

CMS/CAIMS Books in Mathematics

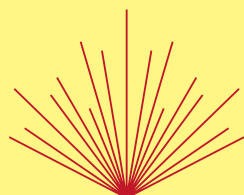


Canadian
Mathematical Society
Société mathématique
du Canada

Alexander Melnikov

A Course of Stochastic Analysis

Second Edition



CAIMS
SCMAI



Springer

CMS/CAIMS Books in Mathematics

Volume 17

Series Editors

Frithjof Lutscher , Department of Mathematics, University of Ottawa, Ottawa, Canada

Nilima Nigam, Department of Mathematics, Simon Fraser University, Burnaby, Canada

Mikhail V. Kotchegov, Dept Mathematics & Statistics, Memorial University of Newfoundland, St. John's, NL, Canada

Marco Merkli, Math & Statistics, Henrietta Harvey Bldg, Memorial University, St. John's, NL, Canada

Associate Editors

Ben Adcock, Department of Mathematics, Simon Fraser University, Burnaby, Canada

Martin Barlow, University of British Columbia, Vancouver, Canada

Heinz H. Bauschke , University of British Columbia, Kelowna, Canada

Matt Davison, Department of Statistical and Actuarial Science, Western University, London, Canada

Leah Keshet, Department of Mathematics, University of British Columbia, Vancouver, Canada

Niky Kamran, Department of Mathematics and Statistics, McGill University, Montreal, Canada

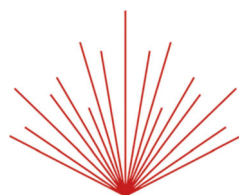
Mikhail Kotchetov, Memorial University of Newfoundland, St. John's, Canada

Raymond J. Spiteri, Department of Computer Science, University of Saskatchewan, Saskatoon, Canada

CMS/CAIMS Books in Mathematics is a collection of monographs and graduate-level textbooks published in cooperation jointly with the Canadian Mathematical Society- Société mathématique du Canada and the Canadian Applied and Industrial Mathematics Society-Société Canadienne de Mathématiques Appliquées et Industrielles. This series offers authors the joint advantage of publishing with two major mathematical societies and with a leading academic publishing company. The series is edited by Karl Dilcher, Frithjof Lutscher, Nilima Nigam, and Keith Taylor. The series publishes high-impact works across the breadth of mathematics and its applications. Books in this series will appeal to all mathematicians, students and established researchers. The series replaces the CMS Books in Mathematics series that successfully published over 45 volumes in 20 years.



CMS
SMC



CAIMS
SCMAI

Alexander Melnikov

A Course of Stochastic Analysis

Second Edition

 Springer

Alexander Melnikov 
Department of Mathematical and Statistical
Sciences
University of Alberta
Edmonton, AB, Canada

ISSN 2730-650X ISSN 2730-6518 (electronic)
CMS/CAIMS Books in Mathematics
ISBN 978-3-032-20481-3 ISBN 978-3-032-20482-0 (eBook)
<https://doi.org/10.1007/978-3-032-20482-0>

Mathematics Subject Classification: 28, 60, 62

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023, 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.

Preface

Preface to First Edition

The foundation of modern probability theory was done by A.N. Kolmogorov in his monograph appeared as “Grundbegriffe der Wahrscheinlichkeitsrechnung” in 1933. Since that time we use the notion of a probability space $(\Omega, \mathcal{F}, P)$, where Ω is an abstract set of elementary outcomes “ ω ” of a random experiment, $\mathcal{F}$ is a family of events, and P is a probability measure, defined on $\mathcal{F}$. In that book there is a special remark about the property of independence as a specific feature of the theory of probability. Due to this property one can use deterministic numerical characteristics like mean and variance to describe the behavior of families of random variables. It was later recognized that dynamics of random events can be exhaustively determined considering t and “ ω ” together and a random process as a function of these two variables on the base of information flow $F_t, t \geq 0$.

