo_aver = exp(mean(log(S(1:L,:))));

V_geo = exp(-r*T)*max(geo_aver-K,0);
price = mean(V_geo)
width = 1.96*std(V_geo)/sqrt(M)
% a half of the width of the 95% confidence interval

```

The following output shows that the pricing formula agrees with the value obtained by the martingale method.

```
V_geo_formula =    4.1915

price =    4.1879

width =    0.0563
```

Simulation 18.2 (Stopped Asset Price Process) We plot five sample paths of the asset price process, some of which are knocked out by an upper barrier before T . See Fig. 18.2.

```
M = 5; % number of sample paths
T = 1.5;
mu = 0.1;
sigma = 0.3;
a = 0;
b = 11;
S0 = 9;
N = 300;
dt = T/N;
t = 0:dt:T;
S = zeros(M,N+1);
S(:,1) = S0;
dW = sqrt(dt)*randn(M,N);
stop = (N+1)*ones(M,1);

for i = 1:M
    for j = 1:N
        S(i,j+1) = S(i,j) + mu*S(i,j)*dt + sigma*S(i,j)*dW(i,j);
        if ( S(i,j+1) >= b )
            S(i,j+1) = b;
            stop(i) = j+1;
            break;
        end
    end
end

for i = 1:M
    plot(0:dt:(stop(i)-1)*dt,S(i,1:stop(i)),'color',hsv2rgb([1-i/M 1 1]));
    hold on
end
```


Simulation 18.3 (Down-and-Out Put) The surface defined by the Black–Scholes–Merton formula for a down-and-out put is given in Fig. 18.5.

```

K = 10;
L = 6; % a lower barrier
r = 0.05;
sigma = 0.3;
M = 75; % number of points on the asset axis including endpoints
N = 40; % number of points on the time axis including endpoints
T = 2;
dt = T/N;
S0 = 10;
S_max = 20;
dS = S_max/M;
S_range = 0:dS:S_max; %Divide the interval into M subintervals.
t_range = 0:dt:T;      %Divide the interval into N subintervals.
Put_do = zeros(M+1,N+1);
index_barrier = ceil(L/dS)

for i=1:M+1
    S = S_range(i);
    for j=1:N
        tau = T - t_range(j);
        d1K = (log(S/K) + (r+0.5*sigma^2)*tau)/sigma/sqrt(tau);
        d2K = (log(S/K) + (r-0.5*sigma^2)*tau)/sigma/sqrt(tau);
        d1L = (log(S/L) + (r+0.5*sigma^2)*tau)/sigma/sqrt(tau);
        d2L = (log(S/L) + (r-0.5*sigma^2)*tau)/sigma/sqrt(tau);
        d3 = (log(L/S) + (r+0.5*sigma^2)*tau)/sigma/sqrt(tau);
        d4 = (log(L/S) + (r-0.5*sigma^2)*tau)/sigma/sqrt(tau);
        d5 = (log(L^2/S/K) + (r+0.5*sigma^2)*tau)/sigma/sqrt(tau);
        d6 = (log(L^2/S/K) + (r-0.5*sigma^2)*tau)/sigma/sqrt(tau);

        I1(i,j)= K*exp(-r*tau)*normcdf(-d2K) - S*normcdf(-d1K);
        I2(i,j)= K*exp(-r*tau)*normcdf(-d2L) - S*normcdf(-d1L);
        I3(i,j)= K*exp(-r*tau)*(L/S)^(2*r/sigma^2-1)*(normcdf(d4)-normcdf(d6));
        I4(i,j)= S*(L/S)^(2*r/sigma^2+1)*(normcdf(d3)-normcdf(d5));

        Put_do(i,j) = I1(i,j) - I2(i,j) -I3(i,j) + I4(i,j);
    end
end
for i=1:index_barrier
    Put_do(i,:) =0;
end

```

```
for i=1:M+1
    Put_do(i,N+1) = max(K-S_range(i),0)*heaviside(S_range(i) - L);
end

[t_grid,S_grid] = meshgrid(t_range,S_range);
surf(S_grid,t_grid,Put_do);
```

Chapter 19

American Options



An option that can be exercised early, i.e., before or on the expiry date, is called an *American* option while a *European* option can be exercised only on the expiry date. Such an early exercise feature makes the pricing of American options harder. In this chapter we show how to compute the price of an American put option, and find the early exercise boundary. The closed form formula of the price of a perpetual American put option is derived. For an introduction, consult (Pascucci and Runggaldier, 2012).

19.1 American Call Options

Since an American option can be exercised at any time, it must be at least as valuable as an otherwise identical European option. For American call options, however, the right to early exercise is worthless. In other words, the right to early exercise does not affect the price of an American call option.

Theorem 19.1 *The prices of an American call option and a European call option are equal if their expiry date and exercise price are equal.*

Proof The holder of an American call option would try to maximize his profit using various strategies. The first strategy is to early exercise the option at time $t < T$. If $S_t > K$, then the holder will gain profit $S_t - K$ at time t . This strategy is no better than the following second strategy: At the time t in the first strategy by short selling the holder receives S_t , and at expiry date T the holder has two choices: (i) either the holder exercises the option and buys a share of the stock at the exercise price K , or (ii) buys a share of the stock in the market at the price S_T . By the second strategy the holder of the option can buy a stock at expiry date paying at most K , and return the stock which was borrowed for short selling.

Since the value of K at time T is not greater than the value of K at time t , the second strategy is better than or as good as the first one. Note that the second strategy

Table 19.1 Viewpoint of the holder of an American call option

Time	0	t	T
Without short sell	Option	Cash $S_t - K$	$e^{r(T-t)}(S_t - K)$
With short sell (I)	Option	Cash S_t and (-1) share of stock	$e^{r(T-t)}S_t - K$
With short sell (II)	Option	Cash S_t and (-1) share of stock	$e^{r(T-t)}S_t - S_T$

can be used for a European option. Hence the prices of an American call option and a European call option are equal if the expiry date and exercise price are identical.

Consult Table 19.1.

See also Simulation 14.6.

Remark 19.1 Consider European call options C_1 and C_2 on a stock with the same exercise price and different expiry dates T_1 and T_2 , $T_1 < T_2$, respectively. Consider an American call option A_2 with expiry date T_2 . It has the same value as C_2 by Theorem 19.1. However, A_2 can be exercised any time before T_2 including at time T_1 . Hence it is worth at least as much as C_1 . Thus C_2 is valuable at least as much as C_1 . This implies that the European call option price is a monotonically increasing function of time to expiry.

19.2 American Put Options

Theorem 19.2 Let $V(S, t)$ denote the price of an American put option with expiry date T and exercise price K . Then for each $t \in [0, T]$ there exists a number $S_t^* \in (0, \infty)$ such that for $0 \leq S \leq S_t^*$ and $0 \leq t \leq T$, we have

$$V(S, t) = K - S \text{ and } \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} < rV ,$$

and for $S_t^* < S$ and $0 \leq t \leq T$,

$$V(S, t) > \max\{K - S, 0\} \text{ and } \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} = rV .$$

The boundary condition at $S = S_t^*$ is that the option price is continuously differentiable with respect to S , is continuous in t , and

$$V(S_t^*, t) = \max\{K - S_t^*, 0\} , \quad \frac{\partial V}{\partial S}(S_t^*, t) = -1 .$$

Proof As in Sect. 15.1 the increment of V is given by

$$\delta V = \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) \delta t + \frac{\partial V}{\partial S} \delta S .$$

Now from the viewpoint of an option seller we construct a self-financing and risk-free portfolio Π by

$$\Pi(S, t) = -V(S, t) + D(S, t) + \Delta(S, t)S . \quad (5)$$

In other words, Π consists of one option sold, and bank deposit D , and Δ units of underlying asset S . We do not know how much S will change as time passes by from t to $t + \delta t$. Thus we fix the hedge ratio Δ from t to $t + \delta t$, and suppose that S changes to $S + \delta S$ and Π changes to $\Pi + \delta \Pi$, respectively. The bank deposit D increases by the amount $\delta D = rD\delta t$. In summary, we have

$$\begin{aligned} \delta \Pi &= -\delta V + rD\delta t + \Delta \delta S \\ &= \left(-\frac{\partial V}{\partial t} - \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rD \right) \delta t + \left(-\frac{\partial V}{\partial S} + \Delta \right) \delta S . \end{aligned}$$

Take $\Delta = \frac{\partial V}{\partial S}$. Then the δS term disappears, and there is no risk in $\delta \Pi$ since there is no risky asset price movement. However, there remains the risk that the holder of an American option may early exercise the option and cause financial loss to the seller, and $\delta \Pi$ should be greater than the risk-free profit to compensate for the risk. Hence we have the inequality $\delta \Pi > r \Pi \delta t$, and hence

$$-\frac{\partial V}{\partial t} - \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rD \geq r \left(-V + D + \frac{\partial V}{\partial S} S \right) .$$

After canceling rD from both sides, we obtain the desired inequality. For the details of the proof, consult (Pascucci, 2011; Shreve, 2004b).

The *optimal exercise boundary* for an American put option is a curve given by $\{(t, S_t^*) : 0 \leq t \leq T\}$. See Fig. 19.1. In the domain given by $\{(t, S) : S > S_t^*\}$ it is better to wait until the asset price goes down further without exercising early. As the expiry date approaches, we have a smaller possibility that the asset price will move down further and the future payoff will increase. Hence we have to exercise the option at a higher price than before to expect the same level of payoff. Thus the function $t \mapsto S_t^*$ is monotonically increasing. (For a simulation based on the binomial tree method, check Simulation 19.3).

In the domain $\{(t, S) : S > S_t^*\}$ the inequality in Theorem 19.2 becomes an equality. The reason is that in the management of the portfolio Π , as in the case of a European option, we do not exercise the American put option before expiry date, and hence we have $d\Pi = r \Pi dt$. In this case, we give a boundary condition

Fig. 19.1 Optimal exercise boundary for an American put option

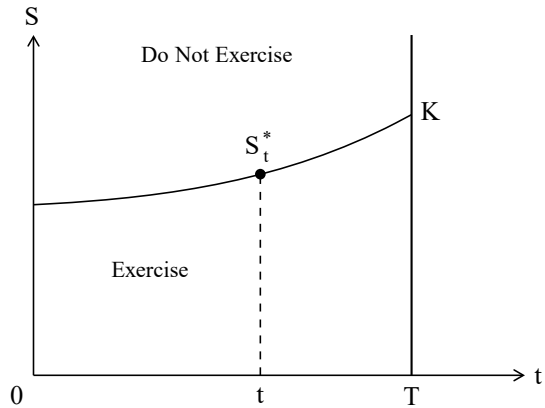
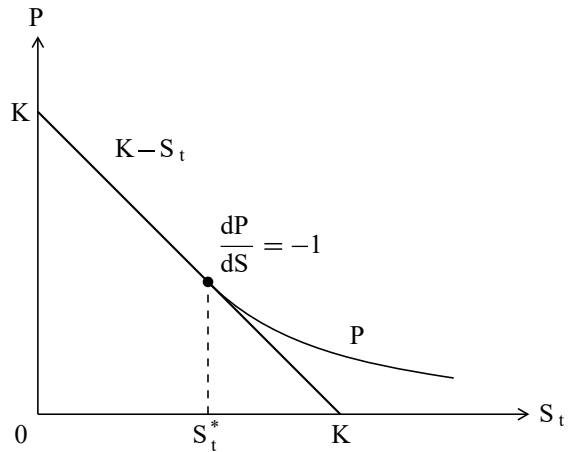


Fig. 19.2 The price P of an American put option when the asset price S_t takes the boundary price S_t^* at time t



$$\frac{\partial P}{\partial S} = -1$$

along the optimal exercise boundary, and solve the Black–Scholes–Merton partial differential equation. (See Fig. 19.2). For such a partial differential equation the boundary of the domain is not explicitly given, thus it is called a free boundary problem.

If $S_t < S_t^*$ at time t , i.e., it is better to early exercise the option, then the price $P(S_t, t)$ of the American put option at time t is given by a line segment $P = K - S_t$ with slope -1 as in Fig. 19.2. If $S_t \geq S_t^*$, then $P(S_t, t) > K - S_t$ as shown in the graph.

For $0 \leq t \leq T$ the price of an American put option at time t with $S_t = x$ is given by

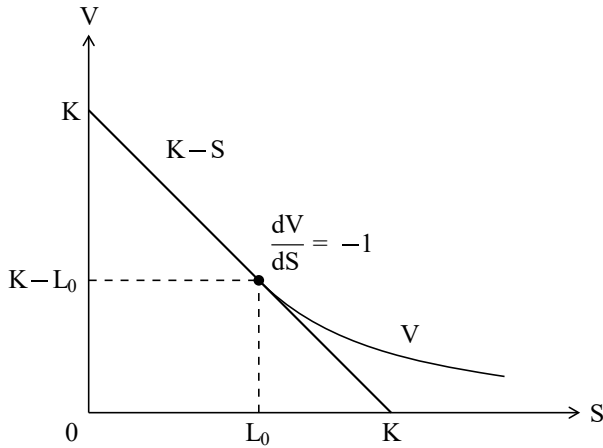


Fig. 19.3 The price V of the perpetual American put option when the asset price S takes the optimal boundary price L_0

$$P(x, t) = \sup_{t \leq \tau \leq T} \mathbb{E} \left[e^{-r(\tau-t)} (K - S_\tau)^+ \middle| S_t = x \right],$$

where τ denotes a stopping time. Continuous time models are theoretically complicated and the explicit pricing formula is not yet known. For more information, consult (Bingham and Kiesel, 2004; Kim, 1990). The binomial tree method is simple to implement in pricing American put options. At every step we compare the American option price with the price of the corresponding European option if the option is not exercised and the profit if it is exercised, and choose the more profitable action. See (Byun and Kim, 1994) for a modified algorithm.

We say that an option is *perpetual* if the time to expiry is infinite. In the following result, note that $L_0 < K$ and $V(S) > 0$ if $\sigma > 0$. (See Fig. 19.3).

Theorem 19.3 (Perpetual American put) *The price of a perpetual American put option is given by*

$$V(S) = \begin{cases} (K - L_0) \left(\frac{L_0}{S} \right)^{2r/\sigma^2}, & L_0 < S < \infty, \\ K - S, & 0 \leq S \leq L_0, \end{cases}$$

where

$$L_0 = \frac{2rK}{2r + \sigma^2}.$$

Proof Since the time to expiry is always infinite, the perpetual option price V does not depend on time t , and is a function of S alone, and the exercise boundary is of the form $S_t^* = L_0$ in Fig. 19.1 for some constant L_0 such that $L_0 < K$ for all $t > 0$. The Black–Scholes–Merton partial differential equation becomes the Cauchy–Euler equation

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0, \quad L < S < \infty.$$

(Consult Remark 15.2). We look for a general solution of the form

$$V(S) = c_1 S^{\lambda_1} + c_2 S^{\lambda_2}, \quad L < S < \infty,$$

for some constants $c_1, c_2, \lambda_1, \lambda_2$ where $\lambda_1 > \lambda_2$. (The boundary value L is not yet known, and will be determined.¹) Note that λ_1 and λ_2 are solutions of the quadratic equation

$$\frac{\sigma^2}{2}\lambda^2 + \left(r - \frac{\sigma^2}{2}\right)\lambda - r = (\lambda - 1)\left(\frac{\sigma^2}{2}\lambda + r\right) = 0.$$

Hence $\lambda_1 = 1$ and $\lambda_2 = -\frac{2r}{\sigma^2}$. From the boundary condition that

$$\lim_{S \rightarrow \infty} V(S) = 0,$$

we have $c_1 = 0$. And from the boundary condition that $V(L) = K - L$, we have

$$K - L = c_2 L^{-2r/\sigma^2},$$

which implies that

$$c_2 = (K - L)L^{2r/\sigma^2}.$$

Finally, from the condition that $\lim_{S \downarrow L} \frac{dV}{dS} = -1$, we have

$$(K - L)L^{2r/\sigma^2} \left(\frac{-2r}{\sigma^2}\right) L^{-2r/\sigma^2 - 1} = -1,$$

which implies that $L_0 = \frac{2rK}{2r + \sigma^2}$.

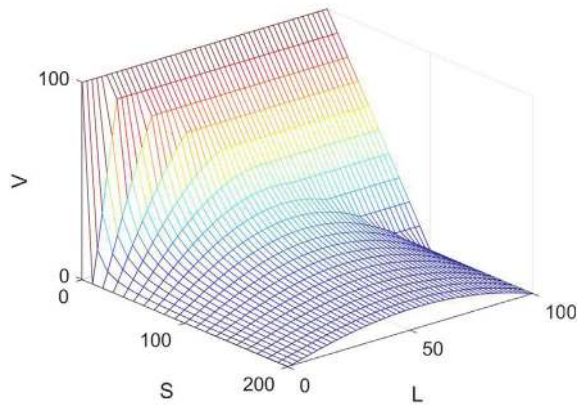
Remark 19.2 (Optimal Exercise Boundary) Since L_0 is optimal among all possible exercise boundary values, and V has maximum at $L = L_0$, we have

$$\frac{\partial V}{\partial L} = S^{-2r/\sigma^2} \frac{\partial}{\partial L} \left((K - L)L^{2r/\sigma^2} \right) = 0$$

and hence

¹ Such a problem is called a free boundary problem.

Fig. 19.4 The graph of the function $V(S, L)$ given in Theorem 19.3 where $K = 100, r = 0.05, \sigma = 0.3$



$$-L^{2r/\sigma^2} + (K - L)\frac{2r}{\sigma^2}L^{2r/\sigma^2-1} = \frac{2r}{\sigma^2}L^{2r/\sigma^2-1} \left(-\frac{\sigma^2}{2r}L + (K - L) \right) = 0$$

at $L_0 = \frac{K}{1 + \sigma^2/(2r)}$ independently of t .

For the three-dimensional graph of $V(S, L)$, see Fig. 19.4 and Simulation 19.4. And for the graph of $V(S, L)$ as a function of S , see Fig. 19.5, where dotted lines indicate the cases when $S < L_0$ and the graphs when $S \geq L_0$ have the maximum at $L = L_0$. In Fig. 19.6 are shown the graph of $V(S, L)$ for various values of L .² Note that as L increases to $L_0 = 52.6316 \dots$ the graphs of $V(S, L)$ (in solid lines) monotonically increase to the maximal function corresponding to $L = L_0$. After L exceeds L_0 and keeps on increasing, the graphs (in dotted lines) monotonically decrease, confirming the fact that L_0 is the optimal exercise price which maximizes the value of the option. The slope at $S = L_0$ of the option price graph corresponding to $L = L_0$ is equal to -1 .

19.3 Computer Experiments

Simulation 19.1 (The Binomial Trees for the Asset and the American Put Price)

We consider an American put option with $S_0 = 110$, exercise price $K = 140$, expiry date $T = 1$, $r = 0.05$, $\sigma = 0.3$ and $N = 5$ time periods, and plot the binomial trees for the asset and the put option price. In Fig. 19.7 for the asset, the exercise price is plotted as a horizontal line. In Fig. 19.8 we calculate the American put prices and print them at the nodes defined by the asset prices. The price at time 0 is given

² Figures 19.5 and 19.6 are obtained by taking cross sections of the graph in Fig. 19.4.

Fig. 19.5 The graph of the function $V(S, L)$, $0 \leq L \leq K$, given in Theorem 19.3, for various values of S where $K = 100$, $r = 0.05$, $\sigma = 0.3$ and $L_0 = 52.63$

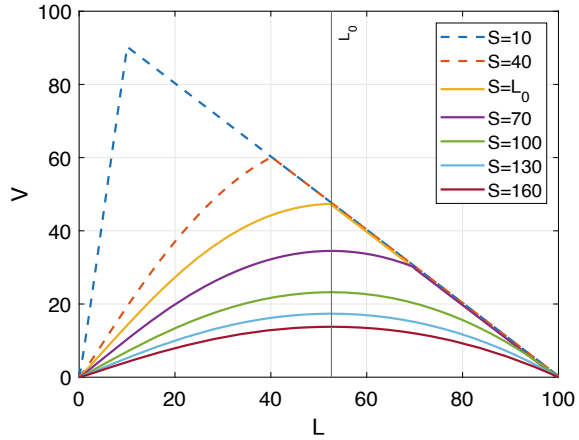
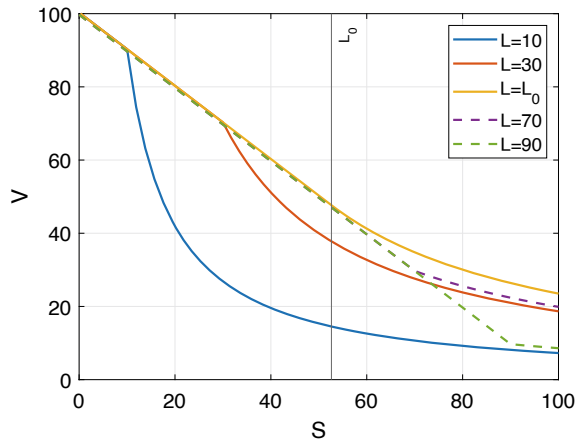


Fig. 19.6 The graph of the function $V(S, L)$, $0 \leq S \leq 200$, given in Theorem 19.3, for various values of L where $K = 100$, $r = 0.05$, $\sigma = 0.3$ and $L_0 = 52.63$



by 31.58. Asset prices are marked by red circles, and the option prices by the blue squares. Compare the result with Simulation 14.2.

```
T = 1; % expiry date
N = 5; % number of time periods
S0 = 110; % asset price at time 0
S = ones(N+1,N+1); % asset price
S(1,:) = S0;
P = zeros(N+1,N+1); % American put price
K = 140; % strike Price
r = 0.05; % interest rate
sigma = 0.3; % volatility
dt = T/N;
```

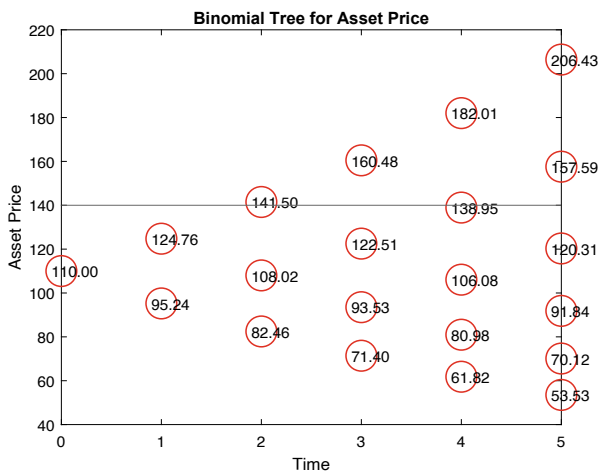


Fig. 19.7 The binomial tree for the asset price with $S_0 = 110$ and exercise price $K = 140$

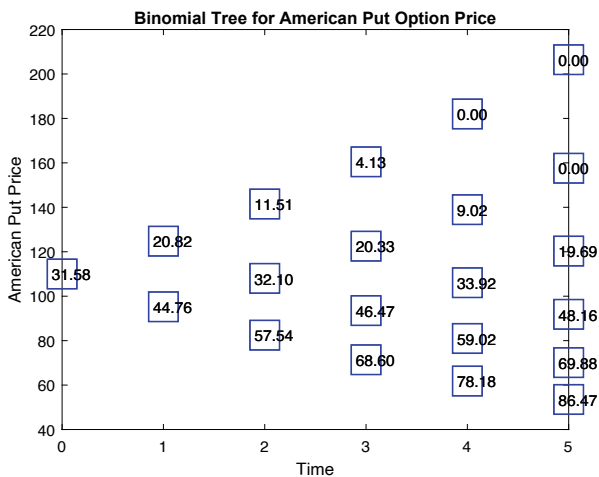


Fig. 19.8 The binomial tree for the American put option price with $S_0 = 110$ and exercise price $K = 140$

```
% risk-neutral probability
u = exp(sigma*sqrt(dt));
d = exp(-sigma*sqrt(dt));
q = (exp(r*dt)-d)/(u-d);

% i for time, j for asset price level
for i = 2:N+1
```

```

    for j = 1:i
        S(i,j) = S0*u^(j-1)*d^(i-j); % asset price
    end
end

for j = 1:N+1
    P(N+1,j) = max(K-S(N+1,j),0); % American put payoff at T
end

for i = N:-1:1 % backward in time
    for j = 1:i
        P(i,j) = max(exp(-r*dt)*(q*P(i+1,j+1)+(1-q)*P(i+1,j)), K-S(i,j));
    end
end

figure(1)
for i = 1:N+1
    plot(i-1,S(i,1:i),'or')
    for j = 1:i
        text(i-1-0.1,S(i,j),num2str(S(i,j),'%.2f'))
    end
    hold on
end

yline(K,'Color','k')
xlabel('Time')
ylabel('Asset')
title('Binomial Tree for Asset Price');

figure(2)
for i = 1:N+1
    plot(i-1,S(i,1:i),'sb')
    for j = 1:i
        text(i-1-0.1,S(i,j),num2str(P(i,j),'%.2f'))
    end
    hold on
end

xlabel('Time')
ylabel('European Put')
title('Binomial Tree for American Put Price');

American_put_price = P(1,1)

```

Simulation 19.2 (American Put Option: the Binomial Tree Method) *We use the binomial tree method to price an American put option. Observe that the price of an American put option is higher than that of a European put when we take the same set of parameter values for both.*

```

S0 = 100;
K = 110;
T = 1;
r = 0.05;
sigma = 0.3;
M = 2000; % number of time steps
dt = T/M;
u = exp(sigma*sqrt(dt));
d = exp(-sigma*sqrt(dt));
q = (exp(r*dt)-d)/(u-d);

% payoff at T
Put_Am = max(K - S0*u.^([M:-1:0]'),.*d.^([0:1:M]'),0);
Put_Eu = max(K - S0*u.^([M:-1:0]'),.*d.^([0:1:M]'),0);

for i = M:-1:1
    S = S0*u.^([i-1:-1:0]').*d.^([0:1:i-1]');
    Put_Am = max(max(K-S,0),exp(-r*dt)*(q*Put_Am(1:i)+(1-q)*Put_Am(2:i+1)));
end

for i = M:-1:1
    S = S0*u.^([i-1:-1:0]').*d.^([0:1:i-1]');
    Put_Eu = exp(-r*dt)*(q*Put_Eu(1:i) + (1-q)*Put_Eu(2:i+1));
end

Put_Am
Put_Eu

% Now we compute the Black-Scholes-Merton price.
d1 = (log(S0/K)+(r+0.5*sigma^2)*T)/(sigma*sqrt(T));
d2 = d1 - sigma*sqrt(T);
Put_Eu_BSM = K*exp(-r*T)*normcdf(-d2) - S0*normcdf(-d1)

```

The output is

```

Put_Am = 15.6177
Put_Eu = 14.6548
Put_Eu_BSM = 14.6553

```

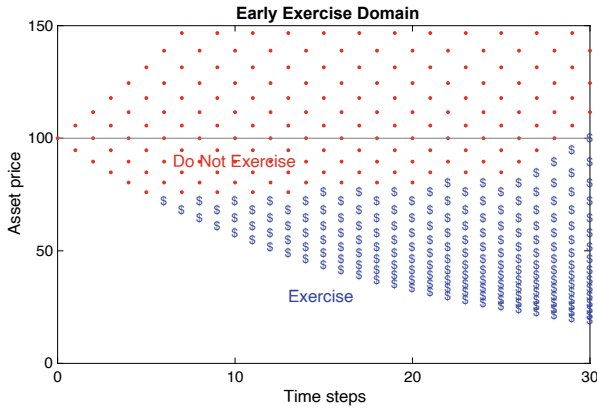


Fig. 19.9 The early exercise domain for an American put option marked by '\$' (in blue) and the no-exercise domain marked by '.' (in red)

Simulation 19.3 (American Put Option: the Early Exercise Domain) *We use the binomial tree method to plot the early exercise domain for an American put option. In Fig. 19.9 the nodes where it is optimal to exercise are marked by the symbol '\$' in blue.*

```
T = 1; % expiry date
N = 30; % number of time periods
S0 = 100; % asset price at time 0
S = ones(N+1,N+1); % asset price
S(1,:) = S0;
P = zeros(N+1,N+1); % American put price
K = 100; % strike Price
r = 0.05; % interest rate
sigma = 0.3; % volatility
dt = T/N;
u = exp(sigma*sqrt(dt)); d = exp(-sigma*sqrt(dt)); q = (exp(r*dt)-d)/(u-d);

% i for time, j for asset price level
for i = 2:N+1
    for j = 1:i
        S(i,j) = S0*u^(j-1)*d^(i-j); % asset price
    end
end

for j = 1:N+1
    P(N+1,j) = max(K-S(N+1,j),0); % American put payoff at T
end
```

```

for i = N:-1:1 % backward in time
for j = 1:i
P(i,j) = max(exp(-r*dt)*(q*P(i+1,j+1)+(1-q)*P(i+1,j)), K-S(i,j));
end
end

% Early Exercise Domain
clf
figure(1)
for j = 1:N+1 % from bottom to top of asset price at time T
    if S(N+1,j) <= K
        text(N-0.1,S(N+1,j),'$', 'Color','blue')
    else
        plot(N,S(N+1,j),'r.')
    end
end
hold on
end
hold on

for i = 1:N
for j = 1:i % from bottom to top of asset price at time i-1
    if exp(-r*dt)*(q*P(i+1,j+1)+(1-q)*P(i+1,j)) < K-S(i,j)
        text(i-1.1,S(i,j),'$', 'Color','blue');
    else
        plot(N,S(N+1,j),'r.')
    end
end
hold on
end
hold on

for i = 1:N
for j = 1:i % from bottom to top of asset price at time i-1
    if exp(-r*dt)*(q*P(i+1,j+1)+(1-q)*P(i+1,j)) < K-S(i,j)
        text(i-1.1,S(i,j),'$', 'Color','blue');
    else
        plot(i-1,S(i,j),'r. ');
    end
end
hold on
end
hold on
end

hold on

yline(K,'LineWidth',1);

```

```

text(6.5,90,'Do Not Exercise','Color','red')
text(13,30,'Exercise','Color','blue')
xlabel('Time steps')
ylabel('Asset price')
box on
set(gca,'xlim',[0 N],'ylim',[0 150]);
title('Early Exercise Domain');

```

Simulation 19.4 (Perpetual American Put Option)

We plot the graph of the function $V(S, L)$ given in Theorem 19.3, where $K = 100$, $r = 0.05$, $\sigma = 0.3$.

```

r = 0.05; % interest rate
sigma = 0.3;
K = 100; % strike price
Smin = 0;
Smax = 200;
M = 21; % number of points for S
N = 51; % number of points for L

S_range = linspace(Smin,Smax,M);
L_range = linspace(0,K,N);
V = zeros(M,N);
for i = 1:M
    S = S_range(i);
    for j = 1:N
        L = L_range(j);
        if S <= L
            V(i,j) = K - S;
        else
            V(i,j) = (K - L)*(L/S)^(2*r/sigma^2);
        end
    end
end

[X,Y] = meshgrid(L_range,S_range);

figure(1)
mesh(X,Y,V)
set (gca, 'xdir', 'reverse')

```

For the output, see Fig. 19.4.

Exercises

19.1 (Continuous Dividend) Consider an American option on an underlying stock with continuous dividend rate q . Define a partial differential operator $\mathcal{L}$ by

$$\mathcal{L}V = \frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + (r - q)S \frac{\partial V}{\partial S} - rV .$$

Show that the price of the American put option satisfies

$$\min\{-\mathcal{L}V, V - h(S)\} = 0$$

and

$$V|_{t=T} = h(S_T) ,$$

where $h(x) = \max\{K - x, 0\}$.

19.2 (Put-Call Symmetry) Consider the American call and put options on the same underlying stock with continuous dividend rate q and the same strike price K and the same expiry date T . Let r denote the risk-free interest rate.

(i) Let $C(S, t, r, q)$ and $P(S, t, r, q)$ be the prices of the American call and put options, respectively. Show that

$$C(S, t; r, q) = \frac{S}{K} P\left(\frac{K^2}{S}, t; q, r\right) .$$

(ii) Let $S_f^C(t; r, q)$ and $S_f^P(t; r, q)$ denote the optimal exercise boundaries of the American call and put options, respectively. Show that

$$S_f^C(t; r, q) S_f^P(t; q, r) = K^2 .$$

(Hint: Consult (Jiang, 2005)).

19.3 (American Binary Call Option) Let $V(S, t)$ be the price of the American binary call option with its payoff at expiry date T given by $V(S_T, T) = 1$ for $S_T \geq K$ and $V(S_T, T) = 0$ for $S_T < K$. The option contract can be exercised to receive \$1 anytime the asset price exceeds the strike price before or at T .

(i) Find the value $V(K, t)$, $0 \leq t \leq T$.

(ii) Show that $V(S, t)$ satisfies the partial differential equation given by

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} = rV , \quad 0 < S < K .$$

(iii) Show that the terminal condition is given by $V(S_T, T) = 0$ for $0 \leq S_T < K$.

(iv) Prove that the solution is given by

$$V(S, t) = \left(\frac{K}{S}\right)^{2r/\sigma^2} N(d_+) + \left(\frac{S}{K}\right) N(d_-),$$

where

$$d_+, d_- = \frac{\log \frac{S}{K} \pm (r - \frac{1}{2}\sigma^2)(T - t)}{\sigma \sqrt{T - t}}.$$

19.4 (Put-Call Parity for American Binary Options) Assume that the risk-free interest rate r is positive.

- (i) Find the price of an American option that pays \$1 at expiry date T .
- (ii) Find the put-call parity for American binary call and put options.

Part VII

Portfolio Management

Chapter 20

The Capital Asset Pricing Model



How can we measure the performance of mutual funds and their investment risk? What is the use of a market index such as S&P 500? The portfolio theory can provide us with the answers. In this chapter we present the Capital Asset Pricing Model (CAPM), which deals with an efficient portfolio management. We consider a single-period model for asset price movements. For a historical introduction see (Bernstein, 1992).

To maximize the return of a portfolio consisting of more than one asset under a given risk level, we find the optimal weight $(w_1, \dots, w_n)$, where w_i is the weight for the i th underlying asset. In this chapter the covariance matrix is denoted by C , and all vectors are column vectors unless otherwise stated. The transpose of a matrix M , including row vectors and column vectors, is denoted by M^t . For example, if $\mathbf{v} = (v_1, \dots, v_n)$ is a row vector then $\mathbf{v}^t$ is a column vector. Throughout the chapter, except when explicitly mentioned, we assume that all the assets under consideration are risky. For an introduction to matrix computation consult Appendix C.

20.1 Return Rate and the Covariance Matrix

Let S_t denote the price of an asset S at time points $t = 0$ and $t = T$. If S_T is constant, then S is said to be risk-free. Otherwise, it is called a risky asset (Fig. 20.1).

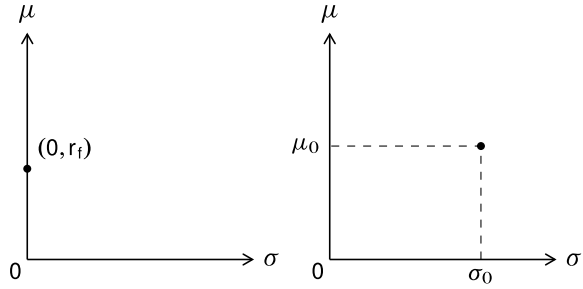
Definition 20.1 (*Return*) The return R_S of an asset S from time 0 to $T > 0$ is defined by

$$R_S = \frac{S_T - S_0}{S_0}.$$

Given several assets $S^1, \dots, S^n$, we consider a portfolio

$$V = x_1 S^1 + \dots + x_n S^n,$$

Fig. 20.1 A risk-free asset (left) and a risky asset (right)



where x_j is a real number. The case when $x_j < 0$ means the short-selling x_j units of the asset S^j .

Let R_j denote the return of asset S^j from time 0 to time $T > 0$. Then its *expected return* and *variance* are given by

$$\mu_j = \mathbb{E}[R_j] \quad \text{and} \quad \sigma_j^2 = \mathbb{E}[(R_j - \mu_j)^2],$$

respectively. By definition, the variance of the return of a *risk-free* asset is zero, while a *risky* asset has positive variance. Unless stated otherwise, all assets in this chapter are assumed to be risky.

Theorem 20.1 (Return of a Portfolio) *The return R_V of a portfolio $V = x_1 S^1 + \cdots + x_n S^n$ from time 0 to time $T > 0$ is given by*

$$R_V = w_1 R_1 + \cdots + w_n R_n,$$

where the weight w_j satisfies

$$w_j = \frac{x_j S_0^j}{V_0}$$

and $w_1 + \cdots + w_n = 1$.

Proof The values of V at time 0 and time T are $V_0 = x_1 S_0^1 + \cdots + x_n S_0^n$ and $V_T = x_1 S_T^1 + \cdots + x_n S_T^n$, respectively, and hence

$$\begin{aligned} R_V &= \frac{V_T - V_0}{V_0} \\ &= \frac{x_1 (S_T^1 - S_0^1) + \cdots + x_n (S_T^n - S_0^n)}{V_0} \\ &= \frac{x_1 S_0^1}{V_0} \frac{S_T^1 - S_0^1}{S_0^1} + \cdots + \frac{x_n S_0^n}{V_0} \frac{S_T^n - S_0^n}{S_0^n}, \end{aligned}$$

which proves the statement. □

Definition 20.2 (*Attainable Portfolio*) The portfolio described in Theorem 20.1 is called an attainable portfolio.

Definition 20.3 (*Covariance*) Define the covariance between the returns R_i and R_j by

$$C_{ij} = \text{Cov}(R_i, R_j) = \mathbb{E}[(R_i - \mu_i)(R_j - \mu_j)]$$

and the covariance matrix by $C = [C_{ij}]$. Note that $C_{ii} = \text{Var}(R_i) = \sigma_i^2$.

Remark 20.1 The covariance between a risky asset and a risk-free asset is 0. Thus if we consider a risk-free asset in defining the covariance matrix then a column and a row of a covariance matrix are zero vectors. In this chapter we assume that every asset is risky and the covariance matrix is invertible.

We define the expectation of a vector composed of random variables $\{X_i\}_{i=1}^n$ or the expectation of a matrix composed of random variables $\{Y_{ij}\}_{i,j=1}^n$ as the vector or the matrix given by the expectations of each component, i.e., $\mathbb{E}[(X_1, \dots, X_n)] = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$ and $\mathbb{E}[[Y_{ij}]_{ij}] = [\mathbb{E}[Y_{ij}]]_{ij}$. Using μ_i , $\boldsymbol{\mu}$ and $\boldsymbol{\mu}$ respectively, we denote the expected return of the i th risky asset, the vector defined by n expected returns, and the expected return of a portfolio consisting of n risky assets. If we let $\mathbf{R} = (R_1, \dots, R_n)^t$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^t$, then

$$C = \mathbb{E}[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^t].$$

Thus C is a product of an $n \times 1$ matrix and a $1 \times n$ matrix, and hence C is an $n \times n$ matrix, and its (i, j) th component is $\mathbb{E}[(R_i - \mu_i)(R_j - \mu_j)]$.

Lemma 20.1 (Positive Semidefiniteness) *The covariance matrix C is symmetric and positive semidefinite. Hence its eigenvalues are nonnegative.*

Proof For $\mathbf{v} = (v_1, \dots, v_n)^t$ we have

$$\mathbf{v}^t C \mathbf{v} = \mathbf{v}^t \mathbb{E}[(\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^t] \mathbf{v} = \mathbb{E}[\mathbf{v}^t (\mathbf{R} - \boldsymbol{\mu})(\mathbf{R} - \boldsymbol{\mu})^t \mathbf{v}] = \mathbb{E}[u^2] \geq 0,$$

where $u = \mathbf{v}^t (\mathbf{R} - \boldsymbol{\mu})$. □

Theorem 20.2 (Return of a Portfolio) *Consider the return of a portfolio V given by*

$$R_V = w_1 R_1 + \dots + w_n R_n = \mathbf{w}^t \mathbf{R},$$

where $\mathbf{w} = (w_1, \dots, w_n)^t$. Its expected return μ_V and variance σ_V^2 are given by

$$\mu_V = w_1 \mu_1 + \dots + w_n \mu_n = \mathbf{w}^t \boldsymbol{\mu},$$

$$\sigma_V^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j c_{ij} = \mathbf{w}^t C \mathbf{w}.$$

Proof Use the fact that

$$\mu_V = \mathbb{E} \left[\sum_{i=1}^n w_i R_i \right] = \sum_{i=1}^n w_i \mu_i$$

and

$$\sigma_V^2 = \text{Cov} \left(\sum_{i=1}^n w_i R_i, \sum_{i=1}^n w_i R_i \right) = \sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij}.$$

□

20.2 Portfolios of Two Assets and More

Consider a portfolio consisting of two risky assets with their expected returns μ_1, μ_2 and their variances σ_1^2, σ_2^2 , respectively.

Definition 20.4 (*Correlation*) The correlation coefficient ρ_{12} between the returns of two risky assets is given by

$$\rho_{12} = \frac{\text{Cov}(R_1, R_2)}{\sigma_1 \sigma_2} = \frac{C_{12}}{\sqrt{C_{11} C_{22}}}.$$

In this case, if we let w_1, w_2 be the weights then the expected return of the portfolio is equal to

$$\mu = w_1 \mu_1 + w_2 \mu_2$$

and its variance is given by

$$\begin{aligned} \sigma^2 &= C_{11} w_1^2 + C_{22} w_2^2 + 2C_{12} w_1 w_2 \\ &= \sigma_1^2 w_1^2 + \sigma_2^2 w_2^2 + 2\rho_{12} \sigma_1 \sigma_2 w_1 w_2, \end{aligned}$$

where C_{ij} is the covariance between the returns. Consult also Definition 4.16.

If $\rho_{12} = 1$ then $\sigma^2 = (\sigma_1 w_1 + \sigma_2 w_2)^2$ and $\sigma = |\sigma_1 w_1 + \sigma_2 w_2|$. Hence the set of the points

$$(\sigma, \mu) = (|\sigma_1 w_1 + \sigma_2 w_2|, w_1 \mu_1 + w_2 \mu_2)$$

on the (σ, μ) -plane is obtained by folding symmetrically along $\sigma = 0$ the left part of the straight line $w_1(\sigma_1, \mu_1) + w_2(\sigma_2, \mu_2)$. See the left graph in Fig. 20.2 where we choose $\mu = \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix}$ and $C = \begin{bmatrix} 0.2 & 0.3 \\ 0.3 & 0.45 \end{bmatrix}$. The points on the thick line segment $0 \leq w_1 \leq 1, 0 \leq w_2 \leq 1$ represent portfolios without short selling.

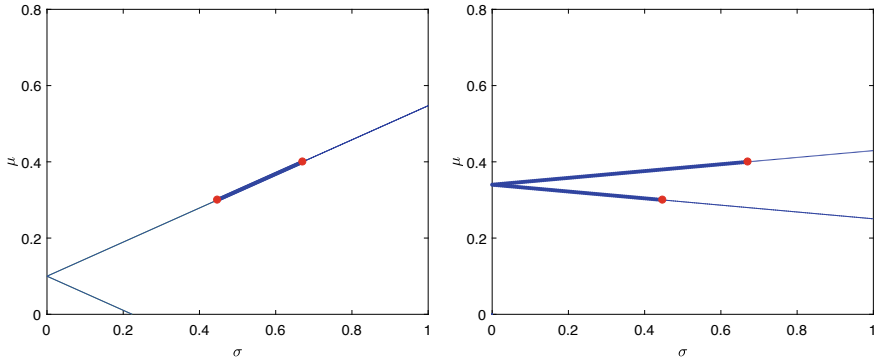


Fig. 20.2 A line representing portfolios for $\rho_{12} = 1$ (left) and a line for $\rho_{12} = -1$ (right)

If $\rho_{12} = -1$ then $\sigma^2 = (\sigma_1 w_1 - \sigma_2 w_2)^2$ and $\sigma = |\sigma_1 w_1 - \sigma_2 w_2|$. Hence the trajectory of the points

$$(\sigma, \mu) = (|\sigma_1 w_1 - \sigma_2 w_2|, w_1 \mu_1 + w_2 \mu_2)$$

on the (σ, μ) -plane is obtained by folding symmetrically along $\sigma = 0$ the left part of the straight line $w_1(\sigma_1, \mu_1) + w_2(-\sigma_2, \mu_2)$. See the right graph in Fig. 20.2, where we choose $\mu = \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix}$ and $C = \begin{bmatrix} 0.2 & -0.3 \\ -0.3 & 0.45 \end{bmatrix}$. The points on the thick line segment $0 \leq w_1 \leq 1, 0 \leq w_2 \leq 1$ represent portfolios without short selling.

For $-1 < \rho_{12} < 1$ the portfolios are represented by the curves in Fig. 20.3, which do not intersect the μ -axis. The portfolios of two assets with $\rho_{12} \approx 1$ are represented by the curves in the left, and the portfolios with $\rho_{12} \approx -1$ in the right. For a numerical example we take $\mu = \begin{bmatrix} 0.3 \\ 0.4 \end{bmatrix}$, $C = \begin{bmatrix} 0.25 & 0.3 \\ 0.3 & 0.45 \end{bmatrix}$ for $\rho_{12} \approx 0.8944$ and $C = \begin{bmatrix} 0.25 & -0.3 \\ -0.3 & 0.45 \end{bmatrix}$ for $\rho_{12} \approx -0.8944$. The points on the thick line segment given by $0 \leq w_1 \leq 1, 0 \leq w_2 \leq 1$ represent portfolios without short selling.

If the second asset is risk-free, i.e., the return R_2 is equal to r_f , then $\sigma_2 = 0$, and hence

$$(\sigma, \mu) = (|w_1| \sigma_1, w_1 \mu_1 + w_2 r_f) = w_1(\sigma_1, \mu_1) + w_2(0, r_f)$$

for $w_1 \geq 0$ and

$$(\sigma, \mu) = (|w_1| \sigma_1, w_1 \mu_1 + w_2 r_f) = w_1(-\sigma_1, \mu_1) + w_2(0, r_f)$$

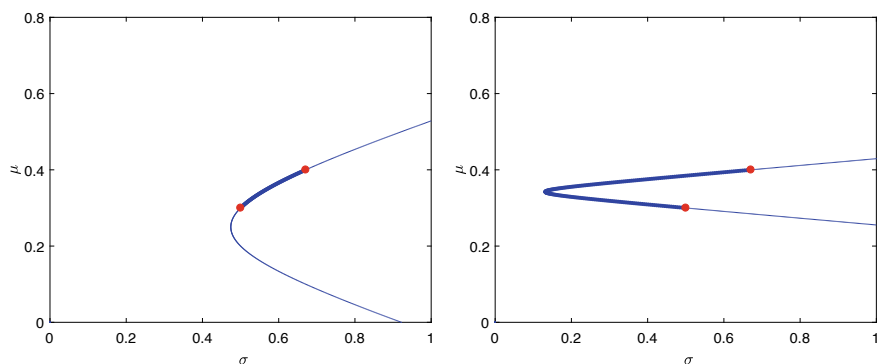


Fig. 20.3 A curve representing portfolios of two risky assets for $0 < \rho_{12} < 1$ (left) and a curve for $-1 < \rho_{12} < 0$ (right)

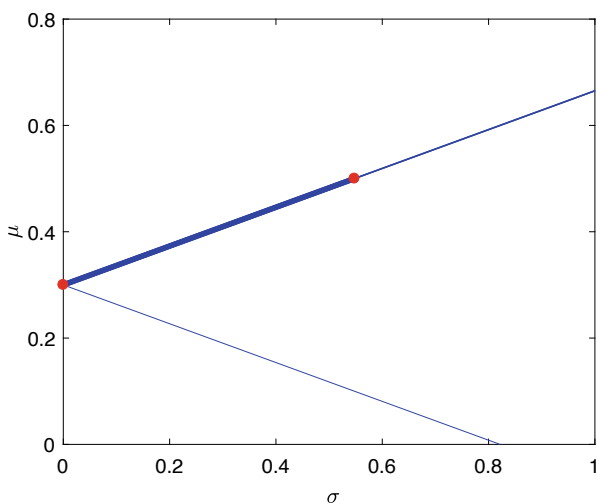
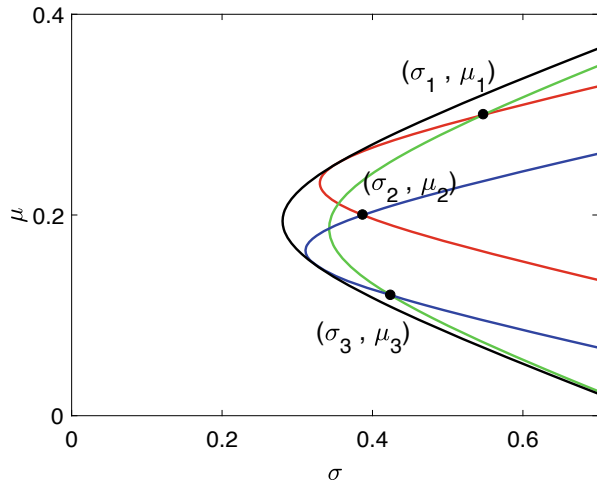


Fig. 20.4 Portfolios consisting of a risk-free asset and a risky asset

for $w_1 < 0$. See Fig. 20.4, where $r_f = 0.3$ and $\sigma_1 = \sqrt{0.3} \approx 0.5477$. The points on the thick line segment for $0 \leq w_1 \leq 1, 0 \leq w_2 \leq 1$ represent portfolios without short selling.

Now we consider portfolios consisting of three or more assets, which are assumed to be risky. For simplicity of notation we only consider portfolios of three assets. Similar conclusions can be obtained for portfolios consisting of more assets. Let the points (σ_1, μ_1) , (σ_2, μ_2) and (σ_3, μ_3) correspond to the given three assets, respectively. We choose any two of them and apply the method in the previous section to plot curves representing the portfolios defined by the chosen two assets. Now we pick one portfolio from each of three curves, and form a new portfolio curve. Such a curve is a set of portfolios defined by a weighted sum of the original three portfolios

Fig. 20.5 The set of attainable portfolios consisting of three risky assets



represented by (σ_1, μ_1) , (σ_2, μ_2) and (σ_3, μ_3) . The set M of all such points is the right side of the curve in Fig. 20.5, which is called the *Markowitz bullet*. In other words, if we let

$$W = \{(w_1, w_2, w_3) : w_1 + w_2 + w_3 = 1\}$$

and take $\mathbf{w}, \mathbf{w}' \in W$ and $c + c' = 1$, then $c\mathbf{w} + c'\mathbf{w}' \in W$. Thus, if $P, P' \in M$ then the new curve determined by P and P' is included in M .

In Fig. 20.6 we have the image of the triangle

$$\{(w_1, w_2, w_3) : w_1 + w_2 + w_3 = 1, w_1, w_2, w_3 \geq 0\},$$

which is the set of weights of three assets in a portfolio constructed without using short selling. We can see that the triangle is mapped into the first quadrant of the (σ, μ) -plane, being folded. (See Simulation 20.2).

20.3 An Application of the Lagrange Multiplier Method

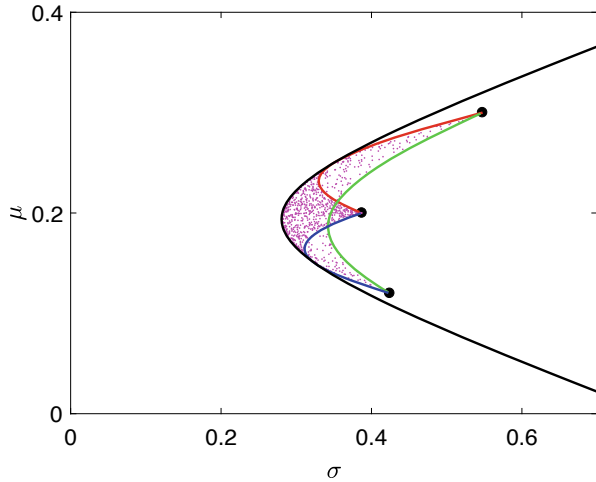
Now we solve the problem of finding the maximum or the minimum of a matrix function using the Lagrange multiplier method. From now on we write

$$\mathbf{1} = (1, \dots, 1)^t$$

and define an $n \times 2$ matrix

$$M = [\mathbf{1}, \boldsymbol{\mu}] \quad (20.1)$$

Fig. 20.6 The image of the triangle $w_1 + w_2 + w_3 = 1$, $w_1, w_2, w_3 \geq 0$



where the column vectors are given by $\mathbf{1}$ and $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^t$. In the following arguments, to eliminate trivial cases we assume that $\boldsymbol{\mu} \neq \alpha \mathbf{1}$ for any arbitrary constant α . If we had $\boldsymbol{\mu} = \alpha \mathbf{1}$ for some α , then all the risky assets would have the same expected return with different risk levels.

Theorem 20.3 (Extremum of a Quadratic Form) *Let A be an invertible $n \times n$ symmetric matrix. A necessary condition that the function $\mathbf{x}^t A \mathbf{x}$ has its extremum at $\mathbf{x}$ under the constraint $\mathbf{1}^t \mathbf{x} - 1 = 0$ is*

$$\mathbf{x} = \frac{1}{\mathbf{1}^t A^{-1} \mathbf{1}} A^{-1} \mathbf{1}.$$

Proof To apply the Lagrange multiplier method, we put

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^t A \mathbf{x},$$

$$g(\mathbf{x}) = \mathbf{1}^t \mathbf{x} - 1,$$

and let

$$F(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x}).$$

By Lemma B.2 we find the partial derivatives of F with respect to x_i , set them equal to 0, and write the equations in a vector notation, and finally obtain

$$\mathbf{0} = \left(\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n} \right) = \nabla F(\mathbf{x}, \lambda) = \nabla f(\mathbf{x}) - \lambda \nabla g(\mathbf{x}) = \mathbf{x}^t A - \lambda \mathbf{1}^t.$$

By taking transposes we have

$$\mathbf{0} = A^t \mathbf{x} - \lambda \mathbf{1} = A \mathbf{x} - \lambda \mathbf{1}$$

and

$$\mathbf{x} = \lambda A^{-1} \mathbf{1}.$$

Substituting the result in $\mathbf{1}^t \mathbf{x} = 1$, we obtain $\lambda = \frac{1}{\mathbf{1}^t A^{-1} \mathbf{1}}$. $\square$

In the above result the symmetry condition can be relaxed. If $B = \frac{1}{2}(A + A^t)$, then B is symmetric and $\mathbf{x}^t A \mathbf{x} = \mathbf{x}^t B \mathbf{x}$. In applications, A is given by the covariance matrix, which is symmetric, and the quadratic form is the variance of the return of a portfolio of n assets. See the next section.

Theorem 20.4 (Extremum of a Quadratic Form) *Let A be an $n \times n$ symmetric matrix. Assume that A^{-1} and $(M^t A^{-1} M)^{-1}$ exist. If $\mathbf{x}^t A \mathbf{x}$ has its minimum at the point $\mathbf{x}$ under the constraints*

$$\begin{cases} \mathbf{1}^t \mathbf{x} - 1 = 0, \\ \boldsymbol{\mu}^t \mathbf{x} - \mu_0 = 0 \end{cases}$$

for some constant μ_0 , then $\mathbf{x}$ satisfies

$$\mathbf{x} = \lambda_1 A^{-1} \mathbf{1} + \lambda_2 A^{-1} \boldsymbol{\mu},$$

where λ_1, λ_2 satisfy the conditions

$$\begin{cases} \lambda_1 \mathbf{1}^t A^{-1} \mathbf{1} + \lambda_2 \mathbf{1}^t A^{-1} \boldsymbol{\mu} = 1, \\ \lambda_1 \boldsymbol{\mu}^t A^{-1} \mathbf{1} + \lambda_2 \boldsymbol{\mu}^t A^{-1} \boldsymbol{\mu} = \mu_0. \end{cases}$$

Proof To apply the Lagrange multiplier method we put

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^t A \mathbf{x}, \quad g_1(\mathbf{x}) = \mathbf{1}^t \mathbf{x} - 1, \quad g_2(\mathbf{x}) = \boldsymbol{\mu}^t \mathbf{x} - \mu_0$$

and let

$$F(\mathbf{x}, \lambda_1, \lambda_2) = f(\mathbf{x}) - \lambda_1 g_1(\mathbf{x}) - \lambda_2 g_2(\mathbf{x}).$$

First, we compute the partial derivatives of F by Lemma B.2 and set them equal to 0, and finally obtain

$$\left(\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n} \right) = \nabla f - \lambda_1 \nabla g_1 - \lambda_2 \nabla g_2 = \mathbf{x}^t A - \lambda_1 \mathbf{1}^t - \lambda_2 \boldsymbol{\mu}^t = \mathbf{0}.$$

Hence

$$\mathbf{x}^t A = \lambda_1 \mathbf{1}^t + \lambda_2 \boldsymbol{\mu}^t.$$

By taking the transpose matrices, we obtain

$$A\mathbf{x} = \lambda_1 \mathbf{1} + \lambda_2 \boldsymbol{\mu}$$

and

$$\mathbf{x} = \lambda_1 A^{-1} \mathbf{1} + \lambda_2 A^{-1} \boldsymbol{\mu} = A^{-1} M \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}. \quad (20.2)$$

We rewrite the constraints

$$\begin{cases} \mathbf{1}' \mathbf{x} = 1 \\ \boldsymbol{\mu}' \mathbf{x} = \mu_0 \end{cases}$$

as

$$M' \mathbf{x} = \begin{bmatrix} 1 \\ \mu_0 \end{bmatrix}.$$

Now using (20.2), we convert the previous equation as

$$M' A^{-1} M \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \mu_0 \end{bmatrix}.$$

□

Remark 20.2 The conclusion of the above theorem can be rewritten as

$$\mathbf{x} = A^{-1} M \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

and

$$M' A^{-1} M \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \mu_0 \end{bmatrix}.$$

Hence

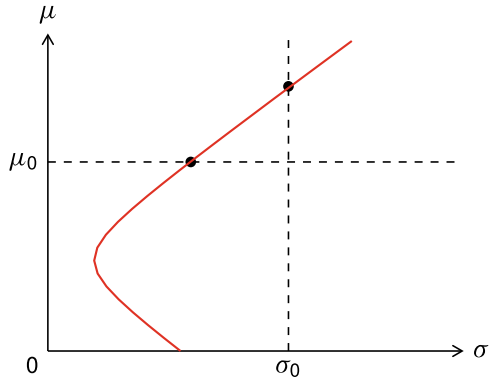
$$\mathbf{x} = A^{-1} M (M' A^{-1} M)^{-1} \begin{bmatrix} 1 \\ \mu_0 \end{bmatrix}.$$

20.4 Minimum Variance Line

In this section we assume that the covariance matrix C is invertible, and that A is given by a covariance matrix in Theorem 20.4. Variance represents the level of risk in our model. Our objective in optimal portfolio management is to find the maximum of expected return when variance is given, or to find the minimum of variance when the expected return is given. See Fig. 20.7.

Remark 20.3 The assumption that C is invertible is a plausible one. For, if C were not invertible, then there would exist $\mathbf{w} = (w_1, \dots, w_n) \neq \mathbf{0}$ such that $C\mathbf{w} = \mathbf{0}$. Now

Fig. 20.7 Two optimization problems: Find the maximum of expected return when σ_0 is given, or find the minimum of variance when μ_0 is given



consider a portfolio

$$V = w_1 S^1 + \cdots + w_n S^n.$$

Then its return R_V is given by

$$R_V = w_1 R_1 + \cdots + w_n R_n$$

and

$$\text{Var}(R_V) = \sum_{ij} w_i w_j \text{Cov}(R_i, R_j) = (C\mathbf{w}, \mathbf{w}) = 0.$$

This implies that R_V is constant, which is nothing but $\sum_i w_i \mu_i$. Since

$$R_V = \sum_i w_i \frac{S_T^i - S_0^i}{S_0^i} = \sum_i \frac{w_i}{S_0^i} S_T^i - 1,$$

we have

$$\sum_i \frac{w_i}{S_0^i} S_T^i - 1 - \sum_i w_i \mu_i = 0,$$

which is a nontrivial linear combination of n random variables S_T^i representing risky assets and one risk-free asset. (The constant $-1 - \sum_i w_i \mu_i$ is regarded as a coefficient $-1 - \sum_i w_i \mu_i$ times the random variable representing a bond that pays 1 at time T .) Hence we would have linear dependence among random variables S_T^i , $1 \leq i \leq n$, and a bond given by a constant function 1. This is not a plausible assumption.

Definition 20.5 Since the covariance matrix C is assumed to be invertible, it has nonzero real eigenvalues. Furthermore, it can easily be shown that eigenvalues are positive, and we have $\mathbf{x}' C \mathbf{x} > 0$ for every $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x} \in \mathbb{R}^n$. The same property holds with C^{-1} .

Define a bilinear form $(\cdot, \cdot)_C$ by

$$(\mathbf{x}, \mathbf{y})_C = \mathbf{y}^t C^{-1} \mathbf{x},$$

then $(\cdot, \cdot)_C$ is a positive definite inner product on $\mathbb{R}^n$.

Lemma 20.2 *Let $M = [\mathbf{1}, \boldsymbol{\mu}]$ be the $n \times 2$ matrix defined by (20.1). Then $M^t C^{-1} M$ is invertible, and*

$$\det(M^t C^{-1} M) > 0.$$

Proof Note that

$$M^t C^{-1} M = \begin{bmatrix} \mathbf{1}^t C^{-1} \mathbf{1} & \mathbf{1}^t C^{-1} \boldsymbol{\mu} \\ \boldsymbol{\mu}^t C^{-1} \mathbf{1} & \boldsymbol{\mu}^t C^{-1} \boldsymbol{\mu} \end{bmatrix}.$$

By the Cauchy–Schwarz inequality we obtain

$$(\mathbf{x}, \mathbf{y})_C^2 \leq (\mathbf{x}, \mathbf{x})_C (\mathbf{y}, \mathbf{y})_C$$

i.e.,

$$|\mathbf{y}^t C^{-1} \mathbf{x}|^2 \leq \mathbf{x}^t C^{-1} \mathbf{x} \mathbf{y}^t C^{-1} \mathbf{y}$$

and again we obtain

$$\mathbf{x}^t C^{-1} \mathbf{x} \mathbf{y}^t C^{-1} \mathbf{y} - \mathbf{y}^t C^{-1} \mathbf{x} \mathbf{x}^t C^{-1} \mathbf{y} \geq 0. \quad (20.3)$$

(Equality holds in (20.3) under the condition given in Theorem C.1.) Substituting $\mathbf{x} = \mathbf{1}$ and $\mathbf{y} = \boldsymbol{\mu}$, we obtain

$$\det(M^t C^{-1} M) = \mathbf{1}^t C^{-1} \mathbf{1} \boldsymbol{\mu}^t C^{-1} \boldsymbol{\mu} - \mathbf{1}^t C^{-1} \boldsymbol{\mu} \boldsymbol{\mu}^t C^{-1} \mathbf{1} \geq 0.$$

Since $\boldsymbol{\mu}$ is not a scalar multiple of $\mathbf{1}$, the inequality is strict. $\square$

If we define a 2×2 invertible matrix L by $L = (M^t C^{-1} M)^{-1}$, then $\det L > 0$, $L^t = L$ and $L^{-1} = M^t C^{-1} M$. Note that

$$L = (\det L) \begin{bmatrix} \boldsymbol{\mu}^t C^{-1} \boldsymbol{\mu} & -\mathbf{1}^t C^{-1} \boldsymbol{\mu} \\ -\boldsymbol{\mu}^t C^{-1} \mathbf{1} & \mathbf{1}^t C^{-1} \mathbf{1} \end{bmatrix}.$$

When the expected return is equal to μ_0 , the weight vector $\mathbf{w}$ that minimizes the variance is given by

$$\mathbf{w} = C^{-1} M L \begin{bmatrix} 1 \\ \mu_0 \end{bmatrix}$$

by Remark 20.2, and the corresponding variance satisfies

$$\begin{aligned}
\mathbf{w}^t C \mathbf{w} &= [1, \mu_0] L^t M^t C^{-1} C C^{-1} M L \begin{bmatrix} 1 \\ \mu_0 \end{bmatrix} \\
&= [1, \mu_0] L M^t C^{-1} M L \begin{bmatrix} 1 \\ \mu_0 \end{bmatrix} \\
&= [1, \mu_0] L \begin{bmatrix} 1 \\ \mu_0 \end{bmatrix} \\
&= L_{11} + 2 L_{12} \mu_0 + L_{22} \mu_0^2,
\end{aligned}$$

where $L = [L_{ij}]$. Since the covariance matrix C is assumed to be invertible, Lemma 20.1 implies $\mathbf{w}^t C \mathbf{w} > 0$ for $\mathbf{w} \neq \mathbf{0}$, and hence $L_{22} > 0$.

Remark 20.4 Note that $L_{22} = (\det L) \mathbf{1}^t C^{-1} \mathbf{1}$.

A point on the (σ, μ) -plane will represent the standard deviation $\sigma = \sqrt{\mathbf{w}^t C \mathbf{w}}$ and the expected return $\mu = \mathbf{w}^t \boldsymbol{\mu}$ of the return of a portfolio consisting of risky assets with a weight vector $\mathbf{w}$ satisfying $\mathbf{1}^t \mathbf{w} = 1$. (In this book μ_i , $\boldsymbol{\mu}$, μ will denote respectively the expected return of the i th risky asset, ordered n -tuples of n expected returns, and the expected return of a portfolio consisting of n risky assets.)

The set of the points on the (σ, μ) -plane corresponding to all the weight vectors is of the form

$$\{ (\sigma, \mu) : \sigma \geq \sqrt{L_{11} + 2 L_{12} \mu + L_{22} \mu^2} \}.$$

The boundary of the domain is given by

$$L_{11} + 2 L_{12} \mu + L_{22} \mu^2 = \sigma^2.$$

Since $L_{22} > 0$, the above equation represents a hyperbola

$$\sigma_0^2 + L_{22}(\mu - \mu_0)^2 = \sigma^2.$$

The asymptotes are given by

$$\mu = \pm \frac{1}{\sqrt{L_{22}}} \sigma + \mu_0,$$

where

$$\mu_0 = -\frac{L_{12}}{L_{22}} = \frac{\mathbf{1}^t C^{-1} \boldsymbol{\mu}}{\mathbf{1}^t C^{-1} \mathbf{1}},$$

$$\sigma_0^2 = L_{11} - \frac{L_{12}^2}{L_{22}} = \frac{\det L}{L_{22}} = \frac{1}{\mathbf{1}^t C^{-1} \mathbf{1}}.$$

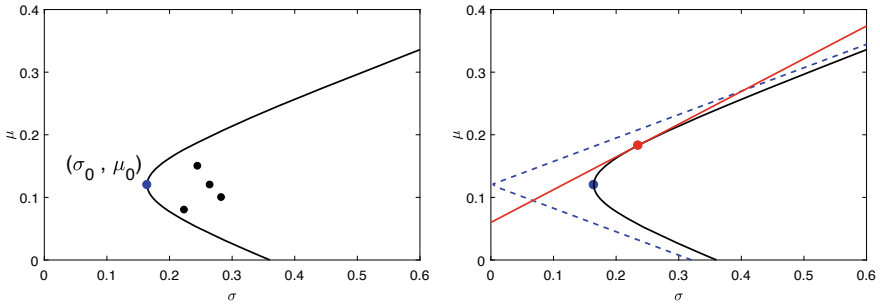


Fig. 20.8 The minimum variance line for risky assets and the point (σ_0, μ_0) representing the portfolio with the minimum variance among all portfolios (*left*), and asymptotes and a tangent line (*right*)

We used the result in Remark 20.4 in finding σ_0 . Since the two asymptotes intersect at $(0, \mu_0)$ on the μ -axis, a tangent line with positive slope to the minimum variance line should have the μ -intercept $(0, r)$ for some $r < \mu_0$. See Fig. 20.8.

Definition 20.6 (*Minimum Variance Line*) The curve thus obtained is called the *minimum variance line*. A point on the minimum variance line represents the portfolio whose variance is the smallest among portfolios with the same expected return.

Among the points on the minimum variance line the point (σ_0, μ_0) represents the investment with the minimal risk. Its weight vector $\mathbf{w}_0$ is given by

$$\begin{aligned} \mathbf{w}_0 &= C^{-1}ML \begin{bmatrix} 1 \\ \mu_0 \end{bmatrix} \\ &= C^{-1}ML \begin{bmatrix} 1 \\ \frac{\mathbf{1}'C^{-1}\boldsymbol{\mu}}{\mathbf{1}'C^{-1}\mathbf{1}} \end{bmatrix} \\ &= \frac{1}{\mathbf{1}'C^{-1}\mathbf{1}}C^{-1}ML \begin{bmatrix} \mathbf{1}'C^{-1}\mathbf{1} \\ \mathbf{1}'C^{-1}\boldsymbol{\mu} \end{bmatrix}. \end{aligned}$$

From the definition of L we have

$$\begin{bmatrix} \mathbf{1}'C^{-1}\mathbf{1} \\ \mathbf{1}'C^{-1}\boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} (L^{-1})_{11} \\ (L^{-1})_{21} \end{bmatrix} = \text{the first column of } L^{-1}$$

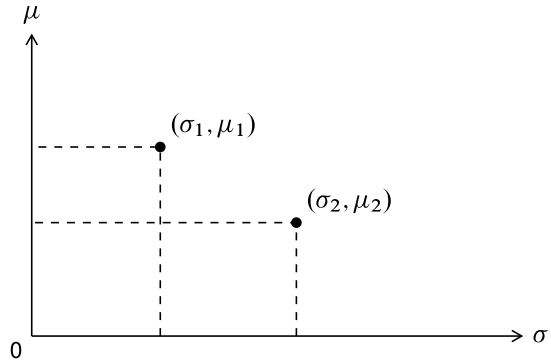
and

$$L \begin{bmatrix} \mathbf{1}'C^{-1}\mathbf{1} \\ \mathbf{1}'C^{-1}\boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Hence

$$\mathbf{w}_0 = \frac{1}{\mathbf{1}'C^{-1}\mathbf{1}}C^{-1}M \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\mathbf{1}'C^{-1}\mathbf{1}}C^{-1}\mathbf{1}.$$

Fig. 20.9 The first asset dominates the second asset



Note that we can apply Theorem 20.3 to obtain the same result.

20.5 The Efficient Frontier

Definition 20.7 (*Dominating Portfolio*) We are given two assets (or portfolios) with expected returns μ_1, μ_2 , and standard deviations σ_1, σ_2 , respectively. If $\mu_1 \geq \mu_2$ and $\sigma_1 \leq \sigma_2$ as in Fig. 20.9, we say that the first asset *dominates* the second asset.

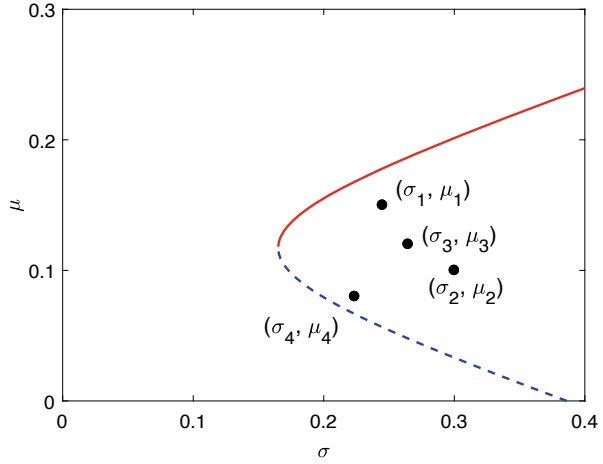
Definition 20.8 (*Efficient Frontier*) A portfolio is said to be *efficient* when there is no other portfolio that dominates it. The set of all efficient portfolios is called the *efficient frontier*. In Fig. 20.10 the minimum variance line is given as the boundary of the set of attainable portfolios constructed by several assets, and the thick upper part is the efficient frontier.

Remark 20.5 (i) When the first asset dominates the second one, it is better to invest in the first one since the first asset has lower risk but higher expected return. Even when one asset dominates the other asset, we can combine two assets and construct a portfolio reducing risk as in Fig. 20.9.

- (ii) For an efficient investment we should construct an efficient portfolio, and set the level of risk, determined by σ , according to the preference of the investor. To gain high return, an investor must accept high risk.
- (iii) The efficient portfolio has the maximum expected return among the portfolios with the same standard deviation. Thus the efficient frontier is a subset of the minimum variance line.

Lemma 20.3 Let $X_i, 1 \leq i \leq n$, be random variables. Then the covariance of two portfolios $\sum_i a_i X_i$ and $\sum_j b_j X_j$ is given by

Fig. 20.10 The efficient frontier



$$\text{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^n b_j X_j\right) = \sum_{i=1}^n \sum_{j=1}^n a_i b_j \text{Cov}(X_i, X_j).$$

In vector notation

$$\text{Cov}(\mathbf{a}^t \mathbf{X}, \mathbf{b}^t \mathbf{X}) = \mathbf{a}^t \mathbf{C} \mathbf{b} = \mathbf{b}^t \mathbf{C} \mathbf{a},$$

where $\mathbf{a} = (a_1, \dots, a_n)^t$, $\mathbf{b} = (b_1, \dots, b_n)^t$, $\mathbf{X} = (X_1, \dots, X_n)^t$ are column vectors and $\mathbf{C} = [\text{Cov}(X_i, X_j)]_{ij}$.

Theorem 20.5 All the efficient portfolios have the same covariance as the minimum variance portfolio represented by the point (σ_0, μ_0) .

Proof Let a point (σ, μ) represent a given efficient portfolio. Then

$$\sigma^2 - L_{22} \mu^2 - 2 L_{12} \mu - L_{11} = 0$$

and the weight vector is given by

$$\mathbf{w} = \mathbf{C}^{-1} \mathbf{M} \mathbf{L} \begin{bmatrix} 1 \\ \mu \end{bmatrix}.$$

In this case we have $\mathbf{w}^t \mathbf{R} = [1, \mu] \mathbf{L} \mathbf{M}^t \mathbf{C}^{-1} \mathbf{R}$ and by Lemma 20.3 we obtain

$$\begin{aligned}
\text{Cov}(\mathbf{w}_0^t \mathbf{R}, \mathbf{w}^t \mathbf{R}) &= \text{Cov}(\mathbf{w}_0^t \mathbf{R}, [1, \mu] L M^t C^{-1} \mathbf{R}) \\
&= [1, \mu] L M^t C^{-1} C \mathbf{w}_0 \\
&= [1, \mu] L M^t \mathbf{w}_0 \\
&= [1, \mu] L \begin{bmatrix} \mathbf{1}^t \\ \boldsymbol{\mu}^t \end{bmatrix} \left(\frac{1}{\mathbf{1}^t C^{-1} \mathbf{1}} C^{-1} \mathbf{1} \right) \\
&= \frac{1}{\mathbf{1}^t C^{-1} \mathbf{1}} [1, \mu] L \begin{bmatrix} \mathbf{1}^t C^{-1} \mathbf{1} \\ \boldsymbol{\mu}^t C^{-1} \mathbf{1} \end{bmatrix} \\
&= \frac{1}{\mathbf{1}^t C^{-1} \mathbf{1}} [1, \mu] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\mathbf{1}^t C^{-1} \mathbf{1}},
\end{aligned}$$

where we used

$$L^{-1} = M^t C^{-1} M = \begin{bmatrix} \mathbf{1}^t C^{-1} \mathbf{1} & \boldsymbol{\mu}^t C^{-1} \mathbf{1} \\ \boldsymbol{\mu}^t C^{-1} \mathbf{1} & \boldsymbol{\mu}^t C^{-1} \boldsymbol{\mu} \end{bmatrix}$$

in the fifth equality. □

20.6 The Market Portfolio

Now we consider a portfolio that contains a risk-free asset such as a bank deposit or bond. If we let $r_f > 0$ denote the risk-free interest rate, then such a risk-free asset is represented by a point $(0, r_f)$ on the (σ, μ) -plane. Consider an investment in which we invest into a risk-free asset by the ratio b , and into a portfolio π of risky assets by the ratio $1 - b$. Assume that the return of the risky portfolio has its standard deviation and average given by (σ_π, μ_π) and that there exists a weight vector $\mathbf{w}$ such that the return is given by $\mathbf{w}^t \mathbf{R}$. Here we do not exclude the case that $b < 0$, which means that borrowed cash is invested more in the risky portfolio. The given risky portfolio is represented by the point (σ_π, μ_π) in Fig. 20.11.

Such a combination of investments has expected return

$$\mathbb{E}[b r_f + (1 - b) \mathbf{w}^t \mathbf{R}] = b r_f + (1 - b) \mu_\pi$$

and standard deviation

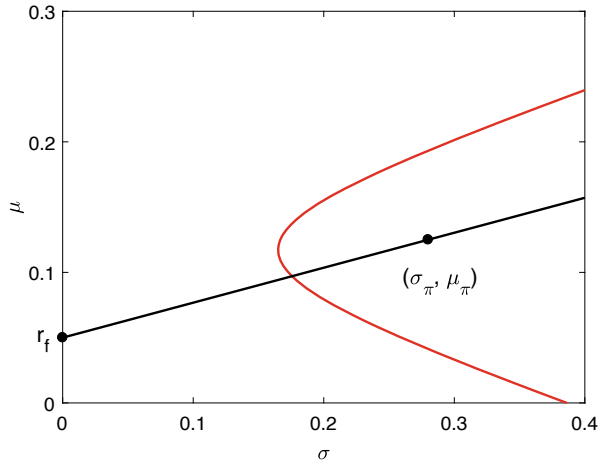
$$\sqrt{\text{Var}(b r_f + (1 - b) \mathbf{w}^t \mathbf{R})} = \sqrt{\text{Var}((1 - b) \mathbf{w}^t \mathbf{R})} = |1 - b| \sigma_\pi,$$

hence the corresponding point on the (σ, μ) -plane lies on the straight line

$$((1 - b) \sigma_\pi, b r_f + (1 - b) \mu_\pi) = b(0, r_f) + (1 - b)(\sigma_\pi, \mu_\pi)$$

for $b \leq 1$. In other words, it is on the line passing through the points $(0, r_f)$ and (σ_π, μ_π) .

Fig. 20.11 A portfolio with a risk-free asset



In Fig. 20.11 the thick line segment corresponds to the range $0 \leq b \leq 1$, and the thin segment to $b < 0$.

The line segment corresponding to $b > 1$ is obtained by reflecting the line in Fig. 20.11 using $\mu = r_f$ as the axis of symmetry. A portfolio represented by a point on the line segment has lower expected return than the case $b \leq 1$ but with the same risk, thus it is not regarded as a possible investment strategy.

When a risk-free asset is included, a necessary condition for an efficient investment is that the portfolio represented by (σ_π, μ_π) must be located on the efficient frontier for the portfolio without a risk-free asset. Furthermore, (σ_π, μ_π) should be the tangent point (σ_M, μ_M) of the tangent line passing through $(0, r_f)$ and tangent to the Markowitz bullet. Therefore, when a risk-free asset is included, the efficient frontier is the tangent line to the Markowitz bullet passing through $(0, r_f)$ as in Fig. 20.12.

Definition 20.9 (*Market Portfolio*) The tangent line given in the above is called the *capital market line*, and its equation is given by

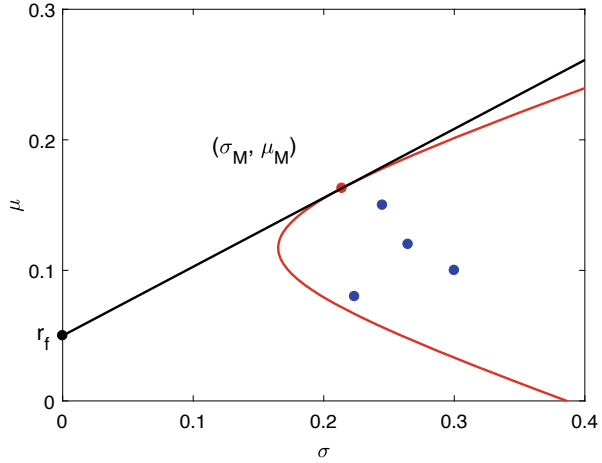
$$\mu = r_f + \frac{\mu_M - r_f}{\sigma_M} \sigma.$$

The portfolio represented by (σ_M, μ_M) is called the *market portfolio* because it is employed by all the market participants for efficient investment. A stock price index is often used as a proxy for the market portfolio.

Definition 20.10 (*Market Price of Risk*) If a portfolio lying on the capital market line has risk σ , then

$$\frac{\mu_M - r_f}{\sigma_M} \sigma$$

Fig. 20.12 The capital market line and the market portfolio



is called the *risk premium* or *market price of risk*. It is a compensation for the investment risk. For more details consult (Capinski & Zastawniak, 2003).

Theorem 20.6 *The market portfolio satisfies the condition*

$$\gamma C\mathbf{w} = \boldsymbol{\mu} - r_f \mathbf{1},$$

where $\mathbf{w}$ is the weight of the portfolio, $\gamma = \frac{\mu_M - r_f}{\sigma_M^2}$, C the covariance matrix, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ the expected return vector, and $\mathbf{1} = (1, \dots, 1)$.

Proof Among the lines passing through $(0, r_f)$ and a point in the interior of the efficient frontier, the capital market line has the maximal slope. See Figs. 20.11 and 20.12. If the market portfolio is represented by the point $(\sigma_M, \mu_M) = (\sqrt{\mathbf{w}^t C \mathbf{w}}, \boldsymbol{\mu}^t \mathbf{w})$ on the (σ, μ) -plane, then its slope is given by

$$\frac{\boldsymbol{\mu}^t \mathbf{w} - r_f}{\sqrt{\mathbf{w}^t C \mathbf{w}}}.$$

To find the maximum under the constraint $\mathbf{1}^t \mathbf{w} = 1$, we apply the Lagrange multiplier method. Let

$$F(\mathbf{w}, \lambda) = \frac{\boldsymbol{\mu}^t \mathbf{w} - r_f}{\sqrt{\mathbf{w}^t C \mathbf{w}}} - \lambda(\mathbf{1}^t \mathbf{w} - 1)$$

and take partial derivatives. Then

$$\frac{\partial F}{\partial w_i} = \frac{\mu_i \sqrt{\mathbf{w}^t C \mathbf{w}} - (\boldsymbol{\mu}^t \mathbf{w} - r_f) \frac{\sum_{j=1}^n C_{ij} w_j}{\sqrt{\mathbf{w}^t C \mathbf{w}}}}{\mathbf{w}^t C \mathbf{w}} - \lambda = 0$$

$$\frac{\partial F}{\partial \lambda} = \mathbf{1}' \mathbf{w} - 1 = 0.$$

The first equation becomes

$$\mu_i \sigma_M^2 - (\mu_M - r_f) \sum_{j=1}^n C_{ij} w_j = \lambda \sigma_M^3,$$

which is equal to

$$\boldsymbol{\mu} - \frac{\mu_M - r_f}{\sigma_M^2} C \mathbf{w} = \lambda \sigma_M \mathbf{1} \quad (*)$$

in a vector form. Now we take the inner product with $\mathbf{w}$ and obtain

$$\mu_M - \frac{\mu_M - r_f}{\sigma_M^2} \sigma_M^2 = \lambda \sigma_M.$$

Hence we have

$$\lambda = \frac{r_f}{\sigma_M},$$

which yields

$$\mathbf{w} = \frac{\sigma_M^2}{\mu_M - r_f} C^{-1} (\boldsymbol{\mu} - r_f \mathbf{1})$$

in combination with (*). Therefore we have $\gamma C \mathbf{w} = \boldsymbol{\mu} - r_f \mathbf{1}$. □

Remark 20.6 (i) As explained in Sect. 20.4, we need the condition $r_f < \mu_0$ for the existence of a tangent line with positive slope.

(ii) If $r_f \geq \mu_0$, the best strategy is to invest everything in the risk-free asset. If risky assets are independent and $\sigma_i = 1$, then we have

$$\mu_0 = \frac{\mathbf{1}' C^{-1} \boldsymbol{\mu}}{\mathbf{1}' C^{-1} \mathbf{1}} = \frac{\mu_1 + \cdots + \mu_n}{n}$$

since $C = \text{diag}(1, \dots, 1)$ is the identity matrix. Thus, if

$$r_f \geq \frac{\mu_1 + \cdots + \mu_n}{n}$$

then it is efficient to invest everything in the risk-free asset. In other words, when r_f is sufficiently high, bank deposit is better than stock.

(iii) For Figs. 20.8, 20.10 and 20.12 we take

$$C = \begin{bmatrix} 0.06 & 0.05 & 0.02 & 0.01 \\ 0.05 & 0.08 & 0.05 & 0.01 \\ 0.02 & 0.05 & 0.07 & 0.01 \\ 0.01 & 0.01 & 0.01 & 0.05 \end{bmatrix} \quad \text{and} \quad \mu = \begin{bmatrix} 0.15 \\ 0.1 \\ 0.12 \\ 0.08 \end{bmatrix}.$$

20.7 The Beta Coefficient

Definition 20.11 (*Beta Coefficient*) The beta coefficient β_V for a portfolio or an individual asset V is defined by

$$\beta_V = \frac{\text{Cov}(R_V, R_M)}{\sigma_M^2}.$$

The beta coefficient measures the sensitivity of a given stock or portfolio price with respect to the stock price index. When we have a bull market, it is better to have a stock with large beta since the change in its price is larger than in other stocks in general. However, with a bear market a stock with smaller beta is better. Sectors with large betas are electronics, finance, medical equipments, and sectors with small betas are chemicals, medicine and food. Consumer-oriented sectors which are not affected during economic downturn have small betas.

Theorem 20.7 For an arbitrary portfolio V with expected return μ_V , its beta coefficient β_V is given by

$$\beta_V = \frac{\mu_V - r_f}{\mu_M - r_f}.$$

Proof By Definition 20.11 we have

$$\beta_V = \frac{\mathbf{w}_V^t C \mathbf{w}_M}{\sigma_M^2}.$$

Theorem 20.6 implies that

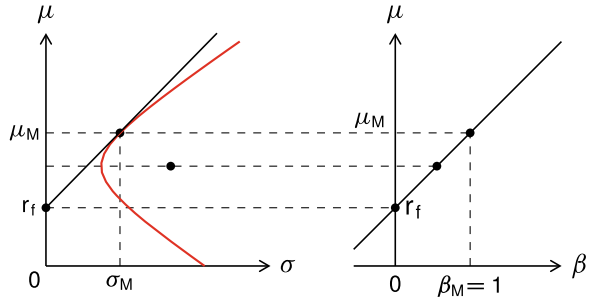
$$\beta_V = \frac{1}{\mu_M - r_f} \mathbf{w}_V^t (\mu - r_f \mathbf{1}) = \frac{1}{\mu_M - r_f} (\mathbf{w}_V^t \mu - r_f \mathbf{w}_V^t \mathbf{1}).$$

Now we use $\mathbf{w}_V^t \mu = \mu_V$ and $\mathbf{w}_V^t \mathbf{1} = 1$. □

For example, the expected return of the i th asset S_i in excess of the risk-free interest rate r_f satisfies

$$\mathbb{E}[R_i] - r_f = \beta_i (\mathbb{E}[R_M] - r_f),$$

Fig. 20.13 The relation between the beta coefficient and expected return



where

$$\beta_i = \frac{\text{Cov}(R_i, R_M)}{\sigma_M^2}.$$

In Fig. 20.13 a portfolio V is represented by a point (β_V, μ_V) on the straight line

$$\mu = r_f + (\mu_M - r_f)\beta,$$

called the *security market line* in the (β, μ) -plane. The term $(\mu_M - r_f)\beta$ is regarded as the risk premium.

Remark 20.7 Let $\mathbf{w} = (w_1, \dots, w_n)$ be the weight vector of the market portfolio. Then $w_i(\mu_i - r_f)$ is the portion of the contribution of the i th asset in the whole return of the portfolio.

20.8 Derivation of the Black–Scholes–Merton Equation

Assume that an asset S_t follows the geometric Brownian motion

$$dS_t = rS_t dt + \sigma S_t dW_t,$$

where r denotes the risk-free interest rate. Then Theorem 20.7 implies that over the time interval $[t, t + dt]$ the expected return of S is given by

$$\mathbb{E}[R_S dt] = \mathbb{E}\left[\frac{dS_t}{S_t}\right] = r dt + \beta_S(\mathbb{E}[R_M] - r) dt, \quad (20.4)$$

where

$$\beta_S = \frac{\text{Cov}(R_S, R_M)}{\sigma_M^2}.$$

Similarly, for the price of a portfolio process V_t , we have

$$\mathbb{E}[R_V] = \mathbb{E}\left[\frac{dV_t}{V_t}\right] = r dt + \beta_V (\mathbb{E}[R_M] - r) dt,$$

and

$$\mathbb{E}[dV_t] = r V_t dt + \beta_V V_t (\mathbb{E}[R_M] - r) dt. \quad (20.5)$$

Now consider a European option with an underlying asset S and let $V_t = V(t, S_t)$ be its price at time t . By Itô's formula in Theorem 10.3, we have

$$dV_t = \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \frac{\partial V}{\partial S} \frac{dS}{S} S_t.$$

Hence

$$\begin{aligned} \mathbb{E}[dV_t] &= \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \frac{\partial V}{\partial S} S \mathbb{E}[R_S dt] \\ &= \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \frac{\partial V}{\partial S} S (r dt + \beta_S (\mathbb{E}[R_M] - r) dt), \end{aligned} \quad (20.6)$$

where we used (20.4) for the second equality.

Since

$$\frac{dV}{V} = \frac{1}{V} \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \frac{\partial V}{\partial S} \frac{dS}{S} \frac{S}{V},$$

we have

$$R_V dt = \frac{1}{V_t} \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \frac{\partial V}{\partial S} \frac{S}{V} R_S dt,$$

and hence

$$R_V = \frac{1}{V_t} \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) + \frac{\partial V}{\partial S} \frac{S_t}{V_t} R_S.$$

By taking the covariance of R_V with R_M , we obtain

$$\text{Cov}(R_V, R_M) = \frac{\partial V}{\partial S} \frac{S_t}{V_t} \text{Cov}(R_S, R_M)$$

since the deterministic component has zero covariance with R_M . Hence

$$\beta_V = \frac{\partial V}{\partial S} \frac{S_t}{V_t} \beta_S,$$

and we rewrite (20.5) as

$$\mathbb{E}[dV_t] = r V dt + \frac{\partial V}{\partial S} S \beta_S (\mathbb{E}[R_M] - r) dt. \quad (20.7)$$

From (20.6) and (20.7), after cancelling common terms, we obtain

$$\left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) dt + \frac{\partial V}{\partial S} S r dt = r V dt.$$

Dividing by dt , we have

$$\left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} \right) + \frac{\partial V}{\partial S} S r = r V,$$

which is the Black–Scholes–Merton equation.

20.9 Computer Experiments

Simulation 20.1 (Portfolios of Two Assets) Take a covariance matrix $C = \begin{bmatrix} 0.2 & -0.3 \\ -0.3 & 0.45 \end{bmatrix}$ which gives the correlation coefficient $\rho_{12} = -1$. See Fig. 20.2.