This fruitful idea was extremely important for a general theory of random processes initiated in the middle of the 20th century. The central notion of this theory is a stochastic basis $(\Omega, \mathcal{F}, (F_t), P)$, i.e. a probability space equipped with an information flow or filtration (F_t) . In such a setting deterministic numerical characteristics induced by the independence property are replaced by their conditional versions with respect to filtration (F_t) . So that a predictability has appeared as a driver of extension of stochastic calculus to the biggest possible class of processes called semimartingales. These processes admit the full description in predictable terms. Moreover, they unify processes with discrete and continuous time. As a result, we arrive to a nice transformation of the theory of probability and stochastic processes to a wider area which is called stochastic analysis.

The primary goal of the book is to deliver basic notions, facts and methods of stochastic analysis using a unified methodology, sufficiently strong and complete, and giving interesting and valuable implementations in mathematical finance and statistics of random processes. There are a lot of examples considered to illustrate theoretical concepts discussed in line with problems for students aligned with material. Moreover, the list of supplementary problems with hints and solutions,

covering both important theoretical statements and purely technical problems intended to motivate deeper understanding of stochastic analysis is provided at the end of the book. The book can be considered as a textbook for both senior undergraduate and graduate courses on stochastic analysis/stochastic processes. It certainly can be helpful for undergraduate and graduate students, instructors as well as for experts in stochastic analysis and its applications.

The book is based on the lectures given by the author in different times in Lomonosov Moscow State University, State University-Higher School of Economics, University of Copenhagen and in the University of Alberta. The author is grateful to his PhD students Ilia Vasilev and Andrey Pak at the Mathematical Finance Program (University of Alberta) for their kind help in preparation of this book.

Preface to Second Edition

The second edition presents a further development of the first edition with the aim to incorporate to the textbook new findings of modern stochastics. As we know, the assumption that the stochastic basis satisfies so-called usual conditions is a foundation concept in the general theory of stochastic processes. This theory has been instrumental in generating many important and even fundamental results in mathematical finance and in the theory of stochastic processes itself. It is difficult to conceive of a theory of stochastic processes without the usual conditions and RCLL (right continuous with finite left limits) processes. However, it is not difficult to give some examples of stochastic basis without the usual conditions. In 1975, Dellacherie started to study the theory of stochastic processes without this standard assumption, termed "unusual conditions." Further developments of this theory were carried on by Lepingle, Horowitz, Lenglart, and mostly by Galtchouk. We believe that the theory optional processes developed in this framework will offer a natural foundation and a versatile set of tools for modeling financial markets, statistics of random processes as well as risk modeling.

The aim of the second edition of the book is to present the elements of the theory of optional processes on unusual stochastic basis and bring its methods to Mathematical Finance, Stochastic Regression Analysis, and Risk Theory. This leading idea was realized in writing three additional Chaps. 12, 13, and 14 with upgrading list of problems as Chap. 15. Chapter 12 presents foundation material on optional processes and optional stochastic calculus. Chapter 13 introduces stochastic exponentials and logarithms and their applications for financial modeling. Chapter 14 describes how optional processes work for Regression Analysis and Risk Theory. Chapter 15 presents the extended list of problems adapted to the main

contents of the book and their solutions. The second edition of the book presents the first exposition of this theory in textbooks. The author thanks his PhD student Pouneh Mohammadi Nejad for her great help in preparation of the second edition of the textbook.

Edmonton, AB, Canada

Alexander Melnikov

Contents

1	Probabilistic Foundations	1
1.1	Classical Theory and the Kolmogorov Axiomatics	1
1.2	Probabilistic Distributions and the Kolmogorov Consistency Theorem	7
2	Random Variables and Their Quantitative Characteristics	13
2.1	Distributions of Random Variables	13
2.2	Expectations of Random Variables	18
3	Expectations and Convergence of Sequences of Random Variables	23
3.1	Limit Behavior of Sequences of Random Variables in Terms of Their Expected Values	23
3.2	Probabilistic Inequalities and Interconnections Between Types of Convergence of Random Variables	28
4	Weak Convergence of Sequences of Random Variables	33
4.1	Weak Convergence and Its Description in Terms of Distributions	33
4.2	Weak Convergence and Central Limit Theorem	37
5	Absolute Continuity of Probability Measures and Conditional Expectations	43
5.1	Absolute Continuity of Measures and the Radon-Nikodym Theorem	43
5.2	Conditional Expectations and Their Properties	45
6	Discrete Time Stochastic Analysis: Basic Results	51
6.1	Basic Notions: Stochastic Basis, Predictability, and Martingales	51
6.2	Martingales on Finite Time Interval	57
6.3	Martingales on Infinite Time Interval	62