```
mu = [0.3; 0.41] ; % returns of assets
% Define a covariance matrix.
C = [ 0.2  -0.3 ; -0.3  0.45 ];
eig(C) % Check the eigenvalues.
rho12 = C(1,2)/sqrt(C(1,1)*C(2,2)) % -1
```

```
N = 1000; % number of plotted points
mu_V = zeros(1,N);
sigma_V = zeros(1,N);
dt=1/N;
for i=1:N
    w = [i*dt; 1-i*dt]; % weight
    mu_V(i) = w' * mu;
    sigma_V(i) = sqrt(w'*C*w);
end;
plot(sigma_V,mu_V, 'o') ;
hold on
```

```
for i=N+1:2*N
    w = [i*dt; 1-i*dt];
    mu_V(i) = w' * mu;
    sigma_V(i) = sqrt(w'*C*w);
end;
plot(sigma_V,mu_V, 'o') ;
hold on
```

```

for i=N+1:2*N
w = [1-i*dt; i*dt];
mu_V(i) = w' * mu;
sigma_V(i) = sqrt(w'*C*w);
end;
plot(sigma_V,mu_V, '.');
hold on

plot(sqrt(C(1,1)),mu(1),'.') ;
plot(sqrt(C(2,2)),mu(2),'.') ;

```

Simulation 20.2 (Portfolios of Three Assets) We find the image of all attainable portfolios of three risky assets without short selling. See Fig. 20.6.

```

mu_vec = [0.3; 0.2; 0.12] ; % returns of three risky assets
% covariance matrix
C = [ 0.3  0.02 0.01 ;
      0.02 0.15 0.03 ;
      0.01 0.03 0.18 ];

N = 1000; % number of plotted points
for i=1:N
w1 = rand;
w2 = 1-w1;
weight = [ w1; w2; 0];
mu12(i) = weight' * mu_vec;
sigma12(i) = sqrt(weight' * C * weight);
end ;

for i=1:N
w1 = rand;
w3 = 1-w1;
weight = [ w1; 0; w3];
mu13(i) = weight' * mu_vec;
sigma13(i) = sqrt(weight' * C * weight);
end ;

for i=1:N
w2 = rand;
w3 = 1-w2;
weight = [0; w2; w3];
mu23(i) = weight' * mu_vec;
sigma23(i) = sqrt(weight' * C * weight);
end ;

```

```

L = 5000;
for i=1:L
w2 = rand;
w1 = (1-w2)*rand;
weight = [ w1; w2; 1-w1-w2 ];
mu(i) = weight' * mu_vec ;
sigma(i) = sqrt(weight' * C * weight);
end ;

plot(sigma,mu,'.') ;
hold on
plot(sigma12,mu12,'.') ;
hold on
plot(sigma13,mu13,'.') ;
hold on
plot(sigma23,mu23,'.') ;

```

Simulation 20.3 (Capital Market Line) Based on the data for four risky assets in Remark 20.6(iii), we plot the minimum variance portfolio, the efficient frontier, the minimum variance line, the market portfolio, and the capital market line.

```

mu = [ 0.15; 0.1; 0.12; 0.08]; % returns of assets
% the covariance matrix.
C = [ 0.06, 0.05, 0.02, 0.01 ;
      0.05, 0.08, 0.05, 0.01 ;
      0.02, 0.05, 0.07, 0.01 ;
      0.01, 0.01, 0.01, 0.05 ] ;

one_vec = ones(4,1); % a vector of 1's
inv_C = inv(C); % inverse of the covariance matrix
A1 = one_vec' * inv_C * one_vec; % 1'*C^(-1)*1
A2 = one_vec' * inv_C * mu; % 1'*C^(-1)*mu = mu'*C^(-1)*1
A3 = mu' * inv_C * mu; % mu'*C^(-1)*mu
L = inv([A1, A2; A2, A3]);

% minimum variance portfolio
mu_0 = - L(1,2)/L(2,2)
sigma_0 = sqrt(1/A1);

N = 200; % number of plotted points
max_sigma = 0.8
%max_mu = 1/sqrt(L(2,2)) * max_sigma + mu_0
max_mu = 0.5;
r = 0.06; % To have a tangent line we take r < mu_0.

```

```

% asymptotes
mu_asym = linspace(0,max_mu,N);
sigma_asym = zeros(1,N);
for i=1:N
    sigma_asym(i) = sqrt(L(2,2)*(mu_asym(i) - mu_0)^2);
end

% minimum variance line
mu_MVL = linspace(0,max_mu,N);
sigma_MVL = zeros(1,N);
for i=1:N
    sigma_MVL(i) = sqrt(L(2,2)*mu_MVL(i)^2 + 2*L(1,2)*mu_MVL(i) + L(1,1));
end

% Market Portfolio
mu_M = -(L(1,1)+L(1,2)*r)/(L(1,2)+L(2,2)*r);
sigma_M = sqrt(L(2,2)*mu_M^2 + 2*L(1,2)*mu_M + L(1,1));

% Capital Market Line
mu_CML = linspace(r,max_mu,N);
sigma_CML = zeros(1,N);
for i=1:N
    sigma_CML(i) = ((mu_CML(i) - r)*sigma_M)/(mu_M - r);
end

plot(sigma_0, mu_0, '.'); % Minimum Variance Portfolio
hold on

plot(sigma_asym, mu_asym, '--'); % Asymptotes
hold on

plot(sigma_MVL, mu_MVL, '-'); % Minimum Variance Line
hold on

plot(sigma_M, mu_M, '.'); % Market Portfolio
hold on

plot(sigma_CML, mu_CML, '-'); % Capital Market Line
xlabel('\sigma');
ylabel('\mu');

```

Exercises

- 20.1** (i) Show that the covariance matrix of the return rates $R^1, \dots, R^n$ of risky assets $S^1, \dots, S^n$ is invertible if and only if there is no redundant asset in the portfolio, in other words, no asset can be a linear combination of the other assets.
- (ii) Show that the invertibility condition given in (i) is equivalent to the positive definiteness of the covariance matrix.

20.2 Let X, Y be random variables, and a, b, c, d constants. Show that the following statements:

- (i) $\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)$.
- (ii) For $ac \neq 0$, $\rho(aX + b, cY + d) = \text{sgn}(ac)\rho(X, Y)$ where $\text{sgn}(x) = \frac{x}{|x|}$ for $x \neq 0$.
- (iii) Let X, Y be discrete random variables, and $a \neq 0$. Then $\rho(X, Y) = \text{sgn}(a)1$ if and only if $Y = aX + b$ with probability 1. What happens if X, Y are continuous random variables?

- 20.3** (i) Consider a portfolio of two assets, and the corresponding covariance matrix $C = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$. Show that the positive definiteness of C is equivalent to the condition that $\sigma_1^2 > 0$ and $\det C = (1 - \rho^2)\sigma_1^2\sigma_2^2 > 0$.
- (ii) Prove that if a real symmetric matrix A is positive definite, then its inverse is also positive definite.

20.4 Consider a hyperbola $x^2 - y^2 = 1$. Find the set of all points P such that there exists a tangent line to the hyperbola passing through P .

Chapter 21

Dynamic Programming



We investigate continuous time models for combined problems of optimal portfolio selection and consumption. An optimal strategy maximizes a given utility, and the solution depends on what to choose as a utility function. Under the assumption that a part of wealth is consumed, we find an explicit solution for optimal consumption and investment under certain assumptions. Utility increases as more wealth is spent, however, less is reinvested and the capability for future consumption decreases, therefore the total utility over the whole period under consideration may decrease. See Bjork (2009), Chang (2004), Merton (1969), Merton (1971), Rogers (2013) for more information.

In some of the previous chapters, we used the Lagrange multiplier method for optimization problems, however, in this chapter we employ the dynamic programming method to deal with stochastic control for optimization. The idea of dynamic programming is to optimize at every time step to achieve overall optimization.

21.1 The Hamilton–Jacobi–Bellman Equation

A classical method for finding minima or maxima is the calculus of variations, which is employed when the future event is deterministic. On the other hand, dynamic programming, which was introduced by the applied mathematician R. Bellman in the 1950s, results in an equation which is a stochastic version of the Hamilton–Jacobi equation in classical mechanics.

Given an Itô stochastic process $\{X_t\}_{0 \leq t \leq T}$, where X_t denotes wealth at time t , we consider an optimization problem defined by

$$\max_u \mathbb{E} \left[\int_0^T f(X_t, t, u_t) dt + H(X_T, T) \right].$$

The Riemann integral represents the total utility from time $t = 0$ to time T . The case that $T = \infty$ is also possible. The term $H(X_T, T)$ represents the utility of the remaining wealth X_T by the time T . The expectation is the average over all possibilities.

In this problem the given constraint is given by the Itô process

$$dX_t = \mu(X_t, t, u_t) dt + \sigma(X_t, t, u_t) dW_t \quad (21.1)$$

and the initial wealth $X_0 = x_0$. The stochastic process u_t is a control variable, in other words, optimal control is achieved by u_t based on the information X_t at time t . In general, the control u_t is of the form

$$u_t = g(X_t, t)$$

for some function $g(x, t)$. (When there is no danger of confusion we write u for simplicity. Even though $g(x, t)$ is a continuous function, it is not differentiable in general.)

In the above problem there might not exist a maximum over all possible choices for u , however the supremum always exists. In the problems presented in this chapter we assume that there exists a solution u for which the maximum exists. Thus we write maximum in place of supremum.

Instead of a rigorous derivation of the Hamilton–Jacobi–Bellman equation, which requires a lot of work, we present the idea in a heuristic manner. For arbitrary $0 \leq t \leq T$ define

$$J(X, t) = \max_u \mathbb{E} \left[\int_t^T f(X_s, s, u_s) ds + H(X_T, T) \middle| X_t = X \right],$$

which is the sum of the total utility from time t to time T under the condition $X_t = X$ and the final remaining value at T given by $H(X_T, T)$. By definition,

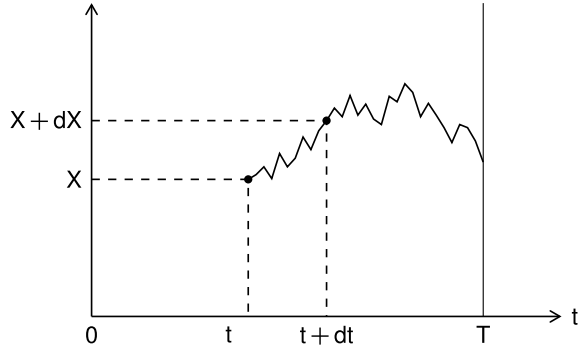
$$J(X, T) = H(X, T).$$

Now we derive a necessary condition under the assumption that the solution of a given optimization problem exists. Instead of presenting a rigorous proof we employ a heuristic method with the assumption that the given functions are sufficiently smooth as needed.

We assume that the solution of an optimization problem over the whole interval is optimal over each subinterval. Hence we have

$$J(X, t) = \max_u \mathbb{E} \left[\int_t^{t+\delta t} f(X_s, s, u_s) ds + J(X + \delta X, t + \delta t) \middle| X_t = X \right], \quad (21.2)$$

where $X + \delta X$, under the condition that $X_t = X$ at time t , denotes $X_{t+\delta t}$ at time $t + \delta t$, which is given by (21.1). See Fig. 21.1. Here the integral is approximated by

Fig. 21.1 Partition of time interval for total utility

$$\int_t^{t+\delta t} f(X_s, s, u_s) ds \approx f(X_t, t, u_t) \delta t. \quad (21.3)$$

By the Itô formula we have

$$\begin{aligned} & J(X + \delta X, t + \delta t) \\ &= J(X, t) + \frac{\partial J(X, t)}{\partial X} \delta X + \frac{\partial J(X, t)}{\partial t} \delta t + \frac{1}{2} \frac{\partial^2 J(X, t)}{\partial X^2} \sigma^2(X, t, u_t) \delta t + o(\delta t) \\ &= J(X, t) + \frac{\partial J(X, t)}{\partial X} \mu(X, t, u) \delta t + \frac{\partial J(X, t)}{\partial X} \sigma(X, t, u_t) \delta W \\ &\quad + \frac{\partial J(X, t)}{\partial t} \delta t + \frac{1}{2} \frac{\partial^2 J(X, t)}{\partial X^2} \sigma^2(X, t, u) \delta t + o(\delta t). \end{aligned} \quad (21.4)$$

Since the conditional expectation of the increment of Brownian motion is 0, we have

$$\mathbb{E} \left[\frac{\partial J}{\partial X} \sigma(X, t, u_t) \delta W \middle| X_t \right] = \mathbb{E} \left[\frac{\partial J}{\partial X} \sigma(X, t, u_t) \middle| X_t \right] \times \mathbb{E} [\delta W | X_t] = 0. \quad (21.5)$$

Using (21.3), (21.4) and (21.5), we compute the right-hand side of (21.2), and obtain

$$\begin{aligned} J(X, t) = \max_u \mathbb{E} \left[& f(X, t, u_t) \delta t + J(X, t) + \frac{\partial J(X, t)}{\partial X} \mu(X, t, u_t) \delta t \right. \\ & + \frac{\partial J(X, t)}{\partial X} \sigma(X, t, u_t) \delta W + \frac{\partial J(X, t)}{\partial t} \delta t \\ & \left. + \frac{1}{2} \frac{\partial^2 J(X, t)}{\partial X^2} \sigma^2(X, t, u) \delta t + o(\delta t) \middle| X_t = X \right] \end{aligned}$$

$$\begin{aligned}
&= \max_u \left(f(X, t, u_t) \delta t + J(X, t) + \frac{\partial J(X, t)}{\partial X} \mu(X, t, u_t) \delta t \right. \\
&\quad \left. + \frac{\partial J(X, t)}{\partial t} \delta t + \frac{1}{2} \frac{\partial^2 J(X, t)}{\partial X^2} \sigma^2(X, t, u_t) \delta t + o(\delta t) \right).
\end{aligned}$$

After subtracting $J(X, t)$ from both sides, then dividing by δt , and finally letting $\delta t \rightarrow 0$, we obtain the following conclusion.

Theorem 21.1 (Hamilton–Jacobi–Bellman Equation) *Suppose that a stochastic process X_t , $t \geq 0$, satisfies*

$$dX_t = \mu(X_t, t, u_t) dt + \sigma(X_t, t, u_t) dW_t.$$

To solve an optimization problem

$$\max_u \mathbb{E} \left[\int_0^T f(X_s, s, u_s) ds + H(X_T, T) \middle| X_t = x_0 \right]$$

define

$$J(X, t) = \max_u \mathbb{E} \left[\int_t^T f(X_s, s, u_s) ds + H(X_T, T) \middle| X_t = X \right].$$

Then J satisfies

$$\begin{aligned}
& - \frac{\partial J}{\partial t}(X_t, t) \\
&= \max_u \left(f(X_t, t, u_t) + \frac{\partial J}{\partial x}(X_t, t) \mu(X_t, t, u_t) + \frac{1}{2} \frac{\partial^2 J}{\partial x^2}(X_t, t) \sigma^2(X_t, t, u_t) \right)
\end{aligned}$$

with boundary condition

$$J(X_T, T) = H(X_T, T).$$

Remark 21.1 (i) In general the necessary condition given in the above is a sufficient condition.

(ii) If an optimization problem is defined for the infinite horizon, i.e., $T = \infty$, then in general the condition $B = 0$ is given, and we find a solution satisfying $\lim_{t \rightarrow \infty} J(X, t) = 0$.

21.2 Portfolio Management for Optimal Consumption

Utility is a measure of satisfaction from the consumption of goods and services, and marginal utility is the amount of satisfaction from the consumption of an additional unit of goods and services. The marginal utility decreases as the quantity of consumed goods increases in general. Let $U(c)$ denote the utility function that measures the

amount of utility where $c \geq 0$ is the consumption rate. That is, $U(c)\delta t$ denotes the total utility from the consumption of $c \delta t$ over a short time span δt . Since more consumption means more satisfaction in general, U is an increasing function.

By the law of diminishing marginal utility, U is a concave function. For example, for some constant $0 < \gamma < 1$ we have

$$U(c) = c^\gamma$$

or

$$U(c) = \log c.$$

The marginal utility is represented as U' , and the law of diminishing marginal utility may be expressed as $U'' < 0$.

First, we assume that the price equation for a risky asset is given by a geometric Brownian motion $dS_t = \alpha S_t dt + \sigma S_t dW_t$. (The symbol μ was already used to define X , and we use a new symbol α to define the asset price S .)

Let X_t denote the value of a portfolio at time t and let u_t be the ratio of investment into a risky asset. Suppose that $c_t dt$ is consumed for a time period of length dt . Assume that the portfolio is self-financing and there is no inflow of new investment from outside. To exclude the trivial case when one consumes as he/she wishes so that utility becomes very large, but the total wealth soon falls below zero, we impose a condition that $X_t \geq 0$. Hence the number of shares of a risky asset is equal to $u_t X_t / S_t$, and the amount invested in the risk-free asset is $(1 - u_t)X_t$. The increment of wealth after time length dt is equal to

$$\begin{aligned} dX_t &= \frac{u_t X_t}{S_t} dS_t + r(1 - u_t)X_t dt - c_t dt \\ &= (au_t X_t + r(1 - u_t)X_t - c_t) dt + \sigma u_t X_t dW_t. \end{aligned}$$

In the right-hand side of the first line $(u_t X_t / S_t) dS_t$ represents the increment of the total stock price, and the second term $r(1 - u_t)X_t dt$ is the interest on bank deposit. The term $-c_t dt$ representing consumption has a negative coefficient since wealth decreases due to consumption.

Now it remains to find suitable u_t and $c_t \geq 0$ to maximize utility. For some constant $\rho > 0$ we consider $e^{-\rho t} U(c_t)$ to discount the future value. In the following problem we take $B = 0$. Consult (Merton, 1969) or (Merton, 1990) for more general cases.

Theorem 21.2 (Power Utility) *For $0 < T < \infty$, we define the maximal utility over the period $0 \leq t \leq T$ by*

$$\max_{u, c} \mathbb{E} \left[\int_0^T e^{-\rho t} U(c_t) dt \right].$$

If there exists $0 < \gamma < 1$ such that $U(c) = c^\gamma$, then we have

$$u_t = \frac{\alpha - r}{\sigma^2(1 - \gamma)} = u^*, \quad (21.6)$$

$$c_t = h(t)^{1/(\gamma-1)} X_t, \quad (21.7)$$

where

$$\begin{aligned} A &= r\gamma + \frac{1}{2} \frac{\gamma}{1 - \gamma} \frac{(\alpha - r)^2}{\sigma^2} - \rho, \\ B &= 1 - \gamma, \\ h(t) &= \left\{ \frac{B}{A} \left(e^{\frac{1}{1-\gamma} A(T-t)} - 1 \right) \right\}^{1-\gamma}. \end{aligned}$$

The maximal utility over the period $0 \leq t \leq T$ is given by

$$\left\{ \frac{B}{A} \left(e^{\frac{1}{1-\gamma} AT} - 1 \right) \right\}^{1-\gamma} X_0^\gamma.$$

Proof For $0 \leq t \leq T$ define

$$J(x, t) = \max_{u, c} \mathbb{E} \left[\int_t^T e^{-\rho s} U(c_s) ds \middle| X_t = x \right].$$

In the Hamilton–Jacobi–Bellman equation we take $f(t, c) = e^{-\rho t} c^\gamma$ and need to maximize

$$\Gamma(u, c) \doteq e^{-\rho t} c^\gamma + \{ux(\alpha - r) + (rx - c)\} \frac{\partial J}{\partial x} + \frac{1}{2} \sigma^2 u^2 x^2 \frac{\partial^2 J}{\partial x^2}$$

by Theorem 21.1. If a maximum is achieved at each fixed interior point (x, t) , then we have

$$\begin{cases} 0 = \frac{\partial \Gamma}{\partial c} = e^{-\rho t} \gamma c^{\gamma-1} - \frac{\partial J}{\partial x} \\ 0 = \frac{\partial \Gamma}{\partial u} = x(\alpha - r) \frac{\partial J}{\partial x} + \sigma^2 u x^2 \frac{\partial^2 J}{\partial x^2}. \end{cases}$$

Hence c and u satisfy the conditions

$$\begin{cases} c^{\gamma-1} = \frac{1}{\gamma} e^{\rho t} \frac{\partial J}{\partial x} \\ u = \frac{\alpha - r}{\sigma^2} \left(-\frac{\partial J}{\partial x} \right) \left(x \frac{\partial^2 J}{\partial x^2} \right)^{-1}. \end{cases}$$

Now we look for J of the form

$$J(x, t) = e^{-\rho t} h(t) x^\gamma.$$

From the final condition $J(x, T) = 0$ we obtain the boundary condition

$$h(T) = 0.$$

Note that

$$\begin{cases} \frac{\partial J}{\partial t} = e^{-\rho t} h'(t) x^\gamma - \rho e^{-\rho t} h(t) x^\gamma \\ \frac{\partial J}{\partial x} = \gamma e^{-\rho t} h(t) x^{\gamma-1} \\ \frac{\partial^2 J}{\partial x^2} = \gamma(\gamma-1) e^{-\rho t} h(t) x^{\gamma-2}. \end{cases}$$

Hence

$$u = \frac{\alpha - r}{\sigma^2} \frac{(-1)\gamma e^{-\rho t} h(t) x^{\gamma-1}}{\gamma(\gamma-1) e^{-\rho t} h(t) x^{\gamma-1}} = \frac{1}{1-\gamma} \frac{\alpha - r}{\sigma^2}.$$

On the other hand,

$$c^{\gamma-1} = \frac{1}{\gamma} e^{\rho t} \gamma e^{-\rho t} h(t) x^{\gamma-1} = h(t) x^{\gamma-1},$$

and hence

$$c = h(t)^{1/(\gamma-1)} x.$$

To find h we substitute the preceding results into the Hamilton–Jacobi–Bellman equation and obtain

$$\begin{aligned} 0 &= \frac{\partial J}{\partial t} + e^{-\rho t} c^\gamma + \{ux(\alpha - r) + (rx - c)\} \frac{\partial J}{\partial x} + \frac{1}{2} \sigma^2 u^2 x^2 \frac{\partial^2 J}{\partial x^2} \\ &= (e^{-\rho t} h'(t) x^\gamma - \rho e^{-\rho t} h(t) x^\gamma) + e^{-\rho t} h(t)^{\frac{\gamma}{\gamma-1}} x^\gamma \\ &\quad + \left\{ \frac{1}{1-\gamma} \frac{(\alpha - r)^2}{\sigma^2} x + (rx - h(t)^{\frac{1}{\gamma-1}} x) \right\} \gamma e^{-\rho t} h(t) x^{\gamma-1} \\ &\quad + \frac{1}{2} \sigma^2 \frac{1}{(1-\gamma)^2} \frac{(\alpha - r)^2}{\sigma^4} x^2 \gamma(\gamma-1) e^{-\rho t} h(t) x^{\gamma-2} \\ &= (e^{-\rho t} h'(t) x^\gamma - \rho e^{-\rho t} h(t) x^\gamma) + e^{-\rho t} h(t)^{\frac{\gamma}{\gamma-1}} x^\gamma \\ &\quad + \left\{ \frac{1}{1-\gamma} \frac{(\alpha - r)^2}{\sigma^2} + r - h(t)^{\frac{1}{\gamma-1}} \right\} \gamma e^{-\rho t} h(t) x^\gamma \\ &\quad + \frac{1}{2} \frac{-1}{(1-\gamma)} \frac{(\alpha - r)^2}{\sigma^2} \gamma e^{-\rho t} h(t) x^\gamma. \end{aligned}$$

After division by $e^{-\rho t} x^\gamma$, the right-hand side becomes

$$h'(t) - \rho h(t) + h(t)^{\frac{\gamma}{\gamma-1}} + \left(r - h(t)^{\frac{1}{\gamma-1}} \right) \gamma h(t) + \frac{1}{2} \frac{1}{(1-\gamma)} \frac{(\alpha - r)^2}{\sigma^2} \gamma h(t).$$

Hence we have

$$h'(t) + Ah(t) + Bh(t)^{-\frac{\gamma}{1-\gamma}} = 0,$$

where the constants A and B are given by

$$A = r\gamma + \frac{1}{2} \frac{\gamma}{1-\gamma} \frac{(\alpha - r)^2}{\sigma^2} - \rho,$$

$$B = 1 - \gamma.$$

To apply the result in Appendix D.2 we define a new variable

$$y(t) = h(t)^{\frac{1}{1-\gamma}},$$

then we obtain a linear differential equation

$$y'(t) + \frac{1}{1-\gamma} Ay(t) + \frac{1}{1-\gamma} B = 0.$$

Using an integrating factor $M(t) = e^{\frac{1}{1-\gamma} At}$, we obtain

$$e^{-\frac{1}{1-\gamma} At} y(t) + \frac{B}{A} = C.$$

From the boundary condition $y(T) = 0$, we find

$$C = \frac{B}{A} e^{\frac{1}{1-\gamma} AT},$$

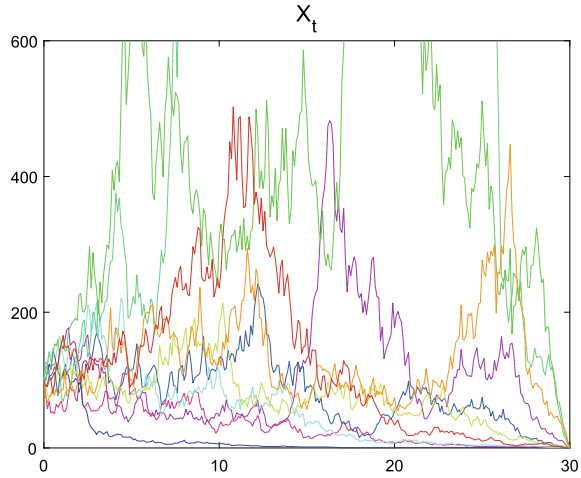
and obtain

$$y(t) = -\frac{B}{A} + Ce^{-\frac{1}{1-\gamma} At} = \frac{B}{A} \left(e^{\frac{1}{1-\gamma} A(T-t)} - 1 \right).$$

Remark 21.2 (i) In Fig. 21.2, we plot ten sample paths of the optimal wealth process X_t , $X_0 = 100$, $0 \leq t \leq T = 30$, $\alpha = 0.1$, $r = 0.05$, $\sigma = 0.3$, $\gamma = 0.5$ and $\rho = 0.06$. In this case, $A = -0.0211$, $B = 0.5000$. We see that $X_t \rightarrow 0$ as $t \rightarrow T$, which agrees with common sense that there should be no remaining wealth after time T to maximize utility over the period $0 \leq t \leq T$.

- (ii) An investor invests wealth into a risky asset with a constant rate u^* , which is proportional to $\alpha - r$, as can be seen in (21.6). For our choice of parameters, $u^* = 1.1111$.
- (iii) As stock price can go up more on average, i.e., with larger values of α , we have to invest more in stocks, as can be seen in (21.6).
- (iv) For larger volatility σ , we should invest less in stocks, as can be seen in (21.6).
- (v) With higher interest rate r , less amount should be invested in stock while saving more in the bank account, as can be seen in (21.6).

Fig. 21.2 Sample paths of optimal portfolio X_t with power utility for $0 \leq t \leq T$

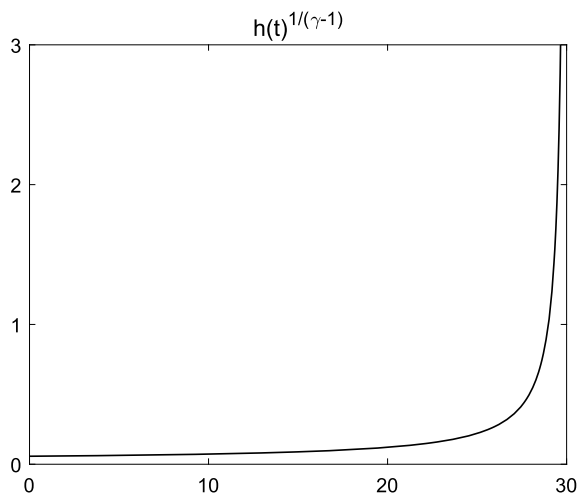


- (vi) An investor spends in proportion to the present wealth X_t . Even though the proportion coefficient $h(t)^{1/(\gamma-1)}$ in (21.7) increases to $+\infty$ as $t \rightarrow T$, the consumption rate

$$c_t = h(t)^{1/(\gamma-1)} X_t$$

can converge to a finite value since $X_t \rightarrow 0$ as $t \rightarrow T$. See Figs. 21.3 and 21.4. (Consult Simulation 21.1.) The wealth process tends to vary widely due to the fact that there is no wage and the income of the investor is generated only by return on stock investment.

Fig. 21.3 The proportion of spending $h(t)^{1/(\gamma-1)}$ for $0 \leq t \leq 30$



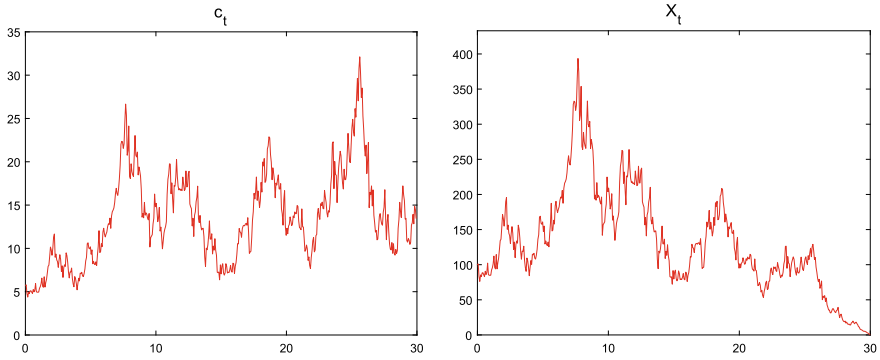


Fig. 21.4 Optimal consumption and portfolio for $0 \leq t \leq 30$

Theorem 21.3 (Log Utility) *Over the period $0 \leq t < \infty$ we define the maximum utility by*

$$\max_{u,c} \mathbb{E} \left[\int_0^\infty e^{-\rho t} U(c_t) dt \right].$$

If $U(c) = \log c$, then the optimal strategy is given by

$$u_t = \frac{\alpha - r}{\sigma^2}, \quad c_t = \rho X_t.$$

The maximum utility over the period $0 \leq t < \infty$ is given by $A \log X_0 + B$, where

$$A = \frac{1}{\rho},$$

$$B = \frac{1}{\rho} \left(\log \rho + \frac{r}{\rho} + \frac{(\alpha - r)^2}{2\rho\sigma^2} - 1 \right).$$

For the sample paths for the optimal portfolio X_t , see Fig. 21.5. Since it is impossible to do a simulation for the infinite time horizon problem, we take sufficiently large T for our simulation. In Simulation 21.2 we choose $T = 100$, which is justified by the fact that the estimate given by the Monte Carlo method with $T = 100$ is close to the closed-form solution.

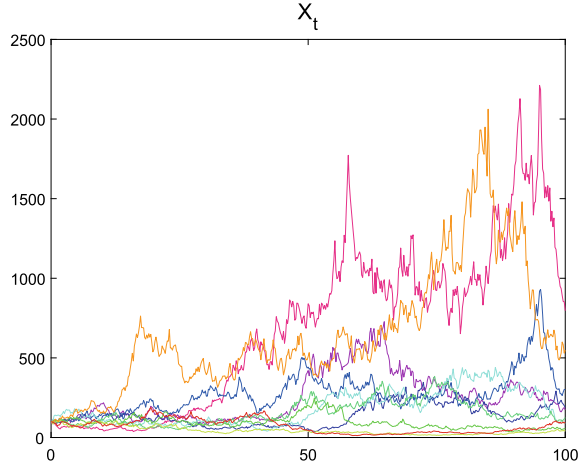
Proof For $t \geq 0$ we define

$$J(x, t) = \max_{u,c} \mathbb{E} \left[\int_t^\infty e^{-\rho s} U(c_s) ds \mid X_t = x \right].$$

Take

$$f(t, c) = e^{-\rho t} \log c$$

Fig. 21.5 Optimal portfolio
 X_t with log utility for
 $0 \leq t < \infty$



in the Hamilton–Jacobi–Bellman equation, and apply Theorem 21.1 to find the maximum of

$$\Gamma(u, c) \doteq e^{-\rho t} \log c + \{ux(\alpha - r) + rx - c\} \frac{\partial J}{\partial x} + \frac{1}{2} \sigma^2 u^2 x^2 \frac{\partial^2 J}{\partial x^2}.$$

Assuming the existence of a maximum for each fixed interior point (x, t) , we obtain the following two equations:

$$\begin{cases} 0 = \frac{\partial \Gamma}{\partial c} = e^{-\rho t} \frac{1}{c} - \frac{\partial J}{\partial x} \\ 0 = \frac{\partial \Gamma}{\partial u} = x(\alpha - r) \frac{\partial J}{\partial x} + \sigma^2 u x^2 \frac{\partial^2 J}{\partial x^2}. \end{cases}$$

Hence c and u satisfy the conditions

$$\begin{cases} c = e^{-\rho t} \left(\frac{\partial J}{\partial x} \right)^{-1} \\ u = \frac{\alpha - r}{\sigma^2} \left(-\frac{\partial J}{\partial x} \right) \left(x \frac{\partial^2 J}{\partial x^2} \right)^{-1}. \end{cases}$$

Now we look for J of the form

$$J(x, t) = e^{-\rho t} (A \log x + B).$$

By taking partial derivatives, we obtain

$$\begin{cases} \frac{\partial J}{\partial t} = -\rho e^{-\rho t} (A \log x + B) \\ \frac{\partial J}{\partial x} = e^{-\rho t} A x^{-1} \\ \frac{\partial^2 J}{\partial x^2} = -e^{-\rho t} A x^{-2}. \end{cases}$$

Hence

$$\begin{cases} c = A^{-1}x, \\ u = \frac{\alpha - r}{\sigma^2} (-e^{-\rho t} A x^{-1}) (-e^{-\rho t} A x^{-1})^{-1} = \frac{\alpha - r}{\sigma^2}. \end{cases}$$

To find constants A and B , we substitute the previous results into the Hamilton–Jacobi–Bellman equation and obtain

$$\begin{aligned} 0 &= \frac{\partial J}{\partial t} + e^{-\rho t} \log c + \{ux(\alpha - r) + (rx - c)\} \frac{\partial J}{\partial x} + \frac{1}{2} \sigma^2 u^2 x^2 \frac{\partial^2 J}{\partial x^2} \\ &= -\rho e^{-\rho t} (A \log x + B) + e^{-\rho t} (-\log A + \log x) \\ &\quad + \left(\frac{(\alpha - r)^2}{\sigma^2} x + (rx - A^{-1}x) \right) e^{-\rho t} A x^{-1} - \frac{1}{2} \sigma^2 \frac{(\alpha - r)^2}{\sigma^4} x^2 e^{-\rho t} A x^{-2} \\ &= -\rho e^{-\rho t} (A \log x + B) + e^{-\rho t} (-\log A + \log x) \\ &\quad + \left(\frac{1}{2} \frac{(\alpha - r)^2}{\sigma^2} + (r - A^{-1}) \right) e^{-\rho t} A. \end{aligned}$$

Dividing by $e^{-\rho t}$ we obtain

$$0 = -\rho (A \log x + B) + (-\log A + \log x) + \left(\frac{1}{2} \frac{(\alpha - r)^2}{\sigma^2} + (r - A^{-1}) \right) A.$$

By comparing the coefficients of $\log x$, we obtain $A = \frac{1}{\rho}$. Thus

$$0 = -\rho B + \log \rho + \left(\frac{1}{2} \frac{(\alpha - r)^2}{\sigma^2} + (r - \rho) \right) \frac{1}{\rho}.$$

For a more general example of a utility function, see spsciteKimOmberg.

21.3 Computer Experiments

Simulation 21.1 (Power Utility) We compare the Monte Carlo method with the closed-form solution for the utility maximization problem with $U(c) = \sqrt{c}$ in Theorem 21.2.

```

T = 30;
N = 500;
dt = T/N;
M = 10^5; % sample size for Monte Carlo
X0 = 100; %initial portfolio

alpha = 0.1;
r = 0.05;
sigma = 0.3;
gamma = 0.5;
rho = 0.06;

u = (alpha-r)/(sigma^2)/(1-gamma) % optimal investment ratio
A = r*gamma + 0.5*(gamma/(1-gamma))*( (alpha-r)^2/sigma^2 ) - rho
B = 1 - gamma

h = zeros(1,N+1);
for i=1:N+1
    h(1,i) = ( (B/A)*(exp(A*(T-(i-1)*dt)/(1-gamma))-1) )^(1-gamma);
end
X = zeros(M,N+1);
c = zeros(M,N+1);
X(:,1) = X0;
c(:,1) = h(1,1)^(1/(gamma-1)) * X(:,1);

dW = sqrt(dt)*randn(M,N);
for j=1:N
    X(:,j+1)=X(:,j)+(alpha*u*X(:,j)+r*(1-u)*X(:,j)-c(:,j))*dt ...
                + sigma*u*X(:,j).*dW(:,j);
    c(:,j+1)=h(j+1)^(1/(gamma-1)) * X(:,j+1);
end

integral = zeros(M,1);
for j=1:N
    integral(:) = integral(:) + exp(-rho*(j-1)*dt) * c(:,j).^gamma * dt;
end

plot(0:dt:T,h(1,1:N+1).^(1/(gamma-1)),'k');
```

```

plot(0:dt:T,c(1,1:N+1),'k');
plot(0:dt:T,X(1,1:N+1),'k');

max_utility_MC = mean(integral) % by Monte Carlo
max_utility_formula = (((B/A)*(exp(A*T/(1-gamma))-1))^(1-gamma))*(X0^gamma)
                    % by the closed-form formula

```

For the plots, see Figs. 21.3 and 21.4. The Monte Carlo method and the closed-form formula give their results as follows:

```

max_utility_MC = 41.3021
max_utility_formula = 41.2441

```

Simulation 21.2 (Log Utility) We compare the Monte Carlo method with the closed-form solution for the utility maximization problem with $U(c) = \log c$ in Theorem 21.3. Since it is impossible to do a simulation for an infinite horizon problem, we take $T = 100$. We take the same set of parameters as in Simulation 21.1. The estimate given by the Monte Carlo method agrees with the theoretical value with an acceptable error as one can see at the end of the paragraph. For the plot of the sample paths, see Fig. 21.5.

```

u = (alpha - r)/(sigma^2) % optimal investment ratio

A = 1/rho;
B = (log(rho) + r/rho + (alpha - r)^2/(2*rho*sigma^2) - 1)/rho;

X = zeros(M,N+1);
c = zeros(M,N+1);
X0 = 100;           % initial portfolio
X(:,1) = X0;
c(:,1) = rho * X(1);

dW = sqrt(dt)*randn(M,N);
for j=1:N
    X(:,j+1)=X(:,j)+(alpha*u*X(:,j)+r*(1-u)*X(:,j)-c(:,j))*dt ...
                +sigma*u*X(:,j).*dW(:,j);
    c(:,j+1)=rho*X(:,j+1);
end

integral = zeros(M,1);
for j=1:N
    integral(:) = integral(:) + exp(-rho*(j-1)*dt)*log(c(:,j))*dt;
end

```

```
max_utility_MC = mean(integral)      % by Monte Carlo  
max_utility_formula = A*log(X0) + B % by the closed-form formula
```

The result of the Monte Carlo method and the solution by the closed-form formula are as follows:

```
max_utility_MC = 30.9302  
max_utility_formula = 30.9429
```

Part VIII

Interest Rate Models

Chapter 22

Bond Pricing



In this chapter we introduce concepts such as coupons of a bond, yield to maturity, duration and convexity. We derive the fundamental equation for bond pricing under a general assumption on interest rate movements. For an introduction to interest rate models see Capinski and Zastawniak (2003), Gianin and Sgarra (2013), Pascucci and Runggaldier (2012) and for a comprehensive introduction to bonds consult (Veronesi, 2010).

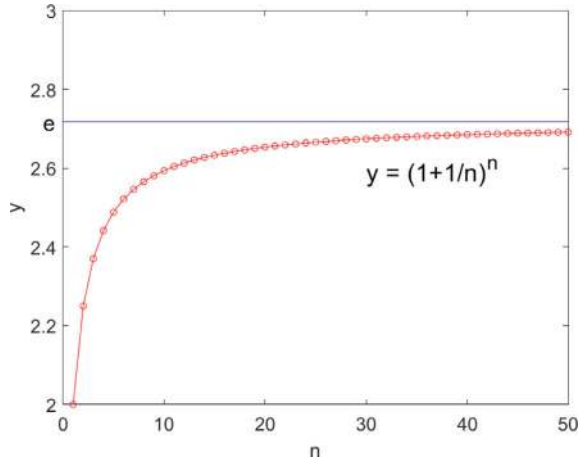
22.1 Periodic and Continuous Compounding

Definition 22.1 The *discount factor* $Z(t, T)$ is defined to be the rate of exchange between a given amount of money at time t and an amount of money at a later date T .

For example, if one hundred dollars at a future date T is worth ninety two dollars and fifty cents today ($t = 0$), then $Z(0, T) = 0.925$. *Compounded interest* means that there is interest on interest while *simple interest* means no interest on interest. A *compounding frequency* of interest accruals is the number of times per year when interest is paid. Throughout the book we consider only compounded interest rates. Suppose that we have the information that the interest rate per year is R without knowing the compounding method. Under the assumption of the periodic compounding let m be the compounding frequency, i.e., the interest is compounded m times per year. In T years the value of the investment of a principal A will be equal to $A[(1 + \frac{R}{m})^m]^T$. Since

$$A = Z(0, T) \times A \left[\left(1 + \frac{R}{m} \right)^m \right]^T,$$

Fig. 22.1 Monotone convergence of $(1 + 1/n)^n$ to the limit $e = 2.71828\dots$ as $n \rightarrow \infty$



we have

$$Z(0, T) = \left(1 + \frac{R}{m}\right)^{-mT}.$$

As $m \rightarrow \infty$ we compound more and more frequently, and obtain

$$\lim_{m \rightarrow \infty} A \left[\left(1 + \frac{R}{m}\right)^m \right]^T = A e^{RT}.$$

Since $(1 + \frac{R}{m})^m$ is monotonically increasing and converges to e^R as $m \rightarrow \infty$ for $R > 0$, a depositor prefers continuous compounding to periodic compounding if the interest rate is positive. (See Fig. 22.1 and Simulation 22.1). When $\$A$ is invested at a constant continuously compounded rate R , then $\$A$ grows to $e^{RT} \times \$A$ at time T , or equivalently, $\$A$ to be received at T is discounted to $e^{-RT} \times \$A$ at $t = 0$, and hence the discount factor for continuous compounding is equal to $Z(0, T) = e^{-RT}$.

Remark 22.1 (Periodic and continuous compounding) Let R_m denote the interest rate periodically compounding m times per year, e.g., $m = 2$ corresponds to semi-annual compounding, and let R_c be the interest rate with continuous compounding. Suppose that we invest the initial amount A and the two compounding methods will produce the same amount at the end of n years. Since

$$A e^{R_c n} = A \left[\left(1 + \frac{R_m}{m}\right)^m \right]^n,$$

we have $e^{R_c} = (1 + \frac{R_m}{m})^m$. Hence $R_c = m \log(1 + \frac{R_m}{m})$ and $R_m = m(e^{R_c/m} - 1)$.

Remark 22.2 The discount factor $Z(t, T)$ is the time t price $P(t, T)$ of a zero coupon bond that pays \$1 with maturity T . In general, the curve $T \mapsto P(t, T)$ is assumed to be sufficiently smooth while $t \mapsto P(t, T)$ is regarded as stochastic.

22.2 Zero Coupon Interest Rates

Definition 22.2 Some bonds pay interest, called a *coupon*, periodically. They are called *coupon bonds*, or *fixed income securities*, because they pay a fixed amount of interest at regular intervals. If a bond does not pay any interest before maturity and provides a payoff only at time T , it is called a zero coupon bond. For a bond without coupon payments the interest and principal is realized at maturity. A *zero rate* (or *spot rate*), is the rate of interest earned on a zero coupon bond. See Definition 22.8 (i) for the details.

Pricing a bond with coupon payments is not easy because a portion of interest is paid in the form of coupons before maturity. Thus we discount each coupon payment at the corresponding zero rate, which is obtained by the *bootstrap method* described as follows:

Consider the bonds traded in the bond market given in Table 22.1. For bonds with time to maturity equal to 1.5 years and 2 years, half of the annual coupon amount is paid every six months, i.e., compounding frequency is given by $m = 2$.

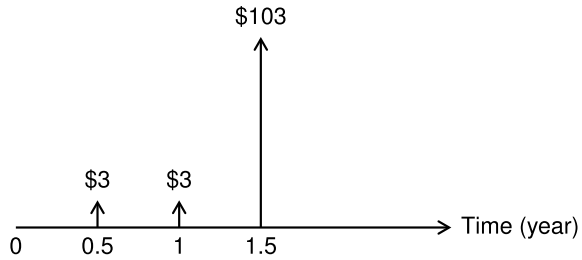
The first three bonds pay no coupons before maturity, and it is easy to calculate their zero rates.

- (i) Let us consider the first bond, which pays \$100, and we make the profit $\$100 - \$98.31 = \$1.69$, in three months. Note that $100 = A \left(1 + \frac{R_m}{m}\right)^{mn}$ with $A = 98.31$, $m = 4$, $n = 0.25$. Hence $100 = 98.31 \left(1 + \frac{R_4}{4}\right)$ and $R_4 = \frac{4 \times (100 - 98.31)}{98.31} = 6.876\%$. From the equation $A e^{R_c n} = A \left(1 + \frac{R_m}{m}\right)^{mn}$, we have $e^{R_c \times 0.25} = 1 + \frac{R_4}{4}$ and $R_c = 6.818\%$.
- (ii) We do similarly for the next two bonds. Let us consider the second bond, which pays \$100, and we make the profit $\$100 - \$96.54 = \$3.46$, in six

Table 22.1 Market data for bonds

Principal (\$)	Time to maturity (years)	Annual coupon (\$)	Bond price (\$)	Compound. freq. (times per year)
100	0.25	0	98.31	4
100	0.50	0	96.54	2
100	1.00	0	93.14	1
100	1.50	6	98.10	2
100	2.00	10	104.62	2

Fig. 22.2 Cash flow of a coupon bond



months. Note that $100 = A \left(1 + \frac{R_m}{m}\right)^{mn}$ with $A = 96.54$, $m = 2$, $n = 0.5$. Hence $100 = 96.54 \left(1 + \frac{R_2}{2}\right)$ and $R_2 = \frac{2 \times (100 - 96.54)}{96.54} = 7.168\%$. From the equation $A e^{R_c n} = A \left(1 + \frac{R_m}{m}\right)^{mn}$, we have $e^{R_c \times 0.5} = 1 + \frac{R_2}{2}$ and $R_c = 7.043\%$.

- (iii) Let us consider the third bond, which pays \$100, and we make the profit $\$100 - \$93.14 = \$6.86$, in a year. Note that $100 = A \left(1 + \frac{R_m}{m}\right)^{mn}$ with $A = 93.14$, $m = 1$, $n = 1$. Hence $100 = 93.14 \left(1 + \frac{R_1}{1}\right)$ and $R_1 = \frac{100 - 93.14}{93.14} = 7.365\%$. From the equation $A e^{R_c n} = A \left(1 + \frac{R_m}{m}\right)^{mn}$, we have $e^{R_c \times 1} = 1 + \frac{R_1}{1}$ and $R_c = 7.107\%$. Using the results from (i), (ii) and (iii) we can present a partially completed table as Table 22.2. Now we consider the remaining cases.
- (iv) Consider the fourth bond, which has coupon payments of \$3 every six months. To calculate the cash price of a bond we discount each cash flow at the appropriate zero rate. The cash flow is given in Fig. 22.2. Now we use the earlier results given in Table 22.2. Note that

$$98.10 = 3e^{-0.07043 \times 0.5} + 3e^{-0.07107 \times 1.0} + 103e^{-R \times 1.5}.$$

Hence $R = 0.07233$, and we can fill a blank space in the previous table.

- (v) Consider the fifth bond, which has coupon payments of \$5 every six months. To calculate the cash price of a bond we discount each cash flow at the appropriate zero rate. The cash flow is given in Fig. 22.3. Now we use the previous results. Since

$$104.62 = 5e^{-0.07043 \times 0.5} + 5e^{-0.07107 \times 1.0} + 5e^{-0.07233 \times 1.5} + 105e^{-R \times 2.0},$$

we have $R = 0.07348$, and finally we obtain Table 22.3.

22.3 Term Structure of Interest Rates

The *term structure of interest rates*, or *spot rate curve*, or *yield curve* at time t is the relation between the level of interest rates and their time to maturity. Usually, long-term interest rates are greater than short term interest rates, and hence the yield

Fig. 22.3 Cash flow of a coupon bond

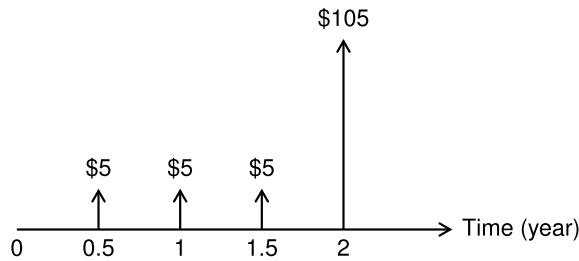


Table 22.2 Continuously compounded zero rates determined by Table 22.1

Maturity (years)	Zero rate (%)
0.25	6.818
0.50	7.043
1.00	7.107
1.50	
2.00	

Table 22.3 Continuously compounded zero rates determined by Table 22.1

Maturity (years)	Zero rate (%)
0.25	6.818
0.50	7.043
1.00	7.107
1.50	7.233
2.00	7.348

curve is upward moving. There are several theories for the term structure of interest rates.

1. Expectations Theory: The interest rate on a long-term bonds will be equal to an average of the short-term interest rates that people expect to occur over the life of the long-term bond. This theory does not explain why the yield curve is upward moving in most cases since the short-term interest rate moves in either direction in general.
2. Segmented Market Theory: Short, medium and long maturity bonds are separate and segmented, and their interest rates are determined by their own supply and demand, and move independently of each other. This theory does not explain why the interest rates on bonds of different maturities tend to move together.
3. Liquidity Premium Theory: The interest rate on a long-term bond will equal an average of short-term interest rates expected to occur over the life of the long-term bonds plus a liquidity premium. Since investors tends to prefer short-

Table 22.4 Present interest rates

Maturity(years)	Deposit rate (%)	Lending rate (%)
1	3	6
5	3	6

Table 22.5 New interest rates

Maturity(years)	Deposit rate (%)	Mortgage rate (%)
1	3	6
5	4	7

term bonds to avoid the liquidity risk, for a long-term bond a positive liquidity premium must be added.

Example 22.1 (*Management of Interest Rate*) Suppose that the interest rates posted by a bank is given in Table 22.4.

Depositors would choose to put their money in the bank only for one year because they want to have more financial flexibility. Otherwise longer term deposits would tie up the fund for a longer period of time. Now suppose that we want a mortgage. We would choose a five-year mortgage at 6% because it fixes the borrowing rate for the next five years and subjects us to less refinancing risk. Therefore, when the bank posts the rates shown in the above table, the majority of its customers would choose one-year deposits and five-year mortgages. This creates an asset/liability mismatch for the bank. There is no problem if interest rates fall. The bank will finance the five-year 6% loans with deposits that cost less than 3% in the future and net interest income will increase. However, if the rates rise, then the deposits that are financing these 6% loans will cost more than 3% in the future and net interest income will decline.

How can we ensure that the maturities of the assets on which interest is earned and the maturities of the liabilities on which interest is paid are matched? We can do it by increasing the five-year rate on both deposits and mortgages. For example, the bank can post new rates given in Table 22.5 where the five-year deposit rate is 4% and the five-year mortgage rate 7%. This would make five-year deposits relatively more attractive and one-year mortgages relatively more attractive. Some customers who chose one-year deposits when the rates were as in Table 22.4 will switch to five-year deposits, and some customers who chose five-year mortgages when the rates were as in Table 22.4 will choose one-year mortgages. This method may lead to the match of the maturities of assets and liabilities. If there is still an imbalance with depositors tending to choose a one-year maturity and borrowers a five-year maturity, five-year deposit and mortgage rates could be increased even further. Eventually the imbalance will disappear.

22.4 Forward Rates

Definition 22.3 The *forward rate* is the future zero rate implied by today's term structure of interest rates. See Definition 22.8(ii) for the details.

For example, consider two different strategies for bond investment. In the first strategy, we buy a bond with maturity equal to two years with interest rate $R_2 = 4.0\%$. In this case, an investment of \$1 will grow to $\$1 \times e^{R_2 \times 2}$ in two years. The second strategy is to buy a bond with maturity of one year which has interest rate $R_1 = 3.0\%$. Using the sum of the principal \$1 and the interest, which is equal to $\$1 \times e^{R_1 \times 1}$, we reinvest at time 1 year, and receive $\$1 \times e^{R_1 \times 1} e^{F_{12} \times (2-1)}$, where F_{12} is the forward rate that is predetermined today. To avoid arbitrage, two investment strategies should yield the same profit, i.e.,

$$e^{R_2 \times 2} = e^{R_1 \times 1} e^{F_{12} \times (2-1)}.$$

Hence $R_2 \times 2 = R_1 \times 1 + F_{12} \times (2 - 1)$. Therefore, we have $F_{12} = 5.0\%$. See Table 22.6.

Remark 22.3 Suppose that the zero rates for time periods T_1 and T_2 are R_1 and R_2 with both rates continuously compounded. The forward rate for the period between times T_1 and T_2 is

$$\frac{R_2 T_2 - R_1 T_1}{T_2 - T_1}.$$

A *forward rate agreement* (FRA) is an agreement that a certain rate, called the *forward rate*, will be exchanged for interest at the market rate to a notional amount during a certain future time period from T_1 to T_2 . The agreement is so structured that neither party needs to make an upfront payment, i.e., the value of the contract is zero when the agreement is entered.

Example 22.2 Suppose that a company has agreed with a bank at time $t = 0$ that it will receive the forward rate 4% on the notional amount \$100 million for 3 months starting in $T_1 = 3$ years. That is, $T_2 = T_1 + \frac{3}{12} = 3.25$. Suppose that the interest rate turns out to be 4.5% (with quarterly compounding) at T_1 . The company has to pay

$$\$125,000 = \$100 \text{ million} \times 0.5\% \times \frac{3}{12}$$

Table 22.6 Calculation of forward rates

Year (n)	Zero rate for an n-year investment (% per annum)	Forward rate for the nth year (% per annum)
1	$R_1 = 3.0$	
2	$R_2 = 4.0$	$F_{12} = 5.0$

to the bank at time T_2 , or equivalently,

$$\$123,630 = (1 + 0.045)^{-0.25} \times \$125,000$$

at T_1 .

22.5 Yield to Maturity

Consider a coupon bond that provides the holder with cash flows $c_1, \dots, c_n$ on dates $0 < t_1 < \dots < t_n = T$. (In most cases, c_i is a coupon payment of fixed amount c for $1 \leq i \leq n-1$, and c_n is a coupon payment plus the principal.) Let $Z(0, t_i)$ be the discount factors for each date t_i . Then the value B of the coupon bond is equal to

$$B = \sum_{i=1}^n c_i Z(0, t_i) . \quad (22.1)$$

Note that B is equal to the present value of the future cash flows.

Definition 22.4 The *yield to maturity*, or just YTM for short, is the single discount rate for all dates t_i that makes the present value of the future cash flows equal to the bond price. It is also called the *internal rate of return*.

(i) In the continuously compounding case, the yield y is defined by

$$Z(0, t_i) = e^{-yt_i} , \quad (22.2)$$

i.e.,

$$B = \sum_{i=1}^n c_i e^{-yt_i} \quad (22.3)$$

by (22.1). Since

$$\frac{d}{dy} B(y) = - \sum_{i=1}^n t_i c_i e^{-yt_i} < 0 ,$$

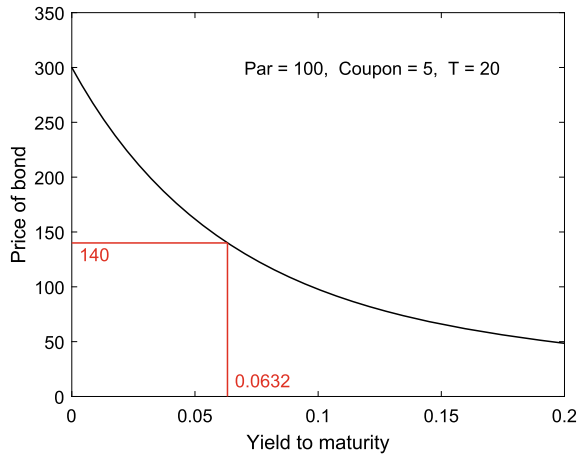
there exists a unique value for y satisfying (22.3). Since

$$\frac{d^2}{dy^2} B(y) = \sum_{i=1}^n t_i^2 c_i e^{-yt_i} > 0 ,$$

the graph of $B(y)$ is monotonically decreasing and convex.

(ii) In the periodic compounding case with the compounding frequency m , the yield is defined by

Fig. 22.4 Computation of the yield to maturity when the present value is given



$$Z(0, t_i) = \left(1 + \frac{y}{m}\right)^{-mt_i}, \quad (22.4)$$

i.e.,

$$B = \sum_{i=1}^n c_i \left(1 + \frac{y}{m}\right)^{-mt_i}. \quad (22.5)$$

The Eqs. (22.4) and (22.5) are discrete time versions of (22.2) and (22.3). Note that the periodically compounded yield converges to the continuously compounded yield.

Example 22.3 (*Continuous Compounding*) Consider a coupon bond, with face value $P = \$100$, that pays $C = \$5$ every six months until maturity $T = 20$. Suppose that its present value is $B = \$140$. What is the yield to maturity? We have to solve

$$B = \sum_{k=1}^{2T} C e^{-0.5ky} + P e^{-Ty} = C e^{-0.5y} \frac{1 - e^{-Ty}}{1 - e^{-0.5y}} + P e^{-Ty}.$$

Using the Newton–Raphson method we find $y = 0.0632$. See Fig. 22.4 and Simulation 22.2. The present bond price $B = \$140$ is represented by the horizontal line, and the vertical line the yield to maturity.

22.6 Duration

The duration of a bond is a measure of how long on average the holder of the bond has to wait before receiving cash payment. A zero-coupon bond that matures in n years has a duration of n years. However, a coupon bond maturing in n years has a

duration of less than n years, because the holder receives some of the payments prior to year n .

Definition 22.5 (*Duration for Continuous Compounding*) Consider a coupon bond that provides the holder with cash flows $c_1, \dots, c_n$ on dates $0 < t_1 < \dots < t_n = T$. Recall that the price B and the (continuously compounded) yield to maturity y satisfy

$$B(y) = \sum_{i=1}^n c_i e^{-yt_i}.$$

The *duration* D of the bond is defined by

$$D = -\frac{B'(y)}{B(y)} = \frac{1}{B} \sum_{i=1}^n t_i c_i e^{-yt_i}.$$

Remark 22.4 (i) Duration was introduced by F. Macaulay in 1938.
(ii) We may rewrite the duration as

$$D = \sum_{i=1}^n \frac{c_i e^{-yt_i}}{B} t_i,$$

which is the weighted average of the coupon payment dates t_i . Note that the sum of the weights equals 1 and that the i th weight is the proportion of the discounted cash flow at t_i .

(iii) When the bond yield increases, the bond price decreases.

Theorem 22.1 Let δy be a small change in the continuously compounded yield to maturity, and let δB the corresponding change in the bond price. Then

$$\frac{\delta B}{B} \approx -D \delta y.$$

Proof Note that

$$\delta B \approx B'(y) \delta y = \sum_{i=1}^n (-t_i) c_i e^{-yt_i} \delta y = -DB \delta y.$$

Example 22.4 Consider a three-year 8% coupon bond with a face value of \$100. Coupon payments of $\$100 \times \frac{0.08}{2} = \$4$ are made every six months. (There are six coupon payments before and on the maturity date $T = 3$.) Suppose that the yield to maturity is 7.5% per annum with continuous compounding, i.e., $y = 0.075$. We will show how to calculate the duration of the bond. First, the cash flow of the coupon bond is given in Table 22.7.

Then we fill the next columns for the present values of the cash flows, the weights, and the payment dates times the weights. The total of the last column in Table 22.8 equals the duration.

Table 22.7 Calculation of duration: a bond with coupons

Time (years)	Cash flow (\$)	Present value (\$)	Weight	Time $\times$ Weight (years)
$t_1 = 0.5$	$c_1 = 4$			
$t_2 = 1.0$	$c_2 = 4$			
$t_3 = 1.5$	$c_3 = 4$			
$t_4 = 2.0$	$c_4 = 4$			
$t_5 = 2.5$	$c_5 = 4$			
$t_6 = 3.0$	$c_6 = 104$			
Total	124			

Table 22.8 Calculation of duration

Time (years)	Cash flow (\$)	Present value (\$)	Weight	Time $\times$ Weight (years)
0.5	4	3.853	0.038	0.019
1.0	4	3.711	0.037	0.037
1.5	4	3.574	0.035	0.053
2.0	4	3.443	0.034	0.068
2.5	4	3.316	0.033	0.082
3.0	104	83.046	0.823	2.468
Total	124	100.943	1.000	$D = 2.727$

Definition 22.6 (*Duration for Periodic Compounding*) Let y be the yield to maturity for the periodic compounding case with a compounding frequency of m times per year. Then the *duration* D is defined by

$$D = \sum_i t_i \frac{c_i \left(1 + \frac{y}{m}\right)^{-mt_i}}{B}$$

and the *modified duration* D^* is defined by

$$D^* = \frac{D}{1 + \frac{y}{m}}.$$

Note that $D^* \rightarrow D$ as $m \rightarrow \infty$. Since

$$B'(y) = - \sum_i t_i c_i \left(1 + \frac{y}{m}\right)^{-mt_i-1},$$

we have

$$\delta B \approx -\frac{BD}{1 + \frac{y}{m}} \delta y = -BD^* \delta y .$$

In other words,

$$\frac{B'(y)}{B(y)} \approx -D^* \delta y .$$

Definition 22.7 (*Convexity*)

- (i) For a continuously compounded case, the *convexity* C of a bond with price B and yield to maturity y is defined by

$$C = \frac{B''(y)}{B(y)} = \frac{1}{B} \sum_i t_i^2 c_i e^{-yt_i} .$$

- (ii) For a periodically compounded case, the *convexity* is defined by

$$C^* = \frac{B''(y)}{B(y)} = \frac{1}{B} \sum_i t_i \left(t_i + \frac{1}{m} \right) c_i \left(1 + \frac{y}{m} \right)^{-mt_i} \frac{1}{\left(1 + \frac{y}{m} \right)^2} .$$

Theorem 22.2 (i) For a continuously compounded case, we have

$$\frac{\delta B}{B} \approx -D \delta y + \frac{1}{2} C \delta y^2 .$$

- (ii) For a periodically compounded case, we have

$$\frac{\delta B}{B} \approx -D^* \delta y + \frac{1}{2} C^* \delta y^2 .$$

Proof By the Taylor series expansion,

$$\delta B \approx B'(y) \delta y + \frac{1}{2} B''(y) \delta y^2 .$$

Dividing both sides by B , we obtain the desired result.

22.7 Definitions of Various Interest Rates

Let the present time be t , and let $P(t, T)$ and $P(t, S)$ denote the zero coupon bond prices with maturities T and S , $T < S$, respectively. Consider a bank that wants to lend money, say \$1, to a borrower for a future period $[T, S]$. Let the interest rate for

lending be $f(t, T, S)$, which is agreed on at time t . The bank shorts $\frac{P(t, T)}{P(t, S)}$ units of the S -maturity zero coupon bond, and long one unit of the T -maturity zero coupon bond. (The price of $\frac{P(t, T)}{P(t, S)}$ units of the S -maturity zero coupon bond at time t equals

$$\frac{P(t, T)}{P(t, S)} \times P(t, S) = P(t, T) ,$$

which is exactly equal to the cost of buying one unit of the T -maturity zero coupon bond. Hence there is no cash transfer.) At time T , the proceeds from the T -maturity bond are lent out for the period $[T, S]$ with the interest rate $f(t, T, S)$. At time S , the loan is paid back from the borrower and the short position of S -bond is recovered, i.e., the S -bond is returned by the bank, resulting in a net cash flow of

$$C = \$1 \times e^{f(t, T, S)(S-T)} - \$1 \times \frac{P(t, T)}{P(t, S)}$$

at time S . Since the bank did not need any money at time t , we have $C = 0$ by the no arbitrage principle. Thus

$$f(t, T, S) = \frac{1}{S - T} \log \frac{P(t, T)}{P(t, S)}$$

and

$$P(t, T)e^{-(S-T)f(t, T, S)} = P(t, S) .$$

Definition 22.8 (i) The *spot rate* $R(t, T)$ at t for maturity T is defined as the yield to maturity of the T -maturity bond, i.e.,

$$P(t, T) = e^{-(T-t)R(t, T)} .$$

Equivalently,

$$R(t, T) = -\frac{\log P(t, T)}{T - t} .$$

(ii) The *forward rate* $f(t, T, S)$ at t for $T < S$ is defined by

$$f(t, T, S) = \frac{1}{S - T} \log \frac{P(t, T)}{P(t, S)} .$$

Since $P(t, t) = 1$, taking $T = t$, we obtain

$$e^{-(S-t)f(t, t, S)} = P(t, S) ,$$

which in turn implies that

$$f(t, t, S) = R(t, S) .$$

- (iii) The *instantaneous forward rate* $f(t, T)$, over the interval $[T, T + dt]$, is defined by

$$f(t, T) = \lim_{S \downarrow T} f(t, T, S) .$$

Note that

$$\begin{aligned} f(t, T) &= \lim_{S \downarrow T} \frac{1}{S - T} \log \frac{P(t, T)}{P(t, S)} \\ &= - \lim_{S \downarrow T} \frac{\log P(t, S) - \log P(t, T)}{S - T} \\ &= - \frac{\partial}{\partial T} \log P(t, T) \\ &= - \frac{1}{P(t, T)} \frac{\partial}{\partial T} P(t, T) . \end{aligned}$$

Hence

$$P(t, T) = e^{-\int_t^T f(t, u) du}$$

and

$$R(t, T) = \frac{1}{T - t} \int_t^T f(t, u) du .$$

- (iv) Define the *short rate* $r(t)$, over the interval $[t, t + dt]$, by

$$r(t) = \lim_{T \downarrow t} R(t, T) = f(t, t) .$$

It is the interest rate for an infinitesimally short period of time.

22.8 The Fundamental Equation for Bond Pricing

A zero coupon bond will pay the investor a risk-free fixed amount at maturity, and it might be regarded as a riskless investment. However, if the interest rate r is stochastic, a bond is a risky asset since a change in the interest rate will cause a change in the present value of the pre-agreed payment at maturity. Especially when the investor wants to sell the bond before maturity, a sudden increase in the short-term interest rate may cause losses.

We assume that the short rate r_t satisfies a stochastic differential equation

$$dr_t = \mu(r_t, t) dt + \sigma(r_t, t) dW_t$$

for some $\mu(r, t)$ and $\sigma(r, t)$. Note that in bond pricing there is no underlying asset with which to hedge the risk. Therefore we construct a portfolio in which one bond is

hedged by another bond with different maturity in the derivation of the fundamental equation for bond pricing. The following result due to O. Vasicek relies on the Delta-hedging idea in the derivation of the Black–Scholes–Merton equation, and it is fundamental in pricing bonds and interest rate derivatives. See Vasicek (1977).

Theorem 22.3 (Fundamental Equation for Bond Pricing) *Let $V(r, t)$ denote the price of a bond with payoff equal to 1 at maturity date T , or an interest rate derivative that pays $g(r_T)$, where r_T denotes the interest rate at T . Then it satisfies*

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2} + m(r, t) \frac{\partial V}{\partial r} = rV \quad (22.6)$$

for some function $m(r, t)$. The final condition is given by $V(r, T) = 1$ for a bond, and $V(r, T) = g(r)$ for a derivative, respectively.

Proof Since a bond is an interest rate derivative with contingent claim given by $g(r_T) = 1$ at maturity, we prove the theorem for general derivatives. We consider two arbitrarily chosen derivatives, and construct a portfolio in which one derivative is hedged by another derivative with different maturity. Let $V_i(r, t)$ denote the price at time t of the i th derivative with maturity T_i where $t \leq \min\{T_1, T_2\}$. Set up a self-financing portfolio Π given by

$$\Pi = V_1 - \Delta V_2,$$

where Δ is the number of units of the second derivative that is shorted. Then the increment of Π after time increment dt is given by

$$d\Pi = \frac{\partial V_1}{\partial t}dt + \frac{\partial V_1}{\partial r}dr + \frac{1}{2}\sigma^2 \frac{\partial^2 V_1}{\partial r^2}dt - \Delta \left(\frac{\partial V_2}{\partial t}dt + \frac{\partial V_2}{\partial r}dr + \frac{1}{2}\sigma^2 \frac{\partial^2 V_2}{\partial r^2}dt \right)$$

by Itô's lemma. To make the coefficient of the risky term dr equal to zero, eliminating randomness in the portfolio, we take

$$\Delta = \frac{\partial V_1}{\partial r} / \frac{\partial V_2}{\partial r}. \quad (22.7)$$

Then

$$d\Pi = \left\{ \frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V_1}{\partial r^2} - \frac{\frac{\partial V_1}{\partial r}}{\frac{\partial V_2}{\partial r}} \left(\frac{\partial V_2}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V_2}{\partial r^2} \right) \right\} dt.$$

Since Π is risk-free, we have $d\Pi = r\Pi dt$, and hence

$$\frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V_1}{\partial r^2} - \frac{\frac{\partial V_1}{\partial r}}{\frac{\partial V_2}{\partial r}} \left(\frac{\partial V_2}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V_2}{\partial r^2} \right) = r \left(V_1 - \frac{\frac{\partial V_1}{\partial r}}{\frac{\partial V_2}{\partial r}} V_2 \right).$$

Hence

$$\frac{1}{\frac{\partial V_1}{\partial r}} \left(\frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V_1}{\partial r^2} \right) - \frac{1}{\frac{\partial V_2}{\partial r}} \left(\frac{\partial V_2}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V_2}{\partial r^2} \right) = r \left(\frac{1}{\frac{\partial V_1}{\partial r}} V_1 - \frac{1}{\frac{\partial V_2}{\partial r}} V_2 \right).$$

By collecting V_1 terms on the left-hand side and V_2 terms on the right, we obtain

$$\frac{\frac{\partial V_1}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V_1}{\partial r^2} - r V_1}{\frac{\partial V_1}{\partial r}} = \frac{\frac{\partial V_2}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V_2}{\partial r^2} - r V_2}{\frac{\partial V_2}{\partial r}}.$$

The left-hand side depends only on V_1 not V_2 , and the right-hand side depends only on V_2 not V_1 , and hence both sides do not depend on the derivative price. Therefore the derivative price V satisfies

$$\frac{\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2} - r V}{\frac{\partial V}{\partial r}} = -m(r, t)$$

for some function $m(r, t)$. The negative sign in front of $m(r, t)$ is for notational convenience in the final statement of the result.

Example 22.5 If V_t denotes the value at time t of an interest rate option with maturity T and strike rate r_K , then its price satisfies the fundamental pricing equation given in Theorem 22.3 with the final condition

$$V(r, T) = N \times \max\{r - r_K, 0\}$$

for some nominal amount N .

Remark 22.5 As in the derivation of the Black–Scholes–Merton partial differential equation presented in Theorem 15.1, we can simulate the self-financing portfolio process

$$\Pi(r, t) = V_1(r, t) - \Delta V_2(r, t) + C$$

where we take Δ given by (22.7) and the price processes V_1 and V_2 of bond 1 and bond 2, respectively, and C denotes the cash account. The fundamental pricing equation depends on our choice of the interest rate model. For example, the Vasicek model

$$dr_t = \alpha(\bar{r} - r_t) dt + \sigma dW_t$$

in Chap. 23 assumes that $m(r, t)$ has the same form as the drift rate of the original interest rate process, although possibly with different parameter values. For the formulas of bond price and related definitions, see Sect. 23.2. In Fig. 22.5 the Vasicek model allows negative interest rate in the first panel. In the third panel two bond price processes are plotted with maturity dates $T_1 = 3$ and $T_2 = 4$, and the price of the bond with maturity T_1 has price equal to 1 at $t = T_1$ by definition. Observe that the bank deposit and the hedged portfolio coincide, as predicted theoretically, even though the interest rate is stochastic. The hedged portfolio is not necessarily exponentially increasing since the interest rate is not constant. Consult Simulation 22.3 for more information.

Remark 22.6 When there is continuous coupon payment, we modify the model as

$$\frac{\partial V}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 V}{\partial r^2} + m(r, t) \frac{\partial V}{\partial r} - rV + K(r, t) = 0 ,$$

where $K(r, t) dt$ is the amount received in a period of length dt . For discrete coupon payment, we introduce the jump condition

$$V(r, t_c-) = V(r, t_c+) + K(r, t_c) ,$$

where t_c is the time of coupon payment and $K(r, t_c)$ is the amount of coupon.

Remark 22.7 (*Risk-neutral measure*) Let P_t be the price at time t of a bond with maturity T such that $P_T = 1$. Let

$$dr_t = \mu(r_t, t) dt + \sigma(r_t, t) dW_t$$

be a model of the movement of the physical interest rate where W_t is a $\mathbb{P}$ -Brownian motion. Let $m(r, t)$ be the function given in Theorem 22.3. Put

$$\theta(r, t) = \frac{\mu(r, t) - m(r, t)}{\sigma(r, t)} \quad (22.8)$$

and $\theta_t = \theta(r_t, t)$. We assume, throughout the rest of the chapter, that the following *Novikov condition* holds:

$$\mathbb{E}^{\mathbb{P}} \left[\exp \left(\frac{1}{2} \int_0^T \theta_t^2 dt \right) \right] < \infty .$$

(Here we are assuming that $\sigma \neq 0$.) Define

$$\tilde{W}_t = W_t + \int_0^t \theta_u du .$$

By Girsanov's theorem, there is a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$, with the *Radon–Nikodym derivative*

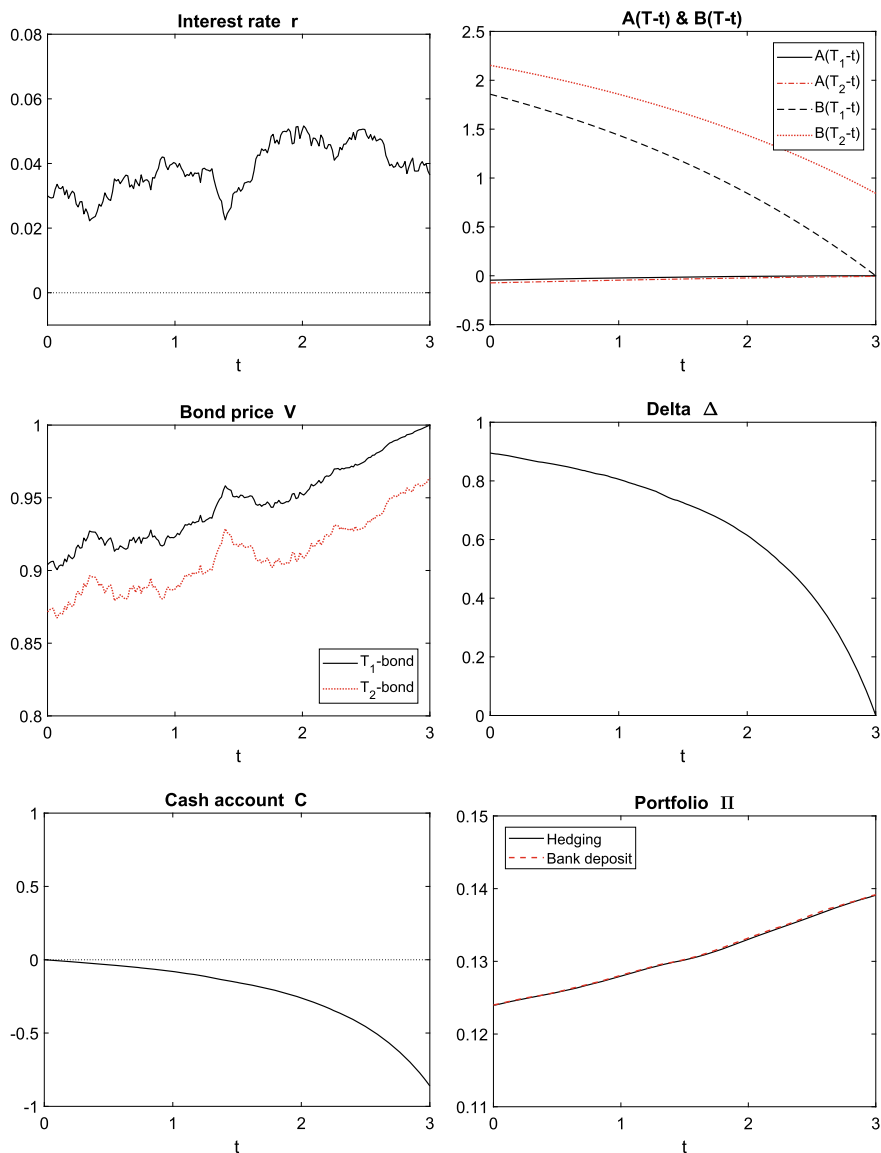


Fig. 22.5 Delta hedging for interest rate derivatives V_1 and V_2 presented in Remark 22.5 using a self-financing portfolio $\Pi(r, t) = V_1(r, t) - \Delta V_2(r, t) + C$, $0 \leq t \leq 3$ under an interest rate model from Sect. 23.2 (from top left to bottom right)

$$\frac{dQ}{dP} = \exp \left(- \int_0^T \theta_t dW_t - \frac{1}{2} \int_0^T \theta_t^2 dt \right)$$

such that $\tilde{W}_t$, $0 \leq t \leq T$, is a $\mathbb{Q}$ -Brownian motion. The measure $\mathbb{Q}$ is called the *risk-neutral measure*. The interest rate now satisfies

$$dr_t = \mu dt + \sigma (d\tilde{W}_t - \theta dt) = m dt + \sigma d\tilde{W}_t ,$$

and it is called the *risk-neutral interest rate*.

Remark 22.8 (*Market price of interest rate risk*) Suppose that the coefficient $m(r, t)$ in Theorem 22.3 is of the form

$$-m(r, t) = \sigma(r, t)\theta(r, t) - \mu(r, t)$$

for some $\theta(r, t)$. (We just take $\theta = \frac{\mu - m}{\sigma}$.) Then (22.6) becomes

$$\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 V}{\partial r^2} + (\mu - \sigma \theta) \frac{\partial V}{\partial r} - rV = 0 . \quad (22.9)$$

By Itô's lemma, we have

$$\begin{aligned} dV &= \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 V}{\partial r^2} \right) dt + \frac{\partial V}{\partial r} dr \\ &= \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 V}{\partial r^2} \right) dt + \frac{\partial V}{\partial r} (\mu dt + \sigma dW) \\ &= \left(\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 V}{\partial r^2} + \mu \frac{\partial V}{\partial r} \right) dt + \sigma \frac{\partial V}{\partial r} dW \\ &= \left(\sigma \theta \frac{\partial V}{\partial r} + rV \right) dt + \sigma \frac{\partial V}{\partial r} dW , \end{aligned}$$

where (22.9) is used for the last equality. Hence

$$dV - rV dt = \sigma \frac{\partial V}{\partial r} (\theta dt + dW) \quad (22.10)$$

and

$$\mathbb{E}^{\mathbb{P}} \left[\frac{dV - rV dt}{\sigma \frac{\partial V}{\partial r}} \right] = \theta dt . \quad (22.11)$$

(Here we assume that $\sigma \frac{\partial V}{\partial r} \neq 0$.) The left-hand side of (22.11) is the risk-adjusted excess return above the risk-free rate, and hence θ defined in (22.8) is called the *market price of interest rate risk*.

Let $m(r, t)$ be the function given in Theorem 22.3. Let θ , $\tilde{W}_t$ and $\mathbb{Q}$ be defined as in Remark 22.7.

Theorem 22.4 (Discounted Feynman–Kac Theorem for Interest Rate Derivatives) *Consider an interest rate derivative with its payoff at maturity S given by X_S . Let V_t denote the price of the interest rate derivative at $0 \leq t \leq S$. Choose a numeraire given by the risk-free cash account process*

$$B_t = \exp \left(\int_0^t r_u du \right)$$

and consider the corresponding discounted derivative price process

$$\tilde{V}_t = B_t^{-1} V_t = \exp \left(- \int_0^t r_u du \right) V_t .$$

Assume that at least one of following conditions holds:

- (i) There exists $Y \geq 0$ such that $\mathbb{E}^{\mathbb{Q}}[Y] < \infty$ and $|\tilde{V}_t| \leq Y$ for every t .
- (ii) $\mathbb{E}^{\mathbb{Q}} \left[\int_0^S \left(B_t^{-1} \sigma \frac{\partial V}{\partial r} \right)^2 dt \right] < \infty$.

$$V_t = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^S r_u du} X_S \middle| \mathcal{F}_t \right] , \quad 0 \leq t \leq S ,$$

where $dr_t = m dt + \sigma d\tilde{W}_t$.

Proof Note that

$$\begin{aligned} d\tilde{V}_t &= e^{-\int_0^t r_u du} dV_t + e^{-\int_0^t r_u du} (-r_t) V_t dt \\ &= B_t^{-1} (dV_t - r_t V_t dt) \\ &= B_t^{-1} \sigma \frac{\partial V}{\partial r} d\tilde{W}_t , \quad (\text{by (22.10)}) \end{aligned}$$

and hence $\tilde{V}_t$ is a local $\mathbb{Q}$ -martingale. Since at least one of two conditions holds, $\tilde{V}_t$ is a $\mathbb{Q}$ -martingale. (See Remark 7.2 in Klebaner (2005) and Lemma 5.1 in Filipović (2009) for the proof.) Hence

$$e^{-\int_0^t r_u du} V_t = \tilde{V}_t = \mathbb{E}^{\mathbb{Q}} \left[\tilde{V}_S \middle| \mathcal{F}_t \right] = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_0^S r_u du} V_S \middle| \mathcal{F}_t \right] .$$

Since $e^{-\int_0^t r_u du}$ is $\mathcal{F}_t$ -measurable, we have

$$V_t = \mathbb{E}^{\mathbb{Q}} \left[e^{\int_0^t r_u du} e^{-\int_0^S r_u du} V_S \middle| \mathcal{F}_t \right] = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^S r_u du} V_S \middle| \mathcal{F}_t \right]$$

for $0 \leq t \leq S$. Now use $V_S = X_S$.

Corollary 22.1 (Discounted Feynman–Kac Theorem for Bonds) *Let $P(t, S)$ denote the price of a zero coupon bond with maturity S at time $0 \leq t \leq S$. Then*

$$P(t, S) = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^S r_u du} \middle| \mathcal{F}_t \right] .$$

Proof Note that the bond price $P(t, S)$ satisfies $P(S, S) = X_S = 1$.

Example 22.6 Consider a forward contract in which a price K will be paid at time T in return for a repayment of \$1 at time S , $T < S$. Equivalently, K is paid at T in return for delivery at T of the S -bond that has a value $P(T, S)$ at T . Now the problem is to find how much the contract is worth at time $t < T$. Since the contract has value

$$X = P(T, S) - K$$

at T , we have

$$\begin{aligned} V(t) &= \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u du} (P(T, S) - K) | \mathcal{F}_t] \\ &= \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u du} \mathbb{E}^{\mathbb{Q}}[e^{-\int_T^S r_u du} | \mathcal{F}_T] | \mathcal{F}_t] - K \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u du} | \mathcal{F}_t] \\ &= \mathbb{E}^{\mathbb{Q}}[\mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u du} e^{-\int_T^S r_u du} | \mathcal{F}_T] | \mathcal{F}_t] - K \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u du} | \mathcal{F}_t] \\ &= \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^S r_u du} | \mathcal{F}_t] - K \mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u du} | \mathcal{F}_t] \\ &= P(t, S) - K P(t, T) . \end{aligned}$$

If K is chosen so that there is no need to exchange any money at time $t = 0$ for the forward contract, then $V(0) = 0$, and hence $K = \frac{P(0, S)}{P(0, T)}$.

Theorem 22.5 *As in the method for pricing an option on a stock given in Sect. 16.1, there exists a self-financing and replicating strategy for interest rate derivatives with maturity S .*

Proof Construct a portfolio composed of two assets: a bond P with maturity T , $S \leq T$, and the risk-free cash deposit B . Let $\mathbb{Q}$ be the risk-neutral measure constructed in Theorem 22.4. Assume that the bond pays \$1 at T and that the risk-free cash account satisfies $dB_t = r_t B_t dt$, $B_0 = 1$, for the interest rate process r_t . Then $B_t = e^{\int_0^t r_u du}$. Let X_S denote the payoff at S of the interest rate derivative under consideration. Define a $\mathbb{Q}$ -martingale

$$M_t = \mathbb{E}^{\mathbb{Q}}[B_S^{-1} X_S | \mathcal{F}_t]$$

for $0 \leq t \leq S$. Let $\tilde{P}_t = B_t^{-1} P_t$ be the discounted bond price. Then, by (22.10),

$$\begin{aligned} d\tilde{P}_t &= -r_t B_t^{-1} P_t dt + B_t^{-1} dP_t \\ &= B_t^{-1} (dP_t - r_t P_t dt) \\ &= B_t^{-1} \sigma \frac{\partial P}{\partial r} d\tilde{W}_t . \end{aligned}$$

Since $\tilde{P}_t$ and M_t are both $\mathbb{Q}$ -martingales, by the Martingale Representation Theorem, there exists a predictable process ϕ_t such that

$$M_t = M_0 + \int_0^t \phi_u d\tilde{P}_u .$$

Define

$$\psi_t = M_t - \phi_t \tilde{P}_t$$

and consider a portfolio V defined by

$$V_t = \phi_t P_t + \psi_t B_t .$$

Since

$$V_t = B_t(\phi_t \tilde{P}_t + \psi_t) = B_t M_t ,$$

we have

$$\begin{aligned} dV_t &= B_t dM_t + M_t dB_t \\ &= B_t \phi_t d\tilde{P}_t + M_t dB_t \\ &= B_t \phi_t (-r_t B_t^{-1} P_t dt + B_t^{-1} dP_t) + (\psi_t + \phi_t B_t^{-1} P_t) dB_t \\ &= \phi_t (-r_t P_t dt + dP_t) + (\psi_t + \phi_t B_t^{-1} P_t) dB_t \\ &= \phi_t dP_t + \psi_t dB_t + \phi_t B_t^{-1} P_t (-r_t B_t dt + dB_t) \\ &= \phi_t dP_t + \psi_t dB_t . \end{aligned}$$

Hence V_t is self-financing. Finally, V_t replicates the contingent claim at time S since

$$V_S = B_S M_S = B_S \mathbb{E}^{\mathbb{Q}}[B_S^{-1} X_S | \mathcal{F}_S] = B_S B_S^{-1} X_S = X_S .$$

Example 22.7 If V itself is a bond with maturity S , and $X_S = V_S = 1$, then holding V is a self-financing and replicating strategy. In this case, $\phi_t = 1$ and $\psi_t = 0$.

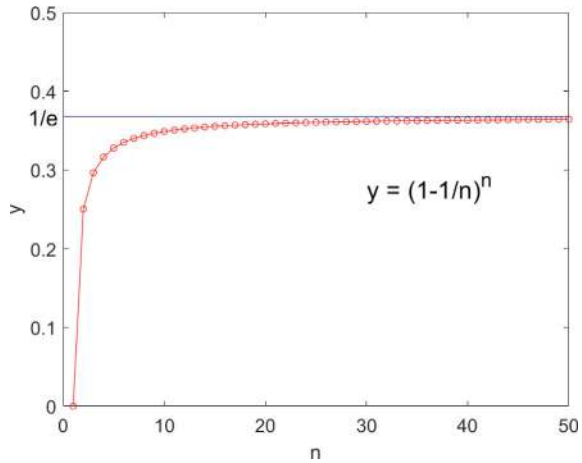
22.9 Computer Experiments

Simulation 22.1 (Monotone Convergence to $e = 2.71828\dots$)

We plot the graph $y = (1 + \frac{1}{n})^n$, $1 \leq n \leq 50$, and check the monotone convergence to the limit e . See Fig. 22.1. See also Exercise 22.1.

```
N = 50;
x = 0:0.1:N;
n = 1:1:N;
```

Fig. 22.6 Monotone convergence of $(1 - 1/n)^n$ to the limit $1/e = 0.367879 \dots$ as $n \rightarrow \infty$



```
y = (1+1./n).^n;
plot(n,y,'ro-');
hold on
plot(x,exp(1)*ones(length(x)));
```

For the convergence of $(1 - \frac{1}{n})^n$ to the limit $1/e = 0.367879 \dots$ see Fig. 22.6.

Simulation 22.2 (Yield to Maturity) Using the Newton–Raphson method we compute the yield to maturity when the present price of a coupon bond is given.

```
B = 140; % current bond price
C = 5; % Coupon is paid every six months.
T = 20; % time to maturity of bond
P = 100; % face value
r_max = 0.2;
L = 100;
dr = r_max/L;
r = 0:dr:r_max; % range of interest rates
PV = zeros(1,L+1);
R = r(1,2:L+1);
PV(1,2:L+1) = C*exp(-0.5*R).*(1-exp(-T*R))./(1-exp(-0.5*R)) + P*exp(-R*T);
PV(1,1) = C*2*T + P;
plot(r(1,:),PV(1,:))
xlabel('Yield to maturity')
ylabel('Price of bond')
```

For the graph of bond price as a function of yield to maturity y , see Fig. 22.4. Now we find the yield to maturity by Newton's method presented in Sect. 16.3, Vol. II.

```

syms y % Create a symbolic variable y.
f = C*exp(-0.5*y)*(1-exp(-T*y))/(1-exp(-0.5*y)) + P*exp(-y*T) - B;
f1 = diff(f,y);
n = 7;
Y = zeros(n+1);
Y(1) = 0.01;
for i = 1:n
    Y(i+1) = Y(i) - subs(f,Y(i)) / subs(f1,Y(i));
end
YTM = Y(n+1) % Yield to maturity

```

Simulation 22.3 (The Fundamental Equation for Pricing) We simulate the Delta hedging used in the derivation of the fundamental equation for pricing interest rate derivatives in the proof of Theorem 22.3. Consider a self-financing portfolio Π_t given by

$$\Pi_t = V_{1,t} - \Delta_t V_{2,t} + C_t, \quad (22.12)$$

where C_t is the risk-free cash account. The self-financing condition implies that

$$\begin{aligned} \delta \Pi_t &= \delta V_{1,t} - \Delta_t \delta V_{2,t} + r_t C_t \delta t \\ &= (V_{1,t+\delta t} - V_{1,t}) - \Delta_t (V_{2,t+\delta t} - V_{2,t}) + r_t C_t \delta t. \end{aligned} \quad (22.13)$$

After discrete time rebalancing, we have

$$\Pi_{t+\delta t} = V_{1,t+\delta t} - \Delta_{t+\delta t} V_{2,t+\delta t} + C_{t+\delta t}. \quad (22.14)$$

If we take

$$\Delta = \frac{\partial V_1}{\partial r} \bigg/ \frac{\partial V_2}{\partial r}$$

in (22.13) as in the proof of Theorem 22.3, then $\delta \Pi_t$ in (22.13) becomes risk-free and Π_t behaves as if it is a risk-free bank account and satisfies

$$\delta \Pi_t = r_t \Pi_t \delta t. \quad (22.15)$$

The simulation result for (22.15), over the time period $0 \leq t \leq T$ where $T = 3$, is represented by the vector `Pi_bank` in the MATLAB code given below. The last panel in Fig. 22.5 displays the risk-free investment

$$\Pi_t = \Pi_0 \exp \left(\int_0^t r_u du \right), \quad 0 \leq t \leq T,$$

with continuous compounding interest r_t .

On the other hand, (22.12), (22.13) and (22.14) together imply that

$$\begin{aligned} & (V_{1,t+\delta t} - \Delta_{t+\delta t} V_{2,t+\delta t} + C_{t+\delta t}) - (V_{1,t} - \Delta_t V_{2,t} + C_t) \\ &= (V_{1,t+\delta t} - V_{1,t}) - \Delta_t (V_{2,t+\delta t} - V_{2,t}) + r_t C_t \delta t, \end{aligned}$$

which is simplified to

$$- \Delta_{t+\delta t} V_{2,t+\delta t} + C_{t+\delta t} - C_t = -\Delta_t V_{2,t+\delta t} + r_t C_t \delta t.$$

Thus

$$C_{t+\delta t} = (1 + r_t \delta t) C_t + (\Delta_{t+\delta t} - \Delta_t) V_{2,t+\delta t}.$$

The simulation result for C_t is given in the fifth panel in Fig. 22.5, and the corresponding portfolio Π_{hedging} is given in the last panel. Observe that the two graphs are almost identical.

In the following simulation for the interest rate r_t we employ the Vasicek model, and consider two bonds with maturity dates $T_1 = 3$ and $T_2 = 4$.

```
T1 = 3; T2 = 4;
N = 200; dt = T1/N; t = 0:dt:T1;
alpha = 0.35;
r_bar = 0.05;
sig = 0.015; % sigma
r0 = 0.03;
r(1) = r0;
dW = sqrt(dt)*randn(1,N);
for i = 1:N
    r(i+1) = r(i) + alpha*(r_bar - r(i))*dt + sig*dW(i);
end

figure(1)
plot(t,r);
xlabel('t');
ylabel('r');
title('Interest rate');

figure(2)
V1 = zeros(N+1,1); % Bond 1 price by Vasicek
Y2 = zeros(N+1,1); % Bond 2 price by Vasicek
% Compute B_k(T_i-t), A_k(T_k-t) and V_k(t) for k=1,2
for i = 1:N+1
    B1(i) = 1/alpha*(1-exp(-alpha*(T1-t(i))));
    A1(i) = (B1(i)-(T1-t(i)))*(r_bar-sig^2/2/alpha^2)-sig^2*B1(i)^2/4/alpha;
    V1(i) = exp(A1(i)-B1(i)*r(i));
```



```

end
for i = 1:N+1
B2(i) = 1/alpha*(1-exp(-alpha*(T2-t(i))));
A2(i) = (B2(i)-(T2-t(i)))*(r_bar-sig^2/2/alpha^2)-sig^2*B2(i)^2/4/alpha;
V2(i) = exp(A2(i)-B2(i)*r(i));
end
plot(t,A1,'-k',t,A2,'--r',t,B1,'-.k',t,B2,'r')
legend('A(T1-t)', 'A(T2-t)', 'B(T1-t)', 'B(T2-t)');
xlabel('t');
ylabel('A & B');
title('A(T-t) & B(T-t)');

figure(3)
plot(t,V1,'-k',t,V2,':r')
legend('T1-bond', 'T2-bond');
xlabel('t');
ylabel('V');
title('Bond price');

figure(4)
Delta = zeros(1,N+1); % hedge ratio
for i = 1:N+1
Delta(i) = B1(i)*V1(i)/(B2(i)*V2(i));
end
plot(t,Delta)
xlabel('t');
ylabel('\Delta');
title('Delta');

figure(5)
% Cash account C with continuously compounding interest
% Choose any amount for C(1).
C = zeros(1,N+1);
% self-financing condition
for i = 1:N
C(i+1) = (1+r(i)*dt)*C(i) + (Delta(i+1)-Delta(i))*V2(i+1);
end
plot(t,C,'Color','k')
xlabel('t');
ylabel('C');
title('Cash account');

figure(6)
Pi_hedging = zeros(1,N+1); % portfolio of V1 and V2 and Cash
for i = 1:N+1

```

```

Pi_hedging(i) = V1(i) - Delta(i)*V2(i) + C(i);
end

Pi_bank = zeros(1,N+1); % risk-free bank deposit
Pi_bank(1) = Pi_hedging(1);
cum_rate = zeros(1,N+1);
for i = 1:N
    cum_rate(i+1) = cum_rate(i) + r(i);
    Pi_bank(i+1) = Pi_hedging(1)*exp(cum_rate(i)*dt);
end

plot(t,Pi_bank,'k',t,Pi_hedging,':r')
xlabel('t');
ylabel('\Pi');
title('Portfolio');
legend('Hedging','Bank deposit');

```

Simulation 22.4 (Discounted Feynman–Kac Theorem) We compute the present ($t = 0$) price of a zero coupon bond with maturity $T = 30$ years using Corollary 22.1 in the Vasicek model

$$dr_t = \alpha(\bar{r} - r_t) dt + \sigma dW_t$$

given in Sect. 23.2. The closed form formula for bond prices in the following code is from Theorem 23.3.

```

T= 30;
N = 1000;
dt = T/N;
t = 0:dt:T;
M = 10^5; % the number of samples for the Monte Carlo method
r = zeros(M,1);
alpha = 0.4;
r_bar = 0.05;
sigma = 0.015;
r0 = 0.03;
r(:) = r0;
Integral = zeros(M,1);
for i = 1:N
    dW = sqrt(dt)*randn(M,1);
    r(:) = r(:) + alpha*(r_bar - r(:))*dt + sigma*dW(:);
    Integral(:) = Integral(:) + r(:)*dt;
end

B = 1/alpha*(1-exp(-alpha*T));

```

```

A = (B-T)*(r_bar - sigma^2/2/alpha^2) - sigma^2*B^2/4/alpha;
Vasicek_formula = exp(A-B*r0)

MC_bond_price = mean(exp(-Integral(:)))
sample_std = std(exp(-Integral(:)))/sqrt(M)

```

The output is given by

```

Vasicek_formula =
    0.2389
MC_bond_price =
    0.2390
sample_std =
    1.4652e-04

```

Note that the theoretical formula under the Vasicek model (Theorem 23.3) and the Monte Carlo method produce the prices very close to each other. If we employ the Cox–Ingersoll–Ross model in Sect. 23.3 given by

$$dr_t = \alpha(\bar{r} - r_t) dt + \sigma \sqrt{r_t} dW_t$$

with MATLAB code

```

r(:) = max(0,r(:) + alpha*(r_bar - r(:))*dt + sigma*sqrt(r(:)).*dW(:));

```

using the same parameters as in the Vasicek model except for

$$\sigma = \frac{0.015}{\sqrt{\bar{r}}} = 0.0671 ,$$

we obtain the bond price 0.2387, which is close to the value obtained by the Vasicek model.

Simulation 22.5 (Distribution of the Discount Factor) To plot the distributions of $\int_0^T r_t dt$ and the discount factor $\exp(-\int_0^T r_t dt)$ considered in Simulation 22.4 under the Vasicek model, we generate 10^6 sample paths. (We observe that all the sample values for the integral $\int_0^T r_t dt$ are positive even though a sample path r_t , $r_0 = 0.03$, can be negative for some $0 < t \leq T$.) The number of sample values belonging to the j th subinterval is given by `bin(j)` in the following code.