7	Discrete Time Stochastic Analysis: Further Results and Applications	69
7.1	Limiting Behavior of Martingales with Statistical Applications	69
7.2	Martingales and Absolute Continuity of Measures. Discrete Time Girsanov Theorem and Its Financial Application	79
7.3	Asymptotic Martingales and Other Extensions of Martingales	82
8	Elements of Classical Theory of Stochastic Processes	89
8.1	Stochastic Processes: Definitions, Properties, and Classical Examples	89
8.2	Stochastic Integrals with Respect to a Wiener Process	100
8.3	The Ito Processes: Formula of Changing of Variables, Theorem of Girsanov, and Representation of Martingales	109
9	Stochastic Differential Equations, Diffusion Processes, and Their Applications	119
9.1	Stochastic Differential Equations	119
9.2	Diffusion Processes and Their Connection with SDEs and PDEs	130
9.3	Applications to Mathematical Finance and Statistics of Random Processes	140
9.4	Controlled Diffusion Processes and Applications to Option Pricing	147
10	General Theory of Stochastic Processes Under “Usual Conditions”	155
10.1	Basic Elements of Martingale Theory	155
10.2	Extension of Martingale Theory by Localization of Stochastic Processes	169
10.3	On Stochastic Calculus for Semimartingales	178
10.4	The Doob-Meyer Decomposition: Proof and Related Remarks	189
11	General Theory of Stochastic Processes in Applications	195
11.1	Stochastic Mathematical Finance	195
11.2	Stochastic Regression Analysis	203
12	Basic Elements of Optional Stochastic Analysis	211
12.1	Building Blocks of Optional Stochastic Analysis	211
12.2	Optional Stochastic Calculus	217
13	Optional Stochastic Differential Equations and Their Applications	225
13.1	Existence and Uniqueness of Solutions of Optional SDEs	225
13.2	Optional Stochastic Exponents and Logarithms	231
13.3	Financial Modeling Using Optional Semimartingales	239

14 Optional Semimartingales for Stochastic Regression

Analysis and Risk Theory 249

 14.1 Optional Regression Analysis 249

 14.2 Optional Risk Model and Estimation of Ruin Probability 257

15 Supplementary Problems 267

References 285

Index 289

Acronyms and Notation

$(\Omega, \mathcal{F}, P)$	Probability space
$(\mathcal{F}_n)_{n=0,1,\dots}$	Filtration
$(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$	Stochastic basis
R^d	d -dimensional Euclidian space
$\mathcal{B}(R^d)$	Borel σ -algebra on R^d
$(R^d, \mathcal{B}(R^d))$	Borel space
$\mathcal{A}^+$	Class of increasing integrable processes
$\mathcal{A}$	Class of processes with integrable variation
$\mathcal{A}_{loc}^+$	Class of increasing locally integrable processes
$\mathcal{A}_{loc}$	Class of processes with locally integrable variation
$\mathcal{V}$	Class of processes with finite variation
L^p	Space of random variables with finite p-moment
$\mathcal{O}$	Optional σ -algebra
$\mathcal{P}$	Predictable σ -algebra
$\mathcal{M}$	Set of uniformly integrable martingales
$\mathcal{M}^2$	Set of square integrable martingales
$\mathcal{M}_{loc}$	Set of local martingales
$\mathcal{M}_{loc}^2$	Set of locally square integrable martingales
$\mathbb{E}X$	Expected value of random variable X
$Var(X)$	Variance of X
$\mathbb{E}(X A)$	Conditional expected value of random variable X with respect to σ -algebra A
$X_n \xrightarrow{P} X$	Convergence in probability
$X_n \rightarrow X (a.s.)$	Convergence almost surely
$X_n \xrightarrow{d} X$	Convergence in distribution
$X_n \xrightarrow{L^p} X$	Convergence in L^p space
$X_n \xrightarrow{w} X$	Weak convergence
$Law_p(X)$	Distribution of X w.r. to P
$Law(X, P)$	Distribution of X with respect to measure P