```

num_bins = 100;
Nb = num_bins - 1;
max_I = max(Integral)
min_I = min(0,min(Integral)) %This checks whether the integral is negative.
width = (max_I - min_I)/Nb;

```

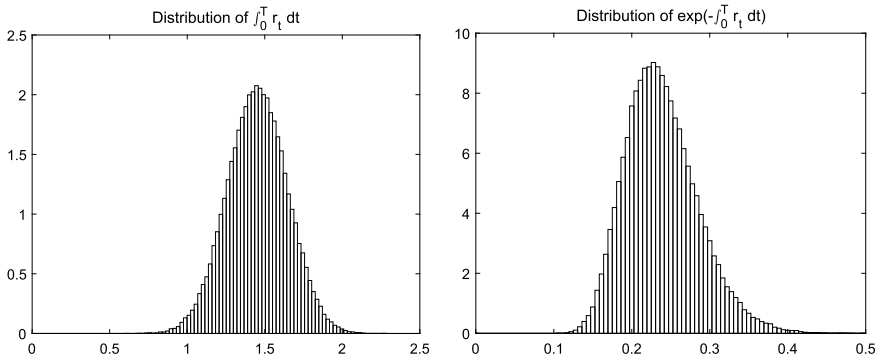


Fig. 22.7 Probability distributions of $\int_0^T r_t dt$ and $\exp(-\int_0^T r_t dt)$ in the Vasicek model

```

min_value = min_I - width/2;
max_value = max_I + width/2;
bin = zeros(1, num_bins);

for i=1:M
    j_bin=ceil((Integral(i)-min_value)/width);
    bin(j_bin)=bin(j_bin)+1;
end

mid_points = min_value + width/2:width:max_value;
bar(mid_points,bin/M/width, 1, 'w')

title('Distribution of \int_0^T r_t dt');
ylabel('Probability');
set(gca,'xlim',[0,3]);

```

For the plots, see Fig. 22.7. Observe that the distributions of $\int_0^T r_t dt$ and $\exp(-\int_0^T r_t dt)$ are approximately normal and lognormal, respectively.

If the Cox–Ingersoll–Ross model in Sect. 23.3 is employed in bond pricing with the same parameters as in the Vasicek model except for

$$\sigma = \frac{0.015}{\sqrt{T}} = 0.0671 ,$$

to match the volatility level, we have the distributions given in Fig. 22.8. Observe that the distribution of $\int_0^T r_t dt$ is skewed to the right and $\exp(-\int_0^T r_t dt)$ is more symmetric than in the Vasicek model case. This is due to the fact that in the CIR model the interest rate is always nonnegative, not symmetrically distributed.

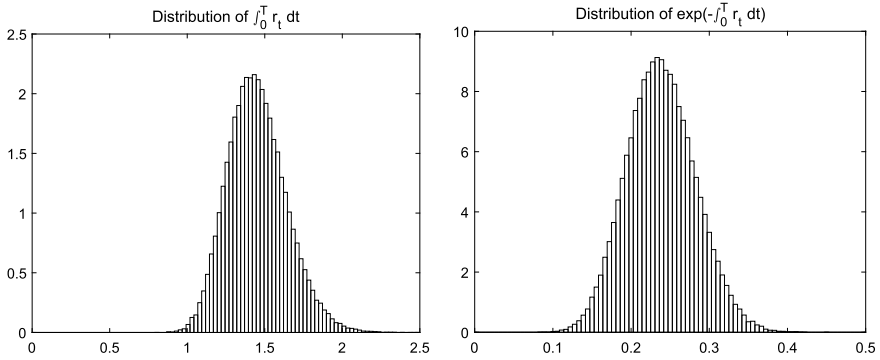


Fig. 22.8 Probability distributions of $\int_0^T r_t dt$ and $\exp(-\int_0^T r_t dt)$ in the CIR model

Exercises

22.1 Let $r > 0$. Show that $(1 + \frac{r}{n})^n$ is monotonically increasing and converges to e^r as $n \rightarrow \infty$. Prove the monotone convergence of $(1 - \frac{r}{n})^n$ to e^{-r} as $n \rightarrow \infty$ under the condition that $1 - \frac{r}{n} > 0$, i.e., $n > r$.

22.2 Let $\Pi = V_1 - \Delta V_2$ as in the proof of Theorem 22.3 and assume that Π is hedged by choosing Δ as in (22.7). Show that the following holds:

$$\frac{1}{\Pi} \frac{\partial \Pi}{\partial t} + \frac{1}{2} \sigma^2 \frac{1}{\Pi} \frac{\partial^2 \Pi}{\partial r^2} = r.$$

22.3 Under the assumption that the bond price at t is given by

$$P(t, S) = \mathbb{E}^{\mathbb{Q}} \left[e^{-\int_t^S r_u du} \middle| \mathcal{F}_t \right]$$

derive the fundamental equation for bond pricing given in Theorem 22.3.

Chapter 23

Short-Rate Models



While the assumption that the interest rate is constant produces reasonable estimates in option pricing, the same assumption would produce less reliable results in pricing bonds and interest rate derivatives. One of the reasons is that the bonds usually have much longer maturity. We investigate interest rate models expressed in terms of stochastic differential equations. In this chapter we introduce models for the short rate r_t , in particular. Recall the definition of the short rate given in Definition 22.8. If not stated otherwise, an interest rate means a short rate in this chapter.

23.1 Affine Term Structure

Definition 23.1 (*Affine Term Structure*) Let $P(t, T)$ denote the price at time t of a bond with maturity T . Assume that $P(T, T) = 1$. If the bond price is given by a formula of the form

$$P(r, t, T) = e^{A(t, T) - B(t, T)r}$$

for some sufficiently smooth deterministic functions $A(t, T)$ and $B(t, T)$, where r denotes the interest rate, then the model is said to have an *affine term structure*. Note that in the exponent the coefficient for r is $-B(t, T)$ so that we have $B(t, T) > 0$ in agreement with the fact that

$$\frac{\partial}{\partial r} P(r, t, T) = -B(t, T)P(r, t, T) < 0.$$

Theorem 23.1 Consider a risk-neutral interest rate process r_t such that

$$dr_t = m(r_t, t) dt + \sigma(r_t, t) d\tilde{W}_t,$$

where $\tilde{W}_t$ is a $\mathbb{Q}$ -Brownian motion. Assume that the interest rate model has an affine term structure. Suppose that its diffusion and drift terms are of the form

$$m(r, t) = a(t) + b(t)r$$

and

$$\sigma^2(r, t) = c(t) + d(t)r$$

for some deterministic functions a, b, c and d . (In other words, m and σ^2 are affine functions in r .) Then A and B given in Definition 23.1 satisfy the system of ordinary differential equations

$$\begin{cases} \frac{\partial}{\partial t} A(t, T) - a(t)B(t, T) + \frac{1}{2}c(t)B(t, T)^2 = 0, & A(T, T) = 0, \\ \frac{\partial}{\partial t} B(t, T) + b(t)B(t, T) - \frac{1}{2}d(t)B(t, T)^2 = -1, & B(T, T) = 0. \end{cases} \quad (23.1)$$

Proof Recall that the bond price $P(t, T)$ satisfies the fundamental equation in Theorem 22.3. By substituting $P(t, T) = e^{A(t, T) - B(t, T)r}$ in place of V in the equation, we obtain the ordinary differential equations in (23.1). The boundary conditions follow since $1 = P(T, T) = e^{A(T, T) - B(T, T)r}$ for every r . $\square$

Remark 23.1 The second equation in (23.1) for $B(t, T)$ is called a Riccati equation. (See Appendix D.2.) Once we solve for B , then we insert it into the first equation, and obtain A .

23.2 The Vasicek Model

Consider the Vasicek model (Vasicek, 1977) for an interest rate given by

$$dr_t = \alpha(\bar{r} - r_t) dt + \sigma d\tilde{W}_t,$$

where r_t is the interest rate at t , and $\alpha, \bar{r}$ and σ are positive constants. (For $\bar{r} = 0$ we obtain the Ornstein–Uhlenbeck process in Example 12.1.) This model has the mean reversion property that is observed in the financial market. That is, if the interest rate r_t is higher than $\bar{r}$, called the long-term mean, then $\alpha(\bar{r} - r_t)$, the coefficient of dt , becomes negative, and the interest rate tends to decrease, and if the level of r_t is lower than $\bar{r}$ then $\alpha(\bar{r} - r_t)$ becomes positive and the interest rate tends to go up. The coefficient α is called the reversion rate. In the long run, the interest rate is distributed around $\bar{r}$. Note that the interest rate can become negative in the Vasicek model.

Theorem 23.2 *The interest rate r_t in the Vasicek model has the normal distribution, and for $0 \leq s \leq t$ we have*

$$r_t = \bar{r} + (r_s - \bar{r}) e^{-\alpha(t-s)} + \sigma \int_s^t e^{-\alpha(t-u)} d\tilde{W}_u.$$

Hence the conditional mean and variance are given by

$$\mathbb{E}^{\mathbb{Q}}[r_t | \mathcal{F}_s] = \bar{r} + (r_s - \bar{r}) e^{-\alpha(t-s)},$$

$$\text{Var}^{\mathbb{Q}}(r_t | \mathcal{F}_s) = \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha(t-s)}).$$

The limiting probability distribution of the interest rate r_t converges to a normal distribution with mean $\bar{r}$ and variance $\frac{\sigma^2}{2\alpha}$ as $t \rightarrow +\infty$.

Proof Note that

$$d(e^{\alpha t} r_t) = \alpha e^{\alpha t} r_t dt + e^{\alpha t} dr_t = \alpha \bar{r} e^{\alpha t} dt + \sigma e^{\alpha t} d\tilde{W}_t.$$

Hence

$$e^{\alpha t} r_t - e^{\alpha s} r_s = \int_s^t \alpha \bar{r} e^{\alpha u} du + \int_s^t \sigma e^{\alpha u} d\tilde{W}_u,$$

which is simplified to the given solution. Recall that an Itô integral of a deterministic function is normally distributed by Corollary 9.1. Since r_t in the above is a sum of a deterministic part

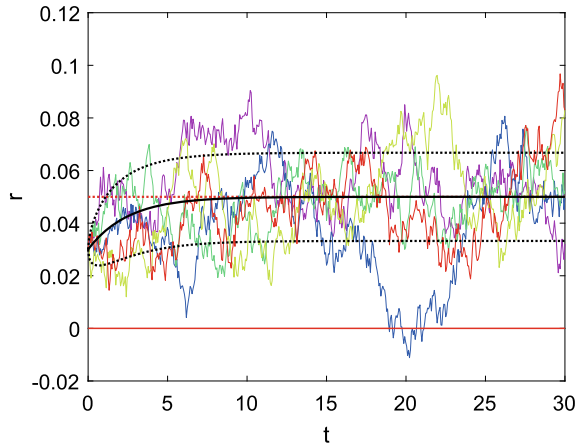
$$\bar{r} + (r_s - \bar{r}) e^{-\alpha(t-s)}$$

and an Itô integral of a deterministic function, it is also normally distributed. Note that

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[r_t | \mathcal{F}_s] &= \mathbb{E}^{\mathbb{Q}} \left[\bar{r} + (r_s - \bar{r}) e^{-\alpha(t-s)} + \sigma \int_s^t e^{-\alpha(t-u)} d\tilde{W}_u \middle| \mathcal{F}_s \right] \\ &= \bar{r} + (r_s - \bar{r}) e^{-\alpha(t-s)} + \sigma \mathbb{E}^{\mathbb{Q}} \left[\int_s^t e^{-\alpha(t-u)} d\tilde{W}_u \middle| \mathcal{F}_s \right] \\ &= \bar{r} + (r_s - \bar{r}) e^{-\alpha(t-s)} + \sigma \mathbb{E}^{\mathbb{Q}} \left[\int_s^t e^{-\alpha(t-u)} d\tilde{W}_u \right] \\ &= \bar{r} + (r_s - \bar{r}) e^{-\alpha(t-s)} + \sigma \times 0, \end{aligned}$$

where the third equality is from the independence property of conditional expectation in Theorem 5.2. Note that

Fig. 23.1 Sample paths of the interest rate in the Vasicek model



$$\begin{aligned}
 \text{Var}^{\mathbb{Q}}(r_t | \mathcal{F}_s) &= \mathbb{E}^{\mathbb{Q}}[(r_t - \mathbb{E}^{\mathbb{Q}}[r_t | \mathcal{F}_s])^2 | \mathcal{F}_s] \\
 &= \mathbb{E}^{\mathbb{Q}} \left[\left(\sigma \int_s^t e^{-\alpha(t-u)} d\tilde{W}_u \right)^2 \middle| \mathcal{F}_s \right] \\
 &= \sigma^2 \int_s^t e^{-2\alpha(t-u)} du \quad (\text{by It\^o identity}) \\
 &= \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha(t-s)}).
 \end{aligned}$$

To find the limiting distribution, we take the limit as $t \rightarrow +\infty$ and note that the limit of normal distributions is also normal. $\square$

In Fig. 23.1 are plotted five sample paths of the interest rate obtained from Simulation 23.1, where we take $\alpha = 0.4$, $\bar{r} = 0.05$, $\sigma = 0.015$, $T = 30$ and $r_0 = 0.3$. Note that r_t can take negative values. The solid line is the graph of the expectation of r_t that converges to the level $\bar{r}$, and the two dotted lines above and below the solid line represent the range of r_t within one standard deviation from the mean.

Theorem 23.3 *The bond price under the Vasicek model is given by*

$$P(r, t, T) = e^{A(T-t) - B(T-t)r},$$

where

$$\begin{aligned}
 B(T-t) &= \frac{1}{\alpha} (1 - e^{-\alpha(T-t)}), \\
 A(T-t) &= (B(T-t) - (T-t)) \left(\bar{r} - \frac{\sigma^2}{2\alpha^2} \right) - \frac{\sigma^2}{4\alpha} B(T-t)^2.
 \end{aligned}$$

Proof The equations in (23.1) now become

$$\begin{cases} \frac{\partial}{\partial t} A(t, T) - \bar{r}\alpha B(t, T) + \frac{\sigma^2}{2} B(t, T)^2 = 0, & A(T, T) = 0, \\ \frac{\partial}{\partial t} B(t, T) - \alpha B(t, T) + 1 = 0, & B(T, T) = 0. \end{cases} \quad (23.2)$$

From the second equation in (23.2) we obtain $B(t, T) = \frac{1}{\alpha}(1 - e^{-\alpha(T-t)})$. By substituting the result in the first equation, we can express $\frac{\partial}{\partial t} A(t, T)$ in terms of constant and exponential functions of t , which can be integrated easily. $\square$

Remark 23.2 (*Calibration*) How can we estimate parameters α , $\bar{r}$ and σ in the interest rate model? First, using the historical data, we estimate σ . Then, using N bond prices beginning at a fixed time, say $t = 0$, with maturities $T_1, \dots, T_N$, we find α and $\bar{r}$ minimizing the sum

$$\sum_{i=1}^N (P^{\text{Vasicek}}(0, T_i) - P^{\text{data}}(0, T_i))^2,$$

where $P^{\text{Vasicek}}(0, T_i)$ is the price by the Vasicek model and $P^{\text{data}}(0, T_i)$ is the market price.

Remark 23.3 (*Term Structure of Interest Rates in the Vasicek Model*) Note that the value of the zero coupon bond at time t depends only on time to maturity $\tau = T - t$ and the spot rate is given by

$$R(t, t + \tau) = -\frac{\log P(r_t, t + \tau)}{\tau} = -\frac{A(\tau)}{\tau} + \frac{B(\tau)}{\tau} r_t.$$

(See Definition 22.8). Thus for fixed τ we have

$$dR(t, t + \tau) = \frac{B(\tau)}{\tau} dr_t = \frac{B(\tau)}{\tau} \gamma (\bar{r} - r_t) dt + \frac{B(\tau)}{\tau} \sigma d\tilde{W}_t.$$

Note that the random component is given by $\frac{B(\tau)}{\tau} \sigma d\tilde{W}_t$ and that

$$\text{Var}(dR(t, t + \tau)) = \left(\frac{B(\tau)}{\tau}\right)^2 \sigma^2 dt = \left(\frac{B(\tau)}{\tau}\right)^2 \text{Var}(dr_t).$$

Since

$$0 < \frac{B(\tau)}{\tau} = \frac{1 - e^{-\alpha\tau}}{\alpha\tau} < 1,$$

financial shocks to short-term interest rates, dr_t , have a milder effect on longer-term bonds, $dR(t, t + \tau)$. Since $\frac{B(\tau)}{\tau}$ is a monotonically decreasing function of τ , which

Fig. 23.2 $B(\tau)/\tau$ as a monotonically decreasing function of τ in the Vasicek model

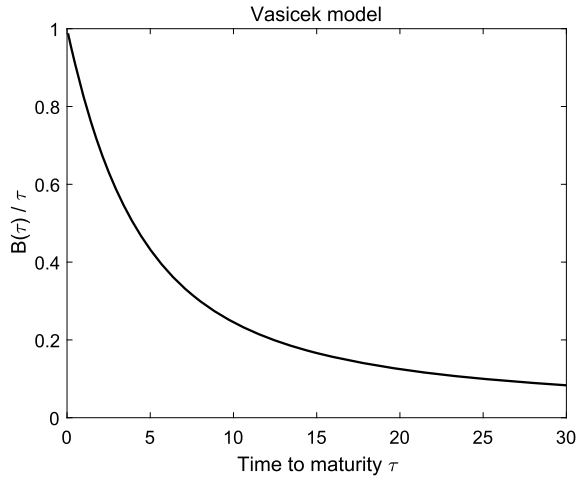
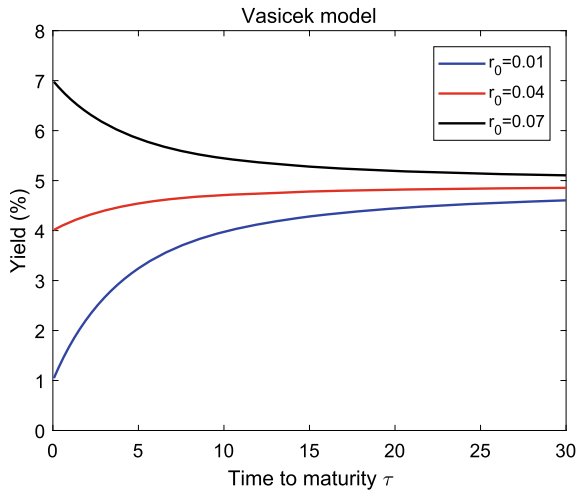


Fig. 23.3 Term structure of interest rates in the Vasicek model

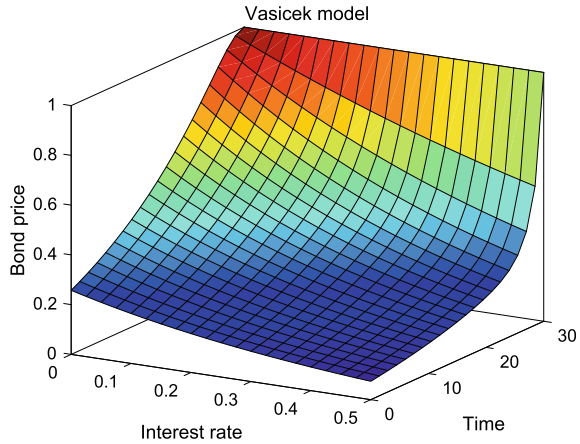


converges to 0 as $\tau \rightarrow \infty$ as shown in Fig. 23.2, we observe that the volatility $\frac{B(\tau)}{\tau} \sigma$ of $dR(t, \tau)$ for long-term spot rates is smaller than the volatility of short-term yields.

As seen in Fig. 23.3, where three curves of $R(0, \tau)$ corresponding to $r_0 = 0.01, 0.04, 0.07$ are given, longer-term rates move less than short-term rates. This does not necessarily imply that the volatility of long-term bond prices is smaller than the volatility of short-term bond prices since the duration of a bond should be multiplied in computing bond price. See Theorem 22.1.

Figure 23.4 shows the surface given by the bond price under the Vasicek model with $\alpha = 0.4, \bar{r} = 0.05, \sigma = 0.015, T = 10$ and $0 \leq r \leq 0.5$.

Fig. 23.4 Bond price by the Vasicek model



23.3 The Cox–Ingersoll–Ross Model

Consider the Cox–Ingersoll–Ross model (Cox et al., 1985), or the CIR model for short, given by

$$dr_t = \alpha(\bar{r} - r_t) dt + \sigma\sqrt{r_t} d\tilde{W}_t.$$

There are two major differences from the Vasicek model. First, due to $\sigma\sqrt{r_t}$, which is the coefficient of dW_t , volatility itself has randomness. Second, the interest rate is always nonnegative. If r_t becomes zero, then the coefficient of dW_t is also zero and the coefficient of dt is positive, and interest rate becomes positive again.

Fig. 23.5 presents sample paths of r_t , where the solid line is the graph of the expectation of r_t that converges to the level $\bar{r}$, and the two dotted lines above and below the solid line represent the range of r_t within one standard deviation from the mean. See Simulation 23.2.

Theorem 23.4 *The interest rate r_t in the CIR model has the noncentral chi-squared distribution. More precisely, the probability density function of the interest rate r at time t , with $r = r_0$ at time $t = 0$, is given by*

$$f(r) = c_t \chi^2(c_t r, d, \lambda_t)$$

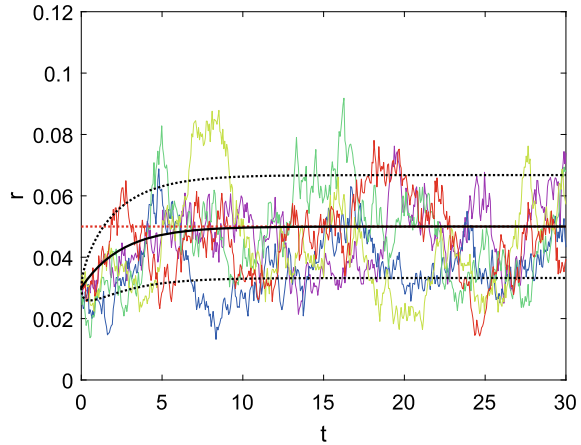
where the values of the degrees of freedom d , the non-centrality parameter λ_t , and normalizing constant c_t are given by

$$d = \frac{4\alpha\bar{r}}{\sigma^2}, \quad \lambda_t = c_t r_0 e^{-\alpha t}, \quad c_t = \frac{4\alpha}{\sigma^2(1 - e^{-\alpha t})}.$$

For $s \leq t$,

$$\mathbb{E}[r_t | r_s] = r_s e^{-\alpha(t-s)} + \bar{r} (1 - e^{-\alpha(t-s)})$$

Fig. 23.5 Sample paths of the interest rate in the CIR model



and

$$\text{Var}(r_t | r_s) = r_s \frac{\sigma^2}{\alpha} (e^{-\alpha(t-s)} - e^{-2\alpha(t-s)}) + \bar{r} \frac{\sigma^2}{2\alpha} (1 - e^{-\alpha(t-s)})^2.$$

Note that $\lim_{t \rightarrow \infty} \mathbb{E}[r_t] = \bar{r}$ and $\lim_{t \rightarrow \infty} \text{Var}(r_t) = \bar{r} \frac{\sigma^2}{2\alpha}$. The limiting distribution itself is given by $\frac{4\alpha}{\sigma^2} \chi^2 \left(\frac{4\alpha}{\sigma^2} \bar{r}, \frac{4\alpha}{\sigma^2} \bar{r}, 0 \right)$.

Proof For the part concerning the noncentral chi-squared distribution, see Cairns (2004). For the part for the mean and variance, consult Brigo and Mercurio (2006). $\square$

Remark 23.4 Let $\tau = \inf\{t > 0 : r_t \leq 0\}$ be the first hitting time that the interest rate r_t becomes zero or possibly negative. (By convention, $\inf \emptyset = +\infty$.) Note that $\{\tau = +\infty\}$ is the event that X never hits zero. If $2\alpha\bar{r} \geq \sigma^2$ then $\Pr(\tau = +\infty) = 1$, i.e., $r_t > 0$ for every $t > 0$ with probability 1. If $2\alpha\bar{r} < \sigma^2$, then $\Pr(\tau < +\infty) = 1$, i.e., the interest rate becomes zero or negative in finite time. For the proof, see Cairns (2004).

See Example F.1 for more information on the noncentral chi-squared distribution. In Fig. 23.6, obtained from Simulation 23.3, the interest rate r_t is nonnegative. The probability distributions for r_t are plotted for $t = 1, 2, 5, 50$ with the initial condition $r_0 = 1\%$ in the left panel, and $r_0 = 30\%$ in the right panel. In both cases the distribution converges to a limiting distribution around $\bar{r} = 0.05$.

Theorem 23.5 The bond price under the CIR model is given by

$$P(r, t, T) = e^{A(T-t) - B(T-t)r},$$

where

$$\gamma = \sqrt{\alpha^2 + 2\sigma^2},$$

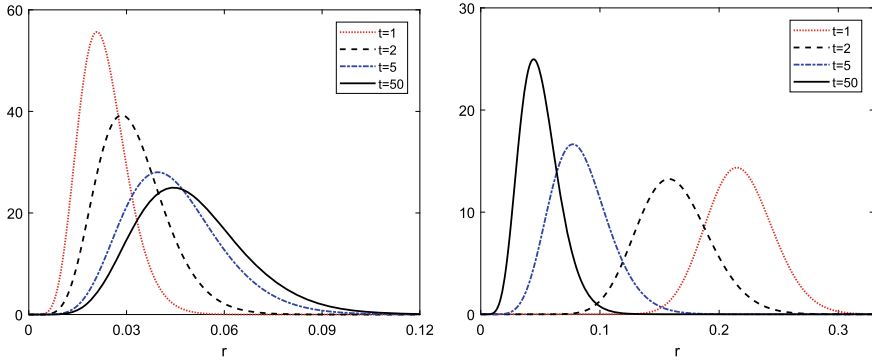


Fig. 23.6 Probability distribution of the interest rate r_t , $t = 1, 2, 5, 50$, in the CIR model: $r_0 = 0.01$ (left) and $r_0 = 0.3$ (right)

$$A(\tau) = \frac{2\alpha\bar{r}}{\sigma^2} \log \left(\frac{2\gamma e^{(\alpha+\gamma)\tau/2}}{(\alpha+\gamma)(e^{\gamma\tau} - 1) + 2\gamma} \right),$$

$$B(\tau) = \frac{2(e^{\gamma\tau} - 1)}{(\alpha+\gamma)(e^{\gamma\tau} - 1) + 2\gamma}.$$

Proof The equations in (23.1) now become

$$\begin{cases} \frac{\partial}{\partial t} A(t, T) = a(t)B(t, T) + \frac{1}{2}c(t)B(t, T)^2, & A(T, T) = 0, \\ \frac{\partial}{\partial t} B(t, T) = \alpha B(t, T) + \frac{\sigma^2}{2}B(t, T)^2 - 1, & B(T, T) = 0. \end{cases} \quad (23.3)$$

The second equation in (23.3), called the Riccati equation, has a solution given in the statement even though it is nonlinear. By substituting it into the first equation, we can express $\frac{\partial}{\partial t} A(t, T)$ in terms of constant and exponential functions of t , which can be integrated easily. $\square$

Fig. 23.7 shows the curves of $R(0, \tau)$ corresponding to $r_0 = 0.01, 0.04, 0.07$, and Fig. 23.8 shows the surface given by the bond price with $\alpha = 0.4$, $\bar{r} = 0.05$, $\sigma = 0.0671$, $T = 10$ and $0 \leq r \leq 0.5$ under the CIR model.

23.4 The Ho–Lee Model

Early term structure models such as the Vasicek and Cox–Ingersoll–Ross models were not arbitrage-free models. Consider the Ho–Lee model Ho and Lee (1986) given by

$$dr_t = \theta(t) dt + \sigma d\tilde{W}_t,$$

Fig. 23.7 Term structure of interest rates in the CIR model

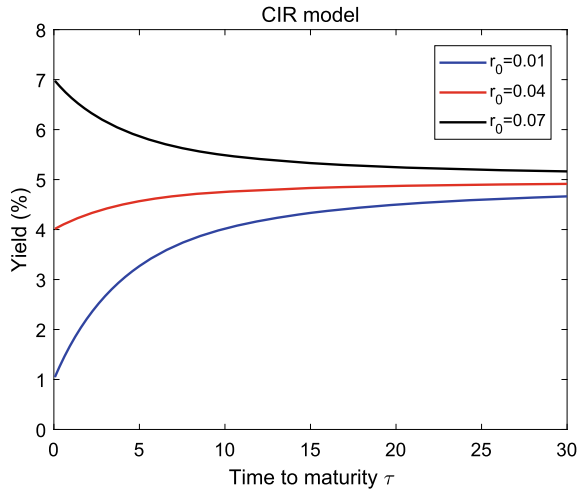
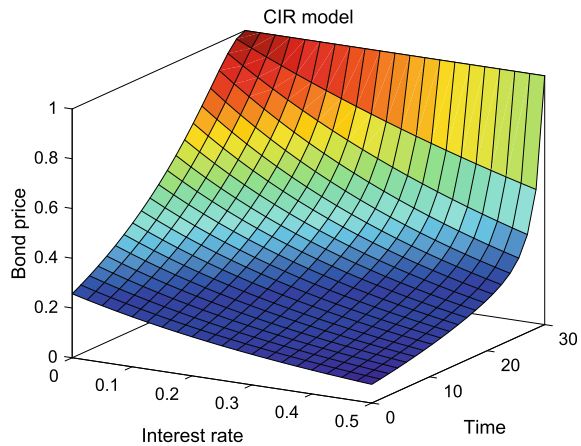


Fig. 23.8 Bond price by the CIR model



where $\tilde{W}_t$ is a $\mathbb{Q}$ -Brownian motion, and $\theta(t)$ and the constant σ are calibrated by the observed market data. The Ho–Lee model was the first arbitrage-free model.

Theorem 23.6 *The bond price under the Ho–Lee model is given by*

$$P(r, t, T) = e^{A(t, T) - (T-t)r},$$

where

$$A(t, T) = \frac{\sigma^2}{6} (T-t)^3 - \int_t^T \theta(s) (T-s) ds.$$

Proof Since $a(t) = \theta(t)$, $b(t) = 0$, $c(t) = \sigma^2$ and $d(t) = 0$ in (23.1), we have

$$\begin{cases} \frac{\partial}{\partial t} A(t, T) = \theta(t)B(t, T) - \frac{1}{2}\sigma^2 B(t, T)^2, & A(T, T) = 0, \\ \frac{\partial}{\partial t} B(t, T) = -1, & B(T, T) = 0. \end{cases}$$

Hence $B(t, T) = T - t$ and

$$\frac{\partial}{\partial t} A(t, T) = \theta(t)(T - t) - \frac{1}{2}\sigma^2(T - t)^2.$$

Thus

$$\begin{aligned} A(t, T) &= A(t, T) - A(T, T) \\ &= - \int_t^T \frac{\partial A(s, T)}{\partial s} ds \\ &= - \int_t^T \theta(s)(T - s) ds + \frac{1}{2}\sigma^2 \int_t^T (T - s)^2 ds, \end{aligned}$$

which completes the proof. $\square$

Remark 23.5 Recall that the instantaneous forward rate is given by

$$\begin{aligned} f(0, T) &= -\frac{\partial}{\partial T} \log P(0, T) \\ &= -\frac{\partial}{\partial T} (A(0, T) - Tr_0) \\ &= -\frac{\partial A(0, T)}{\partial T} + r_0 \\ &= -\frac{\sigma^2}{2}T^2 + \int_0^T \theta(s) ds. \end{aligned}$$

(See Definition 22.8.) We assume that at $t = 0$ for some T_1 , say $T_1 = 30$ years, we have the information of the bond prices $P(0, T)$ at $t = 0$ for every maturity $0 \leq T \leq T_1$ from the market.¹ Hence $f(0, T)$, $0 \leq T \leq T_1$, is also known at $t = 0$, and used as input data into the formula. Since

$$f(0, T) = -\frac{\partial A(0, T)}{\partial T} + r_0 = -\frac{\sigma^2}{2}T^2 + \int_0^T \theta(s) ds,$$

we have

$$\theta(T) = \frac{\partial f(0, T)}{\partial T} + \sigma^2 T.$$

¹ In practice, since $P(0, T)$ is given only for finitely many discrete values of T in real data, we apply some interpolation method by sufficiently smooth curves. Thus we may assume that it is differentiable in T as many times as needed.

Remark 23.6 (Drawbacks of the Ho–Lee Model) The interest rate process in the Ho–Lee model can grow to $\pm\infty$ as $T \rightarrow +\infty$, which is not realistic. Furthermore, the volatility of the changes in long-term interest rates, which is measured by the variance of $dR(t, t + \tau)$, is equal to the volatility of the changes in the short-term interest rate since

$$R(t, t + \tau) = \frac{-A(\tau) + \tau r_t}{\tau} = -\frac{A(\tau)}{\tau} + r_t.$$

That is, for a fixed τ we have $\text{Var}(dR(t, t + \tau)) = \text{Var}(dr_t) = \sigma^2 dt$. This is not in agreement with the market data which shows gradual decline as a function of τ .

23.5 The Hull–White Model

Consider the Hull–White model (Hull & White 1990) given by

$$dr_t = (\theta(t) - \gamma r_t) dt + \sigma d\tilde{W}_t,$$

where $\tilde{W}_t$ is a $\mathbb{Q}$ -Brownian motion, $\theta(t)$ is a deterministic function, and σ is a constant. To calibrate $\theta(t)$ and σ we use the market data. More precisely, $\theta(t)$ is chosen first to match exactly the term structure of interest rates, and after that we choose σ and γ to fit the term structure of spot rate volatilities.

Theorem 23.7 *The bond pricing formula in the Hull–White model is given by*

$$P(r, t, T) = e^{A(t, T) - B(t, T)r},$$

where

$$B(t, T) = \frac{1}{\gamma}(1 - e^{-\gamma(T-t)})$$

and

$$A(t, T) = -\int_t^T B(s, T)\theta(s)ds + \frac{\sigma^2}{2\gamma^2}\left(T - t + \frac{1 - e^{-2\gamma(T-t)}}{2\gamma} - 2B(t, T)\right).$$

Proof Since

$$\frac{\partial P}{\partial t} + (\theta(t) - \gamma r)\frac{\partial P}{\partial r} + \frac{1}{2}\sigma^2\frac{\partial^2 P}{\partial r^2} = rP,$$

we have

$$\begin{cases} \frac{\partial}{\partial t} A(t, T) = \theta(t)B(t, T) - \frac{1}{2}\sigma^2 B(t, T)^2, & A(T, T) = 0, \\ \frac{\partial}{\partial t} B(t, T) = \gamma B(t, T) - 1, & B(T, T) = 0. \end{cases}$$

Now find $B(t, T)$ first, then solve for $A(t, T)$. □

Remark 23.7 1. It can be shown that

$$\theta(t) = \frac{\partial f(0, t)}{\partial t} + \gamma f(0, t) + \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t}).$$

2. Since

$$R(t, t + \tau) = -\frac{A(\tau)}{\tau} + \frac{1}{\gamma} \frac{1 - e^{-\gamma \tau}}{\tau} r_t,$$

we have

$$dR(t, t + \tau) = \frac{1}{\gamma} \frac{1 - e^{-\gamma \tau}}{\tau} dr_t.$$

Let

$$f(\tau) = \frac{1}{\gamma} \frac{1 - e^{-\gamma \tau}}{\tau}, \quad \tau > 0.$$

Then

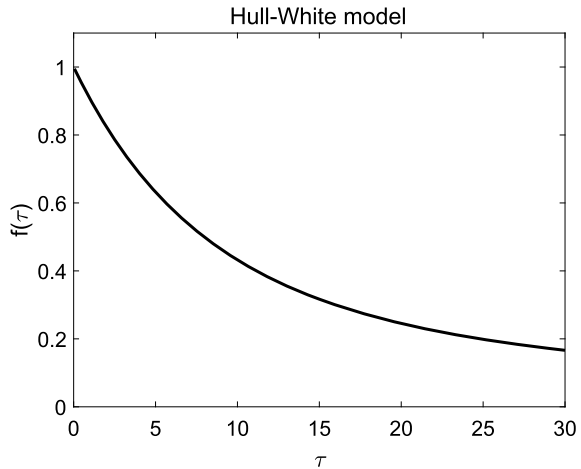
$$\text{Var}(dR(t, t + \tau)) = f(\tau)^2 \text{Var}(dr_t) = f(\tau)^2 \sigma^2 dt,$$

which is decreasing as τ increases since

$$\frac{\partial f}{\partial \tau}(\tau) = \frac{1}{\gamma} \frac{\gamma e^{-\gamma \tau} \tau - (1 - e^{-\gamma \tau})}{\tau^2} = \frac{(\gamma \tau + 1) - e^{\gamma \tau}}{\gamma \tau^2 e^{\gamma \tau}} < 0,$$

in agreement with the real market term structure. See Fig. 23.9 for the plot of the graph of $f(\tau)$, $0 < \tau \leq 30$.

Fig. 23.9 Volatility decreases as time to maturity τ increases in the Hull–White model



23.6 Computer Experiments

Simulation 23.1 (Vasicek Model) We generate $M = 10$ sample paths of the interest rate in the Vasicek model. See Fig. 23.1.

```
T= 30; N = 500; dt = T/N; t = 0:dt:T;
M = 10;
r = zeros(M,N);
alpha = 0.4; r_bar = 0.05; sigma = 0.015; r0 = 0.03;
r(:,1) = r0 ;
dW = sqrt(dt)*randn(M,N);
for i = 1:N
    r(:,i+1) = r(:,i) + a*(r_bar - r(:,i))*dt + sigma*dW(:,i);
end
for j = 1:M
    plot(t,r(j,:), 'color', hsv2rgb([1-j/M 1 1]));
    hold on
end
t = 0:dt:T;
plot(t,0, 'k')
plot(t,r_bar, 'k:')
% mean
plot(t,r_bar+(r0-r_bar)*exp(-alpha*t))

% standard deviation
stdev = sqrt(sigma^2/(2*alpha)*(1-exp(-2*alpha*t)));
plot(t,r_bar+(r0-r_bar)*exp(-alpha*t) + stdev, 'k-.');
plot(t,r_bar+(r0-r_bar)*exp(-alpha*t) - stdev, 'k-.');
```

Simulation 23.2 (Cox–Ingersoll–Ross Model) We generate $M = 10$ sample paths of the interest rate in the CIR model. Except for $\sigma = 0.015/\sqrt{r} = 0.0671$, to match the volatility level in the Vasicek model we take the same set of parameter values for T , N , M , α , $\bar{r}$ and r_0 . See Fig. 23.5.

```
T= 30; N = 500; dt = T/N; t = 0:dt:T;
M = 10;
r = zeros(M,N);
alpha = 0.4; r_bar = 0.05; sigma = 0.015/sqrt(r_bar); r0 = 0.03;
r(:,1) = r0;
dW = sqrt(dt)*randn(M,N);

% Avoid the event that the discretized interest rate becomes negative!
for i = 1:N
    r(:,i+1) = max(0, r(:,i) + alpha*(r_bar-r(:,i))*dt ...
```

```

                                + sigma*sqrt(r(:,i)).*dw(:,i));

end

for j = 1:M
    plot(t,r(j,:), 'color',hsv2rgb([1-j/M 1 1]));
    hold on
end
t = 0:dt:T;
plot(t,0, 'k--')
plot(t,r_bar, 'b')

% expectation
Exp_r = r0*exp(-alpha*t)+r_bar*(1-exp(-alpha*t));
plot(t,Exp_r)

% standard deviation
stdev = sqrt(r0*sigma^2/alpha*(exp(-alpha*t)-exp(-2*alpha*t)) ...
            + r_bar*sigma^2/(2*alpha)*(1-exp(-alpha*t)).^2);
plot(t,Exp_r + stdev, 'k-.');
plot(t,Exp_r - stdev, 'k-.');

```

Simulation 23.3 (CIR Model: Probability Distribution) We plot the probability density functions of r_t in the Cox–Ingersoll–Ross model for several values of t as shown in Fig. 23.6. We start with $r_0 = 1\%$ and observe that the limiting distribution of r_t for $t = 50$ is distributed around $\bar{r} = 5\%$.

```

alpha = 0.4; r_bar = 0.05; r0 = 0.01; sigma = 0.015; r_max = 0.07;
N = 500; dr = r_max/N;

t0 = 1;
c_t = 4*alpha/sigma^2/(1-exp(-alpha*t0));
d = 4*alpha*r_bar/sigma^2;
lambda_t = c_t*r0*exp(-alpha*t0);
r = 0:dr:r_max;
y0=c_t*ncx2pdf(c_t*r,d,lambda_t); % noncentral chi-squared distribution
plot(r,y0)
hold on;

```

We plot the same type of graphs for $t_1 = 2$, $t_2 = 5$ and $t_3 = 50$ in the same panel. See the left panel in Fig. 23.6.

Exercises

23.1 (Merton Model) Suppose that the interest rate r_t is given by $dr = \mu dt + \sigma dW$ for some constants μ and σ and that the market price of risk λ is constant.

1. Show that the price of a zero-coupon bond is given by

$$P(r, t, T) = \exp \left(-r(T-t) - \frac{1}{2}(\mu - \lambda\sigma)(T-t)^2 + \frac{1}{6}\sigma^2(T-t)^3 \right).$$

2. Show that the yield to maturity of a zero coupon bond in Definition 22.8 is given by $R(t, T) = r + \frac{1}{2}(\mu - \lambda\sigma)(T-t) - \frac{1}{6}\sigma^2(T-t)^2$.

23.2 (Vasicek Model) For the Vasicek model prove the following facts:

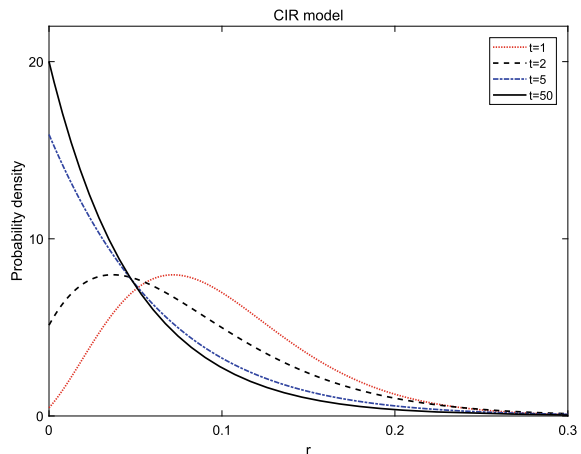
- (i) $\text{Cov}(r_t, r_s) = e^{-\alpha(t+s)} \left(\sigma_0 + \sigma^2 \frac{e^{2\alpha s} - 1}{2\alpha} \right)$.
- (ii) $\mathbb{E}[\int_s^t r_u du | \mathcal{F}_s] = \bar{r}(t-s) + (r_s - \bar{r}) \frac{1 - e^{-\alpha(t-s)}}{\alpha}$.
- (iii) Derive the limiting probability density of the interest rate directly from the Kolmogorov equation.

23.3 (CIR Model) For the CIR model, derive the limiting pdf of the interest rate directly from the Kolmogorov equation. More precisely, show that the pdf $p(x)$ is given by

$$p(x) = Cx^{(2\alpha\bar{r})/\sigma^2 - 1} e^{-(2\alpha/\sigma^2)x}, \quad x \geq 0,$$

where $C > 0$ is the normalizing constant satisfying $\int_{-\infty}^{\infty} p(x)dx = 1$. To ensure that $\lim_{x \rightarrow 0+} p(x) = 0$, we impose the condition that $2\alpha\bar{r} > \sigma^2$. What happens if $2\alpha\bar{r} = \sigma^2$? (Hint: See Fig. 23.10.)

Fig. 23.10 Limit of the pdf of the interest rate r_t in the CIR model with $2\alpha\bar{r} = \sigma^2$



23.4 (Ho–Lee Model) Show that, in the Ho–Lee Model, for every fixed time to maturity $\tau = T - t$ we have $\text{Var}(dR(t, \tau)) = \text{Var}(dr_t) = \sigma^2 dt$.

Chapter 24

Numeraires



A *numeraire* is a reference asset against which all other assets are evaluated. For example, the concept of time value of money is equivalent to discounting assets using the risk-free bond as a numeraire. Sometimes, a suitable choice of a numeraire makes the computation of option prices easier. In this chapter we present examples of how to use numeraires to price financial derivatives including interest rate derivatives.

24.1 Change of Numeraire for a Binomial Tree Model

As an illustration of numeraire change we present a discrete time model. Consider a one period binomial tree model for an asset price movement from time $t = 0$ to $t = T$. The values of all the assets including the risk-free deposit B_t , the underlying asset (for a given option) S_t , and the replicating portfolio V_t are determined by the ratio to a numeraire, which is the reference asset in our computation of option price. Assume that $B_0 = 1$ and $B_T = e^{rT}$, where r is the risk-free interest rate for the time period of length T . Also assume that S_T can have two values uS_0 and dS_0 , where $d < e^{rT} < u$, depending on up and down states at T . See Fig. 24.1. As usual, we take a probability space $\Omega = \{u, d\}$, where u and d represent up and down states respectively, and choose a filtration $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_T = \{\emptyset, \Omega, \{u\}, \{d\}\}$.

Now we choose the underlying asset S_t as a numeraire. Then the discounted assets corresponding to B_t and S_t are denoted by

$$\tilde{B}_t = \frac{B_t}{S_t}$$

and

$$\tilde{S}_t = \frac{S_t}{S_t} = 1,$$

Fig. 24.1 The bank deposit and a risky asset price

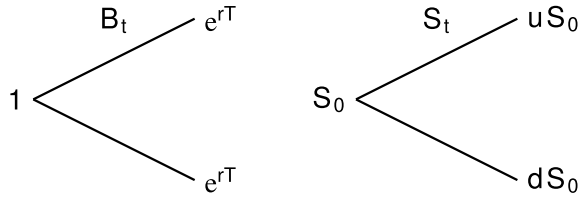
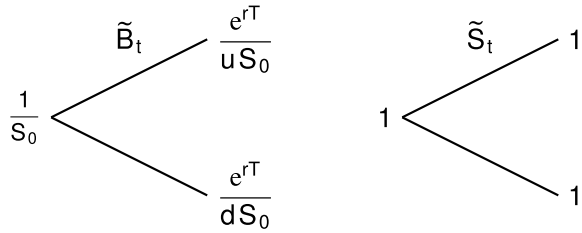


Fig. 24.2 The discounted bank deposit and a discounted risky asset price when the risky asset is a numeraire



respectively. For example, $\tilde{B}_0 = \frac{1}{S_0}$. See Fig. 24.2.

Note that $\tilde{S}_t$ is a martingale with respect to any probability measure $\mathbb{Q} = (q_u, q_d)$, i.e.,

$$\mathbb{E}^{\mathbb{Q}}[\tilde{S}_T | \mathcal{F}_0] = \tilde{S}_0,$$

since $\tilde{S}_T = \tilde{S}_0 = 1$. (See the diagram on the right in Fig. 24.2).

Lemma 24.1 *There exists a probability measure $\mathbb{Q} = (q_u, q_d)$ for which the following statements hold true:*

- (i) *The discounted bank deposit $\tilde{B}_t = \frac{B_t}{S_t}$ in terms of the numeraire S_t is a martingale, i.e.,*

$$\mathbb{E}^{\mathbb{Q}}[\tilde{B}_T | \mathcal{F}_0] = \tilde{B}_0.$$

- (i) *Consider a portfolio $V_t = c_1 S_t + c_2 B_t$. Then $\tilde{V}_t = \frac{V_t}{S_t}$ is a $\mathbb{Q}$ -martingale.*

Proof (i) It suffices to show

$$q_u \left(\frac{e^{rT}}{uS_0} \right) + q_d \left(\frac{e^{rT}}{dS_0} \right) = \frac{1}{S_0} \quad (24.1)$$

together with

$$q_u + q_d = 1 \quad (24.2)$$

for some $q_u > 0$ and $q_d > 0$. Note that the system of linear equations defined by (24.1) and (24.2) is equivalent to

$$\begin{cases} dq_u + uq_d = ude^{-rT} \\ +q_d = 1 \end{cases}$$

which has a solution

$$q_u = \frac{u(1 - de^{-rT})}{u - d}, \quad q_d = \frac{d(ue^{-rT} - 1)}{u - d}.$$

By the no arbitrage principle, we have $dS_0 < e^{rT}S_0 < uS_0$, and hence $q_u > 0$, $q_d > 0$.

(ii) By the linearity of conditional expectation, we have

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[\tilde{V}_T | \mathcal{F}_0] &= \mathbb{E}^{\mathbb{Q}}[c_1 \tilde{S}_T + c_2 \tilde{B}_T | \mathcal{F}_0] \\ &= c_1 \mathbb{E}^{\mathbb{Q}}[\tilde{S}_T | \mathcal{F}_0] + c_2 \mathbb{E}^{\mathbb{Q}}[\tilde{B}_T | \mathcal{F}_0] \\ &= c_1 \tilde{S}_0 + c_2 \tilde{B}_0, \end{aligned}$$

and hence $\mathbb{E}^{\mathbb{Q}}[\tilde{V}_T | \mathcal{F}_0] = \tilde{V}_0$. □

Now we apply the preceding idea to compute the option value at time 0.

Theorem 24.1 *With the underlying asset as a numeraire, we have the same option pricing formula given in Sect. 14.2 as follows:*

$$V_0 = e^{-rT} \left(\frac{e^{rT} - d}{u - d} V^u + \frac{u - e^{rT}}{u - d} V^d \right).$$

Proof By Lemma 24.1 we have

$$\frac{V_0}{S_0} = \mathbb{E}^{\mathbb{Q}} \left[\frac{V_T}{S_T} \right] = q_u \frac{V^u}{uS_0} + q_d \frac{V^d}{dS_0},$$

and hence

$$V_0 = q_u \frac{1}{u} V^u + q_d \frac{1}{d} V^d = \frac{1 - de^{-rT}}{u - d} V^u + \frac{ue^{-rT} - 1}{u - d} V^d.$$

□

24.2 Change of Numeraire for Continuous Time

Using the stock price S_t as a numeraire in the continuous time model, we derive the formula for the price of a European call option. Let $B_t = e^{rt}$ be the risk-free asset price at time t , and take

$$\tilde{B}_t = \frac{B_t}{S_t}.$$

We will find an equivalent probability measure $\mathbb{Q}$ for which $\tilde{B}_t$ is a $\mathbb{Q}$ -martingale. By Ito's lemma,

$$d\left(\frac{1}{S_t}\right) = \left(-\frac{1}{S_t^2}\right) dS_t + \frac{1}{2} \frac{2}{S_t^3} (dS_t)^2 = -\frac{1}{S_t} \mu dt - \frac{1}{S_t} \sigma dW_t + \frac{1}{S_t} \sigma^2 dt.$$

Hence

$$S_t d\left(\frac{1}{S_t}\right) = (\sigma^2 - \mu)dt - \sigma dW_t. \quad (24.3)$$

Then

$$\begin{aligned} d\tilde{B}_t &= d\left(e^{rt} \frac{1}{S_t}\right) \\ &= re^{rt} \frac{1}{S_t} dt + e^{rt} \frac{1}{S_t} ((\sigma^2 - \mu)dt - \sigma dW_t) \\ &= \frac{e^{rt}}{S_t} ((r + \sigma^2 - \mu)dt - \sigma dW_t) \\ &= -\sigma \tilde{B}_t \left(\frac{r + \sigma^2 - \mu}{-\sigma} dt + dW_t \right), \end{aligned}$$

where we used

$$\left[dB_t, d\left(\frac{1}{S_t}\right) \right]_t = 0.$$

Put

$$\theta = \frac{r + \sigma^2 - \mu}{-\sigma} = \frac{\mu - r}{\sigma} - \sigma$$

and

$$X_t = W_t + \theta t.$$

Then

$$dS_t = \mu S_t dt + \sigma S_t dW_t = (r + \sigma^2) S_t dt + \sigma S_t dX_t. \quad (24.4)$$

Let $\mathbb{Q}$ be a probability measure defined by

$$d\mathbb{Q} = \exp\left(-\frac{1}{2}\theta^2 T - \theta W_T\right) d\mathbb{P}.$$

Then X_t is a $\mathbb{Q}$ -Brownian motion by Girsanov's theorem, and $\tilde{B}_t$ is a $\mathbb{Q}$ -martingale since

$$d\tilde{B}_t = -\sigma \tilde{B}_t dX_t. \quad (24.5)$$

Define $\tilde{V}_t$ and V_t by

$$\tilde{V}_t = \frac{V_t}{S_t} = \mathbb{E}^{\mathbb{Q}} \left[\frac{C_T}{S_T} \middle| \mathcal{F}_t \right],$$

where C_T is the contingent claim on expiry date T . Then

$$V_T = S_T \mathbb{E}^{\mathbb{Q}} \left[\frac{C_T}{S_T} \middle| \mathcal{F}_T \right] = S_T \frac{C_T}{S_T} = C_T.$$

Now, to prove that V_t is the option price, it remains to show that V_t is self-financing. Since $\tilde{V}_t$ is a $\mathbb{Q}$ -martingale, by the Martingale Representation Theorem and (24.5) there exists a process α_t such that

$$\tilde{V}_t = \tilde{V}_0 + \int_0^t \alpha_u dX_u = \tilde{V}_0 + \int_0^t \alpha_u \left(-\frac{1}{\sigma \tilde{B}_u} \right) d\tilde{B}_u.$$

Put

$$\phi_t = \alpha_t \left(-\frac{1}{\sigma \tilde{B}_t} \right).$$

Then

$$\tilde{V}_t = \tilde{V}_0 + \int_0^t \phi_u d\tilde{B}_u,$$

i.e.,

$$d\tilde{V}_t = \phi_t d\tilde{B}_t. \quad (24.6)$$

Now let

$$\psi_t = \tilde{V}_t - \phi_t \tilde{B}_t.$$

Then

$$\tilde{V}_t = \phi_t \tilde{B}_t + \psi_t,$$

and

$$V_t = \tilde{V}_t S_t = \phi_t B_t + \psi_t S_t. \quad (24.7)$$

Now let

$$dV_t = a_t dt + b_t dW_t$$

for some a_t and b_t . Then (24.3) implies that the covariation of V_t and $1/S_t$ is given by

$$\left[dV_t, d\left(\frac{1}{S_t}\right) \right]_t = -\frac{\sigma b_t}{S_t} dt. \quad (24.8)$$

From (24.6), we have

$$\frac{1}{S_t} dV_t + V_t d\left(\frac{1}{S_t}\right) + \left[dV_t, d\left(\frac{1}{S_t}\right)\right]_t = \phi_t \left\{ \frac{1}{S_t} dB_t + B_t d\left(\frac{1}{S_t}\right) \right\}.$$

Hence (24.3) and (24.8) imply that

$$dV_t + V_t ((\sigma^2 - \mu)dt - \sigma dW_t) - \sigma b_t dt = \phi_t \{ dB_t + B_t ((\sigma^2 - \mu)dt - \sigma dW_t) \}.$$

Then (24.7) implies that

$$dV_t = \phi_t dB_t - \psi_t S_t ((\sigma^2 - \mu)dt - \sigma dW_t) + \sigma b_t dt. \quad (24.9)$$

Thus, taking covariances with $d(1/S_t)$ of both sides of (24.9), we have

$$b_t \left(-\frac{\sigma}{S_t} \right) dt = (-\psi_t S_t)(-\sigma) \left(-\frac{\sigma}{S_t} \right) dt.$$

Hence

$$b_t = \sigma \psi_t S_t. \quad (24.10)$$

Substituting (24.10) in (24.9), we obtain

$$\begin{aligned} dV_t &= \phi_t dB_t - \psi_t S_t ((\sigma^2 - \mu)dt - \sigma dW_t) + \sigma^2 \psi_t S_t dt \\ &= \phi_t dB_t - \psi_t S_t (-\mu dt - \sigma dW_t) \\ &= \phi_t dB_t + \psi_t dS_t. \end{aligned} \quad (24.11)$$

Now (24.7) and (24.11) show that V_t is self-financing.

Next, since $\tilde{V}_t$ is a $\mathbb{Q}$ -martingale, we have

$$\frac{V_0}{S_0} = \tilde{V}_0 = \mathbb{E}^{\mathbb{Q}}[\tilde{V}_T] = \mathbb{E}^{\mathbb{Q}}\left[\frac{V_T}{S_T}\right] = \mathbb{E}^{\mathbb{Q}}\left[\frac{C_T}{S_T}\right].$$

For a European call option with strike price K and expiry date T , we have

$$C_T = (S_T - K) \mathbb{1}_{\{S_T \geq K\}},$$

and hence

$$\begin{aligned} V_0 &= S_0 \mathbb{E}^{\mathbb{Q}} \left[\frac{(S_T - K) \mathbb{1}_{\{S_T \geq K\}}}{S_T} \right] \\ &= S_0 \mathbb{E}^{\mathbb{Q}} [\mathbb{1}_{\{S_T \geq K\}}] - S_0 K \mathbb{E}^{\mathbb{Q}} \left[\frac{\mathbb{1}_{\{S_T \geq K\}}}{S_T} \right]. \end{aligned} \quad (24.12)$$

Note that

$$\begin{aligned}
& \mathbb{E}^{\mathbb{Q}}[\mathbb{1}_{\{S_T \geq K\}}] \\
&= \mathbb{Q}(\{S_T \geq K\}) \\
&= \mathbb{Q}(\{S_0 e^{(r + \frac{1}{2}\sigma^2)T + \sigma X_T} \geq K\}) \quad (\text{by (24.4)}) \\
&= \mathbb{Q}\left(\left\{X_T \geq \frac{\log \frac{K}{S_0} - (r + \frac{1}{2}\sigma^2)T}{\sigma}\right\}\right) \\
&= \mathbb{Q}\left(\left\{X_1 \geq \frac{\log \frac{K}{S_0} - (r + \frac{1}{2}\sigma^2)T}{\sigma \sqrt{T}}\right\}\right) \quad (\text{since } X_T, \sqrt{T}X_1 \sim N(0, T)) \\
&= \mathbb{Q}\left(\left\{X_1 \leq \frac{\log \frac{S_0}{K} + (r + \frac{1}{2}\sigma^2)T}{\sigma \sqrt{T}}\right\}\right) \quad (\text{since } X_1 \sim N(0, 1)) \\
&= N(d_1), \tag{24.13}
\end{aligned}$$

and that

$$\begin{aligned}
& S_0 \mathbb{E}^{\mathbb{Q}} \left[\frac{\mathbb{1}_{\{S_T \geq K\}}}{S_T} \right] \\
&= \mathbb{E}^{\mathbb{P}} \left[\mathbb{1}_{\{S_T \geq K\}} \exp \left(-(\mu - \frac{1}{2}\sigma^2)T - \sigma W_T \right) \exp \left(-\frac{1}{2}\theta^2 T - \theta W_T \right) \right].
\end{aligned}$$

Now put

$$\tilde{\theta} = \sigma + \theta = \frac{\mu - r}{\sigma}.$$

Then

$$\mu - \frac{1}{2}\sigma^2 + \frac{1}{2}\theta^2 = r + \frac{1}{2}\tilde{\theta}^2.$$

Now define another equivalent probability measure $\mathbb{Q}'$ by

$$\frac{d\mathbb{Q}'}{d\mathbb{P}} = \exp \left(-\frac{1}{2}\tilde{\theta}^2 T - \tilde{\theta} W_T \right).$$

Let

$$Y_t = W_t + \tilde{\theta}t.$$

Then Y_t is a $\mathbb{Q}'$ -Brownian motion, and $\tilde{S}_t = e^{-rt} S_t = \frac{S_t}{B_t}$ is a $\mathbb{Q}'$ -martingale. Then

Table 24.1 Equivalent measures, Brownian motions and martingales

Probability	Brownian motion	Geometric Brownian motion	Martingale
$\mathbb{P}$	W_t	$dS_t = \mu S_t dt + \sigma S_t dW_t$	
$\mathbb{Q}$	$X_t = W_t + \theta t$	$dS_t = (r + \sigma^2) S_t dt + \sigma S_t dX_t$	B_t / S_t
$\mathbb{Q}'$	$Y_t = W_t + \tilde{\theta} t$	$dS_t = r S_t dt + \sigma S_t dY_t$	S_t / B_t

$$\begin{aligned}
S_0 K \mathbb{E}^{\mathbb{Q}} \left[\frac{\mathbb{1}_{\{S_T \geq K\}}}{S_T} \right] &= K e^{-rT} \mathbb{E}^{\mathbb{P}} \left[\mathbb{1}_{\{S_T \geq K\}} \exp \left(-\frac{1}{2} \tilde{\theta}^2 T - \tilde{\theta} W_T \right) \right] \\
&= K e^{-rT} \mathbb{E}^{\mathbb{Q}'} [\mathbb{1}_{\{S_T \geq K\}}] \\
&= K e^{-rT} \mathbb{Q}'(\{S_T \geq K\}) \\
&= K e^{-rT} \mathbb{Q}'(\{S_0 e^{(r - \frac{1}{2}\sigma^2)T + \sigma Y_T} \geq K\}) \\
&= K e^{-rT} N(d_2).
\end{aligned} \tag{24.14}$$

From (24.12), (24.13) and (24.14) we may rewrite the Black–Scholes–Merton formula as

$$V_0 = S_0 \mathbb{Q}(\{S_T \geq K\}) - K e^{-rT} \mathbb{Q}'(\{S_T \geq K\}). \tag{24.15}$$

In Table 24.1 we list three equivalent probability measures $\mathbb{P}$, $\mathbb{Q}$ and $\mathbb{Q}'$ introduced in the preceding discussion, the corresponding Brownian motions W_t , X_t and Y_t , respectively; and the martingales $\frac{B_t}{S_t}$ and $\frac{S_t}{B_t}$ with respect to $\mathbb{Q}$ and $\mathbb{Q}'$, respectively. Note that (24.15) expresses the option price using two in-the-money probabilities.

Remark 24.1 (i) An asset-or-nothing European call option with strike price K and expiry date T has payoff equal to $S_T \mathbb{1}_{\{S_T > K\}}$. Its price V_0^{asset} at time $t = 0$ satisfies

$$\frac{V_0^{\text{asset}}}{S_0} = \mathbb{E}^{\mathbb{Q}} \left[\frac{S_T \mathbb{1}_{\{S_T \geq K\}}}{S_T} \right] = \mathbb{Q}(\{S_T \geq K\}).$$

Hence $V_0^{\text{asset}} = S_0 N(d_1)$. See also Exercise 17.2.

(ii) A cash-or-nothing European call option with strike price K and expiry date T has payoff equal to $\mathbb{1}_{\{S_T \geq K\}}$. Its price V_0^{cash} at time $t = 0$ satisfies

$$\frac{V_0^{\text{cash}}}{B_0} = \mathbb{E}^{\mathbb{Q}'} \left[\frac{\mathbb{1}_{\{S_T \geq K\}}}{B_T} \right] = e^{-rT} \mathbb{Q}'(\{S_T \geq K\}).$$

Hence $V_0^{\text{cash}} = e^{-rT} N(d_2)$ as shown in Theorem 17.2.

24.3 Numeraires for Pricing of Interest Rate Derivatives

Let $P(r, t; T)$ be the price of the (zero coupon) bond with maturity T , which is used as a numeraire to discount an interest rate derivative $V(r, t; T)$, and let

$$\tilde{V}(r, t; T) = \frac{V(r, t; T)}{P(r, t; T)}.$$

Once it is understood that the maturity dates for P and V are both equal to T , and that the time t has the range $0 \leq t \leq T$, we write $P(r, t)$, $V(r, t)$, $\tilde{V}(r, t)$ instead of $P(r, t; T)$, $V(r, t; T)$, $\tilde{V}(r, t; T)$, respectively, when there is no danger of confusion. In the following, as in the case for V , we have the fundamental equation for pricing $\tilde{V}$, too.

Theorem 24.2 (Fundamental Equation for Pricing Discounted Interest Rate Derivatives) *Let*

$$dr_t = m(r_t, t) dt + \sigma(r_t, t) dW_t$$

describe the interest rate process, where W_t is a $\mathbb{Q}$ -Brownian motion for a risk-neutral probability measure $\mathbb{Q}$. Then a discounted interest rate derivative $\tilde{V}$ satisfies

$$\frac{\partial \tilde{V}}{\partial t} + (m(r, t) + \sigma_P(r, t)\sigma(r, t)) \frac{\partial \tilde{V}}{\partial r} + \frac{1}{2} \sigma(r, t)^2 \frac{\partial^2 \tilde{V}}{\partial r^2} = 0, \quad (24.16)$$

where

$$\sigma_P(r, t) = \frac{1}{P(r, t)} \frac{\partial P(r, t)}{\partial r} \sigma(r, t).$$

Proof Since $V = P\tilde{V}$, we have

$$\begin{cases} \frac{\partial V}{\partial t} = \frac{\partial P}{\partial t} \tilde{V} + P \frac{\partial \tilde{V}}{\partial t} \\ \frac{\partial V}{\partial r} = \frac{\partial P}{\partial r} \tilde{V} + P \frac{\partial \tilde{V}}{\partial r} \\ \frac{\partial^2 V}{\partial r^2} = \frac{\partial^2 P}{\partial r^2} \tilde{V} + \frac{\partial P}{\partial r} \frac{\partial \tilde{V}}{\partial r} + \frac{\partial P}{\partial r} \frac{\partial \tilde{V}}{\partial r} + P \frac{\partial^2 \tilde{V}}{\partial r^2}. \end{cases}$$

Since Theorem 22.3 holds for P , we have

$$\frac{\partial P}{\partial t} + m(r, t) \frac{\partial P}{\partial r} + \frac{1}{2} \sigma(r, t)^2 \frac{\partial^2 P}{\partial r^2} = rP,$$

and hence the following equation, obtained from Theorem 22.3,

$$\begin{aligned} \left(\frac{\partial P}{\partial t} \tilde{V} + P \frac{\partial \tilde{V}}{\partial t} \right) + m(r, t) \left(\frac{\partial P}{\partial r} \tilde{V} + P \frac{\partial \tilde{V}}{\partial r} \right) \\ + \frac{1}{2} \sigma(r, t)^2 \left(\frac{\partial^2 P}{\partial r^2} \tilde{V} + 2 \frac{\partial P}{\partial r} \frac{\partial \tilde{V}}{\partial r} + P \frac{\partial^2 \tilde{V}}{\partial r^2} \right) = rV \end{aligned}$$

is simplified to

$$rP\tilde{V} + P \left(\frac{\partial \tilde{V}}{\partial t} + m(r, t) \frac{\partial \tilde{V}}{\partial r} + \frac{1}{2} \sigma(r, t)^2 \frac{\partial^2 \tilde{V}}{\partial r^2} \right) + \sigma(r, t)^2 \frac{\partial P}{\partial r} \frac{\partial \tilde{V}}{\partial r} = rV,$$

which is further reduced to

$$P \left(\frac{\partial \tilde{V}}{\partial t} + m(r, t) \frac{\partial \tilde{V}}{\partial r} + \frac{1}{2} \sigma(r, t)^2 \frac{\partial^2 \tilde{V}}{\partial r^2} \right) + \sigma(r, t)^2 \frac{\partial P}{\partial r} \frac{\partial \tilde{V}}{\partial r} = 0$$

since $rP\tilde{V} = rV$. After dividing by P , we obtain

$$\frac{\partial \tilde{V}}{\partial t} + m(r, t) \frac{\partial \tilde{V}}{\partial r} + \frac{1}{2} \sigma(r, t)^2 \frac{\partial^2 \tilde{V}}{\partial r^2} + \sigma(r, t)^2 \frac{1}{P} \frac{\partial P}{\partial r} \frac{\partial \tilde{V}}{\partial r} = 0.$$

Finally, we take $\sigma_P = \sigma \frac{1}{P} \frac{\partial P}{\partial r}$. □

Remark 24.2 The fundamental equation (24.16) for $\tilde{V}$ in Theorem 24.2 is different from the result (22.6) for V in Theorem 22.3 in two ways: First, the right-hand side is zero, and the second, $\frac{\partial \tilde{V}}{\partial r}$ has a coefficient that has one more term $\sigma_P(r, t)\sigma(r, t)$.

Definition 24.1 Put

$$\theta = \frac{m - (m + \sigma_P \sigma)}{\sigma} = -\sigma_P,$$

and define an equivalent probability measure $\mathbb{Q}^T$, called the *forward risk-neutral* or *T-forward risk-neutral measure*, by

$$\mathbb{E}^{\mathbb{Q}} \left[\frac{d\mathbb{Q}^T}{d\mathbb{Q}} \middle| \mathcal{F}_t \right] = \exp \left(-\frac{1}{2} \int_0^t \theta_s^2 ds - \int_0^t \theta_s dW_s \right),$$

where $\{\mathcal{F}_t\}$ is the filtration defined by the information up to t .

Let

$$\tilde{W}_t = W_t + \int_0^t \theta_s ds.$$

Then, by Theorem 11.2, $\tilde{W}_t$ is a $\mathbb{Q}^T$ -Brownian motion and

$$dr = m dt + \sigma (d\tilde{W} - \theta dt) = (m + \sigma_P \sigma) dt + \sigma d\tilde{W}.$$

Theorem 24.3 If $\tilde{V}(r, T) = g_T$ is the payoff of a discounted interest rate derivative at maturity T , then

$$\tilde{V}(r, t) = \mathbb{E}^{\mathbb{Q}^T} [g_T | \mathcal{F}_t],$$

where the expectation is taken over the interest rate given by

$$dr_t = (m(r_t, t) + \sigma_P(r, t)\sigma(r, t)) dt + \sigma(r_t, t) d\tilde{W}_t.$$

In other words,

$$V(r, t; T) = P(r, t; T) \mathbb{E}^{\mathbb{Q}^T} [g_T | \mathcal{F}_t].$$

Proof Use Theorems 13.1 and 24.2. □

Remark 24.3 (i) The bond price $P(r, t; T)$ is known in the market at time t .
(ii) Recall that Theorem 22.4 states that

$$\begin{aligned} V(r, t; T) &= \mathbb{E}^{\mathbb{Q}} [e^{-\int_t^T r_u du} g_T | \mathcal{F}_t] \\ &= \mathbb{E}^{\mathbb{Q}} [e^{-\int_t^T r_u du} | \mathcal{F}_t] \mathbb{E}^{\mathbb{Q}} [g_T | \mathcal{F}_t] \\ &= P(r, t; T) \mathbb{E}^{\mathbb{Q}} [g_T | \mathcal{F}_t], \end{aligned}$$

since $e^{-\int_t^T r_u du}$ and g_T are not necessarily uncorrelated with respect to $\mathbb{Q}$.

Part IX

Python Programming

Chapter 25

Examples of Python Codes



In this chapter we include some representative Python programs which can be modified for other experiments. For more information on Python programming, consult Johansson (2019), Weiming (2019).

In the beginning of a Python code, we start with the command

```
import math
```

to import the mathematics module. Comments may be inserted between two lines containing `"""`, or after `#`. In a Python code, indentation rules should be respected.

25.1 Basic Commands

25.1.1 Arithmetic

Here are the arithmetic operations:

```
a = 10; b = 2; c = 30; d = 7; e = 22;

a += 1    # a = a + 1
b -= 10   # b = b - 10
c *= 2    # c = c * 2
d /= 3    # d = d / 3
e %= 4    # remainder left over after division

print(a,b,c,d,e)

11 -8 60 2.3333333333333335 2
```

We compute the sum $1 + 2 + 3 + 4$ as follows:

```
sum = 0
for n in range(1, 5):
    sum = sum + n
    print("n =", n, "sum =", sum)
```

Observe that `range(1,5)` means the iteration stops just one step before it reaches 5. The output is given as follows:

```
n = 1 sum = 1
n = 2 sum = 3
n = 3 sum = 6
n = 4 sum = 10
```

The same result can be obtained by using `while`.