$\tilde{P} << P$	Absolute continuity of measure $\tilde{P}$ with respect to P
$\tilde{P} \sim P$	Equivalence of measures $\tilde{P}$ and P
$\binom{n}{k} = C_n^k$	Number of combinations from n by k
$\langle X, Y \rangle$	Joined quadratic characteristic of X and Y
$[X, Y]$	Joined quadratic bracket X and Y
$\llbracket \tau, \sigma \rrbracket$	Stochastic integral with finite limits τ and σ
$\llbracket \tau \rrbracket$	Graph of τ
$\mathcal{V}^+$	Class of increasing processes
$\langle X, X \rangle$	Quadratic characteristic (compensator) of X
$\text{conv}(f_1, f_2, \dots)$	Linear convolution of vectors $f_1, f_2, \dots$
$A \cup B$	Union of sets A and B
$A \cap B$	Intersection of sets A and B
A^c	Complement of set A
$P(A)$	Probability of set A
$\varepsilon(X)$	Stochastic exponent of X
$P(A \mathcal{A})$	Conditional probability of A w.r. to σ -algebra $\mathcal{A}$
$\text{cov}(X, Y)$	Covariance of random variables X and Y
$a \wedge b$	$\min(a, b)$
$a \vee b$	$\max(a, b)$
a^+	$\max(a, 0)$
a^-	$\max(-a, 0)$
$\ \cdot\ _p$	Norm in space L^p
$l_X(t, a)$	Local time of W at level a during $[0, t]$
$\{\omega : X_t \rightarrow\}$	Set of ω for which there exists a finite limit $\lim_{t \rightarrow \infty} X_t(\omega)$

Chapter 1

Probabilistic Foundations



In the first chapter, Kolmogorov's axioms of the theory of probability are presented. The choice of the system of the axioms is explained in context of the Caratheodory theorem. Different probability spaces as well as probability distributions are introduced too. The famous Kolmogorov consistency theorem is formulated, and as a consequence, a brief construction of the Wiener measure is shown (see [7, 12, 13, 17, 28, 33, 55], and [62]).

1.1 Classical Theory and the Kolmogorov Axiomatics

Classical theory of probability deals with a *finite probability space*, i.e., it creates a mathematical model for any *random (stochastic) experiment* with a *finite* number of possible outcomes (elementary events, results) $\omega \in \Omega$, $|\Omega| = N < \infty$, where Ω is the whole set of outcomes.

Any other events A , B , and C are simply called *events*. We cannot be certain of the event A under consideration before a random experiment. Nevertheless, we expect to have a *quantitative measure* of such a possibility. In classical probability, the problem of finding this measure can be solved perfectly.

Denote $\omega_1, \dots, \omega_N$ elementary outcomes and define $p_1, \dots, p_N$ corresponding possibilities for them. Any other and more complicated event A can be reconstructed with the help of $\omega_1, \dots, \omega_N$. Hence, to determine a measure of possibility for event A , one can put

$$P(A) = \sum_{i: \omega_i \in A} p_i.$$

The situation is *different if Ω is not finite*. Consider the well-known random experiment called *infinite coin trial* with outcomes H(head) and T(tail), and with the *probability of success (H)* p , $p \in (0, 1)$.

It is quite clear that in this case the set Ω can be chosen as follows:

$$\Omega = \{\omega : \omega = (a_1, a_2, \dots), a_i = 0 \text{ or } 1\}.$$

The natural question arises: How big is this set? Mathematically speaking, we should find the power $|\Omega|$ of this set.

It is well-known that any real number $a \in [0, 1)$ can be represented uniquely in the form

$$a = \frac{a_1}{2} + \frac{a_2}{2^2} + \dots,$$

and this representation provides a one-to-one correspondence between $[0, 1)$ and Ω :

$$a \in [0, 1) \Leftrightarrow \omega = (a_1, a_2, \dots) \in \Omega.$$

If we follow to the classical probability theory, we should find the way to define the *probability for any elementary event ω* . It is equivalent to a random choice of a point from the semi-interval $[0, 1)$. Taking into account a symmetry of such trials, we must conclude that all elementary events ω are uniformly possible, and hence,

$$p = p(\omega) = \text{const} = c.$$

Further, we have

$$1 = |[0, 1)| = \text{Probability of } \Omega = \sum_{\omega \in \Omega} p(\omega) = \infty$$

because the set Ω is not countable. So, we must define $p(\omega) = c = 0$; however, this choice is not satisfactory because, for example, for $A = [0, 1/2)$ we have

$$1/2 = \text{Probability of } A = P(A) = \sum_{\omega \in A} p(\omega),$$

and we arrive to a contradiction again.