```
sum = 0
n = 1
while n <= 4:
    sum = sum + n
    print("n =", n, "sum =", sum)
    n = n + 1
```

25.1.2 *Logic*

Here are five logical statements with their corresponding logical values:

```
a = True
b = False
print(a, b, a and b, a or b, not a)

True False False True False
```

Here are four logical statements with their corresponding logical values:

```
m = 2
n = 1
print(m == n)
print(m <= n)
print(m != n)
print(m > m)
```

```
False
False
True
False
```

Here are two logical statements with their corresponding logical values:

```
S = ['alpha', 'beta', 30, 40, 50]
print('beta' in S)
print(40 in S)

True
True
```

25.1.3 *Conditions for Loops*

In the following if-else statement, keep the sentences indented.

```
x = 10
if x >= 0:
    print("x is nonnegative.")
else:
    print("x is negative.")

x is nonnegative.
```

We factorize an integer into a product of prime numbers.

```
x = 300; # prime factorization
p = 2

while p <= x:
    if x % p == 0:
        print(p)
        x = x / p
    else:
        p = p + 1

2
2
3
5
5
```

25.1.4 *Random Numbers*

We randomly generate five integers between 0 and 2.

```
import random
for k in range(5):
    a = random.randint(0, 2)
    print(a)

0
2
2
1
```

25.1.5 *Definition of a Function in Python*

We define a function that computes the area of a rectangle.

```
def rect(a, b): # area of a rectangle with its length a and width b
    c = a * b
    return c

print(rect(3,5))

15
```

We define a function that computes the factorial.

```
def factorial(n):
    prod = 1
    for k in range(1, n+1):
        prod = prod * k
    return prod

print(factorial(5))

120
```

We define a recursive function, and add the integers from 1 to $n = 100$.

```
def sum_1_to_n(n):
    if n == 0:
```

```

        return 0
    sum1n = n + sum_1_to_n(n-1)
    return sum1n

print(sum_1_to_n(100))

```

Here is the output.

```
5050
```

25.1.6 Set Operations

First, we define the sets.

```

A = {0, 1, 2}
B = {x for x in range(1, 10) if x % 2 == 0} # B = {2, 4, 6, 8}
print(B)

{8, 2, 4, 6}

```

Next, we check whether $1 \in A$ and $1 \in B$.

```

print(1 in A) # belong to
print(1 in B) # belong to

True
False

```

Find $A \cup B$, $A \cap B$, $A \setminus B$

```

print(A | B) # set union
print(A & B) # set intersection
print(A - B) # set minus

{0, 1, 2, 4, 6, 8}
{2}
{0, 1}

```

Check whether $A \subset B$.

```
print(A < B)    # set inclusion
```

```
False
```

Compute the cardinality of a set.

```
len(A)
```

```
3
```

25.1.7 Sequences

Here are the basic commands for selecting elements of a given sequence.

```
import numpy as np
```

```
v = np.arange(0,10)
```

```
print('v = ', v)
```

```
print('v[0] = ', v[0])
```

```
print('v[9] = ', v[9])
```

```
print('v[-1] = ', v[-1])
```

```
print('v[-2] = ', v[-2])
```

```
print('v[:] = ', v[:])
```

```
print('v[1:3] = ', v[1:3])
```

```
print('v[:3] = ', v[:3])
```

```
print('v[5:] = ', v[5:])
```

```
print('v[5:9] = ', v[5:9])
```

```
print('v[5:-1] = ', v[5:-1])
```

```
print('v[5:-2] = ', v[5:-2])
```

```
print('v[0:8:2] = ', v[0:8:2]) # with increment 2
```

```
print('v[::-1] = ', v[::-1]) # all the elements in reverse order
```

The outputs are given in the following:

```
v = [0 1 2 3 4 5 6 7 8 9]
```

```
v[0] = 0
```

```
v[9] = 9
```

```
v[-1] = 9
```

```
v[-2] = 8
```

```
v[:] = [0 1 2 3 4 5 6 7 8 9]
```



```

v[1:3] = [1 2]
v[:3] = [0 1 2]
v[5:] = [5 6 7 8 9]
v[5:9] = [5 6 7 8]
v[5:-1] = [5 6 7 8]
v[5:-2] = [5 6 7]
v[0:8:2] = [0 2 4 6]
v[::-1] = [9 8 7 6 5 4 3 2 1 0]

```

Manipulations of sequences are given in the following:

```

a = [1, 2, 3, 4]
print(len(a)) # length
print(a[0]) # the first element
print(a[-1]) # the last element

a.append(100) # Add an element at the end.
print(a)

a.pop() # Delete the last element.
print(a)

b = [1000, 2000]
c = a + b # juxtaposition of two sequences
print(c)

```

The output is given as follows:

```

4
1
4
[1, 2, 3, 4, 100]
[1, 2, 3, 4]
[1, 2, 3, 4, 1000, 2000]

```

25.1.8 *Copies of a Sequence*

Copies of a sequence are defined in two different ways with different results:

```

list1 = [1, 2, 3]
print('list1 = ', list1)

```

```

list2 = list1
print('list2 = ', list2)
list2.append(4)
print('list2 after append = ', list2)
print('list1 also changed = ', list1) # list1 is also changed.

list3 = [10, 20, 30]
print('list3 = ', list3)
list4 = list3[:]
print('list4 = ', list4)
list4.append(40)
print('list4 after append = ', list4)
print('list3 not changed = ', list3) # list4 is not changed.

```

The output is given as follows:

```

list1 = [1, 2, 3]
list2 = [1, 2, 3]
list2 after append = [1, 2, 3, 4]
list1 also changed = [1, 2, 3, 4]

list3 = [10, 20, 30]
list4 = [10, 20, 30]
list4 after append = [10, 20, 30, 40]
list3 not changed = [10, 20, 30]

```

25.1.9 Plotting of Graphs

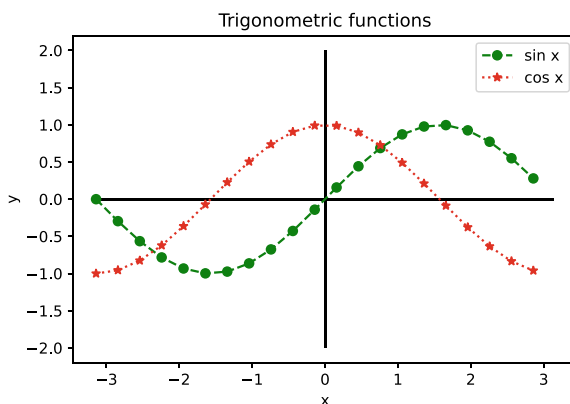
We plot the graphs of trigonometric functions.

```

import matplotlib.pyplot as plt
import numpy as np
import math

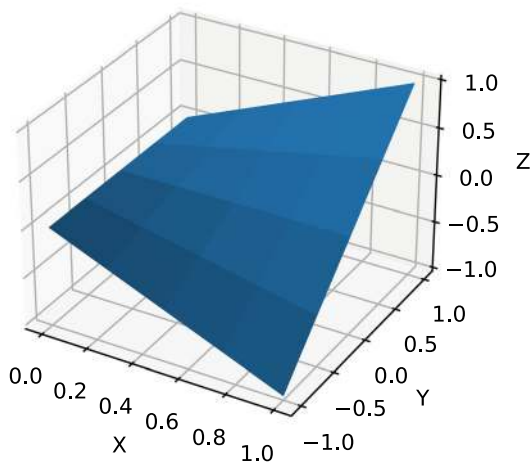
x = np.arange(-math.pi, math.pi, 0.3)
fig, axes = plt.subplots()
axes.vlines(0, -2, 2, color='k')
axes.hlines(0, -math.pi, math.pi, color='k')
plt.plot(x, np.sin(x), 'o--g', label='sin x')
plt.plot(x, np.cos(x), '*:r', label='cos x')
plt.xlabel('x')

```

Fig. 25.1 Graphs of trigonometric functions

```
plt.ylabel('y')
plt.title('Trigonometric functions')
plt.legend()
plt.show()
```

See Fig. 25.1, where different line styles, markers and colors are used for two graphs. For a function $Z = f(X, Y)$ of two variables X, Y , we first define a mesh grid and then plot a surface. See Fig. 25.2.

Fig. 25.2 Plot of $Z = XY$ over a mesh grid

```

np.linspace(10,20,5)

array([10. , 12.5, 15. , 17.5, 20. ])

x = np.linspace(0, 1, 5)
y = np.linspace(-1, 1, 5)
X, Y = np.meshgrid(x, y)
print('X = \n', X)
print('Y = \n', Y)

X =
[[0.   0.25 0.5   0.75 1.   ]
 [0.   0.25 0.5   0.75 1.   ]
 [0.   0.25 0.5   0.75 1.   ]
 [0.   0.25 0.5   0.75 1.   ]
 [0.   0.25 0.5   0.75 1.   ]]
Y =
[[-1.  -1.  -1.  -1.  -1. ]
 [-0.5 -0.5 -0.5 -0.5 -0.5]
 [ 0.   0.   0.   0.   0. ]
 [ 0.5  0.5  0.5  0.5  0.5]
 [ 1.   1.   1.   1.   1.   ]]

Z = X*Y

print(Z)

[[-0.   -0.25 -0.5   -0.75 -1.   ]
 [-0.   -0.125 -0.25 -0.375 -0.5 ]
 [ 0.    0.    0.    0.    0.   ]
 [ 0.    0.125 0.25  0.375 0.5 ]
 [ 0.    0.25  0.5   0.75  1.   ]]

import matplotlib.pyplot as plt
fig = plt.figure()
ax = plt.axes(projection='3d')
ax.plot_surface(X,Y,Z)
ax.set_xlabel("X")
ax.set_ylabel("Y")
ax.set_zlabel("Z")
plt.show()

```

25.1.10 Accumulation of Errors

Even an addition of the same constant c can cause accumulation of round-off error as shown below:

```
from math import fsum

N = 10**7
c = 0.1
list_c = [c] * N
true_value = c * N
error_sum = true_value - sum(list_c)
error_fsum = true_value - fsum(list_c)
print(f'error in summation = {error_sum:.5E}')
print(f'error in fsum = {error_fsum:.5E}')
```

The command `fsum` returns an accurate floating point sum of values.

```
error in summation = 1.61025E-04
error in fsum = 0.00000E+00
```

As a second example, consider the sum $\sum_{k=1}^N r^k = \frac{r(1-r^N)}{1-r}$ and observe that the order of summation matters when we need to pay attention to the numerical errors in floating point arithmetic. (See Kahan (1965).) In the following we take the sum in two different ways: first, we add a monotonically increasing sequence, i.e., add backward from $k = N$ to $k = 1$, and second, we add a monotonically decreasing sequence, i.e., add forward from $k = 1$ to $k = N$.

```
N = 10**7
r = 0.9999
sum_inc = 0 # sum of an increasing sequence
for k in range(N, 0, -1):
    sum_inc += r**k

sum_dec = 0 # sum of a decreasing sequence
for k in range(1, N + 1):
    sum_dec += r**k

true_value = r * (1 - r**N) / (1 - r)
error_inc = abs(true_value - sum_inc)
error_dec = abs(true_value - sum_dec)
print(f'error_inc = {error_inc:.5E}    error_dec = {error_dec:.5E}')
```

Compare the output given below:

```
error_inc = 3.63798E-11    error_dec = 6.38647E-09
```

We observe the same phenomenon with the sum $\sum_{k=1}^N \frac{1}{k(k+1)} = \frac{N}{N+1}$. In general, to significantly reduce numerical error in addition we employ the Kahan summation algorithm defined in the following. (See Kahan (1965).)

```
def kahanSum(list_a):
    summ = 0.0 # initial value for summation
    err = 0.0 # initial value for error
    for aj in list_a:
        y = aj - err
        # A possible error is subtracted from the previous iteration.
        tot = summ + y
        # A new number is added to the running total minus any error.
        err = tot - summ - y
        # Error from the new addition is calculated.
        summ = tot
        # The new total is assigned.
    return summ
```

Now we test the algorithm using the first example.

```
N = 10**7
a = 0.1
error = a * N - kahanSum(list_a=[a] * N)
print(f'error in Kahan algorithm = {error:.5E}')
```

We conclude that practically there is no error.

```
error in Kahan algorithm = 0.00000E+00
```

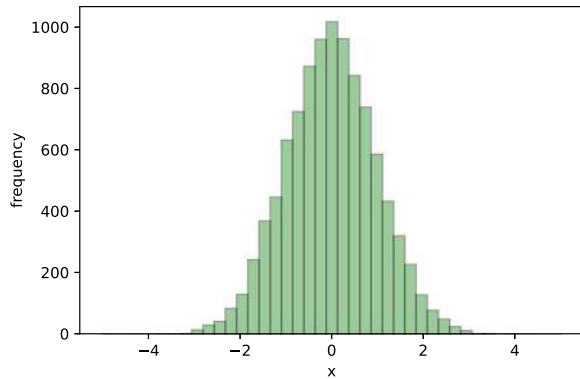
25.2 Statistics

25.2.1 Moments of a Data Set

We compute the mean, variance, skewness and kurtosis of a data set.

```
import numpy
import matplotlib.pyplot as plt
```

Fig. 25.3 Histogram for the standard normal distribution



```
import scipy.stats

np.random.seed(0)
x = np.random.normal(size=10*4) # the standard normal distribution

print('mean =', numpy.mean(x))
print('variance =', numpy.var(x))
print('skew =', scipy.stats.skew(x))
print('kurtosis =', scipy.stats.kurtosis(x, fisher=True))
```

The outcome is given in the following:

```
mean = -0.018433720158265818
variance = 0.9752679753477971
skew = 0.026634616738395577
kurtosis = -0.03095451095565238
```

Now we plot the histogram.

```
plt.hist(x, bins=41, range=(-5, 5), facecolor='g', edgecolor='k', alpha=0.4)
plt.title('Histogram of the Standard Normal Distribution')
plt.xlabel('x')
plt.ylabel('frequency')
plt.show()
```

See Fig. [25.3](#).

25.2.2 Correlation Coefficient and Scatter Plot

We compute the correlation coefficient ρ of two sets of data.

```
import numpy as np
import matplotlib.pyplot as plt

rho = 0.7
n = 1000
X = np.random.randn(n)
Y = np.random.randn(n)
Z = rho*X + np.sqrt(1-rho**2)*Y

print(np.corrcoef(X, Z))
```

The correlation coefficient matrix is given in the following:

```
[[1.          0.69823043]
 [0.69823043 1.          ]]
```

Now we plot the scatter plot.

```
plt.figure(figsize = (5,5))
plt.axis([-4,4,-4,4])
plt.grid()
plt.scatter(X,Z, s=3, color='b', alpha=0.5)
plt.xlabel('X')
plt.ylabel('Y')
plt.savefig('Python_scatterplot.pdf')
```

See Fig. [25.4](#).

25.3 Linear Algebra

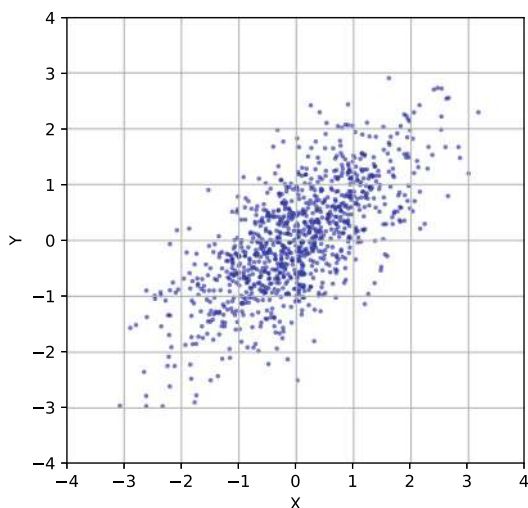
25.3.1 Arrays and Matrices

We present some basic operations with arrays and matrices.

```
import numpy as np

# the inner product of two vectors v1 and v2
v1 = ([2, 4, 2])
```


Fig. 25.4 Scatter plot of two correlated sets of data



```
v2 = ([3, -1, 0])
np.inner(v1, v2)
```

```
2
```

```
# definition of a matrix A
A = np.array([[2, 0, 1], [0, 2, 3], [-1, 0, 2], [0, 0, 1]])
print(np.shape(A)) # number of elements for each dimension
print(np.size(A)) # the total number of elements
print(np.ndim(A)) # number of dimensions
A
```

```
(4, 3)
```

```
12
```

```
2
```

```
array([[ 2,  0,  1],
       [ 0,  2,  3],
       [-1,  0,  2],
       [ 0,  0,  1]])
```

```
A.shape
```

```
(4, 3)
```

```
A.size
```

```
12
```

```
A.ndim
```

```
2
```

Find the diagonal elements of a square matrix B .

```
B = np.array([[0, 2, 3], [-1, 5, 6], [0, 7, 1]])

print('B = \n', B)

B =
[[ 0  2  3]
 [-1  5  6]
 [ 0  7  1]]

B[:, 1] # the second column

array([2, 5, 7])

np.diag(B) # diagonal elements

array([0, 5, 1])

np.diag([-4, 2, 3], k = 1) # an off-diagonal matrix

array([[ 0, -4,  0,  0],
       [ 0,  0,  2,  0],
       [ 0,  0,  0,  3],
       [ 0,  0,  0,  0]])

np.diag([-4, 2, 3], k = -1) # an off-diagonal matrix

array([[ 0,  0,  0,  0],
       [-4,  0,  0,  0],
       [ 0,  2,  0,  0],
       [ 0,  0,  3,  0]])
```

Define the identity matrix, a diagonal matrix, the zero matrix and the matrix of 1's.

```
np.identity(3) # the identity matrix

array([[1., 0., 0.],
```

```
[0., 1., 0.],
[0., 0., 1.])

np.eye(3) # the identity matrix

array([[1., 0., 0.],
       [0., 1., 0.],
       [0., 0., 1.]])

np.diag([-2, 6, 4, 1]) # the definition of a diagonal matrix

array([[ -2,  0,  0,  0],
       [  0,  6,  0,  0],
       [  0,  0,  4,  0],
       [  0,  0,  0,  1]])

b = np.array([-4, -1], dtype = complex) # complex numbers
np.sqrt(b)

array([0.+2.j, 0.+1.j])

np.sqrt(b).real # the real parts

array([0., 0.])

np.sqrt(b).imag # the imaginary parts

array([2., 1.])

np.zeros(shape=(3,6),dtype = int) # the zero matrix

array([[0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0],
       [0, 0, 0, 0, 0, 0]])

np.ones(shape=(4,3),dtype = int) # integers

array([[1, 1, 1],
       [1, 1, 1],
       [1, 1, 1],
       [1, 1, 1]])
```

25.3.2 A System of Linear Equations

We solve a system of linear equations $A\mathbf{x} = \mathbf{b}$.

```
import numpy as np

A = np.array([[2, 0, 1], [0, 2, 3], [-1, 0, 2]])
b = np.array([-5, 2, 7])
print('A = \n', A)
print('b = \n', b)

x = np.linalg.solve(A, b)
print('x = \n', x)
print('Ax = b ?\n', np.dot(A, x))
```

The outcome is given in the following:

```
A =
[[ 2  0  1]
 [ 0  2  3]
 [-1  0  2]]
b =
[-5  2  7]
x =
[-3.4 -1.7  1.8]
Ax = b ?
[-5.  2.  7.]
```

Compute the inverse matrix A^{-1} .

```
B = np.linalg.inv(A)
print(B)

print(np.dot(B,A))
print(np.dot(A,B))
```

The output is given as follows:

```
[[ 0.4  0. -0.2]
 [-0.3  0.5 -0.6]
 [ 0.2  0.  0.4]]

array([[ 1.00000000e+00,  0.00000000e+00,  0.00000000e+00],
       [ 0.00000000e+00,  1.00000000e+00, -2.22044605e-16],
```

```
[ 0.00000000e+00,  0.00000000e+00,  1.00000000e+00]])

array([[1., 0., 0.],
       [0., 1., 0.],
       [0., 0., 1.]])
```

25.3.3 *Eigenvalues and Eigenvectors*

We find eigenvalues and eigenvectors of a square matrix A .

```
import numpy as np
from numpy import linalg

A = np.array([[1, 0, 1],[0, -1, 1],[0, 1, 0]])
print('A =\n',A)

print('determinant =', np.linalg.det(A))
print('trace =', np.trace(A))

eigvals, eigvecs = linalg .eig(A)

print('eigenvalues =\n', eigvals)

print('eigenvectors =\n', eigvecs)

lambda1 = eigvals[0]
print('lambda1 =', lambda1)

v1 = eigvecs[:,0]
print('1st eigenvector v1 =\n', v1)

print('Av1 =\n', np.dot(A,v1))
```

The outcome is given in the following:

```
A =
[[ 1  0  1]
 [ 0 -1  1]
 [ 0  1  0]]

determinant = -1.0 trace = 0
```

```

eigenvalues =
[ 1.          -1.61803399  0.61803399]

eigenvectors =
[[ 1.          -0.19688101  0.91225304]
 [ 0.          -0.83400135 -0.21535373]
 [ 0.          0.51544118 -0.34844965]]

lambda1 = 1.0

1st eigenvector v1 =
[1. 0. 0.]

Av1 =
[1. 0. 0.]

```

25.3.4 *LU-Decomposition*

We find matrices P, L, U such that $A = PLU$ for a given matrix A , where P is a permutation matrix, L is a lower triangular matrix and U is an upper triangular matrix.

```

import numpy as np
import scipy.linalg

A = np.array([[2, -3, 0],[0, 4, -1],[1, 0, -2]])
P, L, U = scipy.linalg.lu(A)

print('P = \n', P)
print('L = \n', L)
print('U = \n', U)

```

The outcome is given in the following:

```

P =
[[1. 0. 0.]
 [0. 1. 0.]
 [0. 0. 1.]]

L =

```

```
[[1.    0.    0.   ]
 [0.    1.    0.   ]
 [0.5   0.375 1.   ]]
```

U =

```
[[ 2.    -3.    0.   ]
 [ 0.     4.   -1.   ]
 [ 0.     0.  -1.625]]
```

25.3.5 Cholesky Decomposition

For a positive-definite matrix A we find a lower triangular matrix L such that $A = LL^t$.

```
import numpy as np

M = np.array([[2, -3, 0],[0, 4, -1],[1, 0, -2]])
A = np.dot(M, np.transpose(M)) # a positive definite matrix
print('A =\n', A)
eig_val, eig_vec = np.linalg.eig(A) # eigenvalues and eigenvectors
print("eigenvalues =", eig_val)

L = np.linalg.cholesky(A)
print("Cholesky matrix L = \n", L)

B = np.dot(L,np.transpose(L))
print('Check B = A ?\n', B)
```

The outcome is given in the following:

```
A =
[[ 13 -12  2]
 [-12  17  2]
 [ 2  2  5]]

eigenvalues = [27.16801685  0.89697844  6.93500471]

Cholesky matrix L =
[[ 3.60555128  0.          0.          ]
 [-3.32820118  2.43373723  0.          ]
 [ 0.5547002   1.58034885  1.48148749]]
```

```

Check B = A ?
[[ 13. -12.  2.]
 [-12.  17.  2.]
 [  2.   2.  5.]]

```

25.3.6 The Jacobi Method

We solve a system of linear equations $A\mathbf{x} = \mathbf{b}$ by the Jacobi method.

```

import numpy as np

def jacobi(A, b, n, tol = 10**(-10)):
    # Initializes x with zeroes with same shape and type as B
    x = np.zeros_like(b)

    for iter_count in range(n):
        x_new = np.zeros_like(x)
        for i in range(A.shape[0]):
            s1 = np.dot(A[i, :i], x[:i])
            s2 = np.dot(A[i, i + 1:], x[i + 1:])
            x_new[i] = (b[i] - s1 - s2) / A[i, i]
        if np.linalg.norm(x - x_new) < tol:
            break
        x = x_new
    return x

A = np.array([
    [10, -1, 2, 0],
    [-1, 11, -1, 3],
    [2, -1, 10, -1],
    [0, 3, -1, 8]])
b = np.array([6., 5, -11, 1]) n = 25

x = jacobi(A, b, n) print('x', '=', x)

```

The solution is given in the following:

```
x = [ 0.89858012  0.48005409 -1.25287356 -0.21162948]
```


25.3.7 *The Gauss–Seidel Method*

We solve a system of linear equations $A\mathbf{x} = \mathbf{b}$ by the Gauss–Seidel method.

```
import numpy as np

def gauss(A, b, n, tol = 10**(-10)):
    L = np.tril(A) # returns the lower triangular matrix of A
    U = A - L # decompose A = L + U
    L_inv = np.linalg.inv(L)
    x = np.zeros_like(b)

    for i in range(n):
        Ux = np.dot(U, x)
        x_new = np.dot(L_inv, b - Ux)
        if np.linalg.norm(x - x_new) < tol:
            break
        x = x_new
    return x

A = np.array([
    [10, -1, 2, 0],
    [-1, 11, -1, 3],
    [2, -1, 10, -1],
    [0, 3, -1, 8]])
b = np.array([6., 5, -11, 1])
n = 50
x = gauss(A, b, n)

print('x =',x)
```

The solution is given in the following:

```
x = [ 0.89858012  0.48005409 -1.25287356 -0.21162948]
```

Compare the result with the one obtained by the Jacobi method.

25.4 Probabilistic Experiments

25.4.1 *The Monty Hall Problem*

We simulate the Monty Hall problem with three doors.

```

import numpy as np

doors = {'Door 1', 'Door 2', 'Door 3'}
def monty_hall(n):
    unchanged = 0
    switched = 0
    for k in range(0,n):
        # The car is randomly placed behind one of three doors.
        door_car = np.random.choice(['Door 1', 'Door 2', 'Door 3'])
        # The player selects a door.
        door_selected = np.random.choice(['Door 1', 'Door 2', 'Door 3'])
        # The host opens a different door that does not have the car.
        door_opened = list(doors - {door_selected, door_car})[0]
        # The third door that remains
        door_remaining = list(doors - {door_selected, door_opened})[0]

        # Add "1" when the player stays with his selection and wins the car.
        unchanged = unchanged + (door_selected == door_car)

        # Add "1" when the player switches his selection and wins the car.
        switched = switched + (door_remaining == door_car)
    print("Probability of winning the car if:")
    print("the player stays with the original selection =", unchanged/n)
    print("the player switches doors =", switched/n)

monty_hall(n=10**4)

```

The output is given as follows:

```

Probability of winning the car if:
the player stays with the original selection = 0.3323
the player switches doors = 0.6677

```

25.4.2 Random Walks

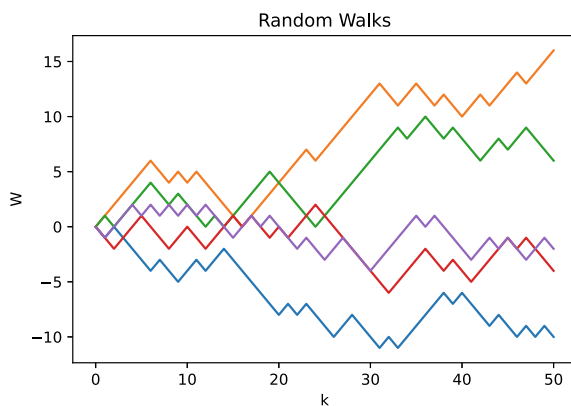
We simulate the random walks depending on the outcome of coin tossing. See Fig. 25.5.

```

import numpy as np
import matplotlib.pyplot as plt

def randomwalk(n):

```

Fig. 25.5 Sample paths of random walks

```

W = 0
positions = [W]
coin_tossings = np.arange(n+1)
possible_outcome = ["heads", "tails"]
for k in range(1, n+1):
    outcome = np.random.choice(possible_outcome)
    if outcome == "heads":
        W += 1
    elif outcome == "tails":
        W -= 1
    positions.append(W)
return coin_tossings, positions

for m in range(0, 5):
    k, W = randomwalk(50)
    plt.plot(k,W)

plt.xlabel('k')
plt.ylabel('W')
plt.title("Random Walks")
plt.savefig('Python_randomwalk.pdf')

```

25.5 Brownian Motion

25.5.1 *Brownian Motion*

We plot the sample paths of a Brownian motion.

```
import numpy as np
from math import sqrt
from matplotlib import pyplot as plt
randn = np.random.normal

T = 1 #time to expiry in years
N = 100 # number of steps
dt = T/N # time step
M = 10**1 # sample size

W = np.zeros([M,N]) # Brownian motion
t = range(0,N,1)

for k in range(0,M-1):
    W[k,0] = 0
    for x in range(0,N-1):
        W[k,x+1] = W[k,x] + sqrt(dt)*randn(0,1)
    plt.plot(t,W[k])

plt.title('Brownian Motion')
plt.xlabel('Time steps')
plt.ylabel('W')
plt.show()
```

25.5.2 *Geometric Brownian Motion*

We plot the sample paths of a geometric Brownian motion.

```
import numpy as np
from math import sqrt
from matplotlib import pyplot as plt
randn = np.random.normal

S0 = 100 # asset price at t=0
mu = 0.1 # risk-free interest rate
```

```

sigma = 0.3 # volatility
T = 1 #time to expiryin years
N = 100 # number of steps
dt = T/N # time step
M = 10**1 # sample size

S = np.zeros([M,N])
t = range(0,N,1)

for k in range(0,M-1):
    S[k,0] = S0
    for x in range(0,N-1):
        S[k,x+1] = S[k,x] + mu*S[k,x]*dt + sigma*S[k,x]*sqrt(dt)*randn(0,1)
    plt.plot(t,S[k])

plt.title('Geometric Brownian Motion')
plt.xlabel('Time steps')
plt.ylabel('S')
plt.show()

```

25.6 Probability Density of Geometric Brownian Motion

We plot the histogram of the values of a geometric Brownian motion.

```

import numpy as np
from math import sqrt
from matplotlib import pyplot as plt
randn = np.random.normal

S0 = 100 # asset price at t=0
mu = 0.1 # risk-free interest rate
sigma = 0.3 # volatility
T = 1 #time to expiryin years
N = 100 # number of steps
dt = T/N # time step
M = 10**4 # sample size

S = np.zeros([M,N]) # geometric Brownian motion
t = range(0,N,1)

for k in range(0,M-1):
    S[k,0] = S0

```

```

for x in range(0,N-1):
    S[k,x+1] = S[k,x] + mu*S[k,x]*dt + sigma*S[k,x]*sqrt(dt)*randn(0,1)

plt.hist(S[:,N-1], bins= 100)
plt.xlabel('S')
plt.ylabel('frequency')

```

25.7 Option Pricing Methods

25.7.1 *The Binomial Tree Model*

We use the binomial tree model to compute a European call option.

```

import numpy as np
from math import sqrt, exp

T = 1 # time to expiry
N = 6 # number of time steps
r = 0.05 # risk_free_rate
sigma = 0.3
K = 100.0 # strike
S0 = 100.0
dt = T/N

u = exp(sigma*sqrt(dt))
d = exp(-sigma*sqrt(dt))
q = (exp(r*dt)-d)/(u-d)

S = np.zeros([N+1,N+1]) # the binomial tree for asset
Eu = np.zeros([N+1,N+1]) # the binomial tree for European put price
Am= np.zeros([N+1,N+1]) # the binomial tree for American put price

S[0,0] = S0
for i in range(1,N+1):
    for j in range(0,i):
        S[i][j] = S[i-1][j]*d
    if(i <= N):
        S[i][i] = S[i-1][i-1]*u
print('Asset price for time step = 0,...,N \n', S.round(2))

for j in range(0,N+1):

```

```

Eu[N][j] = max(K - S[N][j],0)
Am[N][j] = max(K - S[N][j],0)

for i in range(N,0,-1):
    for j in range(0,i):
        Eu[i-1][j] = exp(-r*dt)*(Eu[i][j+1]*q + Eu[i][j]*(1-q))
print('European put price for time step = 0,...,N \n',Eu.round(2))

for i in range(N,0,-1):
    for j in range(0,i):
        aver= exp(-r*dt)*(Am[i][j+1]*q + Am[i][j]*(1-q))
        exercise_value = max(K - S[i-1][j],0)
        Am[i-1][j] = max(aver,exercise_value)
print('American put price for time step = 0,...,N \n',Am.round(2))

```

The output is given below:

```

Asset price for time step = 0,...,N
[[100.    0.    0.    0.    0.    0.    0. ]
 [ 88.47 113.03  0.    0.    0.    0.    0. ]
 [ 78.27 100.   127.76  0.    0.    0.    0. ]
 [ 69.25  88.47 113.03 144.4  0.    0.    0. ]
 [ 61.27  78.27 100.   127.76 163.21  0.    0. ]
 [ 54.21  69.25  88.47 113.03 144.4  184.48  0. ]
 [ 47.96  61.27  78.27 100.   127.76 163.21 208.52]]

```

```

European put price for time step = 0,...,N
[[ 8.88  0.    0.    0.    0.    0.    0. ]
 [13.66  4.31  0.    0.    0.    0.    0. ]
 [20.17  7.46  1.28  0.    0.    0.    0. ]
 [28.28 12.51  2.59  0.    0.    0.    0. ]
 [37.08 20.07  5.27  0.    0.    0.    0. ]
 [44.96 29.92 10.7   0.    0.    0.    0. ]
 [52.04 38.73 21.73  0.    0.    0.    0. ]]

```

```

American put price for time step = 0,...,N
[[ 9.61  0.    0.    0.    0.    0.    0. ]
 [14.8   4.65  0.    0.    0.    0.    0. ]
 [21.9   8.06  1.38  0.    0.    0.    0. ]
 [30.75 13.53  2.79  0.    0.    0.    0. ]
 [38.73 21.73  5.68  0.    0.    0.    0. ]
 [45.79 30.75 11.53  0.    0.    0.    0. ]
 [52.04 38.73 21.73  0.    0.    0.    0. ]]

```

25.7.2 *The Black–Scholes–Merton Formula*

We define the functions for the call and put option prices.

```
import numpy as np
from scipy.stats import norm
from math import log, sqrt, exp
import matplotlib.pyplot as plt
N = norm.cdf

def black_scholes_call(S,K,T,r,sigma):
    d1 = (np.log(S/K) + (r + 0.5*sigma**2)*T)/(sigma*np.sqrt(T))
    d2 = (np.log(S/K) + (r - 0.5*sigma**2)*T)/(sigma*np.sqrt(T))
    return S*N(d1) - K* np.exp(-r*T) * N(d2)

def black_scholes_put(S,K,T,r,sigma):
    d1 = (np.log(S/K) + (r + sigma**2/2)*T) / (sigma*np.sqrt(T))
    d2 = d1 - sigma* np.sqrt(T)
    return K*np.exp(-r*T)*N(-d2) - S*N(-d1)

S0 = 100 # asset price at t=0
K = 100 # strike price
T = 1 # time to expiry
r = 0.05 # risk-free interest rate
sigma = 0.3 # volatility

a1 = black_scholes_call(S0,K,T,r,sigma)
print(a1)
a2 = black_scholes_put(S0,K,T,r,sigma)
print(a2)

# We plot the graphs of call and put option prices for 0<S<160.
S = np.arange(0.001,160,1)

calls = [black_scholes_call(s, K, T, r, sigma) for s in S]
puts = [black_scholes_put(s, K, T, r, sigma) for s in S]

plt.plot(calls, label='Call price')
plt.plot(puts, label='Put price')
plt.xlabel('$S$')
plt.ylabel('Option Price')
plt.legend()
```


The output is given below:

```
14.231254785985826
9.354197236057232
```

25.7.3 Risk-Neutral Method for Option Pricing

```
import numpy as np
from math import sqrt, exp
randn = np.random.normal

S0 = 100 # asset price at t=0
K = 100 # strike price
r = 0.05 # risk-free interest rate
sigma = 0.3 # volatility
T = 1 # time to expiry in years
M = 10**6 # number of samples

S = np.zeros([M,1])
Payoff = np.zeros([M,1])

for j in range(0,M-1):
    S[j,0] = S0*np.exp((r-sigma**2/2)*T + sigma*sqrt(T)*randn(0,1))
    Payoff[j,0] = np.maximum(S[j,0]-K,0)
Price = np.exp(-r*T)*np.average(Payoff)
print('call price =', Price)
```

The price of the European call option is given below:

```
call price = 14.251674238762785
```

It agrees with the price obtained by the Black–Scholes–Merton formula in Sect. [25.7.2](#).

25.8 Bond Pricing

We use discrete time in pricing bonds.

25.8.1 *Zero-Coupon Bond*

We compute the price of a zero-coupon bond by discounting.

```
import numpy as np

def zero_coupon_bond(face_value, year, T):
    return np.exp(-year*T)*face_value

print(zero_coupon_bond(100, 0.05, 3))
```

The bond price is given in the following:

```
86.07079764250578
```

25.8.2 *Yield to Maturity*

We compute the yield to maturity.

```
import numpy as np
import scipy.optimize as optimize

def bond_ytm(price, par, T, coupon_year, freq = 1):
    periods = T*freq
    coupon = coupon_year/freq
    coupon_date = [(k+1)/freq for k in range(int(periods))]
    ytm_func = lambda y: sum([coupon/((1+y/freq)**t) for t in coupon_date]) \
        + par/((1+y/freq)**(freq*T)) - price
    return optimize.newton(ytm_func, 0.1)

ytm = bond_ytm(price = 95, par = 100, T = 3, coupon_year = 5, freq = 12)
print(ytm)
```

The yield to maturity is given in the following:

```
0.0740202164098959
```

25.8.3 *Pricing of Coupon Bonds*

We compute the price of a bond with coupon payments.

```
def bond_price(par, T, y, coupon, freq = 1):
    periods = T*freq
    coupon = coupon/freq
    coupon_date = [(k+1)/freq for k in range(int(periods))]
    price = sum([coupon*(1 + y/freq)**(-t) for t in coupon_date]) \
            + par*((1 + y/freq)**(-freq*T))

    return price

price = bond_price(par= 100, T = 3, y = 0.07402, coupon= 5, freq =
12) print(price)
```

The bond price is given in the following:

```
95.00005212052133
```

25.8.4 *Duration*

We compute the modified duration of a coupon bond with periodic compounding, using the approximation of $B'(y)$ given by the full central difference $(B(y + \delta y) - B(y - \delta y))/(2\delta y)$.

```
def bond_duration(price, par, T, coupon, freq, dy=0.01):
    y = bond_ytm(price, par, T, coupon, freq)
    y_plus = y + dy
    price_plus = bond_price(par, T, y_plus, coupon, freq)
    y_minus = y - dy
    price_minus = bond_price(par, T, y_minus, coupon, freq)
    duration = (price_minus - price_plus)/(2*price*dy)
    return duration

D = bond_duration(price=95, par=100, T=3, coupon=5 , freq=12, dy=0.01)
print(D)
```

The modified duration is approximately given in the following:

```
2.5355790222722048
```

25.8.5 Convexity

We compute the convexity of a coupon bond with periodic compounding, using the approximation of $B''(y)$ given by the second order central difference $(B(y + \delta y) - 2B(y) + B(y - \delta y))/\delta y^2$.

```
def bond_convexity(price, par, T, coupon, freq, dy = 0.01):
    y = bond_ytm(price, par, T, coupon, freq)
    y_plus = y + dy
    y_minus = y - dy
    price_plus = bond_price(par, T, y_plus, coupon, freq)
    price_minus = bond_price(par, T, y_minus, coupon, freq)
    convexity = (price_plus - 2*price + price_minus)/(price*dy**2)
    return convexity
```

```
C = bond_convexity(price=95, par=100, T=3, coupon=5, freq=12,
dy=0.01) print(C)
```

The convexity is approximately given in the following:

```
7.713492148917267
```

25.8.6 The Vasicek Model

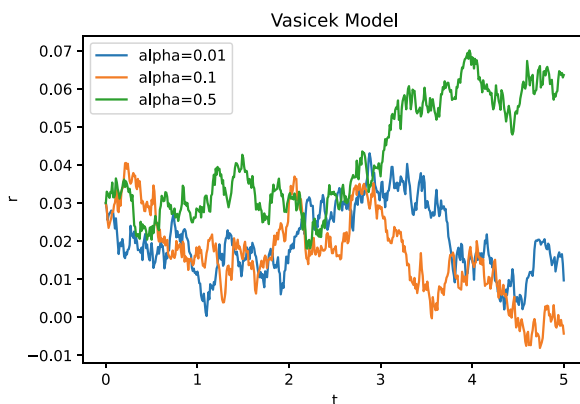
We simulate the interest rate process using the Vasicek model.

```
import numpy as np
from matplotlib import pyplot as plt

def vasicek(r0, alpha, rbar, sigma, T=1, N=10):
    dt = T / N
    r = [r0]
    for i in range(N):
        dr = alpha*(rbar - r[-1])*dt + sigma*np.sqrt(dt)*np.random.normal()
        r.append(r[-1] + dr)
    return np.arange(0, T + dt, dt), r

for alpha in [0.01, 0.1, 0.5]:
    x, r = vasicek(0.03, alpha, rbar=0.05, sigma=0.02, T=5, N=500)
    plt.plot(x, r, label='alpha=%s' %alpha)
    plt.legend(loc='upper left');
```

Fig. 25.6 Interest rates by the Vasicek model



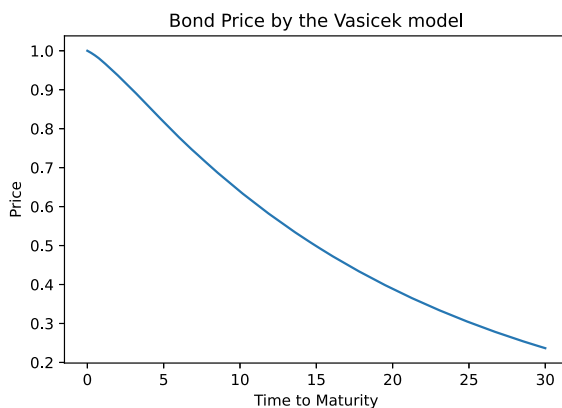
```
plt.legend(loc='upper left')
plt.xlabel('t')
plt.ylabel('r')
plt.title('Vasicek Model');
plt.savefig('Python_Vasicek_r.pdf')
```

See Fig. 25.6. Now we compute the price of a zero coupon bond.

```
def vasicek_price(rbar, alpha, sigma, tau, r0=0.02):
    B1 = (1 - np.exp(-alpha*tau)) / alpha
    A1 = (B1 - tau)*(rbar-sigma**2/(2*alpha**2)) - sigma**2/(4*alpha)*B1**2
    return np.exp(A1 - B1*r0)

x = np.arange(0, 30.25, 0.25)
```

Fig. 25.7 Interest rates by the Vasicek model



```
y = [vasicek_price(0.05, 0.6, 0.015, tau, r0=0.02) for tau in x]

plt.plot(x, y)
plt.xlabel("Time to Maturity");
plt.ylabel("Price")
plt.title("Bond Price by the Vasicek model");
plt.savefig('Python_Vasicek_price.pdf')
```

See Fig. [25.7](#).

Appendix A

MATLAB Programming

We present a brief introduction to the software MATLAB, providing just enough to understand and modify the programs given in this book. MATLAB[®] is a registered trademark of The MathWorks, Inc., and was developed to deal with mathematical problems mostly in applied mathematics. It stands for MATrix LABoratory, and as the name suggests, it is efficient in doing numerical computations involving vectors and matrices. For a more comprehensive introduction to various examples of MATLAB programs, see Moler (2004).

A.1 How to Start

The name of a MATLAB file ends with the extension ‘.m’. The command `clear all` removes all stored variables from the current workspace, freeing up system memory. The command `clc` clears the area on the screen called the Command Window, and the output is displayed in the same starting position on the screen. In the beginning of a MATLAB file it is convenient to include a line consisting of two commands `clear all; clc;`. If the given program plots a graph or graphs, an additional line containing `clf;` clears the current figure window(s). To save space the above three commands are usually not shown in MATLAB codes presented in this book.

If a command or a name of a variable ends without a semicolon (;) then the output is printed in the Command Window. If there is no need to see the output, just type a semicolon at the end of the line, then MATLAB will perform the command, but will not show the result on the screen. The following code lists a few basic mathematical operations.

```
1+4
9-2
5*6
2^7
```

produces

```
ans =
      5
ans =
      7
ans =
     30
ans =
    128
```

The following code

```
p=1/3
format long
p
a=exp(-1.2)
b=log(a)
get(0,'format') % Check the format for display of decimal numbers.
format short
b
```

produces

```
p =
    0.3333
p =
0.3333333333333333
a =
0.301194211912202
b =
-1.2000000000000000

ans =
long

b =
-1.2000
```

For trigonometric functions

```
pi
sin(pi)
cos(pi/2)
tan(pi/4)
```


produces

```
ans =
    3.1416
ans =
    1.2246e-16
ans =
    6.1232e-17
ans =
    1.0000
```

A.2 Random Numbers

To generate 2×5 matrices A , B , C consisting of random numbers with the uniform distribution in $(0, 1)$, with the standard normal distribution, and with the uniform distribution in the set of integers $-1 \leq n \leq 4$, respectively, we use

```
A = rand(2,5)
B = randn(2,5)
C = randi([-1,3],2,5)
```

which produces

```
A =
    0.4363    0.0261    0.4306    0.7624    0.6800
    0.1739    0.9547    0.9616    0.0073    0.7060
B =
    0.0831   -0.5279   -0.8499    0.7253   -0.3864
    0.1578    0.7231   -0.7964    1.6865   -0.5051
C =
     2     3     0     1    -1
     3     1     3     3     1
```

A.3 Vectors and Matrices

MATLAB can process a given group of objects more efficiently if they are expressed as vectors or matrices. Here are examples of some basic commands:

```
u = [1 2 3 4]
```

```

v = transpose(u)
w = zeros(4,1)
v + w
3*v
A = ones(3,4)
A*v
B = [1 2 3; 0 1 4; 2 3 1; -1 0 7]
C = A*B

```

There are two ways to compute the inverse of a matrix.

```

A = rand(3,3);
det(A) % Check the determinant to see whether the matrix is invertible.
I = diag(ones(3,1)); % the identity matrix
I / A;
A \ I;
inv(A);

E = rand(3,3);

E / A;
E*inv(A);

A \ E;
inv(A)*E;

```

It is better to use E/D instead of $E*\text{inv}(D)$ when we solve $XD = E$, and use $D \backslash E$ instead of $\text{inv}(D)*E$ when we solve $DX = E$.

The following defines a tridiagonal matrix.

```

a = -1; b = 2; c = 3;
m = 6;
T = diag(a*ones(m,1)) + diag(b*ones(m-1,1),1) + diag(c*ones(m-1,1),-1)

```

produces

```

T =
    -1     2     0     0     0     0
     3    -1     2     0     0     0
     0     3    -1     2     0     0
     0     0     3    -1     2     0
     0     0     0     3    -1     2
     0     0     0     0     3    -1

```

A.4 Loops for Iterative Algorithms

As an example of an iterative algorithm, we present a method of testing the uniform distribution of M pseudorandom numbers generated by the command `rand`. First, we divide the unit interval into N subintervals of equal length, $[0, \frac{1}{N})$, $[\frac{1}{N}, \frac{2}{N})$, $\dots$, $[\frac{N-1}{N}, 1)$, and count how many times the pseudorandom numbers belong to each subinterval.

```
M = 500;
N = 5;
x = rand(M,1);
width = 1/N;
bin = zeros(1,N); % initial values for each bin

for i = 1:M
    j = ceil(x(i)/width);
    bin(j) = bin(j) + 1;
end

bin
```

produces

```
bin =
    95    104    111    95    95
```

The i th random number $x(i)$ belongs to the bin numbered $\text{ceil}(x(i)/\text{width})$, where $\text{ceil}(x)$ for ‘ceiling’ is the smallest integer greater than or equal to x .

For the second example of an iterative algorithm, consider the Newton method for computing $\sqrt{5}$. Take $f(x) = x^2 - 5$.

```
format long
x0 = 2.0; % Choose a suitable starting point.
f = x0^2 - 5;
fprime = 2*x0;
x = x0 - f/fprime

while abs(x - x0) > 10^(-10)
    x0 = x;
    f = x^2 - 5;
    fprime = 2*x;
    x = x - f/fprime
```

end

produces

```
x =
    2.250000000000000
x =
    2.236111111111111
x =
    2.236067977915804
x =
    2.236067977499790
x =
    2.236067977499790
```

A.5 How to Plot Graphs

As an example, let us plot the graph of the standard normal density function $y = 1/(\sqrt{2\pi}) \exp(-x^2/2)$ on the interval $-5 \leq x \leq 5$. First, take 1001 equally spaced points in the interval where the function is evaluated.

```
x = -5:0.01:5;
```

Without the semicolon (;) at the end, there would be a display of $1001 = \frac{5-(-5)}{0.01} + 1$ points starting from -5 to 5 with equal increment 0.01 because x is regarded as a *vector* by MATLAB, i.e., $x = (-5.00, -4.99, -4.98, -4.97, -4.96, \dots)$. To evaluate the function we use

```
y = 1/(sqrt(2*pi))*exp(-x.^2/2);
```

where

```
x.^2
```

means that x^2 is computed for each entry x of the *vector* x .

```
plot(x,y,'Color','b','LineWidth',1.5)
```

produces a graph where b means the color blue. Other abbreviations for available colors are k for black, r for red, g for green, and y for yellow. The line width can be changed to 1 or 2 , or any positive number.

For a specification of the range of the plot, we use

```
set(gca, 'xlim', [-5 5], 'ylim', [0 0.5], 'xtick', [-4:1:4], 'ytick', [0:0.1:0.5]);
```

where `gca` means the current axes for plotting a graph, and `xlim` and `ylim` specify the ranges for the axes, and `xtick` and `ytick` specify the locations where tick marks are drawn.

A.6 Curve Fitting

For a set of two-dimensional data we find the best fitting polynomial in the least squares sense. The command `polyfit(X,Y,n)` finds the polynomial of degree n that best fits the data. In the following we find the best fitting quadratic polynomial $Y = aX^2 + bX + c$. (See Fig. A.1.)

```
x = 0:0.1:10;
X = [1.2 2.4 3.2 4.7 5.6 7.0 9.0];
Y = [1.9 2.4 2.5 3.0 4.2 5.7 7.3];
p = polyfit(X,Y,2) % Choose n=1 for a linear fit.
f = polyval(p,x);
plot(x,f)
hold on
plot(X,Y,'o')
```

The coefficients a, b, c are given below:

```
p =
    0.0631    0.0722    1.7048
```

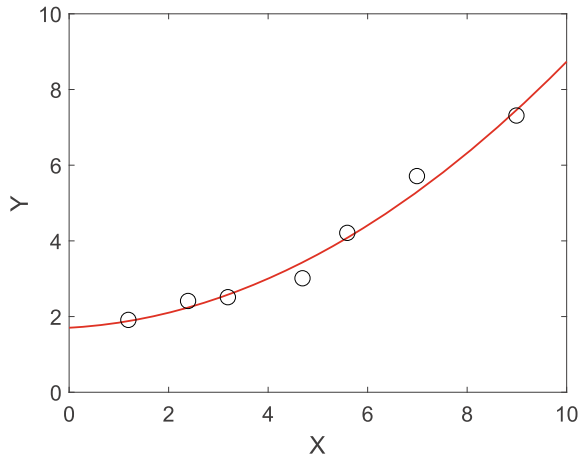
For a linear fit see Fig. C.1.

A.7 How to Define a Function

A simple method of defining a function such as $f(x) = \frac{\sin x}{x}$ is to use the command `inline` as follows:

```
f = inline('sin(x)/x')
x = 0.1;
f(x)
```

Fig. A.1 Polynomial fitting of data



which produces

```
f =
    Inline function:
    f(x) = sin(x)/x
ans =
    0.9983
```

For a complicated definition such as the Black–Scholes–Merton formula we define a separate MATLAB file and call it from the inside of another file. Here is an example. Define a function `BlackScholes()` of five variables by

$$\text{BlackScholes}(S, K, T, r, \sigma) = SN(d_1) - Ke^{-rT}N(d_2)$$

where $d_{1,2} = (\log \frac{S}{K} + (r \pm \frac{1}{2}\sigma^2)T)/(\sigma\sqrt{T})$. (We may choose any function name instead of `BlackScholes`, however, an informative name will be helpful to avoid confusion with other functions. The function name itself does not matter. Only the name of the file containing the definition is called from another file.) Save the definition of `BlackScholes()` as a separate MATLAB file with a title such as `BSM.m` as follows:

```
function [Call_price, Put_price] = BlackScholes(S,K,T,r,sigma)
% an example of a function file
d1 = (log(S/K)+(r + 0.5*sigma^2)*T)/(sigma*sqrt(T));
d2 = d1 - sigma*sqrt(T);
N1 = 0.5*( 1+erf(d1/sqrt(2)) );
N2 = 0.5*( 1+erf(d2/sqrt(2)) );
Call_price = S*N1 - K*exp(-r*T)*N2;
```

```
Put_price = Call_price + K*exp(-r*T) - S;
```

(The command `erf` is defined in Definition 4.2.) Input variables, S, K, T, r, σ , are written inside a pair of parentheses, and outputs, `Call_price`, `Put_price`, are written between a pair of brackets. From another MATLAB file we call the function file name `BSM` as follows:

```
S = 100;
K = 110;
T = 1;
r = 0.05;
sigma = 0.3;
[Call_price, Put_price] = BSM(S,K,T,r,sigma)
```

Then we obtain the following output in the command window:

```
Call_price =
          10.0201

Put_price =
          14.6553
```

A.8 Probability and Statistics

The following code computes the moments of the standard normal distribution to check $\mathbb{E}[Z^{2n}] = 1 \times 3 \times 5 \times \cdots \times (2n - 1)$, $n \geq 1$. (See Exercise 4.8.)

```
N = 10^8;
Z = randn(N,1);
M2 = mean(Z.^2)           % = 1
M4 = mean(Z.^4)           % = 3
M6 = mean(Z.^6)           % = 3 x 5
M8 = mean(Z.^8)           % = 3 x 5 x 7
M10 = mean(Z.^10)         % = 3 x 5 x 7 x 9
factor(3)
factor(15)
factor(105)
factor(945)
```

The output confirms the known fact.

```

M2 =
    0.9998
M4 =
    2.9985
M6 =
   14.9876
M8 =
  104.8641
M10 =
  943.0375
ans =
     3
ans =
     3     5
ans =
     3     5     7
ans =
     3     3     3     5     7

```

Compute the cumulative probability $\Pr(Z < a)$ where $Z \sim N(0, 1)$ using the command `normcdf(a)`.

```

normcdf(0)
normcdf(1)
normcdf(2)
normcdf(3)

```

produces

```

ans =
    0.5000
ans =
    0.8413
ans =
    0.9772
ans =
    0.9987

```

The command `mean(X)` computes the average of sample values of a random variable X while `mean(X > a)` gives the probability of the event $\{X > a\}$. That is, $\text{mean}(X) = \mathbb{E}[X]$ and $\text{mean}(X > a) = \Pr(X > a) = \mathbb{E}[\mathbf{1}_{\{X > a\}}]$.

```

X = randn(100000,1);
mean(X.^2)

```



```

1 - mean(X > 1)
1 - mean(X > 2)
1 - mean(X > 3)

```

produces

```

ans =
    1.0043
ans =
    0.8418
ans =
    0.9770
ans =
    0.9986

```

The following code checks the fact that $\mathbb{E}[e^{-Z^2}] = \frac{1}{\sqrt{3}}$. (See Exercise [4.10](#).)

```

N = 10^7;
Z = randn(N,1);
mean(exp(-Z.^2))
1/sqrt(3)

```

The output confirms the fact.

```

ans =
    0.5774
ans =
    0.5774

```

The following code checks the fact that $\mathbb{E}[|Z|] = \sqrt{\frac{2}{\pi}}$. (See Exercise [4.22](#).)

```

N = 10^7;
Z = randn(N,1);
MonteCarlo = mean(abs(Z))
Exact = vpa(sqrt(2/pi),6)

```

The output confirms the fact.

```

MonteCarlo =
    0.7980
Exact =
    0.797885

```

The command `norminv(u)` gives the inverse cumulative probability of the standard normal distribution. That is, $u = \Pr(Z \leq \text{norminv}(u))$, $0 \leq u \leq 1$.