Based on these considerations, one can conclude that the approach of classical theory of probability must be seriously transformed:

The probability should be well-defined instead of elementary outcomes for a reasonable collection of events of Ω .

The natural candidate for this type of collection is given in the following definition.

Definition 1.1 A collection $\mathcal{A}$ of subsets of given set Ω is called an algebra if the following conditions are fulfilled:

1. $\Omega \in \mathcal{A}$
2. $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}, A \cap B \in \mathcal{A}$
3. $A \in \mathcal{A} \Rightarrow A^c = \bar{A} = \Omega \setminus A \in \mathcal{A}$

Using this definition we can realize the idea described above as follows.

Definition 1.2 A function $\mu : \mathcal{A} \rightarrow [0, \infty]$ is a *finite-additive measure* on $\mathcal{A}$, if for any disjoint sets $A, B \in \mathcal{A}$:

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

In case of $\mu(\Omega) < \infty$, this measure is called *finite*. In particular case $\mu(\Omega) = 1$, such a measure is called a (finite-additive) *probability measure*. If there is a partition of $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$, $\Omega_k \cap \Omega_m = \emptyset, k \neq m, \Omega_n \in \mathcal{A}$ such that $\mu(\Omega_n) < \infty, n = 1, 2, \dots$, then μ is *σ -finite*.

Based on these definitions, we arrive to the first probability model that will be called a *probability space of the first level*:

$(\Omega, \mathcal{A}, P)$, where $\mathcal{A}$ is an algebra, and P is a probability (finite-additive measure with $P(\Omega) = 1$).

Assume that a finite-additive probability P on the probability space of the first level $(\Omega, \mathcal{A}, P)$ is *countably additive* (*σ -additive*). It means that for any sequence of disjoint events $A_n \in \mathcal{A}$ with $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ the following equality holds:

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

In this case, we call this measure a *probability (probability measure)*. Let us count a *list of natural properties of probability*:

1. $\emptyset \in \mathcal{A} \Rightarrow P(\emptyset) = 0$.
2. $A, B \in \mathcal{A} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
3. $A, B \in \mathcal{A}, A \subseteq B \Rightarrow P(A) \leq P(B)$.
4. $A_n \in \mathcal{A}, n = 1, 2, \dots, \bigcup_{n=1}^{\infty} A_n \in \mathcal{A} \Rightarrow P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$.

Problem 1.1 Prove the properties (1) to (4).

The idea of the proof is shown here in case of property (1): For any $A \in \mathcal{A}$, we have $\emptyset \cap A \in \mathcal{A}$. Further, using the finite additivity of P and equality $\emptyset \cup \Omega = \Omega$, we obtain

$$1 = P(\emptyset \cup \Omega = \Omega) = P(\emptyset) + P(\Omega) = P(\emptyset) + 1,$$

and, hence, $P(\emptyset) = 0$.

The following theorem contains conditions under which a finite-additive probability measure is countable additive.

Theorem 1.1 *Let P be a finitely additive set function on the algebra $\mathcal{A}$ with $P(\Omega) = 1$. Then the following sentences are equivalent:*

1. P is a probability (σ -additive).
2. P is continuous from below:

$$P(\cup_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} P(A_n) \text{ for } A_1 \subseteq A_2 \subseteq \dots, \cup_{n=1}^{\infty} A_n \in \mathcal{A}.$$

3. P is continuous from above:

$$P(\cap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} P(A_n) \text{ for } A_1 \supseteq A_2 \supseteq \dots, \cap_{n=1}^{\infty} A_n \in \mathcal{A}.$$

4. P is continuous at $\emptyset$:

$$\lim_{n \rightarrow \infty} P(A_n) = 0 \text{ for } A_1 \supseteq A_2 \supseteq \dots, \cap_{n=1}^{\infty} A_n = \emptyset.$$

Proof We only show the implication (1) $\Rightarrow$ (2) because the other statements can be proved in a similar way.