```
norminv(0)
norminv(0.5)
norminv(0.99)
norminv(1)
normcdf(norminv(0.25))
```

produces

```
ans =
    -Inf
ans =
     0
ans =
    2.3263
ans =
     Inf
ans =
    0.2500
```

Note that `-Inf` and `Inf` represent $-\infty$ and $+\infty$, respectively.

Appendix B

Basic Analysis

We introduce the definitions, notations and facts for sets, functions and metric spaces.

B.1 Sets and Functions

We denote the sets of the natural numbers, integers, rational numbers, real numbers, complex numbers by $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$, respectively.¹ The *difference* of two sets A and B is defined by $A \setminus B = \{x : x \in A, x \notin B\}$, and their *symmetric difference* is defined by $A \triangle B = (A \cup B) \setminus (A \cap B)$. If $A \subset X$, then A^c denotes the complement of A , i.e., $A^c = X \setminus A$.

Given two sets X and Y , a function $f : X \rightarrow Y$ (or a map or a mapping) from X to Y is a rule that assigns for every $x \in X$ a unique element $f(x) \in Y$. The sets X and Y are called the domain and range (or codomain) of f , respectively. For each $x \in X$ the element $f(x)$ is called the image of x under f . Sometimes the range of f means the subset $\{f(x) : x \in X\}$.

If $\{f(x) : x \in X\} = Y$, then f is called an onto function. If $f(x_1) \neq f(x_2)$ whenever $x_1 \neq x_2$, then f is called a one-to-one function. If $f : X \rightarrow Y$ is onto and one-to-one, it is called a bijection or a one-to-one correspondence, and in this case there exists an inverse function f^{-1} .

Even when the inverse function of $f : X \rightarrow Y$ does not exist, the inverse image of $E \subset Y$ under f is defined by $f^{-1}(E) = \{x : f(x) \in E\}$. The operation of taking inverse images satisfies $f^{-1}(E \cup F) = f^{-1}(E) \cup f^{-1}(F)$, $f^{-1}(E \cap F) = f^{-1}(E) \cap f^{-1}(F)$ and $f^{-1}(Y \setminus E) = X \setminus f^{-1}(E)$.

The number of elements of a set A is called the cardinality of A , and denoted by $\text{card} A$. If $f : X \rightarrow Y$ is a one-to-one correspondence, then X and Y have the same cardinality. If X has finitely many elements, or X and $\mathbb{N}$ have the same cardinality,

¹ The symbol $\mathbb{Z}$ is from the German word ‘Zahl’ for number, and $\mathbb{Q}$ from the Italian word ‘quoziente’ for quotient.

then X is called a countable set. Examples of countable sets are $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, etc. A countable union of countable sets is also countable, more precisely, if A_i is a countable set for each $i = 1, 2, \dots$ then $\bigcup_{i=1}^{\infty} A_i$ is also a countable set. If X is not countable, X is called an uncountable set. Examples of uncountable sets are $\mathbb{R}$ and $\mathbb{C}$. If we add uncountably many positive numbers, the sum is always infinite, and that is why we consider only a countable sum of positive numbers and use the notation $\sum$.

For a subset $A \subset X$, the characteristic function (or the indicator function) of A , denoted by χ_A , is defined by

$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

B.2 Metric Spaces

Definition B.1 A metric (or distance) on a set X is a function $d : X \times X \rightarrow \mathbb{R}$ satisfying the following conditions:

- (i) $d(x, y) \geq 0$ for every $x, y \in X$. $d(x, y) = 0$ holds only for $x = y$.
 - (ii) Symmetry holds, i.e., $d(x, y) = d(y, x)$ for every $x, y \in X$.
 - (iii) The triangle inequality holds, i.e., $d(x, z) \leq d(x, y) + d(y, z)$ for $x, y, z \in X$.
- A set X on which a metric d is defined is called a metric space and denoted by (X, d) .

Example B.1 For any two points $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$ in the Euclidean space $\mathbb{R}^n$ define $d(x, y) = (\sum_{i=1}^n (x_i - y_i)^2)^{1/2}$, then $(\mathbb{R}^n, d)$ is a metric space. It is the standard metric on the Euclidean spaces.

Definition B.2 A norm on a vector space V is a function $\|\cdot\| : V \rightarrow \mathbb{R}$ satisfying the following conditions:

- (i) For every $v \in V$, $\|v\| \geq 0$, and $\|v\| = 0$ holds only for $v = 0$.
- (ii) For a scalar c and a vector $v \in V$ we have $\|cv\| = |c| \|v\|$.
- (iii) The triangle inequality $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$ holds.

A vector space X equipped with a norm $\|\cdot\|$ is denoted by $(V, \|\cdot\|)$ and called a normed space.

Fact B.1 On a normed space a metric can be defined by $d(x, y) = \|x - y\|$. Hence a normed space is a metric space.

Example B.2 For $1 \leq p < \infty$ and $\mathbf{x} \in \mathbb{R}^n$ define $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$, and for $p = \infty$ define $\|\mathbf{x}\|_{\infty} = \max_{1 \leq i \leq n} |x_i|$, then $\|\cdot\|_p$ is a norm for every $1 \leq p \leq \infty$.

Given a sequence $\{x_n\}_{n=1}^{\infty}$ of points in a metric space (X, d) , if the sequence of real numbers $\{d(x_n, x)\}_{n=1}^{\infty}$ converges to 0 then we say that the sequence $\{x_n\}_{n=1}^{\infty}$ converges to the limit x in X and write $x = \lim_{n \rightarrow \infty} x_n$ or $x_n \rightarrow x$. More precisely, x_n

converges to x if for any arbitrary $\varepsilon > 0$ there exists an $N \geq 1$ such that $d(x, x_n) < \varepsilon$ for any $n \geq N$.

Given a sequence $x_1, x_2, x_3, \dots$ of points of a metric space (X, d) if for any arbitrary $\varepsilon > 0$ there exists an N such that $d(x_m, x_n) < \varepsilon$ for any $m, n \geq N$, then $\{x_n\}$ is called a Cauchy sequence.

A convergent sequence is a Cauchy sequence. A Cauchy sequence in a space X need not converge to a point in X . For example, a sequence $1, 1.4, 1.41, 1.414, \dots$ in $X = \mathbb{Q}$ does not have its limit $\sqrt{2}$ in X . If any arbitrary Cauchy sequence converges in a metric space X , then it is called a complete space. For example, $\mathbb{R}^n$ is a complete space.

B.3 Continuous Functions

Throughout this section we consider a function $f : X \rightarrow Y$ is given where (X, d_X) and (Y, d_Y) are metric spaces.

Definition B.3 (*Continuity*) If a function f satisfies $\lim_{n \rightarrow \infty} f(x_n) = f(x)$ as $\lim_{n \rightarrow \infty} x_n = x$, then f is said to be continuous at x . If this property holds for every $x \in X$, then f is said to be continuous on X .

Fact B.2 The following statements are equivalent:

- (i) f is continuous at $x \in X$.
- (ii) For any constant $\varepsilon > 0$ there exists a constant $\delta > 0$ such that $d_X(x, x') < \delta$ implies $d_Y(f(x), f(x')) < \varepsilon$. The constant δ depends on x and ε , i.e., $\delta = \delta(\varepsilon, x)$.

Definition B.4 (*Uniform Continuity*) If for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $d_X(x, x') < \delta$ implies $d_Y(f(x), f(x')) < \varepsilon$, and if δ does not depend on x but only on ε , then f is said to be uniformly continuous on X .

Example B.3 (i) Let $X = [0, 1]$ and $X_0 = (0, 1]$. The function $f(x) = \frac{1}{x}$ on X_0 is continuous but not uniformly continuous. It cannot be extended to X . This fact is consistent with Theorem B.2.

- (ii) If $f : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ is Lipschitz continuous, i.e., there exists a constant $0 \leq M < \infty$ such that $|f(x) - f(y)| < M|x - y|$ for every x, y , then f is uniformly continuous.
- (iii) If $f : \mathbb{R}^1 \rightarrow \mathbb{R}^1$ is differentiable and if $|f'(x)| \leq M < \infty$, then by the Mean Value Theorem f satisfies $|f(x) - f(y)| < |f'(z)||x - y| \leq M|x - y|$, and hence f is Lipschitz continuous and uniformly continuous.

Now we consider the metric space consisting of continuous functions.

Example B.4 Let $C([0, 1], \mathbb{R})$ and $C([0, 1], \mathbb{C})$ be the sets of all continuous functions f defined on $[0, 1]$ with their values in $\mathbb{R}$ and $\mathbb{C}$, respectively. They are vector spaces with the norm defined by $\|f\| = \max_{x \in [0, 1]} |f(x)|$, called the uniform norm. We can define a metric d by $d(f, g) = \max_{x \in [0, 1]} |f(x) - g(x)|$ for continuous functions f, g , and hence $C([0, 1], \mathbb{R})$ and $C([0, 1], \mathbb{C})$ are metric spaces. They are complete metric spaces.

Definition B.5 (*Open Ball*) On a metric space (X, d) , for $r > 0$ the subset $B_r(x_0) = \{x \in X : d(x_0, x) < r\}$ is called the *open ball* of radius r centered at x_0 . Let $U \subset X$. If for any point x_0 of U there exists an $r = r(x_0) > 0$ such that $B_r(x_0) \subset U$, then U is said to be *open*. A set is *closed* if its complement is open. Trivial examples of open subsets are $\emptyset$ and X .

Example B.5 (i) If $X = \mathbb{R}^1$, then open intervals (a, b) , $(-\infty, b)$ and (a, ∞) are open sets.
(ii) If X is a normed space with a norm $\|\cdot\|$, then the open unit ball $B_1(\mathbf{0}) = \{x \in X : \|x\| < 1\}$ is open.

If a subset $K \subset X$ is covered by subsets $\{U_\lambda : \lambda \in \Lambda\}$, i.e., $K \subset \bigcup_\lambda U_\lambda$, then $\{U_\lambda : \lambda \in \Lambda\}$ is called a cover of K . If there exists a $\Lambda_0 \subset \Lambda$ such that $K \subset \bigcup_{\lambda \in \Lambda_0} U_\lambda$, then we call $\{U_\lambda : \lambda \in \Lambda_0\}$ a subcover of K , or a subcover of $\{U_\lambda : \lambda \in \Lambda\}$. If every U_λ is an open set, then $\{U_\lambda : \lambda \in \Lambda\}$ is called an open cover. If Λ_0 is a finite set, then $\{U_\lambda : \lambda \in \Lambda_0\}$ is called a finite subcover.

Definition B.6 (*Compact Set*) If, for any open cover of a set K , there exists a finite subcover, then K is said to be *compact*.

Fact B.3 (Heine–Borel Theorem) A subset of $\mathbb{R}^n$ is compact if and only if it is closed and bounded.

Fact B.4 (Bolzano–Weierstrass Theorem) Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.

Example B.6 A metric space (X, d) is given.

- (i) The empty set $\emptyset$ and a finite set are compact.
- (ii) For $X = \mathbb{R}^1$, a closed and bounded interval $[a, b]$ is compact.
- (iii) For $X = \mathbb{R}^1$, closed and unbounded intervals such as $\mathbb{R}^1$, $(-\infty, b]$ and $[a, \infty)$ are not compact.

Fact B.5 Given two metric spaces (X, d_X) , (Y, d_Y) and a continuous function $f : X \rightarrow Y$. If X is a compact set, then the following holds:

- (i) $f(X)$ is a compact subset of Y .
- (ii) For $Y = \mathbb{R}^1$, f assumes its maximum and minimum on X , i.e., there exist $x_1, x_2 \in K$ such that $f(x_1) = \max_{x \in K} f(x)$ and $f(x_2) = \min_{x \in K} f(x)$.
- (iii) f is uniformly continuous.

Definition B.7 (*Dense Subset*) A subset X_0 of a metric space X is *dense* in X if for every $x \in X$ and for every $r > 0$ the open ball $B_r(x)$ satisfies $B_r(x) \cap X_0 \neq \emptyset$. Equivalently, for every $x \in X$ there exists a sequence $x_n \in X_0$ that converges to x .

Example B.7 (i) The set of integers $\mathbb{Z}$ is not dense in $\mathbb{R}$.

(ii) The set of rational numbers $\mathbb{Q}$ is dense in the set of real numbers $\mathbb{R}$.

(iii) The set of irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$.

(iv) The open interval (a, b) is dense in the interval $[a, b]$.

Fact B.6 (Weierstrass) A real-valued trigonometric function of period 1 is a function of the form

$$a_0 + \sum_{n=1}^N a_n \cos(2\pi nx) + \sum_{n=1}^N b_n \sin(2\pi nx), \quad a_n, b_n \in \mathbb{R}, \quad N \in \mathbb{N}.$$

The set of all such functions is dense in $C([0, 1], \mathbb{R})$.

Fact B.7 (Stone–Weierstrass) A complex-valued trigonometric function of period 1 is a function of the form

$$\sum_{n=-N}^N c_n e^{2\pi i n x}, \quad c_n \in \mathbb{C}, \quad N \in \mathbb{N}$$

where $i^2 = -1$. The set of all such functions is dense in $C([0, 1], \mathbb{C})$.

B.4 Bounded Linear Transformations

We are given two normed spaces $(X, \|\cdot\|_X)$, $(Y, \|\cdot\|_Y)$. When there is no danger of confusion we write $\|\cdot\|$ to denote both $\|\cdot\|_X$ and $\|\cdot\|_Y$. If a map $T : X \rightarrow Y$ satisfies $T(x_1 + x_2) = T(x_1) + T(x_2)$ for $x_1, x_2 \in X$ and $T(cx) = cT(x)$ for $x \in X$ and scalar c , then T is called a linear transformation. For a linear transformation T we usually write Tx instead of $T(x)$.

Define a *norm* $\|T\|$ of a linear transformation $T : X \rightarrow Y$ by $\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} = \sup_{\|x\|=1} \|Tx\|$. The transformation T is called a bounded transformation if $\|T\| < \infty$, which is equivalent to the condition that there exists an $M < \infty$ such that

$$\|Tx\| \leq M\|x\|$$

for every x . In fact, $\|T\|$ is the infimum of such constants M .

Example B.8 For $X = \mathbb{R}^n$, $Y = \mathbb{R}^m$ and an $m \times n$ matrix A , the map $T : X \rightarrow Y$ defined by $Tx = Ax$ is linear.

- (i) Define a norm on $\mathbb{R}^k$ by $\|\mathbf{v}\|_1 = \sum_{i=1}^k |v_i|$, $\mathbf{v} = (v_1, \dots, v_k)$. Then the norm of $T : X \rightarrow Y$ is given by

$$\|T\| = \max_{1 \leq j \leq n} \sum_{i=1}^m |A_{ij}|.$$

- (ii) Define a norm on $\mathbb{R}^k$ by $\|\mathbf{v}\|_\infty = \max_{1 \leq i \leq k} |v_i|$, $\mathbf{v} = (v_1, \dots, v_k)$. Then the norm of $T : X \rightarrow Y$ is given by

$$\|T\| = \max_{1 \leq i \leq m} \sum_{j=1}^n |A_{ij}|.$$

- (iii) Let $m = n$ and define a norm on $\mathbb{R}^n$ by $\|\mathbf{v}\|_2 = (\sum_{i=1}^n |v_i|^2)^{1/2}$, $\mathbf{v} = (v_1, \dots, v_n)$. Let λ_i denote the eigenvalues of the symmetric matrix $A^t A$ or the singular values of A . Then the norm of $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is given by

$$\|T\| = \max_{1 \leq i \leq n} \sqrt{\lambda_i}.$$

Fact B.8 Any two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a finite-dimensional vector space X are equivalent, i.e., there exist constants $0 < A \leq B$ such that $A\|x\|_1 \leq \|x\|_2 \leq B\|x\|_1$ for every $x \in X$. In this case, the metrics induced by the all the norms are equivalent since $A\|x - y\|_1 \leq \|x - y\|_2 \leq B\|x - y\|_1$ for every $x, y \in X$, and a convergent sequence with respect to one norm is also convergence with respect to the other norm.

Theorem B.1 Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces, and let $f : X \rightarrow Y$ be a linear mapping. Then the following statements are equivalent:

- (i) f is bounded.
- (ii) f is continuous.
- (iii) f is uniformly continuous.

Proof (i) $\Rightarrow$ (iii) If there exists an $M < \infty$ such that $\|f(x)\| \leq M\|x\|$ for every x , then by the linearity we have $\|f(x_1) - f(x_2)\| \leq M\|x_1 - x_2\|$ for $x_1, x_2 \in X$. For every $\varepsilon > 0$ choose $\delta = \varepsilon/M$, then $\|x_1 - x_2\| < \delta$ implies $\|f(x_1) - f(x_2)\| < \varepsilon$. In other words, f is uniformly continuous.

(iii) $\Rightarrow$ (ii) This is obvious by definition.

(ii) $\Rightarrow$ (i) If $\|f(x)\| \leq M\|x\|$ does not hold for all x , then there exists a sequence $x_n \in X$ such that $\|x_n\| = 1$ and $\|f(x_n)\| > n$. Define a new sequence $z_n = \frac{1}{n}x_n$. Then $\|f(z_n)\| > 1$. Hence $z_n \rightarrow 0$, but $f(z_n)$ does not converge to $0 \in Y$. In other words, f is not continuous, which is a contradiction.

□

B.5 Extension of a Function

We are given two sets X, Y and a function $f : X_0 \rightarrow Y$, where X_0 is a subset of X . If there exists a function $F : X \rightarrow Y$ such that $F(x) = f(x)$ for $x \in X_0$, then F is called an *extension* of f to X . In this case f is called the *restriction* of F to X_0 .

Now we consider the extension of a continuous function defined on a dense subset to the whole space.

Lemma B.1 *Let (X, d_X) and (Y, d_Y) be metric spaces. If $f : X \rightarrow Y$ is a uniformly continuous function, then f maps a Cauchy sequence $\{x_n\}$ to a Cauchy sequence $\{f(x_n)\}$.*

Proof For every $\varepsilon > 0$ the uniform discontinuity of f implies that there exists a $\delta > 0$ such that $d_X(z, z') < \delta$ implies $d_Y(f(z), f(z')) < \varepsilon$. If $\{x_n\}$ is a Cauchy sequence, then there exists a sufficiently large N such that $d_X(x_m, x_n) < \delta$ for $m, n \geq N$. Hence $d_Y(f(x_m), f(x_n)) < \varepsilon$. $\square$

Theorem B.2 *Let (X, d_X) and (Y, d_Y) be metric spaces, and let $X_0 \subset X$ be a dense subspace of X . Suppose that Y is a complete space. If $f : X_0 \rightarrow Y$ is uniformly continuous, then f can be extended to X as a continuous and linear map.*

Proof First, we will define $f(x)$ for any $x \in X$. Take a sequence $x_n \in X_0$ converging to x . By Lemma B.1, $f(x_n)$ is a Cauchy sequence. Since Y is a complete space, $f(x_n)$ converges to a limit, which is defined to be $f(x)$. To show that the preceding definition is well-defined and does not depend on any particular sequence $x_n \in X_0$, take another sequence $x'_n \in X_0$ and show $\lim_{n \rightarrow \infty} f(x_n) = \lim_{n \rightarrow \infty} f(x'_n)$. Arranging the elements of two sequences alternately we obtain a new sequence $x_1, x'_1, x_2, x'_2, x_3, x'_3, \dots$, which also converges to x , hence is a Cauchy sequence. By Lemma B.1, the sequence $f(x_1), f(x'_1), f(x_2), f(x'_2), f(x_3), f(x'_3), \dots$ is also Cauchy. Since Y is a complete space, there exists a unique limit, which is the common limit of two subsequences $f(x_1), f(x_2), \dots$ and $f(x'_1), f(x'_2), \dots$, and hence they have the same limit. $\square$

Corollary B.1 *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed spaces, and let $X_0 \subset X$ be a dense vector subspace of X . Assume that Y is a complete space. If a linear transformation $f : X_0 \rightarrow Y$ is continuous, then f is uniformly continuous, and hence by Theorem B.2 f can be extended to X as a continuous and linear transformation.*

B.6 Differentiation of a Function

Definition B.8 (*Differentiation*) If a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ satisfies the condition that for every point $\mathbf{x}$ there exists an $m \times n$ matrix $Df(\mathbf{x})$ such that

$$\lim_{\|\mathbf{h}\| \rightarrow 0} \frac{\|f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - Df(\mathbf{x})\mathbf{h}\|}{\|\mathbf{h}\|} = 0 ,$$

we say that f is differentiable and call the matrix $Df(\mathbf{x})$, or the associated linear transformation, the *derivative* of f .

Remark B.1 The derivative of a differentiable function $f = (f_1, \dots, f_m) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is defined by

$$Df(\mathbf{x}) = \left[\frac{\partial f_i}{\partial x_j} \right]_{ij}.$$

For $m = 1$ we have

$$Df(\mathbf{x}) = \nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right),$$

where $f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) \approx \nabla f(\mathbf{x})\mathbf{h}$ and the right-hand side is a product of a $1 \times n$ matrix and an $n \times 1$ matrix. Throughout the book, the expression $G_1 \approx G_2$ means that G_1 and G_2 are ‘approximately equal’ in a suitable sense, and in this section the norm $\|G_1 - G_2\|$ is close to zero.

Now we consider the partial derivatives of a function defined by a matrix.

Lemma B.2 (i) For a column vector $\mathbf{a} \in \mathbb{R}^n$ define $g : \mathbb{R}^n \rightarrow \mathbb{R}^1$ by $g(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} = \mathbf{a}^t \mathbf{x}$. Then $\nabla g(\mathbf{x}) = \mathbf{a}^t$.

(ii) For an $n \times n$ matrix A , define a function $f : \mathbb{R}^n \rightarrow \mathbb{R}^1$ by $f(\mathbf{x}) = \mathbf{x}^t A \mathbf{x}$. Then $\nabla f(\mathbf{x}) = \mathbf{x}^t (A + A^t)$.

Proof (i) The proof is a direct application of the definition. Intuitively, if $\mathbf{h} \approx \mathbf{0}$ then $g(\mathbf{x} + \mathbf{h}) - g(\mathbf{x}) \approx \nabla g(\mathbf{x})\mathbf{h}$, and hence $\mathbf{a}^t(\mathbf{x} + \mathbf{h}) - \mathbf{a}^t \mathbf{x} = \mathbf{a}^t \mathbf{h} \approx \nabla g(\mathbf{x})\mathbf{h}$. Thus $\nabla g(\mathbf{x}) = \mathbf{a}^t$.

(ii) As in the first part, for $\mathbf{h} \approx \mathbf{0}$ we have

$$\begin{aligned} f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) &= (\mathbf{x} + \mathbf{h})^t A (\mathbf{x} + \mathbf{h}) - \mathbf{x}^t A \mathbf{x} \\ &= \mathbf{x}^t A \mathbf{x} + \mathbf{x}^t A \mathbf{h} + \mathbf{h}^t A \mathbf{x} + \mathbf{h}^t A \mathbf{h} - \mathbf{x}^t A \mathbf{x} \\ &= \mathbf{x}^t A \mathbf{h} + \mathbf{h}^t A \mathbf{x} + \mathbf{h}^t A \mathbf{h} \\ &\approx \mathbf{x}^t A \mathbf{h} + \mathbf{h}^t A \mathbf{x} \\ &= \mathbf{x}^t A \mathbf{h} + (A \mathbf{x})^t \mathbf{h} \\ &= \mathbf{x}^t A \mathbf{h} + \mathbf{x}^t A^t \mathbf{h} \\ &= \mathbf{x}^t (A + A^t) \mathbf{h} \end{aligned}$$

and $\nabla f(\mathbf{x}) = \mathbf{x}^t (A + A^t)$. □

Appendix C

Linear Algebra

Linear algebra deals with operations of vectors and matrices, and it is widely used in many areas because it is an essential tool in solving a system of linear equations. The goal of this appendix is to introduce some basic concepts and terminology in linear algebra.

C.1 Vectors

Theorem C.1 (Cauchy–Schwarz Inequality) *Let V be a vector space with an inner product $(\cdot, \cdot)$. Then we have $(v, w)^2 \leq (v, v)(w, w)$ for $v, w \in V$. Equality holds if and only if one of the vectors v, w is a constant multiple of the other one.*

Proof First, consider the case when the scalar field is real. If $v = 0$ or $w = 0$, then the inequality clearly holds. It suffices to prove the theorem for the case $v \neq 0$ and $w \neq 0$. Since

$$f(t) = (tv + w, tv + w) = t^2(v, v) + 2(v, w)t + (w, w) \geq 0$$

for every real t , $f(t)$ has nonnegative minimum

$$f(t_0) = \frac{(v, v)(w, w) - (v, w)^2}{(v, v)}$$

at $t_0 = -\frac{(v, w)}{(v, v)}$. The quadratic function $f(t)$ is equal to 0 if and only if there exists a real number t such that $tv + w$ is the zero vector, and otherwise $f(t) > 0$. $\square$

Given a linear map $T : V \rightarrow V$, if there exists a nonzero vector v such that $Tv = \lambda v$ for some scalar λ , then v is called a characteristic vector or an eigenvector, and λ is called a characteristic value or an eigenvalue.

C.2 Matrices

Let V and W be two vector spaces with bases $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$, respectively. A convenient way of expressing a linear transformation $T : V \rightarrow W$ is to compute $T(v_j) = \sum_i a_{ij} w_i$. Then the nm scalars a_{ij} have all the essential information on T . Here are some basic facts on matrix multiplication.

- Fact C.1** (i) Put $\mathbf{1} = (1, \dots, 1)^t \in \mathbb{R}^n$. Then $\mathbf{1}^t \mathbf{1} = n$, and $\mathbf{1} \mathbf{1}^t$ is an $n \times n$ matrix all of whose n^2 components are 1. Note that $\mathbf{1}^t \mathbf{x} = x_1 + \dots + x_n$.
(ii) Let L and M be matrices of sizes $m \times n$ and $n \times \ell$, respectively. Then

$$LM = [LM^1 \dots LM^\ell],$$

where M^j denotes the j th column of M . Note that LM^j is an $m \times 1$ matrix, i.e., an m -dimensional column vector, since it is the product of matrices of sizes $m \times n$ and $n \times 1$, respectively.

- (iii) Let M and L be matrices of sizes $n \times \ell$ and $\ell \times k$. Then

$$ML = \begin{bmatrix} M_1 L \\ \vdots \\ M_n L \end{bmatrix},$$

where M_i is the i th row of M . Note that $M_i L$ is a k -dimensional row vector since it is a product of matrices of sizes $1 \times \ell$ and $\ell \times k$.

If a basis $\mathcal{B}$ of V consists of eigenvectors $v_1, \dots, v_n$ of T , then the matrix $A = [a_{ij}]$ of T with respect to $\mathcal{B}$ satisfies $a_{ij} = 0$ for $i \neq j$ and $T(v_i) = a_{ii} v_i$ for every i . Thus A is a diagonal matrix. One of the goals of linear algebra is to take a suitable basis so that the corresponding matrix of T is close to a diagonal matrix. The *trace* of A is defined by $\sum_{i=1}^n A_{ii}$ and is equal to the sum of eigenvalues.

The *column rank* of a matrix A is the dimension of the column space of A , and the *row rank* of A is the dimension of the row space of A . The column rank and the row rank are equal, and $\text{rank}(A) \leq \min(m, n)$. If $\text{rank}(A) = \min\{m, n\}$, then A is said to have full rank.

The linear transformation $T(\mathbf{x}) = A\mathbf{x}$ is injective if and only if $\text{rank}(A) = n$, i.e., A has full column rank. The linear transformation $T(\mathbf{x}) = A\mathbf{x}$ is surjective if and only if $\text{rank}(A) = m$, i.e., A has full row rank. If $m = n$, then A is invertible if and only if $\text{rank}(A) = n$.

If B is an $n \times k$ matrix, then $\text{rank}(AB) \leq \min\{\text{rank } A, \text{rank } B\}$. If B is an $n \times k$ matrix of rank n , then $\text{rank}(AB) = \text{rank}(A)$. If C is an $\ell \times m$ matrix of rank m , then $\text{rank}(CA) = \text{rank}(A)$.

C.3 The Method of Least Squares

Let V be a vector space with an inner product $(\cdot, \cdot)$. The case when V is infinite-dimensional is not excluded in the following argument. Consider the problem of approximating a vector $v \in V$ by a subspace V_0 . We introduce the least squares method, which finds the minimum

$$\min_{w \in V_0} \|v - w\|^2 = (v - w, v - w) .$$

We consider the case when V_0 is finite-dimensional. There are two methods. For the first method we choose $V_0 = \text{span}\{w_1, \dots, w_k\}$ and define

$$f(c_1, \dots, c_k) = \|v - (c_1 w_1 + \dots + c_k w_k)\|^2 ,$$

then take partial derivatives with respect to c_j , and set them equal to 0.

The second method is based on the fact that the distance minimizing vector is the orthogonal projection of v onto V_0 , and it solves the system of equations $(w_j, v - (c_1 w_1 + \dots + c_k w_k)) = 0$.

For $\Omega = \{\omega_1, \dots, \omega_n\}$ with the normalized counting measure $\mathbb{P}$, i.e., $\mathbb{P}(\{\omega_i\}) = \frac{1}{n}$ for $1 \leq i \leq n$, we may regard $L^2(\Omega)$ as $\mathbb{R}^n$. Table C.1 lists the corresponding concepts in Lebesgue integral theory and linear algebra. For example, the constant function $1 = \mathbb{1}_\Omega$ is identified with the vector $(1, \dots, 1)$.

For $V = L^2(\Omega)$ let V_0 be a two-dimensional subspace of V spanned by the constant function $1 = \mathbb{1}_\Omega$ and a nonconstant function X , i.e., $V_0 = \{a1 + bX : a, b \in \mathbb{R}\}$. Fix $Y \in V$. If a and b solve the minimization problem

$$\min_{a, b} \|Y - (a1 + bX)\|_{L^2} .$$

To solve for a and b , either we set

$$f(a, b) = \|Y - (a1 + bX)\|_{L^2}^2 = (Y - (a1 + bX), Y - (a1 + bX))$$

and take the partial derivatives with respect to a and b , set the derivatives equal to 0, or we use the orthogonality relations

$$\begin{cases} (1, Y - (a1 + bX)) = 0 \\ (X, Y - (a1 + bX)) = 0 , \end{cases}$$

Table C.1 Comparison of Lebesgue integral on a finite space and linear algebra

Lebesgue integral	Linear algebra
A random variable X	A vector $\mathbf{v}$
The constant function 1	The vector $(1, \dots, 1)$
$\int_{\Omega} XY d\mathbb{P} = \mathbb{E}[XY]$	$\frac{1}{n} \mathbf{v} \cdot \mathbf{w}$
$\int_{\Omega} X^2 d\mathbb{P} = \mathbb{E}[X^2] = \ X\ _{L^2}^2$	$\frac{1}{n} \ \mathbf{v}\ ^2$
$L^2(\Omega, \mathbb{P})$	$\mathbb{R}^n$

or, equivalently

$$\begin{bmatrix} 1 & \mathbb{E}[X] \\ \mathbb{E}[X] & \mathbb{E}[X^2] \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \mathbb{E}[Y] \\ \mathbb{E}[XY] \end{bmatrix}. \quad (\text{C.1})$$

Since 1 and X are linearly independent, the determinant $\mathbb{E}[X^2] - \mathbb{E}[X]^2$ is nonzero by the Cauchy–Schwarz inequality. (In fact, it is positive.) Hence the solution of (C.1) is given by $b = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$ and $a = \mathbb{E}[Y] - b \mathbb{E}[X]$.

Example C.1 (*Linear Regression*) If we try to approximately express data

$$\{(x_1, y_1), \dots, (x_n, y_n)\}, \quad n \geq 2$$

using a line of best fit $Y = a + bX$, we need to solve the least squares problem defined by

$$\min_{a,b} \sum_{i=1}^n |y_i - (a + bx_i)|^2.$$

If we put $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_n)$, $\mathbf{1} = (1, \dots, 1)$, then the above problem is converted into

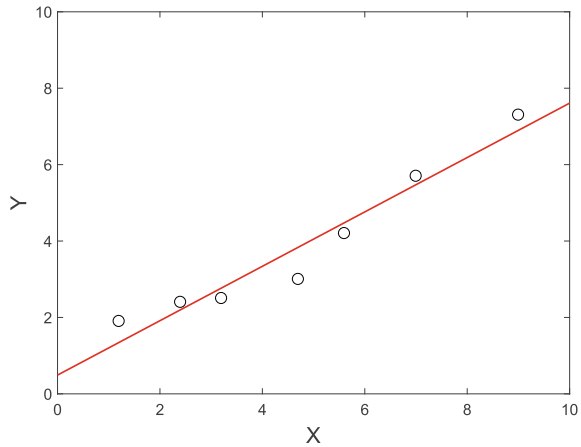
$$\min_{a,b} \|\mathbf{y} - (a\mathbf{1} + b\mathbf{x})\|^2,$$

which has a solution

$$b = \frac{\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

by the preceding computation. The hidden assumption in this problem is the condition that $\mathbf{1}$ and $\mathbf{x}$ are linearly independent, which is equivalent to the condition that not all x_i are equal. See Fig. C.1.

Fig. C.1 The line of best fit and the least squares method



C.4 Symmetric Matrices

Theorem C.2 *Let A be a symmetric matrix, then the following statements are equivalent:*

- (i) A is positive definite.
- (ii) There exists a symmetric positive definite matrix B such that $A = B^2$.
- (iii) There exists an invertible matrix B such that $A = B^t B$.

Proof To prove that the statement (i) implies the statement (ii), recall that A is orthogonally diagonalizable. Hence $P^t A P = D$, where P is orthogonal and D is diagonal. Note that the diagonal entries of D are eigenvalues of A , which are positive. Hence we have $D = D_0^2$, where D_0 is the diagonal matrix whose entries are positive square roots of eigenvalues of A . Thus

$$A = P D_0^2 P^t = P D_0 D_0 P^t = P D_0 P^t P D_0 P^t = (P D_0 P^t)(P D_0 P^t) = B^2,$$

where $B = P D_0 P^t$.

To prove (ii) $\Rightarrow$ (iii), note that $A = B B = B^t B$.

To prove (iii) $\Rightarrow$ (i), note that $\mathbf{x}^t A \mathbf{x} = \mathbf{x}^t B^t B \mathbf{x} = (B \mathbf{x})^t (B \mathbf{x}) = \|B \mathbf{x}\|^2 > 0$, $\mathbf{x} \neq \mathbf{0}$. □

Theorem C.3 *Let A be a nonnegative definite matrix. Then there exists a lower triangular matrix L having nonnegative diagonal elements such that $A = L L^t$. Furthermore, if A is positive definite, then the matrix L is unique and has positive diagonal elements.*

Proof See p. 147, (Schott, 2005). □

Definition C.1 If A is symmetric and positive definite, then there exists a unique lower triangular matrix L with positive diagonal entries such that $A = LL^t$. This is called the *Cholesky decomposition* or *Cholesky factorization* of A .

Example C.2 Here is how to transform a random vector so that the resulting random vector is normalized, i.e., its mean is zero and the covariance matrix is the identity matrix. Suppose that an $m \times 1$ random vector $\mathbf{X}$ has a mean vector $\boldsymbol{\mu} = \mathbb{E}[\mathbf{X}]$ and a positive-definite covariance matrix Σ . Choose a matrix L satisfying $\Sigma = LL^t$, and put $\mathbf{Z} = L^{-1}(\mathbf{X} - \boldsymbol{\mu})$. Then

$$\mathbb{E}[\mathbf{Z}] = \mathbb{E}[L^{-1}(\mathbf{X} - \boldsymbol{\mu})] = L^{-1}\mathbb{E}[\mathbf{X} - \boldsymbol{\mu}] = L^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}) = \mathbf{0}$$

and

$$\begin{aligned} \text{Var}(\mathbf{Z}) &= \text{Var}(L^{-1}\mathbf{X} - L^{-1}\boldsymbol{\mu}) \\ &= \text{Var}(L^{-1}\mathbf{X}) \\ &= \mathbb{E}[L^{-1}\mathbf{X}(L^{-1}\mathbf{X})^t] \\ &= \mathbb{E}[L^{-1}\mathbf{X}\mathbf{X}^t(L^{-1})^t] \\ &= L^{-1}\mathbb{E}[\mathbf{X}\mathbf{X}^t](L^{-1})^t \\ &= L^{-1}\text{Var}(\mathbf{X})(L^{-1})^t \\ &= L^{-1}\Sigma(L^{-1})^t = I_m. \end{aligned}$$

Appendix D

Ordinary Differential Equations

An ordinary differential equation is an equation defined by derivatives of a function of one variable. In this appendix we present basic facts on ordinary differential equations. For a general introduction to ordinary differential equations, consult (Hirsh & Smale, 1974, Simmons, 1991).

D.1 Linear Differential Equations with Constant Coefficients

A differential operator L is said to be linear if it satisfies $L(c_1 f_1 + c_2 f_2) = c_1 L(f_1) + c_2 L(f_2)$ and $L(cf) = cL(f)$ for arbitrary constants c_1, c_2 and c . To solve a linear ordinary differential equation

$$a_n \frac{\partial^n f}{\partial x^n} + \cdots + a_1 \frac{\partial f}{\partial x} + a_0 f = g(x) , \quad (\text{D.1})$$

where $a_n, \dots, a_1, a_0$ are constants, we first solve the homogeneous equation

$$a_n \frac{\partial^n f}{\partial x^n} + \cdots + a_1 \frac{\partial f}{\partial x} + a_0 f = 0 \quad (\text{D.2})$$

and find a homogeneous solution f_H , then take the sum with a particular solution f_P of (D.1). To find a general solution of (D.2) we solve a polynomial equation

$$a_n \lambda^n + \cdots + a_1 \lambda + a_0 = 0 . \quad (\text{D.3})$$

- (i) If there exist distinct real solutions $\lambda_1, \dots, \lambda_n$ for (D.3), then a general solution is of the form $c_1 e^{\lambda_1 x} + \cdots + c_n e^{\lambda_n x}$.

- (ii) If some of them are double roots, say $\lambda_1 = \lambda_2$, then the corresponding solution is of the form

$$c_1 e^{\lambda_1 x} + c_2 x e^{\lambda_1 x} + c_3 e^{\lambda_3 x} + \cdots + c_n e^{\lambda_n x} .$$

If some of them are triple roots, say $\lambda_1 = \lambda_2 = \lambda_3$, then the corresponding solution is of the form

$$c_1 e^{\lambda_1 x} + c_2 x e^{\lambda_1 x} + c_3 x^2 e^{\lambda_1 x} + \cdots + c_n e^{\lambda_n x}$$

and so on.

- (iii) If there are complex roots $\lambda = a + b\sqrt{-1}$, $b \neq 0$, then there exist solutions of the form $e^x \cos(bx)$ and $e^x \sin(bx)$. A general solution is a linear combination of such functions. The set of general solutions of (D.1) is an n -dimensional vector space V , and the set of solutions of (D.2) is a parallel translate of the form $f_P + V$.

Example D.1 To solve a differential equation

$$f'' - 2af' + a^2 f = (D - a)^2 f = 0 ,$$

where $D = \frac{d}{dx}$, we solve $(t - a)^2 = 0$ and find a double root $t = a$. Put $g(x) = e^{-ax} f(x)$. Then $D^2 g = 0$, and hence $g(x) = c_1 + c_2 x$. Thus $f(x) = c_1 e^{-x} + c_2 x e^{-x}$. The same idea can be used for a triple root as in a differential equation $(D - 1)^3 f = 0$, and so on.

Example D.2 To solve a differential equation

$$((D - \alpha)^2 + \beta^2) f = 0 ,$$

where α, β are real constants, we put $g(x) = e^{-\alpha x} f(x)$. Then $D^2 g = -\beta^2 g$, $g(x) = c_1 \cos \beta x + c_2 \sin \beta x$, and finally

$$f(x) = c_1 e^{\alpha x} \cos \beta x + c_2 e^{\alpha x} \sin \beta x .$$

Example D.3 To solve an ordinary differential equation

$$f'' + f = 1$$

we first solve the homogeneous equation $f'' + f = 0$, find the two-dimensional vector space $V = \{c_1 \sin x + c_2 \cos x : c_1, c_2 \text{ constant}\}$ consisting of homogeneous solutions, translate the space by $f_P = 1$, and find the desired solution set

$$V + f_P = \{c_1 \sin x + c_2 \cos x + 1 : c_1, c_2 \text{ constant}\} .$$

D.2 Linear Differential Equations with Nonconstant Coefficients

The linear equation

$$\frac{dy}{dx} + P(x)y = Q(x)$$

can be solved by multiplying by an integrating factor $e^{\int P(x)dx}$. Then we have

$$\frac{d}{dx} \left(e^{\int P(x)dx} y \right) = Q(x) e^{\int P(x)dx} ,$$

which produces the solution

$$y = e^{-\int P(x)dx} \left(\int Q(x) e^{\int P(x)dx} dx + C \right) .$$

Example D.4 An ordinary differential equation need not have a global solution. For example, consider $y' = y^2$, which has a trivial solution $y = 0$ and a nontrivial solution of the form $y = \frac{1}{-x+C}$ where C is a constant.

Example D.5 The initial value problem $y' = 2\sqrt{xy}$, $x \geq 0$, $y(0) = 1$, has a solution $y = (1 + \frac{2}{3}x^{\frac{3}{2}})^2$.

Example D.6 Consider *Laguerre's equation* defined by

$$xy'' + (1-x)y' + py = 0 ,$$

where p is a constant. The solutions are polynomials if p is a nonnegative integer, and they are given by

$$L_n(x) = e^{-x/2} \frac{1}{n!} e^x \frac{d^n}{dx} (x^n e^{-x}) , \quad n \geq 1 .$$

For example, $L_0(x) = e^{-x/2}$, $L_1(x) = e^{-x/2}(1-x)$, $L_2(x) = e^{-x/2}(1-2x+\frac{x^2}{2})$. Consult (Simmons, 1991) for more details.

Lemma D.1 (Gronwall's Inequality) *Let $g(t) : [0, a] \rightarrow \mathbb{R}$ be continuous and non-negative. Suppose that there exist constants $C \geq 0$, $K \geq 0$ such that*

$$g(t) \leq C + K \int_0^t g(s) ds ,$$

for every $t \in [0, a]$. Then $g(t) \leq Ce^{Kt}$ for every $t \in [0, a]$.

Proof First, suppose that $C > 0$. Define $G(t)$ by

$$G(t) = C + K \int_0^t g(s) \, ds .$$

Then $G(t) \geq g(t)$ and $G(t) \geq C > 0$. Since

$$\frac{d}{dt} G(t) = K g(t) \leq K G(t) ,$$

we have $\frac{d}{dt} (\log G(t)) \leq K$, and $\log G(t) \leq \log G(0) + Kt$. Since $G(0) = c$, we conclude that $g(t) \leq G(t) \leq G(0)e^{Kt} = Ce^{Kt}$. $\square$

D.3 The Cauchy–Euler Equation

The ordinary differential equation of the form

$$a_n x^n y^{(n)} + \cdots + a_1 x y' + a_0 y = 0$$

is called the Cauchy–Euler equation, and its general solution is a linear combination of solutions of the form $y = x^\lambda$ if the associated polynomial in λ has distinct roots.

Example D.7 Consider

$$x^2 y'' + x y' + y = 0 .$$

From $\lambda(\lambda - 1) + \lambda + 1 = 0$, we have $\lambda^2 + 1 = 0$ with two complex roots and obtain solutions

$$y = x^{\pm i} = \exp(\pm i \log x) = \cos \log x \pm i \log \sin x .$$

By taking the real and imaginary parts, we obtain a real general solution

$$y = c_1 \cos \log x + c_2 \log \sin x .$$

We used the definition $b^z = e^{z \log b}$ for $b \in \mathbb{R}$, $b > 0$ and $z \in \mathbb{C}$.

Example D.8 If the second order Cauchy–Euler equation

$$a_2 x^2 y'' + a_1 x y' + a_0 y = 0$$

gives rise to the quadratic equation

$$a_2 \lambda(\lambda - 1) + a_1 \lambda + a_0 = 0$$

with a double root λ , then the general solution of the differential equation is of the form

$$c_1 x^\lambda + c_2 x^\lambda \log x ,$$

where c_1 and c_2 are arbitrary constants.

A heuristic argument to derive the formula is to assume that there exist two real solutions λ and $\lambda + \Delta\lambda$, $\Delta\lambda \approx 0$, and note that

$$\begin{aligned} c_1 x^\lambda + c_2 x^{\lambda+\Delta\lambda} &= c_1 x^\lambda + c_2 x^\lambda x^{\Delta\lambda} \\ &= c_1 x^\lambda + c_2 x^\lambda e^{\Delta\lambda \log x} \\ &\approx c_1 x^\lambda + c_2 x^\lambda (1 + \Delta\lambda \log x) \quad (\text{Taylor approximation}) \\ &= (c_1 + c_2) x^\lambda + c_2 \Delta\lambda x^\lambda \log x \\ &= \alpha_1 x^\lambda + \alpha_2 x^\lambda \log x , \end{aligned}$$

where α_1 and α_2 are arbitrary constants. Now we let $\Delta\lambda \rightarrow 0$ without affecting the general form of the solution.

If the Cauchy–Euler equation of higher order has the associated polynomial in λ with a multiple root then we suppose that there exist distinct real roots λ , $\lambda + \Delta\lambda$, $\dots$, $\lambda + k\Delta\lambda$ and the general solution is of the form

$$c_1 x^\lambda + c_2 x^{\lambda+\Delta\lambda} + \dots + c_k x^{\lambda+k\Delta\lambda}$$

and use the Taylor polynomials of degree up to k , which produces

$$\alpha_1 x^\lambda + \alpha_2 x^\lambda \log x + \dots + \alpha_k x^\lambda (\log x)^k$$

where $\alpha_1, \dots, \alpha_k$ are arbitrary constants.²

D.4 Nonlinear Differential Equations

Definition D.1 An ordinary differential equation of the form

$$\frac{dy}{dx} + P(x)y = Q(x)y^n \tag{D.4}$$

is called a Bernoulli equation.

For $n \neq 0$ and $n \neq 1$, multiplying both sides of (D.4) by $(1 - n)y^{-n}$, we obtain

² A similar technique can be used for the linear differential equations with constant coefficients, using the Taylor approximation $e^x = 1 + x + \frac{1}{2}x^2 + \dots$. See Sect. D.1.

$$(1-n)y^{-n}\frac{dy}{dx} + (1-n)P(x)y^{1-n} = (1-n)Q(x) .$$

By the change of variable $z = y^{1-n}$, we have

$$\frac{dz}{dx} = (1-n)y^{-n}\frac{dy}{dx} .$$

Hence the nonlinear equation given by (D.4) is converted into the linear equation

$$\frac{dz}{dx} + (1-n)P(x)z = (1-n)Q(x) .$$

Definition D.2 An ordinary differential equation of the form

$$\frac{dy}{dx} = p(x) + q(x)y + r(x)y^2 \tag{D.5}$$

is called a Riccati equation.

There is no closed-form solution for (D.5) and we have to resort to numerical solutions, in general. However, if we can find a particular solution $y_1(x)$, then the general solution is of the form

$$y(x) = y_1(x) + z(x) ,$$

where $z(x)$ is the general solution of

$$\frac{dz}{dx} - \{q(x) + 2r(x)y_1(x)\} z(x) = r(x)z(x)^2 ,$$

which is the Bernoulli equation with $n = 2$.

D.5 Ordinary Differential Equations Defined by Vector Fields

A vector field on $\mathbb{R}^n$ is a vector-valued function $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$. If a curve $\phi : (a, b) \rightarrow \mathbb{R}^n$ satisfies the differential equation $\frac{d\phi}{dt} = \mathbf{F}(\phi(t))$, then ϕ is called an integral curve of the vector field $\mathbf{F}$.

Definition D.3 (*Exponential of a Matrix*) Let A denote an $n \times n$ matrix. Define e^A by

$$\exp(A) = e^A = \sum_{k=0}^{\infty} \frac{1}{k!} A^k ,$$

where the infinite sum is defined by the convergence of finite sums of matrices in the n^2 -dimensional Euclidean space. To show that the sum converges, first we regard $A = (A_{ij})_{i,j=1}^n$ as a point in the Euclidean space $\mathbb{R}^{n^2}$ with the norm $\|\cdot\|$ defined by $\|A\| = \max_{ij} |A_{ij}|$. Recall that $\mathbb{R}^{n^2}$ is a complete metric space. Then we use the fact that the partial sums $\sum_{n=1}^N \frac{1}{n!} A^n$ form a Cauchy sequence in $\mathbb{R}^{n^2}$, and converges to a limit as $N \rightarrow \infty$.

Theorem D.1 (Differential Equation Defined by a Matrix) *Let $\mathbf{x} : \mathbb{R}^1 \rightarrow \mathbb{R}^n$ be a differentiable curve. Consider an ordinary differential equation*

$$\frac{d}{dt}\mathbf{x}(t) = A\mathbf{x}(t), \quad \mathbf{x}(0) = \mathbf{x}_0$$

defined by an $n \times n$ matrix A . Then its solution is given by $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$.

Example D.9 (Ordinary Differential Equation) Consider the differential equation $x'' = x$. To convert it into a first order differential equation we put $y = x'$. Then we have a system of equations $x' = y$, $y' = x$. Let $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then

$$\frac{d}{dt}\mathbf{x} = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix} = A\mathbf{x}.$$

Its solution is given by $\mathbf{x}(t) = e^{tA}\mathbf{x}_0$.

Appendix E

Solution of the Heat Equation

An equation consisting of partial derivatives of a function of several variables is called a partial differential equation. When one of the variables represents time, it is usually denoted by t . In this appendix we present examples of second order linear partial differential equations such as the Laplace equation, the heat equation and the wave equation, and introduce the Fourier transform and the Laplace transform, focusing on the solution of the heat equation.

E.1 Examples of Partial Differential Equations

The following are typical examples of second order linear partial differential equations with constant coefficients.

Example E.1 The partial differential equation

$$\frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2} = 0$$

is called the *Laplace equation*, and its solutions are called harmonic functions. For the existence and uniqueness of the solution, we impose a boundary condition along the boundary of the domain.

Example E.2 The partial differential equation

$$\frac{\partial f}{\partial t} = a \frac{\partial^2 f}{\partial x^2}, \quad a > 0,$$

is called the *heat equation* or *diffusion equation*, and represents dissipation of heat or chemicals. The Black–Scholes–Merton partial differential equation in Chap. 15 can be converted to the heat equation through changes of variables.

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2026

G. H. Choe, *Quantitative Methods for Finance with Simulations I*, Springer Texts in Business and Economics, <https://doi.org/10.1007/978-3-032-12327-5>

Example E.3 The partial differential equation

$$\frac{\partial^2 f}{\partial t^2} - \left(\frac{\partial^2 f}{\partial x_1^2} + \cdots + \frac{\partial^2 f}{\partial x_n^2} \right) = 0$$

is called the *wave equation*, and describes propagation of waves in n -dimensional space. For the existence and uniqueness of the solution, we need an initial condition, and in some cases a boundary condition.

E.2 The Fourier Transform

For a function $f(x)$ such that $|f(x)| \rightarrow 0$ rapidly as $x \rightarrow \pm\infty$, we define the Fourier transform $\hat{f}(\xi)$ by

$$\hat{f}(\xi) = \mathcal{F}[f(x)](\xi) = \int_{-\infty}^{\infty} e^{-i\xi x} f(x) dx .$$

(Often, the right-hand side of the definition is multiplied by a constant $\frac{1}{\sqrt{2\pi}}$. See Rudin (1986), for example.)

Fact E.1 Let f and g be rapidly decreasing functions, that is, f and g and their derivatives converge to 0 rapidly as $x \rightarrow \pm\infty$. Then the following statements hold:

- (i) (Multiplication by x corresponds to i times differentiation) If $g(x) = xf(x)$, then

$$\hat{g}(\xi) = i \frac{d}{d\xi} \hat{f}(\xi) .$$

- (ii) (Differentiation corresponds to multiplication by $i\xi$)

$$\widehat{f'}(\xi) = i\xi \hat{f}(\xi) .$$

- (iii) (Fubini's theorem)

$$\int \hat{f}(\xi) g(\xi) d\xi = \int f(\xi) \hat{g}(\xi) d\xi .$$

- (iv) (Parallel translation) For a constant a let $f_a(x) = f(x + a)$. Then

$$\hat{f}_a(\xi) = e^{ia\xi} \hat{f}(\xi) .$$

- (v) (Inverse Fourier transform) If f is a rapidly decreasing function, then so is $g = \hat{f}$, and

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(\zeta) e^{ix\zeta} d\zeta .$$

For more information consult (Adams & Guillemin, 1996).

Example E.4 Let

$$f(x) = \begin{cases} 0, & x < 0, \\ e^{-x}, & x \geq 0. \end{cases}$$

Its Fourier transform is given by $\hat{f}(\zeta) = \frac{1 - i\zeta}{1 + \zeta^2}$.

Example E.5 Let

$$f(x) = \begin{cases} 1, & -1 \leq x \leq 1, \\ 0, & |x| > 1. \end{cases}$$

Its Fourier transform is given by $\hat{f}(\zeta) = \frac{2 \sin \zeta}{\zeta}$.

Example E.6 The Fourier transform of the function $f(x) = e^{-x^2/2}$ is given by $\hat{f}(\zeta) = \sqrt{2\pi} e^{-\zeta^2/2}$. In other words, up to constant multiple, the Fourier transform of f is itself. To see why, first note that $f'(x) + xf(x) = 0$. Hence, by Fact E.1 (i), (ii), we have $i\zeta \hat{f}(\zeta) + i\hat{f}'(\zeta) = 0$. Thus $\hat{f}(\zeta) = Ce^{-\zeta^2/2}$ and $C = \hat{f}(0) = \int_{-\infty}^{\infty} f(x) dx = \sqrt{2\pi}$.

Example E.7 (*Solution of heat equation*) Consider the heat equation given by

$$\frac{\partial f}{\partial t} = a^2 \frac{\partial^2 f}{\partial x^2}, \quad -\infty < x < \infty, \quad t > 0 \quad (\text{E.1})$$

together with the initial condition $f(x, 0) = g(x)$, $-\infty < x < \infty$, and the boundary condition $f(x, t) \rightarrow 0$, $\frac{\partial f}{\partial x}(x, t) \rightarrow 0$ as $x \rightarrow \pm\infty$, $t > 0$. Apply the Fourier transform to both sides of (E.1), and derive and solve the resulting equation in terms of $F(s)$, and apply the inverse Fourier transform. Then we have

$$f(x, t) = \frac{1}{2a\sqrt{\pi t}} \int_{-\infty}^{\infty} g(s) \exp\left(-\frac{(x-s)^2}{4a^2 t}\right) ds .$$

E.3 The Laplace Transform

Define the Laplace transform $F(s)$ of a function $f(t)$, $t \geq 0$, by

$$F(s) = \mathcal{L}[f(t)](s) = \int_0^{\infty} e^{-st} f(t) dt .$$

For some functions the Laplace transform is not defined. For example, for $s > 0$,

$$\mathcal{L}\left[\frac{1}{t}\right] = \int_0^1 \frac{e^{-st}}{t} dt + \int_1^{\infty} \frac{e^{-st}}{t} dt > \int_{\varepsilon}^1 \frac{e^{-s}}{t} dt \rightarrow +\infty$$

as $\varepsilon \downarrow 0$. If $f(t)$ is piecewise continuous on $t \geq 0$ and satisfies the condition that $|f(t)| \leq M e^{at}$, $a \leq 0$, then $\mathcal{L}[f(t)](s)$ exists for all $s > a$. It is easy to see that $\mathcal{L}$ is a linear transformation.

Fact E.2 The following basic facts are known. Consult (Strauss, 2008) for the proofs.

- (i) $\mathcal{L}[f(at)] = \frac{1}{a} F\left(\frac{s}{a}\right)$, $a \neq 0$.
- (ii) $\mathcal{L}[t^{\alpha}] = \frac{\Gamma(\alpha+1)}{s^{\alpha+1}}$, $\alpha \geq 0$. In particular, $\mathcal{L}[1] = \frac{1}{s}$, $s > 0$.
- (iii) $\mathcal{L}[e^{at}] = \frac{1}{s-a}$, $s > a$.
- (iv) $\mathcal{L}[f'(t)] = s\mathcal{L}[f] - f(0)$.
- (v) $\mathcal{L}[f''(t)] = s^2\mathcal{L}[f] - sf(0) - f'(0)$.
- (vi) $\mathcal{L}[tf(t)] = -F'(s)$.
- (vii) $\mathcal{L}\left[\int_0^t f(u)du\right] = \frac{1}{s}\mathcal{L}[f(t)]$.
- (viii) $\mathcal{L}\left[\frac{f(t)}{t}\right] = \int_s^{\infty} F(v)dv$.
- (ix) $\mathcal{L}[e^{at}f(t)] = F(s-a)$.
- (x) $\mathcal{L}[f(t-a)H(t-a)] = e^{-as}\mathcal{L}[f]$, $a > 0$, where $H(x)$ is the Heaviside function defined by $H(x) = 1$, $x \geq 0$, and $H(x) = 0$, $x < 0$.
- (xi) $\mathcal{L}[\delta(t-a)] = e^{-as}$, $a > 0$.

Example E.8

$$\begin{aligned} \mathcal{L}\left[t^{-\frac{1}{2}}\right] &= \int_0^{\infty} e^{-st} t^{-\frac{1}{2}} dt \quad (\text{Take } st = u.) \\ &= s^{-\frac{1}{2}} \int_0^{\infty} e^{-u} u^{-\frac{1}{2}} du \quad (\text{Take } u = z^2.) \\ &= 2s^{-\frac{1}{2}} \int_0^{\infty} e^{-z^2} dz = 2s^{-\frac{1}{2}} \frac{\sqrt{\pi}}{2} . \end{aligned}$$

Fact E.3 The following facts are known:

- (i) Define the convolution of f and g by

$$(f * g)(t) = \int_0^t f(t-u) g(u) du .$$

Then $\mathcal{L}[f * g] = \mathcal{L}[f] \mathcal{L}[g]$.

- (ii) The inverse of the Laplace transform is given by

$$\mathcal{L}^{-1}[F(s)] = f(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} F(s) e^{st} ds .$$

Consider the error function in Definition 4.2 given by

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du , \quad -\infty < x < \infty .$$

Let $N(\cdot)$ denote the cumulative distribution function of the standard normal distribution. Then

$$\frac{1 + \operatorname{erf}(x)}{2} = \int_{-\infty}^{\sqrt{2}x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = N(\sqrt{2}x) . \quad (\text{E.2})$$

Fact E.4 The error function satisfies

$$\mathcal{L}[\operatorname{erf}(\sqrt{t})] = \frac{1}{s\sqrt{s+1}} .$$

To see why, note that

$$\operatorname{erf}(\sqrt{t}) = \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{t}} e^{-x^2} dx = \frac{1}{\sqrt{\pi}} \int_0^t u^{-\frac{1}{2}} e^{-u} du .$$

Hence

$$\mathcal{L}[\operatorname{erf}(\sqrt{t})] = \frac{1}{\sqrt{\pi}} \mathcal{L}\left[\int_0^t u^{-\frac{1}{2}} e^{-u} du\right] = \frac{1}{\sqrt{\pi}} \frac{1}{s} \mathcal{L}\left[t^{-\frac{1}{2}} e^{-t}\right] .$$

Now use $\mathcal{L}[t^{-\frac{1}{2}} e^{-t}] = \frac{\sqrt{\pi}}{\sqrt{s+1}}$.

Fact E.5 For $C \geq 0$ the error function satisfies

$$\mathcal{L}\left[\operatorname{erf}\left(\frac{C}{\sqrt{t}}\right)\right] = \frac{1}{s}(1 - e^{-2C\sqrt{s}}) .$$

Example E.9 (*Heat Equation*) Using the Laplace transformation we solve

$$\frac{\partial f}{\partial t} = a^2 \frac{\partial^2 f}{\partial x^2} \quad (\text{E.3})$$

for $a > 0$. The initial condition is given by the Dirac delta function $f(x, 0) = \delta_0(x)$. For $t > 0$, $f(x, t)$ represents a probability density and satisfies

$$\lim_{x \rightarrow \pm\infty} f(x, t) = 0$$

and

$$\lim_{x \rightarrow \pm\infty} \frac{\partial f}{\partial x}(x, t) = 0 .$$

This problem is not easy to solve because the initial condition is given by a measure. We make it smoother in the following way: First, put

$$u(y, t) = \int_{-\infty}^y f(x, t) dx .$$

From (E.3) we obtain

$$\int_{-\infty}^y \frac{\partial f}{\partial t}(x, t) dx = a^2 \int_{-\infty}^y \frac{\partial^2 f}{\partial x^2}(x, t) dx .$$

Then the left- and the right-hand sides are respectively equal to $\frac{\partial u}{\partial t}(y, t)$ and

$$a^2 \int_{-\infty}^y \frac{\partial^2 f}{\partial x^2}(x, t) dx = a^2 \left(\frac{\partial f}{\partial x}(y, t) - \frac{\partial f}{\partial x}(-\infty, t) \right) = a^2 \frac{\partial f}{\partial x}(y, t) .$$

Hence we obtain

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} \quad (\text{E.4})$$

with $u(x, 0) = H(x)$, where $H(x)$ is the Heaviside function. Recall that from the distribution theory, established by Laurent Schwartz, we have $H'(x) = \delta_0(x)$.

Let $U(x, s)$ denote the Laplace transform of $u(x, t)$ with respect to t . Since

$$\mathcal{L} \left[\frac{\partial u}{\partial t} \right] (x, s) = s \mathcal{L}[u](x, s) - u(x, 0) ,$$

by taking the Laplace transforms of both sides of (E.4), we obtain

$$a^2 \frac{\partial^2 U}{\partial x^2} - s U = -H(x) . \quad (\text{E.5})$$

To solve it, we first find a solution of the corresponding homogeneous equation, for which there exist two linearly independent solutions $e^{(\sqrt{s}/a)x}$ and $e^{-(\sqrt{s}/a)x}$. Now, for $x > 0$, (E.5) has a particular solution $U_P(x, s) = \frac{1}{s}$, and hence (E.5) has a general solution of the form

$$U(x, s) = c_1 e^{(\sqrt{s}/a)x} + c_2 e^{-(\sqrt{s}/a)x} + \frac{1}{s}.$$

Since $\lim_{x \rightarrow \infty} u(x, t) = 1$, we have

$$\lim_{x \rightarrow \infty} U(x, s) = \frac{1}{s}.$$

Hence $c_1 = 0$ and the general solution of (E.5) is of the form

$$U(x, s) = c_2 e^{-(\sqrt{s}/a)x} + \frac{1}{s}.$$

Recall that there still remains an unused condition. The initial data is symmetric and the differential equation is also symmetric, and hence the solution, which is a probability density function $f(x, t)$, is also symmetric with respect to $x = 0$. Hence $u(0, t) = \frac{1}{2}$ and $U(0, s) = \frac{1}{2s}$. Since

$$c_2 e^{-(\sqrt{s}/a) \times 0} + \frac{1}{s} = \frac{1}{2s},$$

we have $c_2 = -\frac{1}{2s}$, and the general solution of (E.5) is of the form

$$U(x, s) = -\frac{1}{2s} e^{-(\sqrt{s}/a)x} + \frac{1}{s}. \quad (\text{E.6})$$

Finally, taking the inverse Laplace transform of (E.6), we obtain

$$\begin{aligned} u(x, t) &= \mathcal{L}^{-1} \left(-\frac{1}{2s} e^{-(\sqrt{s}/a)x} + \frac{1}{s} \right) \\ &= \frac{1}{2} \mathcal{L}^{-1} \left(\frac{1}{s} (1 - e^{-(\sqrt{s}/a)x}) \right) + \mathcal{L}^{-1} \left(\frac{1}{2s} \right) \\ &= \frac{1}{2} \operatorname{erf} \left(\frac{x}{2a\sqrt{t}} \right) + \frac{1}{2} \\ &= N \left(\frac{x}{a\sqrt{2t}} \right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x/a\sqrt{2t}} \exp \left(-\frac{y^2}{2} \right) dy. \end{aligned}$$

Therefore

$$\begin{aligned} f(x, t) &= \frac{\partial u}{\partial x}(x, t) \\ &= \frac{1}{2a\sqrt{\pi t}} \exp\left(-\frac{x^2}{4a^2 t}\right) . \end{aligned}$$

Such a solution is called a fundamental solution of the heat equation. For a slightly different approach, consult (Gianin & Sgarra, [2013](#)).

Appendix F

Statistical Facts

In this appendix we introduce χ^2 -distributions and noncentral χ^2 -distributions.

F.1 Sums of Independent Normal Variables

Definition F.1 (χ^2 -Distribution) For independent standard normal random variables $Z_1, \dots, Z_n$, the distribution of the sum

$$X = Z_1^2 + \dots + Z_n^2$$

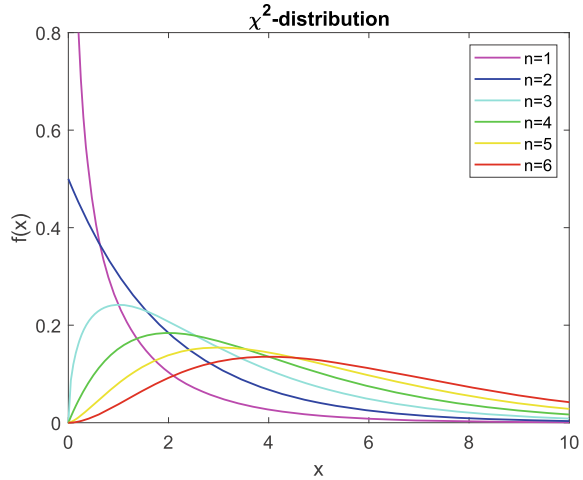
is called the *chi-squared distribution*. In this case, we write

$$X \sim \chi^2(n) .$$

Example F.1 The probability density function of the chi-squared distribution for $n = 2$ is given by $f(x) = \frac{1}{2}e^{-x/2}$, $x \geq 0$. To see why, note that

$$\begin{aligned} f_{Z_1^2+Z_2^2}(x) &= \int_{-\infty}^{\infty} f_{Z_1^2}(x-y)f_{Z_2^2}(y)dy \\ &= \int_0^x f_{Z_1^2}(x-y)f_{Z_2^2}(y)dy \\ &= \int_0^x \frac{1}{\sqrt{2\pi(x-y)}}e^{-(x-y)/2} \frac{1}{\sqrt{2\pi y}}e^{-y/2}dy \\ &= \frac{1}{2\pi}e^{-x/2} \int_0^x \frac{1}{\sqrt{y(x-y)}}dy \\ &= \frac{1}{2}e^{-x/2} . \end{aligned}$$

Fig. F.1 Probability density functions of $\chi^2(n)$ for $1 \leq n \leq 6$



For the last equality we used the fact that

$$\int_0^x \frac{1}{\sqrt{y(x-y)}} dy = \int_0^1 \frac{1}{\sqrt{u(1-u)}} du = \pi. \quad (\text{Substitute } y = xu.)$$

The pdf for any $n \geq 1$ is given by

$$f(x) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} e^{-x/2} x^{n/2-1}, \quad x > 0,$$

where Γ denotes the gamma function. The mean and variance of the $\chi^2(n)$ -distribution are n and $2n$, respectively. For $n = 1$ see Example 4.19, and note that

$$\text{Var}(Z^2) = \mathbb{E}[(Z^2)^2] - (\mathbb{E}[Z^2])^2 = 3 - 1 = 2.$$

For the graph of $f(x)$ for $1 \leq n \leq 6$, see Fig. F.1 and Simulation F.1.

Definition F.2 (*Noncentral χ^2 -Distribution*) For independent normal variables $Y_1, \dots, Y_n$ with mean μ_i and variance 1, $1 \leq i \leq n$, the sum

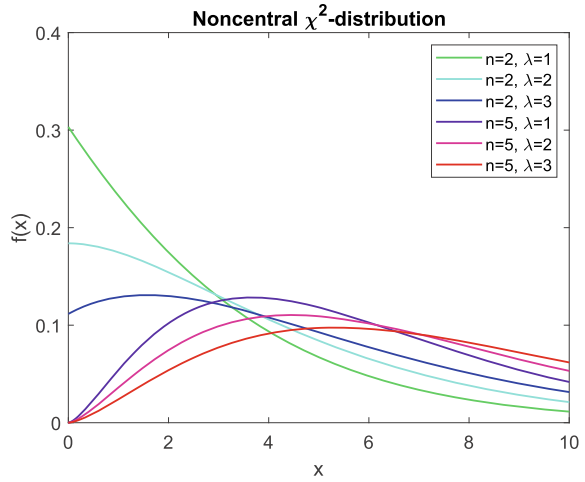
$$X = Y_1^2 + \dots + Y_n^2$$

is said to have the *noncentral chi-squared distribution* with noncentrality parameter

$$\lambda = \mu_1^2 + \dots + \mu_n^2.$$

Its probability density function is given by

Fig. F.2 Probability density functions of noncentral $\chi^2(n)$ for $n = 2, 5$ and $\lambda = 1, 2, 3$



$$f(x) = e^{-\lambda/2} {}_0F_1 \left(; \frac{n}{2}; \frac{\lambda x}{4} \right) \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} e^{-x/2} x^{n/2-1}, \quad x > 0,$$

where ${}_pF_q$ is the hypergeometric function defined by

$${}_pF_q(a_1, \dots, a_p; b_1, \dots, b_q; x) = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \frac{x^k}{k!}$$

where

$$(a)_k = a(a+1) \cdots (a+k-1).$$

The mean and variance are given by $n + \lambda$ and $2n + 4\lambda$. Consult (Muirhead, 2005). For the graph of $f(x)$, see Fig. F.2 and Simulation F.2. For an example in interest rate modeling, see Theorem 23.4.

F.2 Computer Experiments

Simulation F.1 (χ^2 -Distribution) We plot the pdf of the χ^2 -distribution for various values of n . For the output, see Fig. F.1.

```
M = 6;
for n = 1:M
    x = 0:0.05:10;
    y = x.^(n/2 - 1) .* exp(-x/2) / (2^(n/2) * gamma(n/2));
```

```

plot(x,y,'color',hsv2rgb([1-n/M 1 1]))
hold on;
end

legend('n=1','n=2','n=3','n=4','n=5','n=6');
title('\chi^2-distribution')
xlabel('x')
ylabel('f(x)')

```

Simulation F.2 (Noncentral χ^2 -Distribution) We plot the pdf of the noncentral χ^2 -distribution for various values of k and λ . For the output, see Fig. F.2.

```

for n = 2:3:5
    for lambda = 1:1:3
        x = 0:0.05:10;
        y = ncx2pdf(x,n,lambda);
        plot(x,y,'color',hsv2rgb([(n+lambda)/8 1 1]))
        hold on;
    end
end

legend('n=2, \lambda=1','n=2, \lambda=2','n=2, \lambda=3', ...
       'n=5, \lambda=1','n=5, \lambda=2','n=5, \lambda=3');
title('Noncentral \chi^2-distribution')
xlabel('x')
ylabel('f(x)')

```

Solutions for Selected Problems

Problems of Chap. 1

1.1 Let K be the closed cone given by all convex linear combinations of $\mathbf{d}_1, \dots, \mathbf{d}_M$ with nonnegative coefficients. Then either $\mathbf{d}_0 \in K$ or $\mathbf{d}_0 \notin K$. If $\mathbf{d}_0 \in K$, then (i) has a solution. In this case, $\boldsymbol{\pi} \cdot \mathbf{d}_0 = \lambda_1 \boldsymbol{\pi} \cdot \mathbf{d}_1 + \dots + \lambda_M \boldsymbol{\pi} \cdot \mathbf{d}_M$, and (ii) does not hold. If $\mathbf{d}_0 \notin K$, then there exists a hyperplane H in $\mathbb{R}^N$ that separates $\mathbf{d}_0$ and K . Choose a normal vector $\boldsymbol{\pi} \neq \mathbf{0}$ to H and pointing in the direction where K is located. Then $\boldsymbol{\pi}$ satisfies the second statement, but not the first statement.

1.2 Take $\mathbf{d}_0 = \mathbf{S}_0$ in Farkas' Lemma. Then case (ii) does not happen and case (i) holds.

1.3

- (i) Consider the equation $\begin{bmatrix} C^u \\ C^d \end{bmatrix} = D^t \begin{bmatrix} a \\ b \end{bmatrix}$. Unless $(S_T^2(\omega_u), S_T^2(\omega_d))$ is a constant multiple of $(S_T^1(\omega_u), S_T^1(\omega_d))$, there is a unique solution for every (C^u, C^d) . That is, unless S^1 and S^2 are essentially the same asset with a different number of units, a unique solution exists and the market is complete.
- (ii) The given system of equations becomes

$$\begin{cases} C^u = auS_0 + b(1+r)B_0 \\ C^d = adS_0 + b(1+r)B_0 \end{cases}.$$

Since $d \neq u$, the determinant of $D^t = \begin{bmatrix} uS_0 & b(1+r) \\ dS_0 & b(1+r) \end{bmatrix}$ is nonzero, and hence there exists a unique solution.

Problems of Chap. 2

2.1 The payoff of the portfolio is given by $-S_T + (S_T - K)^+$, which can be written as a piecewise linear function $f(S_T)$ defined by $f(S_T) = -S_T$, $0 \leq S_T \leq K$, and $f(S_T) = -K$, $S_T \geq K$. The maximum loss is equal to $-K$.

2.2 A call option provides insurance if the holder of the call option has a short position in the asset on which the option is written. A put option provides insurance if the holder of the put option owns the asset on which the option is written.

2.3 A strangle is a sum of a put option with payoff $(K_1 - S_T)^+$ and a call option with payoff $(S_T - K_2)^+$.

Problems of Chap. 3

3.3 Since $A = \bigcup_{k=1}^{\infty} \left\{x = \sum_{n=1}^k a_n 2^{-n} : a_n = 0, 1\right\}$, A has countably many points, and its Lebesgue measure is 0.

3.4 If we take $\phi(x) = \frac{1}{x}$ in Jensen's inequality, then $\frac{1}{\int X d\mathbb{P}} \leq \int \frac{1}{X} d\mathbb{P}$ and hence $\int X d\mathbb{P} \int \frac{1}{X} d\mathbb{P} \geq 1$. Since $Y \geq \frac{1}{X}$, we conclude that $\int X d\mathbb{P} \int Y d\mathbb{P} \geq 1$.

3.5 For any partition P of $[0, 1]$, we have the lower Riemann sum $L(f, P) = 0$ and the upper Riemann sum $U(f, P) = \frac{1}{2}$. Thus f is not Riemann integrable. Since $f(x) = x$ almost everywhere, $\int_{[0,1]} f dm = \frac{1}{2}$.

3.6 Let $A_n = \{\omega : n \leq X(\omega) < n+1\}$, $n \geq 0$, then Ω is a pairwise disjoint union of A_n . Hence

$$\begin{aligned} \sum_{n=0}^{\infty} \mu(\{\omega : X(\omega) \geq n\}) &= \sum_{n=0}^{\infty} \sum_{m=n}^{\infty} \mu(\{m \leq X(\omega) < m+1\}) \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^m \mu(\{m \leq X(\omega) < m+1\}) \\ &\geq \sum_{m=0}^{\infty} \int_{A_m} X d\mu = \int X d\mu. \end{aligned}$$

3.7 Note that $\frac{\sin x}{x} < 1$ for $0 < x < \infty$ and

$$\int_0^{\infty} \left(\frac{\sin x}{x}\right)^n dx \leq \int_0^1 1 dx + \int_1^{\infty} \frac{1}{x^2} dx < \infty.$$

By the Lebesgue Dominated Convergence Theorem,

$$\lim_{n \rightarrow \infty} \int_0^\infty \left(\frac{\sin x}{x} \right)^n dx = \int_0^\infty \lim_{n \rightarrow \infty} \left(\frac{\sin x}{x} \right)^n dx = 0.$$

Problems of Chap. 4

4.1 Note that

$$\sum_{j=0}^{\infty} \mathbb{P}(\{X \geq j\}) = \sum_{j=0}^{\infty} \sum_{k=j}^{\infty} \mathbb{P}(\{k \leq X < k+1\}).$$

4.2 Since $\phi(X) = 1$, we have $0 = \phi(\mathbb{E}[X]) < \mathbb{E}[\phi(X)] = 1$.

$$4.3 \quad F(x) = \begin{cases} 0, & x < a, \\ 1, & x \geq a. \end{cases}$$

4.4 Substituting $t = \sqrt{x}$, we have

$$\int_0^1 e^{\sqrt{x}} dx = \int_0^1 e^t 2t dt = [e^t 2t]_0^1 - \int_0^1 e^t 2 dt = 2e - 2(e - 1) = 2.$$

4.5 Let $f(x)$ denote the pdf of the standard normal distribution. Since $f(-x) = f(x)$, we have $N(a) + N(-a) = \int_{-\infty}^a f(x) dx + \int_{-\infty}^{-a} f(x) dx = \int_{-\infty}^a f(x) dx + \int_a^\infty f(x) dx = \int_{-\infty}^\infty f(x) dx = 1$.

4.6 Since $\Pr(F^{-1}(U) \leq x) = \Pr(U \leq F(x)) = F(x)$, we see that $F^{-1}(U)$ and X have the same distribution function F .

4.7

$$\begin{aligned} \mathbb{E}[e^X] &= \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx \quad (\text{Take } z = \frac{x-\mu}{\sigma}.) \\ &= \int_{-\infty}^{\infty} e^{\sigma z + \mu} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= \int_{-\infty}^{\infty} e^{\mu + \sigma^2/2} \frac{1}{\sqrt{2\pi}} e^{-(z-\sigma)^2/2} dz \\ &= e^{\mu + \sigma^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = e^{\mu + \sigma^2/2}. \end{aligned}$$

Since $2X$ has mean 2μ and variance $4\sigma^2$, we have $\mathbb{E}[e^{2X}] = e^{2\mu + 2\sigma^2}$. Thus

$$\begin{aligned}
\text{Var}(e^X) &= \mathbb{E}[(e^X - e^{\mu+\sigma^2/2})^2] \\
&= \mathbb{E}[e^{2X} - 2e^{\mu+\sigma^2/2}e^X + e^{2\mu+\sigma^2}] \\
&= e^{2\mu+2\sigma^2} - 2e^{\mu+\sigma^2/2}e^{\mu+\sigma^2/2} + e^{2\mu+\sigma^2} \\
&= e^{2\mu+\sigma^2}(e^{\sigma^2} - 1) .
\end{aligned}$$

4.8 Let Z be a standard normal variable and take $u = a + ib \in \mathbb{C}$, $a, b \in \mathbb{R}$. Compute $\mathbb{E}[e^{uZ}]$. Then

$$\begin{aligned}
\mathbb{E}[e^{uZ}] &= \int_{-\infty}^{\infty} e^{uz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\
&= e^{a^2/2} \int_{-\infty}^{\infty} e^{ibz} \frac{1}{\sqrt{2\pi}} e^{-(z-a)^2/2} dz \quad (\text{Take } y = z - a.) \\
&= e^{a^2/2} e^{iab} \int_{-\infty}^{\infty} e^{iby} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\
&= e^{a^2/2} e^{iab} e^{-b^2/2} = e^{u^2/2} .
\end{aligned}$$

To find the moments of Z compare the coefficients of the Taylor series expansions

$$\sum_{k=0}^{\infty} \frac{1}{k!} \mathbb{E}[Z^k] u^k = \mathbb{E}[e^{uZ}] = e^{\frac{1}{2}u^2} = \sum_{n=1}^{\infty} \frac{1}{n!} \frac{1}{2^n} u^{2n} .$$

4.9 Since αZ is normal with mean 0 and variance α^2 , by Exercise 4.7 we have

$$\|e^{\alpha Z}\|_p^p = \mathbb{E}[e^{\alpha p Z}] = e^{\alpha^2 p^2/2} < \infty .$$

4.10

$$\begin{aligned}
\mathbb{E}[e^{\alpha Z^2}] &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\alpha x^2} e^{-x^2/2} dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1-2\alpha)x^2/2} dx \quad (\text{Take } z = \sqrt{1-2\alpha} x .) \\
&= \frac{1}{\sqrt{1-2\alpha}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{1}{\sqrt{1-2\alpha}} .
\end{aligned}$$

4.11 By Cauchy's theorem, we have

$$0 = \oint_{\Gamma} f(z) dz = \int_{\Gamma_1} f(z) dz + \int_{\Gamma_2} f(z) dz + \int_{\Gamma_3} f(z) dz + \int_{\Gamma_4} f(z) dz . \quad (\text{F.1})$$

Since

$$\int_{\Gamma_1} f(z) dz = - \int_{-R}^R e^{-u^2/(2t)} du ,$$

we have

$$\lim_{R \rightarrow +\infty} \int_{\Gamma_1} f(z) dz = -\sqrt{2\pi t} .$$

Since

$$\int_{\Gamma_2} f(z) dz = \int_0^t e^{-(R-iu)^2/(2t)} du = \int_0^t e^{(-R^2+u^2-2iRu)/(2t)} du$$

and

$$\int_{\Gamma_4} f(z) dz = - \int_0^t e^{-(R-iu)^2/(2t)} du = \int_0^t e^{(-R^2+u^2+2iRu)/(2t)} du ,$$

we have

$$\left| \int_{\Gamma_2} f(z) dz \right| \leq \int_0^t e^{(-R^2+u^2)/(2t)} du \leq t e^{(-R^2+t^2)/(2t)} \rightarrow 0$$

as $R \rightarrow +\infty$, and similarly

$$\lim_{R \rightarrow +\infty} \int_{\Gamma_4} f(z) dz = 0 .$$

Hence, by taking limits of both sides of (F.1), we obtain

$$0 = -\sqrt{2\pi t} + \lim_{R \rightarrow +\infty} \int_{\Gamma_3} f(z) dz . \quad (\text{F.2})$$

Thus

$$\begin{aligned} \mathbb{E}[e^{iX}] &= \int_{-\infty}^{\infty} e^{ix} \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-(x-it)^2/(2t)} e^{-t/2} dx \\ &= e^{-t/2} \frac{1}{\sqrt{2\pi t}} \lim_{R \rightarrow +\infty} \int_{-R}^R e^{-(x-it)^2/(2t)} dx \\ &= e^{-t/2} \frac{1}{\sqrt{2\pi t}} \lim_{R \rightarrow +\infty} \int_{\Gamma_3} f(z) dz = e^{-t/2} , \end{aligned}$$

where we used (F.2) for the last equality.

4.12 Note that

$$\mathbb{E}[Y^4] = \mathbb{E} \left[\sum_i \sum_j \sum_k \sum_\ell (U_i - \frac{1}{2})(U_j - \frac{1}{2})(U_k - \frac{1}{2})(U_\ell - \frac{1}{2}) \right] .$$

There are five cases: (i) all of i, j, k, ℓ are distinct, (ii) there exists only one equal pair, (iii) there are two equal pairs, (iv) one equal triple, and (v) all are equal. The cases (i), (ii), (iv) produce zero, while (iii), (v) yield $\frac{1}{144} \binom{12}{2} \binom{4}{2} \binom{2}{2} = \frac{11}{4}$ and $\frac{1}{80} \binom{12}{1} \binom{4}{4} = \frac{3}{20}$, respectively. Hence $\mathbb{E}[Y^4] = \frac{11}{4} + \frac{3}{20} = \frac{29}{10}$.

4.13 Since $h(x) = x^2$, we have $h^{-1}(y) = \pm\sqrt{y}$ for $0 \leq y \leq 1$ and

$$f_Y(y) = \frac{1}{|-2\sqrt{y}|} f_X(-\sqrt{y}) + \frac{1}{|2\sqrt{y}|} f_X(\sqrt{y}) = \frac{1}{2\sqrt{y}}, \quad 0 < y \leq 1.$$

4.14 Since X and Y are independent, the joint pdf is the product of f_X and f_Y , i.e., $f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$. Hence for $-\frac{\pi}{2} < \theta_1 < \theta_2 < \frac{\pi}{2}$,

$$\begin{aligned} \Pr(\theta_1 < \tan^{-1} \frac{Y}{X} < \theta_2) &= \Pr(\tan \theta_1 < \frac{Y}{X} < \tan \theta_2) \\ &= \iint_D \frac{1}{2\pi} e^{-(x^2+y^2)/2} dx dy \\ &= \frac{1}{\pi} \int_0^\infty \int_{\theta_1}^{\theta_2} e^{-r^2/2} r dr d\theta = \frac{\theta_2 - \theta_1}{\pi}, \end{aligned}$$

where $D = \{(x, y) : \theta_1 < \tan^{-1} \frac{y}{x} < \theta_2\}$. (Note that D has two pieces!)

4.15 By the chain rule and the invariance of probability, we have

$$\begin{aligned} \iint_Q f_{V,W}(\Phi(x, y)) |J_\Phi(x, y)| dx dy &= \iint_{\Phi(Q)} f_{V,W}(v, w) dv dw \\ &= \iint_Q f_{X,Y}(x, y) dx dy \end{aligned}$$

for any square $Q \subset \mathbb{R}^2$. Thus $f_{V,W}(\Phi(x, y)) |J_\Phi(x, y)| = f_{X,Y}(x, y)$.

4.16 Take $X \sim N(0, 1)$ and $Y = X^2$. Then X and Y are not independent, and $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] = 0$.

4.17 Define $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$ where the vectors $\mathbf{v}_1 = (a_{11}, a_{12})$, $\mathbf{v}_2 = (a_{21}, a_{22})$ are of length 1 so that X_1, X_2 have variances equal to 1. A necessary and sufficient condition for $\mathbb{E}[X_1 X_2] = \rho$ is that $\mathbf{v}_1 \cdot \mathbf{v}_2 = \rho$. Now we choose $\mathbf{v}_1 = (1, 0)$ and find $\mathbf{v}_2$.

4.18 Note that

$$\begin{aligned}
f_X(x) &= \int_{-|x|}^{\infty} f_{X,Y}(x, y) dy \\
&= \int_{-|x|}^{\infty} \frac{2|x| + y}{\sqrt{2\pi}} \exp\left(-\frac{(2|x| + y)^2}{2}\right) dy \\
&= \left[-\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(2|x| + y)^2}{2}\right)\right]_{-|x|}^{\infty} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)
\end{aligned}$$

and

$$\begin{aligned}
f_Y(y) &= \int_{\{y \geq -|x|\}} f_{X,Y}(x, y) dx \\
&= \int_{\{y \geq -|x|\}} \frac{2|x| + y}{\sqrt{2\pi}} \exp\left(-\frac{(2|x| + y)^2}{2}\right) dx \\
&= \begin{cases} 2 \int_0^{\infty} \frac{2x+y}{\sqrt{2\pi}} \exp\left(-\frac{(2x+y)^2}{2}\right) dx, & y \geq 0 \\ \int_{-\infty}^y \frac{-2x+y}{\sqrt{2\pi}} \exp\left(-\frac{(-2x+y)^2}{2}\right) dx + \int_{-y}^{\infty} \frac{2x+y}{\sqrt{2\pi}} \exp\left(-\frac{(2x+y)^2}{2}\right) dx, & y < 0 \end{cases} \\
&= \begin{cases} \left[-\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(2x+y)^2}{2}\right)\right]_0^{\infty}, & y \geq 0 \\ \left[\frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{(-2x+y)^2}{2}\right)\right]_{-\infty}^y + \left[-\frac{1}{2\sqrt{2\pi}} \exp\left(-\frac{(2x+y)^2}{2}\right)\right]_{-y}^{\infty}, & y < 0 \end{cases} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right).
\end{aligned}$$

Since $\mathbb{E}[X] = \mathbb{E}[Y] = 0$, we have

$$\begin{aligned}
\text{Cov}(X, Y) &= \mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{X,Y}(x, y) dy dx \\
&= \int_{-\infty}^{\infty} \int_{-|x|}^{\infty} xy \frac{2|x| + y}{\sqrt{2\pi}} \exp\left(-\frac{(2|x| + y)^2}{2}\right) dy dx.
\end{aligned}$$

Note that in the above double integral the inside integral

$$\int_{-|x|}^{\infty} xy \frac{2|x| + y}{\sqrt{2\pi}} \exp\left(-\frac{(2|x| + y)^2}{2}\right) dy$$

is an odd function of x . Hence the double integral is zero, and $\text{Cov}(X, Y) = 0$. Finally, since $f_X(x)f_Y(y) \neq f_{X,Y}(x, y)$, X and Y are not independent.

4.19

(i) Let $f_X(x)$ denote the pdf of X . Then

$$\begin{aligned}
 \int_0^\infty x^2 f_X(x) dx &= \int_0^\infty \left(\int_0^x 2t dt \right) f_X(x) dx \\
 &= \int_0^\infty \left(\int_0^\infty 2t \mathbf{1}_{[0,x]}(t) dt \right) f_X(x) dx \\
 &= \int_0^\infty \int_0^\infty 2t \mathbf{1}_{[0,x]}(t) f_X(x) dx dt \quad (\text{by Fubini's theorem}) \\
 &= \int_0^\infty \int_0^\infty 2t \mathbf{1}_{[t,\infty)}(x) f_X(x) dx dt \\
 &= 2 \int_0^\infty t \left(\int_0^\infty \mathbf{1}_{[t,\infty)}(x) f_X(x) dx \right) dt \\
 &= 2 \int_0^\infty t \mathbb{P}(X > t) dt .
 \end{aligned}$$

The formula has a geometric interpretation: Consider the volume obtained by rotating the graph $y = 1 - F_X(x)$, $x \geq 0$, around the y -axis. The left-hand side is obtained by adding horizontal disks of radius x with thickness $|dy| = f_X(x) dx$, while the right-hand side is obtained by adding the volumes of infinitesimally thin vertical cylinders.

(ii) Note that

$$\begin{aligned}
 \sum_{n=0}^\infty n^2 \mathbb{P}(X = n) &= \sum_{n=1}^\infty \left(\sum_{k=1}^n (2k-1) \right) \mathbb{P}(X = n) \\
 &= \sum_{k=1}^\infty (2k-1) \sum_{n=k}^\infty \mathbb{P}(X = n) = \sum_{k=2}^\infty (2k-1) \mathbb{P}(X \geq k) .
 \end{aligned}$$

4.20 Since the joint pdf of X and Y is equal to $f_X(x)f_Y(y)$, we have

$$\mathbb{P}(XY \leq a) = \int_0^\infty \int_{-\infty}^{a/y} f_X(x)f_Y(y) dx dy + \int_{-\infty}^0 \int_{a/y}^\infty f_X(x)f_Y(y) dx dy .$$

(See Fig. F.3.) Now we differentiate the both sides with respect to a , and obtain

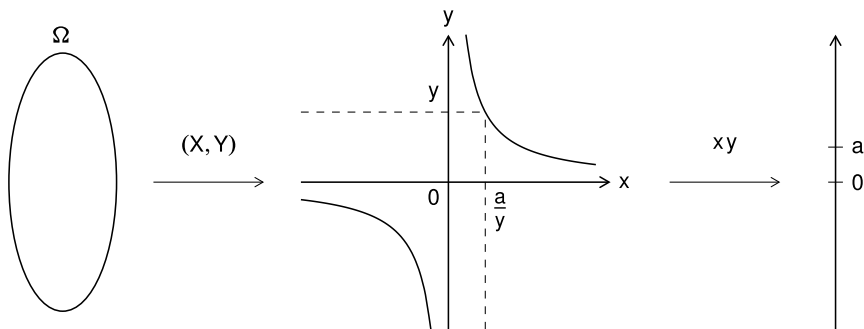


Fig. F.3 The product of two independent random variables X and Y

$$\begin{aligned}
 f_{XY}(a) &= \int_0^\infty f_X\left(\frac{a}{y}\right) \frac{1}{y} f_Y(y) dy - \int_{-\infty}^0 f_X\left(\frac{a}{y}\right) \frac{1}{y} f_Y(y) dy \\
 &= \int_0^\infty f_X\left(\frac{a}{y}\right) \frac{1}{|y|} f_Y(y) dy + \int_{-\infty}^0 f_X\left(\frac{a}{y}\right) \frac{1}{|y|} f_Y(y) dy \\
 &= \int_{-\infty}^\infty f_X\left(\frac{a}{y}\right) \frac{1}{|y|} f_Y(y) dy .
 \end{aligned}$$

4.21 By Exercise 4.20, we have

$$f_{UV}(a) = \int_0^1 \frac{1}{y} f_U\left(\frac{a}{y}\right) dy = \int_a^1 \frac{1}{y} dy = -\log a$$

for $0 \leq a \leq 1$, and $f_{UV}(a) = 0$ elsewhere.

4.22

$$\mathbb{E}[|Z|] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty |x| e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}} \int_0^\infty x e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}} .$$

Problems of Chap. 5

5.2

- (i) $\mathcal{G} = \{\emptyset, \{a, b\}, \{c, d\}, \Omega\}$.
- (ii) For a Borel subset $A \subset \mathbb{R}^1$, we have $Z^{-1}(A) = \Omega \in \mathcal{G}$ if $1 \in A$. Otherwise, $Z^{-1}(A) = \emptyset \in \mathcal{G}$. Hence Z is $\mathcal{G}$ -measurable.
- (iii) Since $Y^{-1}(\{5\})$ is $\mathcal{G}$ -measurable and contains a , $Y^{-1}(\{5\}) = \{a, b\}$ or $Y^{-1}(\{5\}) = \Omega$. In either case, $Y(b) = 5$.
- (iv) Since $\mathcal{H}$ contains $X^{-1}(\{0\}) = \{a\}$, $X^{-1}(\{3\}) = \{b, c\}$, $X^{-1}(\{1\}) = \{d\}$, we have $\mathcal{H} = \sigma(X) = \{\emptyset, \{a\}, \{b, c\}, \{d\}, \{a, b, c\}, \{a, d\}, \{b, c, d\}, \Omega\}$.

(v) Note that W is $\mathcal{H}$ -measurable since $W^{-1}(\{10\}) = \{a\} \in \mathcal{H}$, $W^{-1}(\{20\}) = \{b, c, d\} \in \mathcal{H}$. Hence $\mathbb{E}[W|X] = W$.

5.5 Define $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ by $T(x_1, x_2, x_3) = (\frac{1}{2}(x_1 + x_2), \frac{1}{2}(x_1 + x_2), x_3)$ with its

associated matrix $M = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Note that the conditional expectation operator

can be identified with T . Since M is real symmetric, it has real eigenvectors which are orthogonal. More precisely, there exist eigenvalues 0 and 1 with corresponding eigenspaces H_0 and H_1 respectively, where H_0 is spanned by $(1, -1, 0)$, and H_1 is spanned by $(1, 1, 0)$ and $(0, 0, 1)$. The conditional expectation $\mathbb{E}[\cdot | \mathcal{G}]$ can be identified with T , which is the orthogonal projection to H_1 .)

5.6 For $\delta y \approx 0$, we have

$$\begin{aligned} \mathbb{E}[X|Y] &\approx \mathbb{E}[X|y \leq Y \leq y + \delta y] \\ &= \frac{\iint_{[0,1] \times [y, y+\delta y]} X \, d\mathbb{P}}{\mathbb{P}([0,1] \times [y, y+\delta y])} \\ &= \frac{\iint_{[0,1] \times [y, y+\delta y]} x(x+y) \, dx dy}{\iint_{[0,1] \times [y, y+\delta y]} (x+y) \, dx dy} \\ &\approx \frac{\int_{[0,1]} x(x+y) \, dx \, \delta y}{\int_{[0,1]} (x+y) \, dx \, \delta y} \rightarrow \frac{\int_{[0,1]} x(x+y) \, dx}{\int_{[0,1]} (x+y) \, dx} = \frac{\frac{1}{2}y + \frac{1}{3}}{y + \frac{1}{2}}, \end{aligned}$$

where the symbol ‘ $\approx$ ’ represents approximation and ‘ $\rightarrow$ ’ denotes convergence in a suitable sense as $\delta y \rightarrow 0$. Hence $\mathbb{E}[X|Y] = \frac{3Y+2}{6Y+3}$.

5.7 Note that for B such that $\mathbb{P}(B) > 0$, we have $\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A)}$. Now use $\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A|B_k)\mathbb{P}(B_k)$.

5.8 Let A denote the event of being infected, and B represent the positive reaction by the test. The superscript c represents complement, e.g., A^c means the event of being healthy. We summarize all the given information as follows: $\mathbb{P}(A) = 0.001$, $\mathbb{P}(B|A) = 0.95$, $\mathbb{P}(B|A^c) = 0.01$. The given exercise problem is to find $\mathbb{P}(A|B)$. Since

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A) = 0.95 \times 0.001 = 0.00095$$

and

$$\mathbb{P}(A^c \cap B) = \mathbb{P}(B|A^c)\mathbb{P}(A^c) = 0.01 \times 0.999 = 0.00999,$$

we have

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{0.00095}{0.01094} \approx 8.68\%.$$

In Table [F.1](#) the necessary probabilities are printed in bold.

Table F.1 Bayes' rule and probabilities

1	$\mathbb{P}(A) = \mathbf{0.00100}$	$\mathbb{P}(A^c) = \mathbf{0.99900}$
$\mathbb{P}(B) = \mathbf{0.01094}$	$\mathbb{P}(A \cap B) = \mathbf{0.00095}$	$\mathbb{P}(A^c \cap B) = \mathbf{0.00999}$
$\mathbb{P}(B^c) = 0.98906$	$\mathbb{P}(A \cap B^c) = 0.00005$	$\mathbb{P}(A^c \cap B^c) = 0.98901$

Problems of Chap. 6

6.1 We need to show that

$$P(X_u \leq y \mid \sigma(X_s), \sigma(X_t)) = P(X_u \leq y \mid \sigma(X_t)) \quad (\text{F.3})$$

for a Markov process $\{X_t\}$. Recall that $P(A|\mathcal{G}) = \mathbb{E}[\mathbf{1}_A|\mathcal{G}]$ for a measurable subset A and a sub- σ -algebra $\mathcal{G}$. Hence (F.3) is equivalent to

$$\mathbb{E}[\mathbf{1}_{\{X_u \leq y\}} \mid \sigma(X_s), \sigma(X_t)] = \mathbb{E}[\mathbf{1}_{\{X_u \leq y\}} \mid \sigma(X_t)] . \quad (\text{F.4})$$

The left-hand side of (F.4) is equal to $\mathbb{E}[\mathbb{E}[\mathbf{1}_{\{X_u \leq y\}}|\mathcal{F}_t] \mid \sigma(X_s), \sigma(X_t)]$ by the tower property since $\sigma(X_s), \sigma(X_t) \subset \mathcal{F}_t$. Since $\mathbb{E}[\mathbf{1}_{\{X_u \leq y\}}|\mathcal{F}_t] = \mathbb{E}[\mathbf{1}_{\{X_u \leq y\}} \mid \sigma(X_t)]$ by the Markov property, the left-hand side is equal to

$$\mathbb{E}[\mathbb{E}[\mathbf{1}_{\{X_u \leq y\}}|\sigma(X_t)] \mid \sigma(X_s), \sigma(X_t)] = \mathbb{E}[\mathbf{1}_{\{X_u \leq y\}}|\sigma(X_t)] .$$

6.2 If $\omega \in \Omega$ is of the form $\omega = a_1 \dots a_n \dots$ for $a_i \in \{0, 1\}$, then $S_n(\omega) = k - (n - k)$ where k is the number of 1's among $a_1, \dots, a_n$. Note that

$$\begin{aligned} \mathbb{E}[S_{n+1}|\mathcal{F}_n](\omega) &= \frac{1}{\mathbb{P}([a_1, \dots, a_n])} \int_{[a_1, \dots, a_n]} S_{n+1} d\mathbb{P} \\ &= \frac{1}{\mathbb{P}([a_1, \dots, a_n])} \left(\int_{[a_1, \dots, a_n, 0]} S_{n+1} d\mathbb{P} + \int_{[a_1, \dots, a_n, 1]} S_{n+1} d\mathbb{P} \right) \\ &= (k - (n - k) + 1)p + (k - (n - k) - 1)q \\ &= k - (n - k) + p - q \neq S_n(\omega) , \end{aligned}$$

since $p \neq q$. More formally, $\mathbb{E}[S_{n+1}|\mathcal{F}_n] = \mathbb{E}[S_n + Z_{n+1}|\mathcal{F}_n] = \mathbb{E}[S_n|\mathcal{F}_n] + \mathbb{E}[Z_{n+1}|\mathcal{F}_n] = S_n + \mathbb{E}[Z_{n+1}|\mathcal{F}_n] = S_n + \mathbb{E}[Z_{n+1}] = S_n + p - q \neq S_n$.

6.4 Use $\mathbb{E}[\mathbb{1}_{\{|X_t| > n\}} \mid X_t] \leq \mathbb{E}[\mathbb{1}_{\{|Y| > n\}} \mid Y]$.

6.5

(i) For $n \geq 2$, we have

$$\begin{aligned}
 \Pr(N > n) &= \Pr(\min\{k : U_1 \leq U_2 \leq \cdots \leq U_{k-1} > U_k\} > n) \\
 &= \Pr(U_1 \leq U_2 \leq \cdots \leq U_n) \\
 &= \int_0^1 \int_0^{u_n} \cdots \int_0^{u_2} du_1 \cdots du_{n-1} du_n \\
 &= \int_0^1 \int_0^{u_n} \cdots \int_0^{u_3} u_2 du_2 \cdots du_{n-1} du_n \\
 &= \int_0^1 \int_0^{u_n} \cdots \int_0^{u_4} \frac{1}{2} u_3^2 du_3 \cdots du_{n-1} du_n \\
 &= \vdots \\
 &= \int_0^1 \frac{1}{(n-1)!} u_n^{n-1} du_n = \frac{1}{n!}.
 \end{aligned}$$

This is clear since $\Pr(U_{i_1} \geq U_{i_2} \geq \cdots \geq U_{i_n}) = \frac{1}{(n-2)!}$ for any permutation $(i_1, \dots, i_n)$ of $(1, \dots, n)$, and $\Pr(N > n) = \Pr(U_1 \geq U_2 \geq \cdots \geq U_n) = \frac{1}{n!}$.

(ii) Since

$$\Pr(M = n) = \Pr(M > n-1) - \Pr(M > n) = \frac{1}{(n-1)!} - \frac{1}{n!} = \frac{1}{n!}(n-1)$$

for $n \geq 2$, we have

$$\mathbb{E}[M] = \sum_{n=2}^{\infty} n \Pr(M = n) = \sum_{n=2}^{\infty} n \frac{1}{n!} (n-1) = \sum_{n=2}^{\infty} \frac{1}{(n-2)!} = \sum_{n=0}^{\infty} \frac{1}{n!} = e.$$

6.7 Hint: Since $\mathcal{G}_t \subset \mathcal{H}_t$, we have $\mathbb{E}[X_t | \mathcal{G}_s] = \mathbb{E}[\mathbb{E}[X_t | \mathcal{H}_s] | \mathcal{G}_s] = \mathbb{E}[X_s | \mathcal{G}_s] = X_s$, $s < t$.

6.8 Since $X_{n-1} = \mathbb{E}[X_n | \mathcal{F}_{n-1}] = X_n$ for every n , we have $X_n = X_{n-1} = \cdots = X_1 = X_0 = \mathbb{E}[X_1 | \mathcal{F}_0]$, and X_n is constant.

6.9 Note that $\{\tau_1 \vee \tau_2 = n\} = \{\tau_1 = n, \tau_2 \leq n\} \cup \{\tau_1 \leq n, \tau_2 = n\} \in \mathcal{F}_n$. Thus $\tau_1 \vee \tau_2$ is a stopping time.

Problems of Chap. 7

7.1 Since $f(\alpha) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} \cos(\alpha x) e^{-x^2/(2t)} dx$, we have

$$\begin{aligned} \frac{df}{d\alpha}(\alpha) &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} (-x \sin(\alpha x)) e^{-x^2/(2t)} dx \\ &= \frac{1}{\sqrt{2\pi t}} \left\{ \left[\sin(\alpha x) t e^{-x^2/(2t)} \right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \alpha \cos(\alpha x) t e^{-x^2/(2t)} dx \right\} \\ &= \frac{1}{\sqrt{2\pi t}} \left\{ 0 - t\alpha \int_{-\infty}^{\infty} \cos(\alpha x) e^{-x^2/(2t)} dx \right\} = -t\alpha f(\alpha). \end{aligned}$$

Hence $\log f(\alpha) = \int (-\alpha t) d\alpha = -\frac{1}{2}\alpha^2 t + \log f(0)$ and $f(\alpha) = f(0)e^{-\frac{1}{2}\alpha^2 t} = e^{-\frac{1}{2}\alpha^2 t}$. Thus $\mathbb{E}[\cos W_t] = f(1) = e^{-\frac{1}{2}t}$.

7.3 Let $f(t, x) = x^3 - 3tx$. Then $df(t, W_t) = (f_t + \frac{1}{2}f_{xx})dt + f_x dW_t = -3tdW_t$, and hence $f(t, W_t) = \int_0^t (3W_s^2 - 3s) dW_s$, which is a martingale.

7.6 $\mathbb{E}[X_t^2] = \mathbb{E}[W_t^2 + t^2 W_1^2 - 2t W_t W_1] = t + t^2 - 2t^2 = t - t^2$.

7.8

$$\begin{aligned} \mathbb{E}[W_t e^{\theta W_t}] &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} x e^{\theta x} e^{-\frac{x^2}{2t}} dx = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\theta t)^2 - \theta^2 t^2}{2t}} dx \\ &= e^{\frac{1}{2}\theta^2 t} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} (y + \theta t) e^{-\frac{y^2}{2t}} dy = e^{\frac{1}{2}\theta^2 t} (0 + \theta t). \end{aligned}$$

7.9

$$\begin{aligned} \mathbb{E}[W_t^2 e^{\theta W_t}] &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} x^2 e^{\theta x} e^{-\frac{x^2}{2t}} dx = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} x^2 e^{-\frac{(x-\theta t)^2 - \theta^2 t^2}{2t}} dx \\ &= e^{\frac{1}{2}\theta^2 t} \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} (y + \theta t)^2 e^{-\frac{y^2}{2t}} dy = e^{\frac{1}{2}\theta^2 t} (t + 0 + \theta^2 t^2). \end{aligned}$$

7.10 We check the conditions in Definition 7.1. All the processes start at 0 and have mean 0.

- (i) The increments are independent, and we have $\text{Var}(X_t - X_s) = \alpha^2 \text{Var}(W_{t/\alpha^2} - W_{s/\alpha^2}) = \alpha^2 \left(\frac{t}{\alpha^2} - \frac{s}{\alpha^2} \right) = t - s$.
- (ii) Take $\alpha = a^{-1/2}$ in Part (i).
- (iii) The increments are independent, and $\text{Var}(X_t - X_s) = \text{Var}((W_{T+t} - W_T) - (W_{T+s} - W_T)) = \text{Var}(W_{T+t} - W_{T+s}) = t - s$.

- (iv) For $s < t \leq T$ and $T < s < t$ we have increments $X_t - X_s$ given by $W_t - W_s$ or $-(W_t - W_s)$. For $s \leq T < t$ we have $X_t - X_s = 2W_T - W_t - W_s = (W_T - W_t) + (W_T - W_s) = -(W_t - W_T) + (W_T - W_s)$. Hence for every partition $0 \leq t_1 < t_2 \leq \dots \leq t_{2n-1} < t_{2n}$ the increments $X_{t_2} - X_{t_1}, \dots, X_{t_{2n}} - X_{t_{2n-1}}$ are independent. Furthermore,

$$\text{Var}(X_t - X_s) = \text{Var}(W_t - W_T) + \text{Var}(W_T - W_s) = (t - T) + (T - s) = t - s.$$

7.11 We will find a function f such that $\mathbb{E}[W_s | W_t] = f(W_t)$. For such a function we have

$$\int_{\{W_t \in B\}} W_s \, d\mathbb{P} = \int_{\{W_t \in B\}} \mathbb{E}[W_s | W_t] \, d\mathbb{P} = \int_{\{W_t \in B\}} f(W_t) \, d\mathbb{P}$$

for every interval B . Note that

$$\begin{aligned} \int_{\{W_t \in B\}} W_s \, d\mathbb{P} &= \int W_s \mathbf{1}_{\{W_t \in B\}} \, d\mathbb{P} \\ &= \int_{-\infty}^{\infty} \int_B x p(s; 0, x) p(t - s; x, y) \, dy \, dx \\ &= \int_B \int_{-\infty}^{\infty} x p(s; 0, x) p(t - s; x, y) \, dx \, dy \end{aligned}$$

and $\int_{\{W_t \in B\}} f(W_t) \, d\mathbb{P} = \int_B f(y) p(t; 0, y) \, dy$. Hence $f(y) = \frac{s}{t}y$ since

$$\begin{aligned} &f(y)p(t; 0, y) \\ &= \int_{-\infty}^{\infty} x p(s; 0, x) p(t - s; x, y) \, dx \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{x^2}{2s}\right) \frac{1}{\sqrt{2\pi}(t-s)} \exp\left(-\frac{(x-y)^2}{2(t-s)}\right) \, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}s(t-s)} \exp\left(-\frac{(t-s)x^2 + sx^2 + sy^2 - 2sxy}{2s(t-s)}\right) \, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}s(t-s)} \exp\left(-\frac{(\sqrt{t}x - \frac{s}{\sqrt{t}}y)^2}{2s(t-s)}\right) \exp\left(-\frac{y^2}{2t}\right) \, dx \\ &= p(t; 0, y) \int_{-\infty}^{\infty} x p\left(\frac{s(t-s)}{t}; \frac{s}{t}y, x\right) \, dx = p(t; 0, y) \frac{s}{t}y. \end{aligned}$$

7.12 Choose $u = \frac{1}{t}$ and $v = \frac{1}{s}$ for $0 < s < t$. Then $u < v$, and $\mathbb{E}[X_v | X_u] = X_u$. In other words, $\mathbb{E}[vW_{1/v} | uW_{1/u}] = uW_{1/u}$. Hence $\mathbb{E}[vW_s | uW_t] = uW_t$. Note that $\sigma(uW_t) = \sigma(W_t)$ and that $\mathbb{E}[vW_s | uW_t] = \mathbb{E}[vW_s | W_t]$.

7.13 Since $\mathbb{E}[e^{at} \cos W_t | \mathcal{F}_s] = e^{at} \mathbb{E}[\cos W_t | \mathcal{F}_s] = e^{at} e^{-(t-s)/2} \cos W_s$, we need $e^{at} e^{-(t-s)/2} = e^{as}$, which would imply $(a - \frac{1}{2})t + \frac{1}{2}s = as$, and $a = \frac{1}{2}$.

7.14 Since $Z_1 = \frac{1}{\sqrt{T}} W_T$ and $Z_2 = \frac{1}{\sqrt{T}} (W_{2T} - W_T)$ are independent standard normal variables, we have

$$\begin{aligned} \Pr(W_T > 0, W_{2T} > 0) &= \Pr(W_T > 0, W_T + (W_{2T} - W_T) > 0) \\ &= \Pr(\sqrt{T} Z_1 > 0, \sqrt{T} Z_1 + \sqrt{T} Z_2 > 0) \\ &= \Pr(Z_1 > 0, Z_1 + Z_2 > 0) \\ &= \frac{1}{2\pi} \iint_A e^{-(z_1^2 + z_2^2)/2} dz_1 dz_2 = \Pr(A), \end{aligned}$$

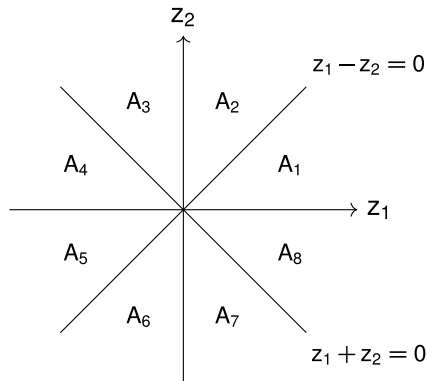
where $A = \{(z_1, z_2) : z_1 > 0, z_1 + z_2 > 0\}$. In Fig. F.4 $A_1, \dots, A_8$ all have the same probability equal to $\frac{1}{8}$ due to the rotational symmetry of the normal density function $\frac{1}{2\pi} e^{-(z_1^2 + z_2^2)/2}$. Since $A = A_1 \cup A_2 \cup A_8$, we have $\Pr(A) = \frac{3}{8}$. Similarly,

$$\begin{aligned} \Pr(W_T > 0, W_{2T} < 0) &= \Pr(W_T > 0, W_T + (W_{2T} - W_T) < 0) \\ &= \Pr(\sqrt{T} Z_1 > 0, \sqrt{T} Z_1 + \sqrt{T} Z_2 < 0) \\ &= \Pr(Z_1 > 0, Z_1 + Z_2 < 0) \\ &= \frac{1}{2\pi} \iint_B e^{-(z_1^2 + z_2^2)/2} dz_1 dz_2 = \Pr(B), \end{aligned}$$

where $B = \{(z_1, z_2) : z_1 > 0, z_1 + z_2 < 0\} = A_7$. Hence $\Pr(B) = \frac{1}{8}$. (Here is an intuitive interpretation of the result: For sufficiently large N , let $dt = T/N$. As a discrete approximation of a Brownian motion, we consider a game in which a fair coin is tossed every dt seconds. If the coin lands heads up, the player receives $\sqrt{dt}$ dollars, and if it comes up tails he/she gets $-\sqrt{dt}$ dollars. Let p_N^+ denote the probability that after N tosses the player is in the deficit and after the N more tosses (i.e., total of $2N$ tosses later) he/she will have net gain, and let p_N^- be the probability that after N tosses the player is in the deficit and after the next N more tosses he/she will have net loss. The result for Brownian motion implies that $p_N^+ \approx \frac{1}{8}$ and $p_N^- \approx \frac{3}{8}$. Once you are behind, life becomes three times tougher.)

7.15 By Exercise 4.22 we have $\mathbb{E}[|W_t|] = \mathbb{E}[\sqrt{t} W_1] = \sqrt{t} \mathbb{E}[|W_1|] = \sqrt{\frac{2t}{\pi}}$.

Fig. F.4 A partition of the plane where the subsets $A_1, \dots, A_8$ have the same probability $\frac{1}{8}$ with respect to the two-dimensional normal density due to the rotational symmetry



Problems of Chap. 8

8.1 $(k - \ell)/(k + \ell)$.

Problems of Chap. 9

9.6 Note that $\text{Cov}(X_t, X_{t+u})$ is equal to

$$\begin{aligned}
 & \text{Cov} \left(\int_0^t f(s) dW_s, \int_0^{t+u} f(s) dW_s \right) \\
 &= \mathbb{E} \left[\left(\int_0^t f(s) dW_s - 0 \right) \left(\int_0^{t+u} f(s) dW_s - 0 \right) \right] \\
 &= \mathbb{E} \left[\int_0^t f(s) dW_s \left(\int_0^t f(s) dW_s + \int_t^{t+u} f(s) dW_s \right) \right] \\
 &= \mathbb{E} \left[\int_0^t f(s) dW_s \int_0^t f(s) dW_s \right] + \mathbb{E} \left[\int_0^t f(s) dW_s \int_t^{t+u} f(s) dW_s \right] \\
 &= \int_0^t \mathbb{E} [f(s)^2] ds + \mathbb{E} \left[\int_0^t f(s) dW_s \right] \mathbb{E} \left[\int_t^{t+u} f(s) dW_s \right],
 \end{aligned}$$

which is equal to $\int_0^t f(s)^2 ds + 0 \times 0$, where we used the fact that $\int_0^t f(s) dW_s$ and $\int_t^{t+u} f(s) dW_s$ are independent in the last equality.

9.7 By the result of Exercise 4.10 we have $\mathbb{E} [X_t^2] = \mathbb{E} [e^{2W_t^2}] = \mathbb{E} [e^{2tW_1^2}] = \frac{1}{\sqrt{1-4t}}$, $0 \leq t < \frac{1}{4}$. Hence the identity in Theorem 9.1 does not hold.

9.8 By the Martingale Representation Theorem, $f(t, W_t) = f(0, W_0) + \int_0^t \alpha_s dW_s$. Now use the Itô formula.

9.9 The function $f(t, x)$ satisfies $f_t + \frac{1}{2}f_{xx} = 0$ and $f_x = f$. From the second equation, we see that $f(t, x) = e^x g(t)$ for some g . From the first equation we have $g' + \frac{1}{2}g = 0$, and hence $g(t) = g(0)e^{-\frac{1}{2}t}$. Thus $X_t = X_0 e^{W_t - \frac{1}{2}t}$.

9.11

- (i) $\int_0^t f(s, W_s) dW_s$ is equal to X_t .
- (ii) To prove that X_t is a martingale, either use the definition of a martingale or apply the Martingale Representation Theorem since X_t is an Itô integral.

9.13 Suppose that $X_t = e^{W_t + g(t)}$ is a martingale.

Solution 1. Since

$$dX_t = \left\{ g'(t)e^{W_t + g(t)} + \frac{1}{2}e^{W_t + g(t)} \right\} dt + e^{W_t + g(t)} dW_t,$$

we have $g'(t)e^{W_t + g(t)} + \frac{1}{2}e^{W_t + g(t)} = 0$ by the Martingale Representation Theorem. Hence $g(t) = -\frac{1}{2}t + C$ for some C .

Solution 2. Since $\mathbb{E}[X_t] = e^{\frac{1}{2}t + g(t)}$, $t \geq 0$, is constant, we have $g(t) = -\frac{1}{2}t + C$.

9.14 Let $f(t, x) = (x + \theta t)e^{-\theta x - \frac{1}{2}\theta^2 t}$. Then $f_t + f_{xx} = 0$. Hence $df(t, W_t)$ has no dt -term. Hence $f(t, W_t) = (W_t + \theta t)e^{-\theta W_t - \frac{1}{2}\theta^2 t}$ is a martingale.

9.16 $dX_t = (2tW_t + 0)dt + t^2dW_t - 2tW_tdt = t^2dW_t$.

9.18 Since $W_t^6 = W_t^6 - W_0^6 = \int_0^t (15W_s^4)ds + \int_0^t 6W_s^5dW_s$, we have $\mathbb{E}[W_t^6] = \mathbb{E}[\int_0^t (15W_s^4)ds] + \mathbb{E}[\int_0^t 6W_s^5dW_s] = 15 \int_0^t \mathbb{E}[W_s^4]ds + 0 = 15 \int_0^t 3s^2ds = 15t^3$.

Problems of Chap. 10

10.1 If δ denotes the continuous dividend yield, then the modified model for the stock price movement is given by $dS_t = (\mu - \delta)S_tdt + \sigma S_tdW_t$.

10.2

- (i) Let $f(t, x) = e^x$. Then $e^{X_t} = f(t, X_t)$. Hence

$$d(e^{X_t}) = e^{X_t}dX_t + \frac{1}{2}e^{X_t}(dX_t)^2 = e^{X_t}(a_t + \frac{1}{2}b_t^2)dt + e^{X_t}b_tdW_t.$$

- (ii) Take $X_t = -\int_0^t r_sds$. Then $dX_t = -r_tdt$, $a_t = -r_t$, $b_t = 0$. Since $Z_t = e^{X_t}$, we have $dZ_t = Z_t(-r_t)dt$.

10.3 Since

$$\begin{aligned} df(t, W_t) &= \left(\frac{1}{2} e^{\frac{1}{2}t} \cos W_t - \frac{1}{2} e^{\frac{1}{2}t} \cos W_t \right) dt - e^{\frac{1}{2}t} \sin W_t dW_t \\ &= -e^{\frac{1}{2}t} \sin W_t dW_t, \end{aligned}$$

$f(t, W_t) - f(0, W_0) = \int_0^t (-e^{\frac{1}{2}s} \sin W_s) dW_s$. Since $f(0, W_0) = 1$, we conclude that

$$\mathbb{E}[e^{\frac{1}{2}t} \cos W_t] = \mathbb{E}[f(t, W_t)] = 1 + \mathbb{E} \left[\int_0^t (-e^{\frac{1}{2}s} \sin W_s) dW_s \right] = 1,$$

where we used Corollary 9.2 for the last equality.

10.4 $X_t = (1 + W_t)^2$.

10.5 For $t = 0$, we have $\int_0^0 W_s dW_s = f(0, W_0)$. Hence $f(0, 0) = 0$. By Itô's lemma we have

$$\begin{cases} f_t + \frac{1}{2} f_{xx} = 0, \\ f_x = x. \end{cases}$$

From the second equation we have $f(t, x) = \frac{1}{2}x^2 + \phi(t)$. Combining the result with the first equation, we obtain $\phi(t) = -\frac{1}{2} + C$ for some C . Thus $f(t, x) = \frac{1}{2}x^2 - \frac{1}{2}t + C$. Since $f(0, 0) = 0$, we have $\int_0^t W_s dW_s = f(t, W_t) = \frac{1}{2}W_t^2 - \frac{1}{2}t$.

10.6

- (i) Suppose that there exists $f(t, x)$ such that $\int_0^t W_s^2 dW_s = f(t, W_t)$. This assumption is wrong and here is why. Suppose that the assumption were right. Taking $t = 0$, we find $0 = f(0, 0)$. By Itô's lemma, we have $W_t dW_t = (f_t + \frac{1}{2} f_{xx}) dt + f_x dW_t$. Hence

$$\begin{cases} f_t + \frac{1}{2} f_{xx} = 0, \\ f_x = x^2. \end{cases}$$

Hence $f = \frac{1}{3}x^3 + \phi(t)$ for some ϕ . Thus $\phi'(t) + x = 0$, which is impossible.

- (ii) Now let us modify the assumption! Assume that there exists $f(t, x)$ and $g(t, x)$ such that $\int_0^t W_s^2 dW_s = f(t, W_t) + \int_0^t g(s, W_s) ds$. Then

$$\begin{cases} f_t + \frac{1}{2} f_{xx} + g = 0, \\ f_x = x^2. \end{cases}$$

From the second equation we have $f(t, x) = \frac{1}{3}x^3 + \phi(t)$. Combining the result with the first equation, we obtain $\phi'(t) + x + g(t, x) = 0$, and hence

$$\int_0^t W_s^2 dW_s = \frac{1}{3}W_t^3 + \phi(t) + \int_0^t (-\phi'(s) - W_s) ds = \frac{1}{3}W_t^3 - \int_0^t W_s ds.$$

10.7 Assume that $\int_0^t s \, dW_s = f(t, W_t) + \int_0^t g(s, W_s) \, ds$ for some f and g . By Itô's lemma, we have

$$\begin{cases} f_t + \frac{1}{2}f_{xx} + g = 0, \\ f_x = t. \end{cases}$$

Hence $f(t, x) = tx + \phi(t)$ for some $\phi(t)$. Note that $0 = f(0, 0) = \phi(0)$. Since $x + \phi'(t) + g = 0$, we have $g(t, x) = -\phi'(t) - x$. Thus

$$\int_0^t s \, dW_s = tW_t + \phi(t) + \int_0^t (-\phi'(s) - W_s) \, ds = tW_t - \int_0^t W_s \, ds.$$

10.9 Assume that there exist f and g such that $\int_0^t e^{W_s} \, dW_s = f(t, W_t) + \int_0^t g(s, W_s) \, ds$. By Itô's lemma, we have $e^{W_t} \, dW_t = (f_t + \frac{1}{2}f_{xx}) \, dt + f_x \, dW_t + g \, dt$. Hence

$$\begin{cases} f_t + \frac{1}{2}f_{xx} + g = 0, \\ f_x = e^x. \end{cases}$$

Hence $f(t, x) = e^x + \phi(t)$ for some $\phi(t)$. Note that $0 = f(0, 0) = 1 + \phi(0)$. Since $\phi'(t) + \frac{1}{2}e^x + g = 0$, we have $g(t, x) = -\phi'(t) - \frac{1}{2}e^x$. Thus

$$\int_0^t e^{W_s} \, dW_s = e^{W_t} + \phi(t) + \int_0^t (-\phi'(s) - \frac{1}{2}e^{W_s}) \, ds = e^{W_t} - 1 - \frac{1}{2} \int_0^t e^{W_s} \, ds.$$

10.10

(i) Let $f(t, x) = x^k$, $k \geq 2$. Then

$$W_t^k = \int_0^t \frac{1}{2}k(k-1)W_s^{k-2} \, ds + \int_0^t kW_s^{k-1} \, dW_s.$$

Hence

$$a_k(t) = \mathbb{E}[W_t^k] = \int_0^t \frac{1}{2}k(k-1)\mathbb{E}[W_s^{k-2}] \, ds = \int_0^t \frac{1}{2}k(k-1)a_{k-2}(s) \, ds.$$

(ii) Since the pdf $f(x)$ of W_t is an even function, we have $\mathbb{E}[W_t^{2k+1}] = \int_{-\infty}^{\infty} x^{2k+1} f(x) \, dx = 0$. Or, we may use induction on the recursive relation $a_{2k+1}(t) = \frac{1}{2}k(k-1) \int_0^t a_{2k-1}(s) \, ds$ together with $a_1 = 0$. Similarly, we have by induction,

$$a_{2k}(t) = \frac{1}{2}(2k)(2k-1) \int_0^t \frac{(2k-2)!s^{k-1}}{2^{k-1}(k-1)!} \, ds = \frac{(2k)!t^k}{2^k k!}.$$

10.12 First, by taking expectations on both sides, we obtain $\alpha(t) = \alpha(0) + \int_0^t (a(s)\alpha(s) + b(s))ds$, and hence $\alpha'(t) = a(t)\alpha(t) + b(t)$. For $\beta(t)$, note that

$$\begin{aligned} X_t^2 &= X_0^2 + \int_0^t [2X_s\{a(s)X_s + b(s)\} + \{c(s)X_s + d(s)\}^2] ds \\ &\quad + \int_0^t [2X_s\{c(s)X_s + d(s)\}] dW_s \end{aligned}$$

and take expectations on both sides, then take derivatives.

Problems of Chap. 11

11.2

- (i) Since $\mathbb{E}^{\mathbb{Q}}[W_t] = \mathbb{E}^{\mathbb{P}}[W_t e^{-\frac{1}{2}\theta^2 t - \theta W_t}] = e^{-\frac{1}{2}\theta^2 t} \mathbb{E}^{\mathbb{P}}[W_t e^{-\theta W_t}]$, we have that $\mathbb{E}^{\mathbb{Q}}[W_t] = e^{-\frac{1}{2}\theta^2 t} (-\theta t) e^{\frac{1}{2}\theta t} = -\theta t$.
- (ii) Since $\mathbb{E}^{\mathbb{Q}}[W_t^2] = \mathbb{E}^{\mathbb{P}}[W_t^2 e^{-\frac{1}{2}\theta^2 t - \theta W_t}] = e^{-\frac{1}{2}\theta^2 t} \mathbb{E}^{\mathbb{P}}[W_t^2 e^{-\theta W_t}]$, we have that $\mathbb{E}^{\mathbb{Q}}[W_t^2] = e^{-\frac{1}{2}\theta^2 t} (\theta^2 t^2 + t) e^{\frac{1}{2}\theta t} = \theta^2 t^2 + t$.

11.3 By the tower property, $\mathbb{E}^{\mathbb{Q}}[e^{\alpha X_t}] = \mathbb{E}^{\mathbb{P}}[e^{\alpha X_t} L] = \mathbb{E}^{\mathbb{P}}[\mathbb{E}^{\mathbb{P}}[e^{\alpha X_t} L | \mathcal{F}_t]] = \mathbb{E}^{\mathbb{P}}[e^{\alpha X_t} \mathbb{E}^{\mathbb{P}}[L | \mathcal{F}_t]] = \mathbb{E}^{\mathbb{P}}[e^{\alpha X_t} L_t] = \mathbb{E}[e^{\alpha(W_t + \theta t)} e^{-\frac{1}{2}\theta^2 t - \theta W_t}]$, which is equal to $e^{a\theta t - \frac{1}{2}\theta^2 t} \mathbb{E}[e^{(\alpha - \theta)W_t}] = e^{a\theta t - \frac{1}{2}\theta^2 t} e^{\frac{1}{2}(\alpha - \theta)^2 t} = e^{\frac{1}{2}\alpha^2 t} = \mathbb{E}[e^{\alpha W_t}]$.

11.4 Since $W_t - W_s$ and $\mathcal{F}_s$ are independent, we have

$$\begin{aligned} \mathbb{E}[(W_t - W_s) e^{\theta(W_t - W_s)} | \mathcal{F}_s] &= \mathbb{E}[(W_t - W_s) e^{\theta(W_t - W_s)}] \\ &= \mathbb{E}[W_{t-s} e^{\theta W_{t-s}}] = \theta(t-s) e^{\frac{1}{2}\theta^2(t-s)}. \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E}[W_t e^{\theta W_t} | \mathcal{F}_s] - W_s \mathbb{E}[e^{\theta W_t} | \mathcal{F}_s] &= \mathbb{E}[(W_t - W_s) e^{\theta W_t} | \mathcal{F}_s] \\ &= \theta(t-s) e^{\frac{1}{2}\theta^2(t-s)} e^{\theta W_s}. \end{aligned}$$

11.5 Since $W_t - W_s$ and $\mathcal{F}_s$ are independent, we have

$$\begin{aligned} \mathbb{E}[(W_t - W_s)^2 e^{\theta(W_t - W_s)} | \mathcal{F}_s] &= \mathbb{E}[(W_t - W_s)^2 e^{\theta(W_t - W_s)}] \\ &= \mathbb{E}[W_{t-s}^2 e^{\theta W_{t-s}}] \\ &= ((t-s) + \theta^2(t-s)^2) e^{\frac{1}{2}\theta^2(t-s)}. \end{aligned}$$

On the other hand,

$$\begin{aligned} & e^{\theta W_s} \mathbb{E}[(W_t - W_s)^2 e^{\theta(W_t - W_s)} | \mathcal{F}_s] \\ &= \mathbb{E}[(W_t - W_s)^2 e^{\theta W_t} | \mathcal{F}_s] \\ &= \mathbb{E}[W_t^2 e^{\theta W_t} | \mathcal{F}_s] - 2W_s \mathbb{E}[W_t e^{\theta W_t} | \mathcal{F}_s] + W_s^2 \mathbb{E}[e^{\theta W_t} | \mathcal{F}_s]. \end{aligned}$$

Hence

$$\begin{aligned} & \mathbb{E}[W_t^2 e^{\theta W_t} | \mathcal{F}_s] \\ &= e^{\theta W_s} \mathbb{E}[(W_t - W_s)^2 e^{\theta(W_t - W_s)} | \mathcal{F}_s] + 2W_s \mathbb{E}[W_t e^{\theta W_t} | \mathcal{F}_s] - W_s^2 \mathbb{E}[e^{\theta W_t} | \mathcal{F}_s] \\ &= \{(t - s) + \theta^2(t - s)^2 + W_s^2 + 2\theta(t - s)W_s\} e^{\frac{1}{2}\theta^2(t-s)} e^{\theta W_s}. \end{aligned}$$

11.6 Take $s < t$. From Exercise 11.4,

$$\mathbb{E}[W_t L_t | \mathcal{F}_s] = e^{-\frac{1}{2}\theta^2 t} \mathbb{E}[W_t e^{-\theta W_t} | \mathcal{F}_s] = (W_s - \theta(t - s)) L_s.$$

From Exercise 11.5,

$$\mathbb{E}[W_t^2 L_t | \mathcal{F}_s] = e^{-\frac{1}{2}\theta^2 t} \mathbb{E}[W_t^2 e^{-\theta W_t} | \mathcal{F}_s] = \{(t - s) + (W_s - \theta(t - s))^2\} L_s.$$

Hence we have

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}}[X_t^2 - t | \mathcal{F}_s] &= \mathbb{E}[X_t^2 L_t L_s^{-1} | \mathcal{F}_s] - t \quad (\text{by Lemma 11.1}) \\ &= L_s^{-1} \mathbb{E}[(W_t + \theta t)^2 L_t | \mathcal{F}_s] - t \\ &= L_s^{-1} (\mathbb{E}[W_t^2 L_t | \mathcal{F}_s] + 2\theta t \mathbb{E}[W_t L_t | \mathcal{F}_s] + \theta^2 t^2 \mathbb{E}[L_t | \mathcal{F}_s]) - t \\ &= \{(t - s) + (W_s - \theta(t - s))^2\} + 2\theta t \{W_s - \theta(t - s)\} + \theta^2 t^2 - t \\ &= (W_s + \theta s)^2 - s = X_s^2 - s. \end{aligned}$$

11.7 Let $A = \{\lim_{t \rightarrow \infty} \frac{W_t}{t} = 0\}$. Then $\mathbb{P}(A) = 1$ by the Law of Large Numbers (Corollary 7.2). Note that $\lim_{t \rightarrow \infty} \frac{X_t}{t} = \lim_{t \rightarrow \infty} \frac{W_t + \theta t}{t} = \theta \neq 0$ on A . Hence $\mathbb{Q}(A) = 0$ since $\mathbb{Q}(\{\lim_{t \rightarrow \infty} \frac{X_t}{t} = 0\}) = 1$. Thus $\mathbb{P}$ and $\mathbb{Q}$ are not equivalent.

11.8 To have a fair game the expected reward must be equal to \$0, and hence $pA - (1 - p)B = 0$, which implies that $B = \frac{p}{1-p}A > A$. In other words, if we interpret the event of having heads and tails as upward and downward movements respectively, we need to give more weight to the downward movement than the upward movement to compensate the bias.

Problems of Chap. 12

12.2 For $a \geq 0$, let $X_t = 0$, $0 \leq t \leq a$, and $X_t = (W_t - a)^3$, $t > a$.

12.6 Take $\alpha_t = \gamma_t = 0$. Then $dX_t = \beta_t X_t dt + \delta_t X_t dW_t$. Let Y_t be an Itô process defined by $dY_t = \beta_t dt + \delta_t dW_t$. Then $d[Y, Y]_t = \delta_t^2 dt$ and $dX_t = X_t dY_t$, and hence

$$\begin{aligned} X_t &= X_0 \exp \left(Y_t - Y_0 - \frac{1}{2} [Y, Y]_t \right) \\ &= X_0 \exp \left(\int_0^t \beta_s ds + \int_0^t \delta_s dW_s - \frac{1}{2} \int_0^t \delta_s^2 ds \right) \\ &= X_0 \exp \left(\int_0^t (\beta_s ds - \frac{1}{2} \delta_s^2 ds) + \int_0^t \delta_s dW_s \right). \end{aligned}$$

For arbitrary α_t, γ_t , we look for a solution of the form $X_t = G_t H_t$, where

$$\begin{aligned} dG_t &= \beta_t G_t dt + \delta_t G_t dW_t, \quad G_0 = 1, \\ dH_t &= \alpha_t dt + b_t dW_t, \quad H_0 = X_0. \end{aligned}$$

12.7 Assume that $X_t = f(t, W_t)$ for some $f(t, x)$. Then

$$(f_t + \frac{1}{2} f_{xx}) dt + f_x dW_t = f^3 dt - f^2 dW_t.$$

Hence

$$f_t + \frac{1}{2} f_{xx} = f^3 \quad \text{and} \quad f_x = -f^2.$$

From the second equation, we have $-\frac{df}{f^2} = dx$, $\frac{1}{f} = x + g(t)$ and $f(t, x) = \frac{1}{x+g(t)}$. Substituting the result into the first equation, we obtain $g(t) = C$ for some constant C . Since $X_0 = 1$, we have $C = 1$. Thus $X_t = 1/(W_t + 1)$.

12.8 Note that

$$\begin{aligned} dY_t &= 2X_t dX_t + \frac{1}{2} \times 2(dX_t)^2 \\ &= 2X_t(-\alpha X_t dt + \sigma dW_t) + (-\alpha X_t dt + \sigma dW_t)^2 \\ &= (\sigma^2 - 2\alpha X_t^2) dt + 2\sigma X_t dW_t. \end{aligned}$$

Thus, by substituting (12.3), we obtain

$$dY_t = (\sigma^2 - 2\alpha Y_t) dt + 2\sigma \left(e^{-\alpha t} x_0 + e^{-\alpha t} \sigma \int_0^t e^{\alpha s} dW_s \right) dW_t.$$

12.9 Let $X_t = f(t, W_t)$ for some $f(t, x)$. Then $\frac{1}{2}f_{xx} + f_t = \frac{3}{4}f^2$ and $f_x = -f^{3/2}$. From the second equation, $f(t, x) = 4/(C + x)^2$, and $X_t = 4/(C + W_t)^2$ for some C . Since $X_0 = 4/(C + 0)^2$, $C = \pm 2/\sqrt{X_0}$. Thus $X_t = 4/(\pm 2/\sqrt{X_0} + W_t)^2$. Note that both solutions have the same distribution.

Problems of Chap. 14

14.1 Hint: Consult Example 14.1.

14.2 It suffices to prove that the American call price C_A is equal to the European call price C_E in the single period model given in Fig. F.5. Since $C_A = \max\{C_E, \max\{S_0 - K, 0\}\}$ by the possibility of early exercise, it is enough to prove that $C_E \geq \max\{S_0 - K, 0\}$. Note that

$$C_E = e^{-rT}(q \max\{S^u - K, 0\} + (1 - q) \max\{S^d - K, 0\})$$

for some $0 < q < 1$ by (14.1) and (14.2).

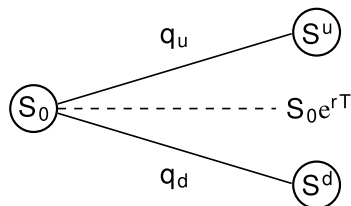
Let $f(x) = (x - K)^+$, $x \geq 0$. Then $f(x)$ is convex, and satisfies Jensen's inequality $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$ for every $0 \leq \alpha \leq 1$ and $x, y \geq 0$. Since $f(0) = 0$, if $y = 0$ in particular, then

$$f(\alpha x) \leq \alpha f(x), \quad 0 \leq \alpha \leq 1 \text{ and } x \geq 0. \quad (\text{F.5})$$

See Fig. F.6, where it is shown that $f(\frac{1}{2}x) \leq f(\frac{1}{2}x)$. Note that $0 \leq e^{-rT} \leq 1$ and $q = (e^{rT}S_0 - S^d)/(S^u - S^d)$. Then $S_0 = qe^{-rT}S^u + (1 - q)e^{-rT}S^d$, i.e., the discounted asset price process $e^{-r \times 0}S_0, e^{-rT}S_T$ is a $(q, 1 - q)$ -martingale. Hence

$$\begin{aligned} C_E &= e^{-rT}(qf(S^u) + (1 - q)f(S^d)) \\ &\geq qf(e^{-rT}S^u) + (1 - q)f(e^{-rT}S^d) \quad (\text{by (G.5)}) \\ &\geq f(qe^{-rT}S^u + (1 - q)e^{-rT}S^d) \quad (\text{by Jensen's inequality}) \\ &= f(S_0). \end{aligned}$$

Fig. F.5 A single-period binomial tree and the risk-neutral probability



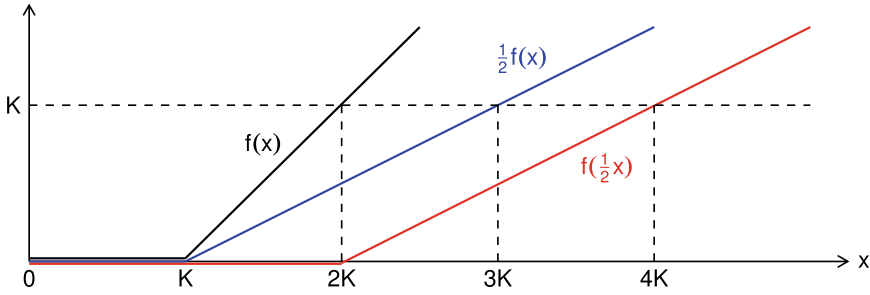


Fig. F.6 The sublinearity of the payoff function $f(x) = (x - K)^+$ of a European call

14.4 Recall that Θ is the instantaneous rate of change with respect to time while the asset price is fixed. Since $S_{i+2}^{j+1} = u S_{i+1}^j = u d S_i^j = S_i^j$, i.e., the asset price does not change, we see that the change $V_{i+2}^{j+1} - V_i^j$ is due to time only.

Problems of Chap. 15

15.3 (i) 0 (ii) 0 or $K e^{-r(T-t)} - S_t$.

15.5 Observe first

$$\begin{aligned} \frac{\partial d_1}{\partial K} &= \frac{K}{S_0} \left(-\frac{S_0}{K^2} \right) \frac{1}{\sigma \sqrt{T}} = -\frac{1}{K \sigma \sqrt{T}}, \\ \frac{\partial d_2}{\partial K} &= \frac{\partial}{\partial K} (d_1 - \sigma \sqrt{T}) = \frac{\partial d_1}{\partial K}. \end{aligned}$$

Hence

$$\begin{aligned} \frac{\partial C}{\partial K} &= S_0 N'(d_1) \frac{\partial d_1}{\partial K} - e^{-rT} N(d_2) - K e^{-rT} N'(d_2) \frac{\partial d_2}{\partial K} \\ &= -S_0 \frac{1}{\sqrt{2\pi}} e^{-d_1^2/2} \frac{1}{K \sigma \sqrt{T}} - e^{-rT} N(d_2) + K e^{-rT} \frac{1}{\sqrt{2\pi}} e^{-d_2^2/2} \frac{1}{K \sigma \sqrt{T}} \\ &= -S_0 \frac{1}{K \sigma \sqrt{2\pi T}} e^{-d_1^2/2} - e^{-rT} N(d_2) + \frac{1}{\sigma \sqrt{2\pi T}} e^{-d_2^2/2 - rT}. \end{aligned}$$

Now we note that

$$\frac{d_2^2}{2} + rT = \frac{(d_1 - \sigma \sqrt{T})^2}{2} + rT = \frac{d_1^2}{2} - \sigma d_1 \sqrt{T} + \frac{\sigma^2 T}{2} + rT = \frac{d_1^2}{2} + \log \frac{S_0}{K}.$$

Therefore

$$\begin{aligned}
\frac{\partial C}{\partial K} &= -S_0 \frac{1}{\sigma K \sqrt{2\pi T}} e^{-d_1^2/2} - e^{-rT} N(d_2) + \frac{1}{\sigma \sqrt{2\pi T}} e^{-d_1^2/2 + \log(S_0/K)} \\
&= -\frac{S_0}{K} \frac{1}{\sigma \sqrt{2\pi T}} e^{-d_1^2/2} - e^{-rT} N(d_2) + \frac{S_0}{K} \frac{1}{\sigma \sqrt{2\pi T}} e^{-d_1^2/2} \\
&= -e^{-rT} N(d_2) .
\end{aligned}$$

15.8

(i) Since $\Delta = \frac{\partial V}{\partial S} \approx \frac{\delta V}{\delta S}$, we have

$$\Omega \approx \frac{\frac{\delta V}{V}}{\frac{\delta S}{S}} = \frac{\delta V}{\delta S} \frac{S}{V} \approx \Delta \frac{S}{V} .$$

- (ii) Since $V \leq SN(d_1)$ and $\Delta = N(d_1)$, we have $\Omega \geq 1$.
(iii) Note that a European put option loses its value if the asset price increases, i.e., $\delta V \leq 0$ if $\delta S > 0$. Hence $\Omega \leq 0$. Or, we may use the fact that the deltas of the call and put options with the same strike price, denoted by Δ_C and Δ_P respectively, satisfy the relation $\Delta_P = \Delta_C - 1 \leq 0$.

Problems of Chap. 16

16.1 The price V_0 at $t = 0$ is given by the expectation with respect to a martingale measure $\mathbb{Q}$, and $V_0 = e^{-rT} \mathbb{E}^{\mathbb{Q}}[1] = e^{-rT}$, which is nothing but the risk-free bond price. If μ is replaced by r in the expectation with respect to a physical measure $\mathbb{P}$, we would have the same answer since $e^{-rT} \mathbb{E}^{\mathbb{P}}[1] = e^{-rT}$.

Problems of Chap. 17

17.1 Using $C_T = S_T - F$, we find F such that $\mathbb{E}^{\mathbb{Q}}[e^{-rT}(S_T - F)|\mathcal{F}_0] = 0$. Hence

$$F = \mathbb{E}^{\mathbb{Q}}[S_T] = \mathbb{E}^{\mathbb{Q}}[Ae^{(r - \frac{1}{2}\sigma^2)T + \sigma X_T}] = Ae^{rT} \mathbb{E}^{\mathbb{Q}}[e^{-\frac{1}{2}\sigma^2 T + \sigma X_T}] = Ae^{rT} ,$$

where by Theorem 6.1 we have $\mathbb{E}^{\mathbb{Q}}[e^{-\frac{1}{2}\sigma^2 T + \sigma X_T}] = \mathbb{E}^{\mathbb{Q}}[e^{-\frac{1}{2}\sigma^2 0 + \sigma X_0}] = 1$ since $e^{-\frac{1}{2}\sigma^2 t + \sigma X_t}$ is a $\mathbb{Q}$ -martingale.

17.2 Since an asset-or-nothing call can be decomposed into the sum of a European call and a cash-or-nothing call as can be seen in the identity

$$\underbrace{S_T \mathbb{1}_{\{S_T > K\}}}_{\text{asset-or-nothing}} = \underbrace{(S_T - K)^+}_{\text{European call}} + \underbrace{K \mathbb{1}_{\{S_T > K\}}}_{\text{cash-or-nothing}} ,$$

the asset-or-nothing call price is equal to the European call price plus the cash-or-nothing call price, which is equal to $S_0 N(d_1)$.

Problems of Chap. 19

19.3 (i) Since the amount that the holder of the option can receive from the early exercise is equal to \$1, we have $V(K, t) = 1, 0 \leq t \leq T$. (iv) Consult (Jiang, 2005).

19.4 (i) Since we can exercise the American option at any time, it is better to exercise it early and receive risk-free interest. Hence the option price satisfies $V(S, t) = 1$.

(ii) Let C and P respectively denote the prices of the American binary put and call options with the same exercise price K and the same expiry date T . Since $V = C + P$ pays \$1 at T , Part (i) implies that $C(S, t) + P(S, t) = 1$.

Problems of Chap. 22

22.1 We present four different proofs.

Proof 1. Define $f(x) = (1 + \frac{r}{x})^x, 0 < x < +\infty$. To show that $f(x)$ is monotonically increasing, we will prove that $g(x) = \log f(x)$ is monotonically increasing. Since

$$g'(x) = \log(1 + \frac{r}{x}) + x \frac{-\frac{r}{x^2}}{1 + \frac{r}{x}} = \log(1 + \frac{r}{x}) - \frac{r}{x+r},$$

observe that

$$g''(x) = \frac{1}{x+r} - \frac{1}{x} - \frac{r}{(x+r)^2} = -\frac{r^2}{x(x+r)^2} < 0$$

and $\lim_{x \rightarrow +\infty} g'(x) = 0$. Hence $g'(x) > 0$, and $g(x)$ is monotonically increasing. For the second case we consider $f(x) = (1 - \frac{r}{x})^x, r < x < +\infty$.

Proof 2. For $0 \leq a < b$, we have

$$(n+1)a^n < \frac{b^{n+1} - a^{n+1}}{b-a} = b^n + b^{n-1}a + \cdots + a^n < (n+1)b^n.$$

Hence

$$b^n \{b - (n+1)(b-a)\} < a^{n+1} \quad (\text{F.6})$$

and

$$a^n \{a + (n+1)(b-a)\} < b^{n+1}. \quad (\text{F.7})$$

Take $b = 1 + \frac{r}{n}$ and $a = 1 + \frac{r}{n+1}$ for $n \geq 1$ in (F.6). Then $b - a = \frac{r}{n(n+1)}$ and

$$\left(1 + \frac{r}{n}\right)^n = \left(1 + \frac{r}{n}\right)^n \left\{ \left(1 + \frac{r}{n}\right) - (n+1) \frac{r}{n(n+1)} \right\} < \left(1 + \frac{r}{n+1}\right)^{n+1}.$$

For the second case, let $b = 1 - \frac{r}{n+1}$ and $a = 1 - \frac{r}{n} > 0$ for $n > r$ in (F.7). Then $b - a = \frac{r}{n(n+1)}$ and

$$\left(1 - \frac{r}{n}\right)^n = \left(1 - \frac{r}{n}\right)^n \left\{ \left(1 - \frac{r}{n}\right) + (n+1) \frac{r}{n(n+1)} \right\} < \left(1 - \frac{r}{n+1}\right)^{n+1}.$$

Proof 3. Note that $\frac{x_1 + \dots + x_{n+1}}{n+1} \geq (x_1 \times \dots \times x_{n+1})^{\frac{1}{n+1}}$ where $x_i \geq 0$ for every $1 \leq i \leq n+1$. Take $x_1 = 1$ and $x_2 = \dots = x_{n+1} = 1 + \frac{r}{n}$. (We assume that n is sufficiently large so that $1 + \frac{r}{n} > 0$. Since $x_1 \neq x_2 = \dots = x_{n+1}$, the arithmetic average is strictly greater than the geometric average.) Then

$$\frac{1 + n(1 + \frac{r}{n})}{n+1} > \left(1 + \frac{r}{n}\right)^{\frac{n}{n+1}},$$

or, equivalently $\left(1 + \frac{r}{n+1}\right)^{n+1} > \left(1 + \frac{r}{n}\right)^n$.

Proof 4. We will use the Bernoulli inequality, that is, $(1+x)^n \geq 1+nx$ for every integer $n \geq 0$ and every real number $x \geq -1$. The strict version of the inequality states that $(1+x)^n > 1+nx$ for every integer $n \geq 2$ and every real number $x > -1$ with $x \neq 0$. (For the proof, use the mathematical induction on n as follows: Prove the inequality for $n = 0$ and $n = 1$. Then show that if the inequality holds for $n = k$ then it holds for $n = k+2$.) Note that

$$\begin{aligned} \frac{\left(1 + \frac{r}{n+1}\right)^{n+1}}{\left(1 + \frac{r}{n}\right)^n} &= \left(\frac{n(n+1+r)}{(n+1)(n+r)}\right)^n \left(1 + \frac{r}{n+1}\right) \\ &= \left(1 - \frac{r}{(n+1)(n+r)}\right)^n \left(1 + \frac{r}{n+1}\right) \\ &\geq \left(1 - \frac{nr}{(n+1)(n+r)}\right) \left(1 + \frac{r}{n+1}\right) \\ &= 1 + \frac{r^2}{(n+1)^2(n+r)} > 1. \end{aligned}$$

For the second case, note that for $n > r$ we have

$$\begin{aligned}
\frac{(1 - \frac{r}{n+1})^{n+1}}{(1 - \frac{r}{n})^n} &= \left(\frac{n(n+1-r)}{(n+1)(n-r)} \right)^n \left(1 - \frac{r}{n+1} \right) \\
&= \left(1 + \frac{r}{(n+1)(n-r)} \right)^n \left(1 - \frac{r}{n+1} \right) \\
&\geq \left(1 + \frac{nr}{(n+1)(n-r)} \right) \left(1 - \frac{r}{n+1} \right) \\
&= 1 + \frac{r^2}{(n+1)^2(n-r)} > 1 .
\end{aligned}$$

22.2 Since $\frac{\partial \Pi}{\partial r} = 0$ after Π is hedged by choosing Δ as given by (22.7), we obtain $\frac{1}{\Pi} \frac{\partial \Pi}{\partial t} + \frac{1}{2} \frac{1}{\Pi} \frac{\partial^2 \Pi}{\partial r^2} \sigma^2 = r$ from the fundamental equation for bond pricing.

Problems of Chap. 23

23.1 Recall that the pricing equation for P is given by

$$\frac{\partial P}{\partial t} + (\mu - \lambda\sigma) \frac{\partial P}{\partial r} + \frac{1}{2} \sigma^2 \frac{\partial^2 P}{\partial r^2} = rP .$$

The boundary condition for a zero-coupon bond is given by

$$P(r, T, T) = 1 . \quad (\text{F.8})$$

(i) We try to find a solution of the form

$$P(r, t, T) = A(T-t) e^{-rB(T-t)} \quad (\text{F.9})$$

for some functions A and B . Then

$$\frac{\partial P}{\partial r} = -A(T-t)B(T-t) e^{-rB(T-t)} ,$$

$$\frac{\partial^2 P}{\partial r^2} = A(T-t)B(T-t)^2 e^{-rB(T-t)} ,$$

$$\frac{\partial P}{\partial t} = -A'(T-t) e^{-rB(T-t)} + rA(T-t)B'(T-t) e^{-rB(T-t)} .$$

After cancellation of common factors, the pricing equation becomes

$$-A' + rAB' - (\mu - \lambda\sigma)AB + \frac{1}{2} \sigma^2 AB^2 = rA ,$$

which must hold for every r . Hence $-A' - (\mu - \lambda\sigma)AB + \frac{1}{2}\sigma^2 AB^2 = 0$ and

$$A(B' - 1) = 0. \quad (\text{F.10})$$

From (F.8) and (F.9) we have $A(0)e^{-rB(0)} = 1$ for every r . Hence $A(0) = 1$ and $B(0) = 0$. Thus we have $B(T - t) = T - t$ from (F.10). Then A satisfies

$$\frac{A'(T - t)}{A(T - t)} = -(\mu - \lambda\sigma)(T - t) + \frac{1}{2}\sigma^2(T - t)^2$$

with $A(0) = 1$. Hence $A(T - t) = \exp(-\frac{1}{2}(\mu - \lambda\sigma)(T - t)^2 + \frac{1}{6}\sigma^2(T - t)^3)$.

- (ii) The result is not realistic because the yield is negative for large T .

23.2

- (i) Use $\text{Cov}(r_t, r_s) = \sigma_0 e^{-\alpha(t+s)} + \sigma^2 \int_0^s e^{-\alpha(s-u)} e^{-\alpha(t-u)} du$.
(ii) Use $\mathbb{E}[\int_s^t r_u du | \mathcal{F}_s] = \int_s^t \mathbb{E}[r_u | \mathcal{F}_s] du$.
(iii) To find the limiting distribution, we may take $t \rightarrow \infty$, or if we want to find it directly, then we take $\frac{\partial p}{\partial t} = 0$ in the Kolmogorov forward equation, and solve $0 = -\frac{\partial}{\partial x}(\alpha(\bar{r} - x)p) + \frac{\partial^2}{\partial x^2}(\frac{1}{2}\sigma^2 p)$. Hence $C = -\alpha(\bar{r} - x)p + \frac{\partial}{\partial x}(\frac{1}{2}\sigma^2 p)$ for some constant C . If we assume that a solution satisfies the condition that $\lim_{x \rightarrow +\infty} xp(x) = 0$ and $\lim_{x \rightarrow +\infty} p'(x) = 0$, then $C = 0$. From

$$\frac{p'(x)}{p(x)} = \frac{\alpha(\bar{r} - x)}{\frac{1}{2}\sigma^2},$$

we obtain

$$p(x) = \frac{1}{\sqrt{\pi \frac{\sigma^2}{\alpha}}} \exp\left(-\frac{(x - \bar{r})^2}{\frac{\sigma^2}{\alpha}}\right).$$

The coefficient is chosen so that $\int_{-\infty}^{\infty} p(x) dx = 1$.

23.3 For the limiting probability density of the interest rate, we take $\frac{\partial p}{\partial t} = 0$ in the Kolmogorov forward equation, and solve

$$0 = -\frac{\partial}{\partial x}(\alpha(\bar{r} - x)p(x)) + \frac{\partial^2}{\partial x^2}\left(\frac{1}{2}\sigma^2 xp(x)\right).$$

As in Exercise 23.2(iii), we solve $C_1 = -\alpha(\bar{r} - x)p(x) + \frac{\partial}{\partial x}(\frac{1}{2}\sigma^2 xp(x))$. Hence $C_1 = -\alpha(\bar{r} - x)p(x) + \frac{1}{2}\sigma^2 p(x) + \frac{1}{2}\sigma^2 xp'(x)$. If we assume that $p(0) = 0$ and $\lim_{x \rightarrow 0} xp'(x) = 0$, then $C_1 = 0$. From

$$\frac{p'(x)}{p(x)} = \frac{\alpha\bar{r} - \frac{1}{2}\sigma^2 - \alpha x}{\frac{1}{2}\sigma^2 x},$$

we obtain $\log p(x) = \left(\frac{2\alpha\bar{r}}{\sigma^2} - 1\right) \log |x| - \frac{2\alpha}{\sigma^2}x + C_2$ for x such that $p(x) > 0$. Thus $p(x) = Cx^{(2\alpha\bar{r})/\sigma^2 - 1}e^{-(2\alpha/\sigma^2)x}$, $x \geq 0$. If $2\alpha\bar{r} = \sigma^2$, then the limit is an exponential function.

23.4 Let $\tau = T - t$ be fixed. For the sake of notational convenience, we shall write $R(t, t + \tau)$ and $P(t, t + \tau)$ in place of $R(t, T)$ and $P(t, T)$, respectively. From Definition 22.8(i),

$$R(t, t + \tau) = -\frac{\log P(t, t + \tau)}{\tau} = -\frac{A(t, t + \tau)}{\tau} + r_t.$$

Hence $dR(t, t + \tau) = dr_t$ and $\text{Var}(dR(t, t + \tau)) = \text{Var}(dr_t) = \text{Var}(\sigma d\tilde{W}_t) = \sigma^2 dt$.

Glossary

- Accrued interest** Interest on a bond that has accrued but has not yet been paid.
- Agent** Someone authorized to do business on behalf of a client.
- American option** An option that can be exercised at any time before or on expiry.
- Arbitrage** An opportunity to make risk-free profit by buying and selling in different markets utilizing the disparity between market prices. When more broadly used, arbitrage means an opportunity to make a profit without risking any future loss.
- Ask price** The price at which a dealer or market-maker offers to sell an asset. Offer price.
- Asset** A financial claim or a piece of property that has monetary value such as cash, shares and bonds.
- At-the-money option** An option for which the current market price of the underlying asset is equal to the strike price.
- Basis point** 0.01%, that is, 0.0001.
- Bid price** The price at which a dealer or market-maker buys an asset. The interest rate a dealer will pay to borrow funds in the money market.
- Bid-ask spread** The difference between the bid price and the ask price.
- Binary option** An option whose payoff is either nothing or a certain amount of cash or shares of the underlying asset. Also called a digital option.
- Binomial tree** A diagram for a possible movement of asset price over time in which asset price can move up or down in each time period.
- Bond** A debt security that promises to make payments periodically for a specified period of time.
- Bund** Bond issued by the Federal German government.
- Broker** An agent, either an individual or firm, who executes buy and sell orders submitted by an investor.
- Call option** An financial contract that gives the holder of the option the right, but not the obligation, to buy the underlying asset at a specified price.
- Cash flows** Cash payments to the holder of a security.

- Collateral** Cash or securities pledged to the lender to guarantee payment in the event that the borrower is not able to make payments. In the futures market a trader has to put up collateral called the initial margin against the possibility that he will lose money on the trade.
- Common stock** A security that is a claim on the earnings and assets of a company with voting right. An ordinary share.
- Convexity** A measure of the curvature of the graph for the price of a bond plotted against the interest rate.
- Contingent claim** A claim that can be made depending on whether one or more specified outcomes occur or do not occur in the future.
- Counterparty** The other party to a trade or contract.
- Coupon** Interest payment from a bond.
- Coupon bond** A bond that periodically pays the owner fixed interest payments, called coupons, until the maturity date at which time a specified final amount, called the principal, is repaid.
- Coupon rate** The amount of the yearly coupon payment expressed as a percentage of the face value of a bond.
- Credit risk** The risk that arises from the possibility that the borrower might default.
- Currency** Paper money and coins.
- Dealer** A person who buys and sells securities, acting as principal in trading for its own account, as opposed to a broker who acts as an agent in executing orders on behalf of its clients. Most dealers act as brokers.
- Default** An event in which the issuer of a debt security is unable to make interest payments or pay off the amount owed when the security matures.
- Delivery** The act of the seller to supply the underlying asset to the buyer. Some derivatives contracts include the physical delivery of the underlying commodity or security.
- Delta-neutral** An option is said to be delta-neutral if it is delta-hedged. It is protected against small changes in the price of the underlying asset.
- Derivative** A financial instrument whose value is determined or derived by the price of another asset.
- Digital option** Another name for a binary option.
- Discount bond** A bond that is bought at a price below its face value and whose face value is repaid at the maturity date. There is no interest payment, and hence it is also called a zero-coupon bond.
- Discount factor** The rate of exchange between a given amount of money at time t and an amount of money at a later date T .
- Diversification** Investing in a portfolio of assets whose returns move in different ways so that the combined risk is lower than for the individual assets.
- Dividend** The payment per share made by a company to its shareholders. On ordinary shares the amount varies with the profitability of the firm. On preferred shares it is usually fixed.
- Down-and-in option** A knock-in option that pays the specified payoff only if the underlying asset price hits a barrier.

Down-and-out option A knock-out option that becomes worthless once the underlying asset price hits a barrier.

Duration A measure of the value-weighted average life of a bond.

Equity Common stock. An equity holder is a part-owner of a company, and receives dividend payment if the company makes a profit.

Exchange Stock exchange. A marketplace in which securities are traded.

Exchange rate The price of one currency expressed in terms of another currency.

European option An option that can be exercised only on expiry date.

Exercise The purchase or sale of the underlying asset at the strike price specified in the option contract.

Exercise price Another name for the strike price of an option. The amount that is paid for the underlying asset.

Exotic option A derivative contract with a nonstandard characteristic.

Expiry The last date of a contract beyond which it becomes worthless if not exercised.

Face value The principal or par value of a bond, which is repaid at maturity.

Fixed income security A security that pays a fixed amount on a regular basis until maturity. A generic term for bonds.

Forward contract A financial contract on the price and quantity of an asset or commodity to sell or buy at a specified time in the future. An agreement between two parties which is not traded on an exchange.

Forward price The price specified in advance in a forward contract for a given asset or commodity. The forward price makes the forward contract have zero value when the contract is written.

Futures contract A financial contract similar to a forward contract that is traded publicly in a futures exchange. The buyer and the seller have to place a margin to avoid a default on the contract, and the futures contract is marked-to-market periodically.

Gilt A bond issued by the UK government. The bond was gilt-edged in older days.

Greeks The rates of changes, denoted by Greek letters except Vega, in the price of a financial derivative when the parameters such as the price of the underlying asset, time, interest rate and volatility change by one unit.

Hedge ratio The ratio of the quantity of the hedge instrument to be bought or sold such as the futures contracts to the quantity of the asset to be hedged.

Hedge To protect oneself by eliminating the risk of loss in an investment.

Implied volatility The volatility implied by the market price of an option. The price computed by applying the Black–Scholes–Merton formula using the implied volatility equals the market price.

Instrument A share, bond or some other tradable security.

Interest rate The cost of borrowing money.

Interest rate risk The possibility of loss associated with changes in interest rates.

In-the-money option An option that would have positive value if exercised now. For example, if the asset price is above the strike price then a call option is in-the-money.

Intrinsic value The value of an option if it can be exercised immediately.

Lognormal distribution A probability distribution of a random variable whose logarithm is normally distributed.

Long position A position that can profit from an increase in the price of an asset under consideration.

Margin A deposit required for both buyers and sellers of a futures contract. It eliminates or reduces the level of loss in case that the buyer or seller fails to meet the obligations of the contract.

Margin call The required amount to be added to the margin account by the owner of a margined position. This can be due to a loss on the position or an increase in the margin requirement.

Market capitalization The total value of a company on the stock market, which is equal to the current share price times the number of shares issued.

Market maker A trader in stocks, bonds, commodities or derivatives who offers to buy at the bid price (lower) and sell at the offer price (higher) at the same time, thereby making a market.

Mark-to-market Updated evaluation of a portfolio or a position based on the current market prices.

Maturity Time to the expiration date of a debt instrument.

No arbitrage principle The principle that one cannot create positive value out of nothing, i.e., there is no free lunch.

Numeraire The units in which a payoff is denominated.

Offer price The price at which a trader offers to sell an asset. The interest rate a dealer asks for lending funds in the money market. Ask price.

Open outcry A trading system in which traders gather in one physical location and convey their offers to sell and buy by gesturing and shouting.

Out-of-the-money option An option that would have zero value if it is exercised now. For example, if the asset price is below the strike price then a call option is out-of-the-money.

Over-the-counter transactions Trades and deals that occur without the involvement of a regulated exchange. Over-the-counter is usually abbreviated as OTC.

Par The face or nominal value of a bond, normally repaid at maturity.

Par bond A bond that is trading at par value.

Payoff The value of a position usually at expiry or maturity.

Physical measure The probability measure defined by the movement of the asset price in the real market.

Position An investor who buys (sells) securities has a long (short) position.

Practitioner A person who works in finance industry in contrast to an academician who teaches at a college.

Preference shares Preference stock. Preference shares pay a fixed dividend and do not carry voting rights.

Premium In the options market the premium is the price of an option. Sometimes it means additional amount above a certain reference level.

Present value Today's value of a cash flow in the future.

Put-call parity A relationship between the prices of European call and put options.

Put option A contract giving the holder the right, but not the obligation, to sell the underlying asset at a specified price.

Rebalancing Resetting of the weights of assets in a portfolio. It involves periodically buying or selling assets in the portfolio to maintain the desired level of asset allocation. For example, consider a portfolio consisting of assets A and B. Suppose that the original target asset allocation was 70% for the asset A and 30% for the asset B. If the asset A performed well in a given period and the weight for the asset A is now 80%, then we sell some units (or shares) of the asset A and buy the asset B in order to maintain the original target allocation.

Replication When a financial derivative can be duplicated by constructing a portfolio consisting of a suitable combination of the underlying asset and the risk-free asset, we say that the portfolio replicates the derivative.

Risk The uncertainty associated with the return on an investment.

Risk-neutral measure The probability measure for which the expected return on the asset is the risk-free interest rate.

Risk premium The difference between the expected return on an asset and the risk-free interest rate. The spread between the interest rate on bonds with default risk and the risk-free interest rate.

Rollover To delay the payment of a debt. To roll a position over one expiry or delivery month to a later month.

S&P 500 Standard and Poor's 500. An index of five hundred shares traded on the New York Stock Exchange (NYSE).

Security A tradable financial instrument such as a stock and bond.

Self-financing portfolio A portfolio without additional investment or withdrawal to manage.

Sharpe ratio The ratio of the risk premium to the standard deviation of the return.

Short position A position that can profit from a decrease in the price of an asset under consideration.

Short selling A transaction in which an investor borrows stock from a broker, sells it in the market, then buys it back and returns it to the lender. A profit will be made if the price has fallen.

Sovereign bond A bond issued by a government.

Split In stock split one share is divided into multiple shares without raising new capital.

Spot price The price of a security for immediate delivery. However, the actual settlement and delivery may take place a few days later.

Spot rate Zero coupon rate.

Spread The difference between two prices or two interest rates. For example, the difference between the bid and offer prices for a dealer, or the difference between lending and funding interest rates for a commercial bank.

Stock A security that is a claim on the earnings and assets of a company.

Stock exchange An organized and regulated market for trading securities. Some stock exchanges use open outcry trading method, while others use telecommunication systems to connect dealers.

Stock index An average of the prices of a collection of stocks. The average is either equally weighted, or weighted with weights proportional to market capitalization. The most commonly used quoted index in the US are the Dow Jones Industrial Average (DJIA) and the S&P 500.

Straddle The purchase of a call and a put with the same strike price and time to expiry.

Strangle The purchase of a call with a higher strike and a put with a lower strike and the same time to expiry.

Systematic risk Undiversifiable or market risk.

Systemic risk The risk that an event such as a bank failure might have a domino effect on the whole financial system.

Tenor Time to maturity.

Term structure of interest rates Behavior of interest rates for a range of maturities.

Theta The rate of change in the value of a derivative due to the passage of time.

Tick size The minimum amount of asset price movement allowed in a price quotation.

Tick value The value of one tick movement in the quoted price on the whole contract size.

Time value The difference between an option's price and its intrinsic value.

Time value of money This expression refers to the fact that money to be received in the future is worth less than the same amount of money today if the interest rate is positive.

Trader A person who buys and sells securities, either for herself, or on behalf of someone else.

Transaction cost The money spent in exchanging financial assets or goods.

Treasury bill A short-term debt security issued by a government.

Treasury note US Treasury notes have maturities between 1 and 10 years when issued.

Treasury bond US Treasury bonds have maturities between 10 and 30 years when issued.

Underlying The asset that is referred to in a derivative contract to define the payoff. Its price determines the price of the derivative.

Up-and-in option A knock-in option that is activated only when the asset price exceeds the barrier.

Up-and-out option A knock-out option that becomes worthless when the asset price exceeds the barrier.

Volatility The standard deviation of the continuously compounded return on an asset. A key parameter in option pricing.

Volatility skew The difference in implied volatility between in-the-money and out-of-the-money options, which makes the implied volatility graph skewed.

Volatility smile The phenomenon that the graph of the implied volatility for a range of strike prices looks like a smile when both in-the-money and out-of-the-money options have higher values of volatility than at-the-money options.

Volatility surface A three-dimensional graph in which volatility is plotted against strike price and time to expiry.

Warrant A long-dated option as a security that can be freely traded.

Writer The seller of an option.

Yield A generic term for the return on an investment, or more specifically, the yield to maturity of a bond.

Yield curve A plot of the interest rates for a given type of bond against time to maturity. It is upward-sloping if the rates on shorter maturity bonds are lower than those on longer maturity bonds, and downward-sloping when short-term yields are higher.

Yield to maturity The interest rate that equates the present value of the payments received from a bond in the future with its current market price.

Zero coupon bond A discount bond without coupon payments.

Zero coupon rate Also spot rate. The interest rate that is applicable to a specific future date.

References

- Adams, M., & Guillemin, V. (1996). *Measure theory and probability*. Birkhäuser.
- Avellaneda, M., & Laurence, P. (1999). *Quantitative modeling of derivative securities: From theory to practice*. Chapman & Hall/CRC.
- Baldi, P. (2017). *Stochastic calculus: An introduction through theory and exercises*. Springer.
- Baz, J., & Chacko, G. (2004). *Financial derivatives: Pricing, applications, and mathematics*. Cambridge University Press.
- Bernstein, P. L. (1992). *Capital ideas: The improbable origins of modern wall street*. Free Press.
- Bernstein, P. L. (1996). *Against the gods: The remarkable story of risk*. John Wiley & Sons.
- Billingsley, P. (2012). *Probability and measure* (Anniversary). John Wiley & Sons.
- Bingham, N. H., & Kiesel, R. (2004). *Risk-neutral valuation: Pricing and hedging of financial derivatives* (2nd ed.). Springer.
- Bjork, T. (2009). *Arbitrage theory in continuous time* (3rd ed.). Oxford University Press.
- Black, F., & Scholes, M. (1973). The Pricing of options and corporate liabilities. *Journal of Political Economy*, 81, 637–654.
- Brigo, D., & Mercurio, F. (2006). *Interest rate models-theory and practice* (2nd ed.). Springer.
- Bruce, C. (2002). *Conned again, Watson! Cautionary tales of logic, math, and probability*. Random House.
- Brzeźniak, Z., & Zastawniak, T. (1999). *Basic stochastic processes*. Berlin: Springer.
- Byun, S. J., & Kim, I. J. (1994). Optimal exercise boundary in a binomial option pricing model. *Journal of Financial Engineering*, 3, 137–158.
- Cairns, A. J. G. (2004). *Interest rate models: An introduction*. Princeton University Press.
- Capinski, M., & Kopp, E. (2012). *The black-Scholes model*. Cambridge University Press.
- Capinski, M., & Zastawniak, T. (2003). *Mathematics for finance: An introduction to financial engineering*. Springer.
- Chang, F. (2004). *Stochastic optimization in continuous time*. Cambridge University Press.
- Choe, G. (2005). *Computational Ergodic theory*. Springer.
- Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53, 385–408.
- Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics*, 7, 229–263.
- Cvitanic, J., & Zapatero, F. (2004). *Introduction to the economics and mathematics of financial markets*. MIT Press.

- Davis, M., & Etheridge, A. (2006). *Louis Bachelier's theory of speculation: The origins of modern finance*. Princeton University Press.
- Durrett, R. (2019). *Probability: Theory and examples* (5th ed.). Cambridge University Press.
- Etheridge, A. (2002). *A course in financial calculus*. Cambridge University Press.
- Filipović, D. (2009). *Term-structure models: A graduate course*. Springer.
- Folland, G. B. (1999). *Real analysis: Modern techniques and their applications* (2nd ed.). John Wiley & Sons.
- Gianin, E. R., & Sgarra, C. (2013). *Mathematical finance: Theory review and exercises-from binomial model to risk measures*. Springer.
- Harrison, J. M., & Kreps, D. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20, 381–408.
- Harrison, J. M., & Pliska, S. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications*, 11, 215–260.
- Haug, E. G. (2007). *The complete guide to option pricing formulas* (2nd ed.). McGraw-Hill.
- Higham, D. (2004). *An introduction to financial option valuation: Mathematics, stochastics and computation*. Cambridge University Press.
- Hirsh, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra*. Academic Press.
- Ho, T., & Lee, S. B. (1986). Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, 41, 1011–1029.
- Hoek, J., & Elliott, R. J. (2006). *Binomial models in finance*. Springer.
- Hull, J. (2011). *Options, futures, and other derivatives* (8th ed.). Upper Saddle River: Prentice Hall.
- Hull, J., & White, A. (1990). Pricing interest-rate-derivative securities. *The Review of Financial Studies*, 3, 573–392.
- Jiang, L. (2005). *Mathematical modeling and methods of option pricing*. World Scientific.
- Johansson, R. (2019). *Numerical Python: Scientific computing and data science applications with Numpy, SciPy and Matplotlib* (2nd ed.). Apress.
- Kahan, W. (1965). Further remarks on reducing truncation errors. *Communications of the ACM*, 8, 40.
- Karatzas, I., & Shreve, S. E. (1998). *Brownian motion and stochastic calculus* (2nd ed.). Springer.
- Kim, I. J. (1990). The analytic valuation of American options. *The Review of Financial Studies*, 3, 547–572.
- Kim, T. S., & Omberg, E. (1996). Dynamic non-myopic portfolio behavior. *The Review of Financial Studies*, 9, 141–161.
- Klebaner, F. C. (2005). *Introduction to stochastic calculus with applications* (2nd ed.). Imperial College Press.
- Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer.
- Korn, R., & Korn, E. (2001). *Option pricing and portfolio optimization: Modern methods of financial mathematics*. American Mathematical Society.
- Kritzman, M. P. (2000). *Puzzles of finance: Six practical problems and their remarkable solutions*. John Wiley & Sons.
- Kumar, A. (2015). *Short selling*. Columbia University Press.
- Kumar, S. S. S. (2007). *Financial derivatives*. Prentice Hall India.
- Kuo, H. (2006). *Introduction to stochastic integration*. Springer.
- McDonald, R. L. (2013). *Derivatives markets* (3rd ed.). Pearson.
- Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: the continuous-time case. *Review of Economics and Statistics*, 51, 247–257.
- Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3, 373–413.
- Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4, 141–183.
- Merton, R. C. (1990). *Continuous-time finance*. Blackwell Publishing.

- Moler, C. B. (2004). *Numerical computing with MATLAB*. Society for Industrial and Applied Mathematics.
- Mörters, P., & Peres, Y. (2010). *Brownian motion*. Cambridge University Press.
- Muirhead, R. J. (2005). *Aspects of multivariate statistical theory* (2nd ed.). Wiley.
- Oksendal, B. (2003). *Stochastic differential equations: An introduction with applications* (6th ed.). Springer.
- Pascucci, A. (2011). *PDE and martingale methods in option pricing*. Springer.
- Pascucci, A., & Runggaldier, W. J. (2012). *Financial mathematics: Theory and problems for multi-period models*. Springer.
- Pfeiffer, P. E. (1990). *Probability for applications*. Springer.
- Pliska, S. R. (1997). *Introduction to mathematical finance: Discrete time models*. Blackwell Publishers.
- Resnick, S. I. (1998). *A probability path*. Birkhäuser.
- Roberts, A. J. (2009). *Elementary calculus of financial mathematics*. Society for Industrial and Applied Mathematics.
- Rogers, L. C. G. (2013). *Optimal investment*. Springer.
- Rogers, L. C. G., & Williams, D. (2000a). *Diffusions, Markov processes and martingales* (Vol. 1). Cambridge University Press.
- Rogers, L. C. G., & Williams, D. (2000b). *Diffusions, Markov processes and martingales* (Vol. 2). Cambridge University Press.
- Royden, H. L. (1989). *Real analysis* (3rd ed.). Macmillan.
- Rudin, W. (1986). *Real and complex analysis* (3rd ed.). McGraw-Hill.
- Schott, J. R. (2005). *Matrix analysis for statistics*. John Wiley & Sons.
- Shreve, S. E. (2004a). *Stochastic calculus for finance I: The binomial asset pricing model*. Springer.
- Shreve, S. E. (2004b). *Stochastic calculus for finance II: Continuous time models*. Springer.
- Shukla, R., Tomas, M.: Black-Scholes option pricing formula (2015). (unpublished).
- Simmons, G. F. (1991). *Differential Equations with applications and historical notes* (2nd ed.). McGraw-Hill.
- Steele, J. M. (2001). *Stochastic calculus and financial applications*. Springer.
- Strauss, W. A. (2008). *Partial differential equations: An introduction* (2nd ed.). John Wiley & Sons.
- Tretyakov, M. (2013). *Introductory course on financial mathematics*. Imperial College Press.
- Vasicek, O. A. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5, 177–188.
- Veronesi, P. (2010). *Fixed income securities: Valuation, risk, and risk management*. John Wiley & Sons.
- Weiming, J. M. (2019). *Mastering Python for finance* (2nd ed.). Packt.
- Wiersema, U. (2008). *Brownian motion calculus*. John Wiley & Sons.
- Woo, G. (1999). *The mathematics of natural catastrophes*. Imperial College Press.

Index

A

Adapted process, 96
Adjoint operator, 241
Affine term structure, 475
Algebra, 27, 118
 Borel σ -algebra, 29
 σ -algebra, 28, 118
Almost surely, 30
American option
 call, 279, 381
 put, 382
Arbitrage, 4, 5
arbitrage-free market, 7
Arbitrage portfolio, 6
Arrow–Debreu security, 7, 9
Asian option, 365
Asset
 risk-free, 400
 risky, 400
Asset-or-nothing option, 342, 364, 500
At-the-money option, 222, 227, 285

B

Bachelier, Louis, 107
Barrier option, 368
 down-and-out put, 372
 up-and-out call, 371
Bayes' theorem, 91
Bernoulli equation, 573
Bertrand's ballot problem, 146
Beta coefficient, 419
Bid-ask spread, 282

Binary expansion, 29, 81
Binary option, 342
Binary representation, 97
Binomial distribution, 50, 263
Binomial tree method
 American put option, 385
 pricing by hedging, 250
 pricing by replication, 252
Black–Scholes–Merton equation
 barrier option, up-and-out call, 371
 derivation by CAPM, 420
 derivation by the binomial tree, 266
 European option, 282
 Feynman–Kac Theorem, 238
 martingale method, 328
 with dividend, 309
Black–Scholes–Merton formula, 289, 328, 367, 550
 with dividend, 309
Black–Scholes–Merton model, 281
Bolzano–Weierstrass theorem, 558
Bootstrap method, 447
Borel
 measurable set, 29
 σ -algebra, 29
Borel–Cantelli lemma, 64, 116
Brownian bridge, 131, 192, 220, 222, 223, 226, 229, 285, 354
Brownian motion, 235
 definition, 108
 first hitting time, 136
 Lévy's characterization, 126
 maximum, 138, 141

multidimensional, 213
 nondifferentiability, 119
 time inversion, 119
 with drift, 236
 Butterfly, 20

C
 Calibration, 479
 Capital Asset Pricing Model (CAPM), 399
 Capital market line, 416
 CAPM, *See* Capital Asset Pricing Model (CAPM)
 Cash-or-nothing option, 342, 500
 Cauchy–Euler equation, 294, 306, 386, 572
 Cauchy–Schwarz inequality, 39, 66
 Cauchy sequence, 561
 Central Limit Theorem, 65, 71
 Change of variables, 108
 Characteristic function, 44, 50, 53
 of a random vector, 55
 Chebyshev’s inequality, 39, 64
 Chi-squared distribution, 585
 noncentral, 481, 586
 Cholesky decomposition, 568
 CIR model, *See* Cox–Ingersoll–Ross (CIR) model
 Closed set, 558
 Coin tossing, 45, 46
 Compact set, 558
 Compact support, 241
 Complete market, 9, 255
 Complete space, 561
 Compounding
 continuous, 446
 periodic, 446
 semi-annual, 446
 Conditional expectation, 90
 given an event, 77
 given a random variable, 85
 given a σ -algebra, 79
 properties of, 81
 Conditional measure, 30
 Conditional probability, 77
 Conditional variance, 87
 Contingent claim, 3
 Continuous measure, 39
 Continuous process, 96
 Convergence
 in distribution, 65
 in measure, 36
 in probability, 36, 63, 101
 of a sequence, 557

Convex function, 37
 Convexity, 456
 Convolution, 57, 581
 Correlated normal variables, 67
 Correlation, 66, 74
 Coupon, 447
 Covariance, 66, 158
 Covariance matrix, 56
 Covariation, 186, 497
 quadratic, 101
 Cox–Ingersoll–Ross (CIR) model, 481
 Cumulative distribution function, 49, 51, 75

Cylinder subset, 29, 81, 117

D

Delta, 298
 Delta hedging, 284, 285, 329, 462
 simulation, 312, 468
 Dense subset, 559
 Derivative, 19
 Diffusion equation, 291
 Digital option, 342
 Dimensionless parameter, 290
 Dirac
 delta measure, 192, 240, 241
 Discounted Feynman–Kac Theorem, 464
 Discount factor, 445
 Discrete measure, 35, 39
 Distribution
 binomial, 263
 chi-squared, 585
 jointly normal, 56
 lognormal, 62
 noncentral chi-squared, 586
 normal, 48, 49, 68
 Dividend, 195, 281, 308
 continuous, 337
 periodic, 339
 Dominating portfolio, 413
 Doob’s L^2 -inequality, 214
 Duration, 454, 480
 modified, 455

E

Early exercise, 279, 383
 Efficient frontier, 413
 Einstein, Albert, 107
 Entropy, 38
 Equivalent measure, 40, 197
 Equivalent metric, 560
 Equivalent norm, 560

Error function, 49, 581
 European option
 call, 19, 381
 put, 19
 Exotic option, 21, 365
 Asian, 365
 barrier, 368
 binary, 342
 digital, 342
 Expectation, 45
 Extension
 of a function, 561

F
 Fair game, 211
 Farkas' Lemma, 14
 Feynman–Kac Theorem, 233, 237, 328, 329
 discounted, 464
 Feynman, Richard, 233
 Filtration, 95
 natural, 96
 Fixed income security, 447
 Fokker–Planck equation, 241
 Forward, 18, 465
 Forward rate, 451
 Forward rate agreement, 451
 Forward risk-neutral measure, 502
 Fourier inversion formula, 50
 Fourier transform, 44, 50, 51, 55, 113, 200, 578
 Free boundary problem, 384
 Fubini's theorem, 153
 Function
 convex, 37
 measurable, 30
 simple, 31
 Fundamental equation for pricing
 discounted interest rate derivatives, 501
 interest rate derivatives, 459
 Futures, 18

G
 Game of twenty questions, 95
 Gamma, 298
 Gamma function, 586
 Gaussian process, 220
 Geometric Brownian motion, 62, 237, 238, 431
 Girsanov's theorem, 203, 211, 226, 461, 496
 generalized, 203

Greeks
 Delta, 298
 Gamma, 298
 Kappa, 302
 Rho, 298
 Theta, 298
 Vega, 298
 Gronwall's inequality, 216, 571

H

Hamilton–Jacobi–Bellman equation, 430
 Heat equation, 290
 Heaviside function, 580, 582
 MATLAB command, 245
 Hedge ratio, 283, 352
 Hedging, 9, 250, 284
 Heine–Borel theorem, 558
 Hölder's inequality, 38
 Hypergeometric function, 587

I

Independence
 of events, 53
 of random variables, 54
 of σ -algebras, 53, 54
 Independence property
 of conditional expectation, 81, 477
 Indicator function, 31, 43
 Inequality
 Cauchy–Schwarz, 39, 66
 Chebyshev, 39
 Gronwall, 216, 571
 Hölder, 38
 Jensen, 37, 83
 Minkowski, 39
 Infinitesimal generator, 237, 241
 Information, 84
 Instantaneous forward rate, 458, 485
 Insurance, 19, 24
 Integrable function, 34, 35
 Integrating factor, 571
 Integrator, 149, 162
 Interest rate
 compounded, 445
 simple, 445
 Internal rate of return, 452
 In-the-money option, 284
 Intrinsic value, 19
 Itô
 formula, 178, 234, 242
 identity, 477

integral, 153
 isometry, 151, 153, 158
 lemma, 178
 process, 185

J

Jensen's inequality, 37, 611
 for conditional expectation, 83
 Joint probability density function, 55, 57, 58
 Jointly normal distribution, 56

K

Kahan summation algorithm, 518
 Kappa, 302
 Kolmogorov
 backward equation, 239
 Extension Theorem, 29, 118
 forward equation, 241, 242, 617
 Kolmogorov, Andrey Nikolaevich, 32
 Kurtosis, 68, 74

L

Lagrange multiplier method, 417
 Laguerre
 equation, 571
 Laplace transform, 53, 580
 Law of Large Numbers, 63, 609
 Law of Large Numbers for Brownian motion, 123
 Least squares method, 565
 Lebesgue
 measure, 29, 97
 Lebesgue integral, 32
 Lévy, Paul, 126
 Lévy's theorem, 126, 219
 Lim sup of subsets, 64
 Line of best fit, 566
 Lipschitz continuous, 557
 Local martingale, 105, 464
 Local time, 192
 Lognormal distribution, 62
 Long position, 6, 17

M

Margin, 18
 Market portfolio, 416
 Market price of interest rate risk, 463
 Market price of risk, 417, 463
 Markov

 process, 96
 property, 96, 255
 Markowitz bullet, 405, 416
 Martingale, 98, 121, 125, 163, 233, 328, 370, 494
 game, 104
 local, 105
 stopped, 370
 Martingale measure, 11
 Martingale Representation Theorem, 164, 233, 466, 497
 MATLAB, 543
 Mean, 45
 Mean reversion, 476
 Measurable
 function, 30
 space, 28
 subset, 28
 Measure, 28
 absolutely continuous, 40
 conditional, 30
 continuous, 39
 discrete, 35, 39
 equivalent, 40, 197
 Lebesgue, 29, 97
 martingale, 11
 monotonicity of, 29
 probability, 28, 35
 singular, 97
 singular continuous, 40
 Minimum variance line, 412
 Minkowski's inequality, 39
 Model, 13
 Black–Scholes–Merton, 281
 Modified duration, 455
 Moment generating function, 53
 Monte Carlo method, 65, 316
 Monty Hall problem, 88

N

Newton–Raphson method, 453, 467
 No arbitrage principle, 5, 22, 250, 252, 279, 284, 457, 495
 Noncentral chi-squared distribution, 481, 586
 Noncentrality parameter, 586
 Norm
 of a linear transformation, 559
 Normal distribution, 48, 49, 53, 68, 476
 Novikov condition, 204, 461
 Numeraire, 464, 493

O

Open ball, 558
 Open set, 558
 Optimal exercise boundary, 383
 Option
 American call, 279, 381
 American put, 382
 asset-or-nothing, 342, 364, 500
 at-the-money, 222, 227, 285
 barrier, 368, 372
 butterfly, 20
 cash-or-nothing, 342, 500
 European call, 19, 381
 European put, 19
 exotic, 21, 365
 in-the-money, 284
 out-of-the-money, 285
 straddle, 19
 strangle, 19
 vanilla, 21
 Option pricing
 by hedging, 322
 by replication, 324, 326
 martingale method, 327
 pdf of underlying asset, 326
 Ornstein–Uhlenbeck process, 216, 240, 242
 Orthogonal projection, 83
 Out-of-the-money option, 285
 Over-The-Counter (OTC), 19

P

Parallelogram law, 101
 Pascal's triangle, 257
 Portfolio, 3, 6
 dominating, 413
 self-financing, 260
 Position
 long, 17
 short, 17
 Predictable process, 98, 106, 162, 251
 Principal, 445
 Probability
 density function, 44, 49
 distribution, 226
 measure, 28, 35
 preserving transformation, 45
 product, 55
 Probability Density Function (PDF), 44, 586

 joint, 55

Process

 adapted, 96
 Itô, 185
 Markov, 96
 predictable, 98
 stochastic, 96
 Put-call parity, 22
 asset-or-nothing option, 343
 cash-or-nothing option, 342
 dividend paying stock, 311
 Put-call symmetry
 American option, 395
 Python, 507

Q

Quadratic
 covariation, 101, 186
 variation, 101, 186

R

Radon–Nikodym
 derivative, 40, 199, 461
 Theorem, 199, 40
 Random variable, 30, 44
 Random vector, 57
 Random walk, 97, 99
 Rank, 564
 Rebalancing, 468
 Recombining tree, 257
 Reflection, 134–136
 Replication, 9, 250, 252, 324
 Restriction
 of a function, 561
 Return, 399
 Rho, 298
 Riccati equation, 476, 483, 574
 Riemann–Stieltjes integral, 149
 Risk
 market price of, 417
 premium, 417, 420
 Risk-free
 asset, 400
 interest rate, 282
 Risk-neutral
 interest rate, 463
 measure, 463
 pricing, 11, 252
 probability, 11, 252
 Risky asset, 400

S

Sample path, 96

construction, 115
 SDE, it See stochastic differential equation (SDE)
 Security, 17
 Security market line, 420
 Self-financing portfolio, 260, 283, 468, 498
 Set, 555
 Short position, 6, 17
 Short selling, 17, 282
 Simple function, 31
 Singular value, 560
 Skewness, 68
 Space
 measurable, 28
 Spot rate, 447
 Spot rate curve, 448
 Spread option, 19
 Standard deviation, 46
 Standard normal distribution, 48, 49
 State price vector, 7
 Stochastic differential equation (SDE), 213, 233, 238
 Stochastic integral
 with respect to a martingale, 163
 Stochastic integration by parts, 186
 Stochastic process, 96
 continuous, 96
 Stock split, 281
 Stone–Weierstrass theorem, 559
 Stopped martingale, 370
 Stopping time, 102, 385
 Straddle, 19
 Strangle, 19, 24
 Submartingale, 98, 100
 Subset
 measurable, 28
 Supermartingale, 98
 Symmetric matrix, 560

T

Tanaka's formula, 192
 Tangent line, 416, 418

Tangent point, 416
 Taylor series expansion, 52
 Term structure of interest rates, 448, 479
 Theta, 298
 Time to expiry, 22
 Tower property, 81, 199, 201, 202, 599, 608
 Trace, 564
 Transition probability, 96, 239, 241

U

Uncorrelated normal variables, 67
 Underlying asset, 17
 Uniform continuity, 561
 Uniform distribution, 46

V

Vanilla option, 21
 Vanishing point, 123
 Variance
 conditional, 87
 Variation, 110, 149
 first order, 111
 quadratic, 101, 110, 111
 Vasicek model, 476
 Vega, 298
 Volatility, 298

W

Weierstrass theorem, 559
 With probability 1, 30

Y

Yield curve, 448
 Yield to maturity, 452, 454

Z

Zero coupon bond, 457
 Zero rate, 447