Let us represent $\cup_{n=1}^{\infty} A_n = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \dots$ as a union of disjoint events. Therefore,

$$P(\cup_{n=1}^{\infty} A_n) = P(A_1) + P(A_2) - P(A_1) + \dots = \lim_{n \rightarrow \infty} P(A_n).$$

□

The first probability model $(\Omega, \mathcal{A}, P)$ looks very constructive. But in the framework of this model, we *cannot operate with events which are combinations of countable numbers of these events*. To avoid the disadvantages of the probability space of the first level, we introduce the following notions.

Definition 1.3 A system $\mathcal{F}$ of subsets of Ω is called a σ -algebra (*sigma-algebra*) if $\mathcal{F}$ is an algebra, and for any sequence $A_n \in \mathcal{F}$, $n = 1, 2, \dots$, their unions and intersections belong to $\mathcal{F}$:

$$A_n \in \mathcal{F} \Rightarrow \cup_{n=1}^{\infty} A_n \text{ and } \cap_{n=1}^{\infty} A_n \in \mathcal{F}.$$

The couple $(\Omega, \mathcal{F})$ is called a *measurable space*. Let us investigate *interconnections* of algebras and σ -algebras. *First*, we note that

$$\mathcal{F}_* = \{\emptyset, \Omega\} \text{ and } \mathcal{F}^* = \{A, A \subseteq \Omega\} = \{\text{all subsets of } \Omega\}$$

are both algebras and σ -algebras. *Second*, we can fix a set $A \subseteq \Omega$ and define another algebra (and a σ -algebra also) $\mathcal{F}_A = \{A, A^c, \emptyset, \Omega\}$, which is generated by event A . This way of construction of algebras and σ -algebras admits a natural generalization.

Lemma 1.1 *Let $\mathcal{E}$ be a collection of sets in Ω , and then there are the smallest algebra and the smallest σ -algebra containing $\mathcal{E}$.*

Idea of the Proof It is clear that $\mathcal{E} \subseteq \mathcal{F}^*$. Therefore, there exist at least one algebra and one σ -algebra with the desirable property. Define $\alpha(\mathcal{E})$ and $\sigma(\mathcal{E})$ as a collection of sets that belong to every algebra (correspondently, σ -algebra) containing $\mathcal{E}$. $\square$

Definition 1.4 A collection of subsets of Ω is called a *monotonic class* if for any increasing (decreasing) sequence of events their union (intersection) belongs to this class.

Denote $\mu(\mathcal{E})$ a minimal monotonic class containing $\mathcal{E}$. It turns out *an algebra $\mathcal{A}$ is a σ -algebra if and only if it is a monotonic class*.

Idea of the Proof The direct implication of this statement is clear because σ -algebra is a monotonic class. The inverse implication can be proved as follows. If $(A_n)_{n=1,2,\dots}$ is an arbitrary sequence of subsets Ω from a monotonic class $\mathcal{A}$, we can define a monotonic class $\mathcal{A}$, and a monotonic sequence $(B_n)_{n=1,2,\dots}$ with $B_n = \bigcup_{i=1}^n A_i \in \mathcal{A}$, and then $B_n \uparrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$. Using this statement, we can easily prove that $\mu(\mathcal{A}) = \sigma(\mathcal{A})$ if $\mathcal{A}$ is an algebra. $\square$

Example 1.1 (Borel Space) Let us take $R = R^1 = (-\infty, \infty)$ as Ω , and consider subintervals

$$(a, b] = \{x \in R : a < x \leq b\} \text{ for } -\infty \leq a < b < \infty.$$

Let $\mathcal{A}$ be a system of subsets A of the form:

$$A = \sum_{i=1}^n (a_i, b_i] = \bigcup_{i=1}^n (a_i, b_i], \quad n = 1, 2, \dots$$

Including to $\mathcal{A}$ the empty set $\emptyset$, we can see that $\mathcal{A}$ is an algebra. Moreover, we can easily define a measure on $\mathcal{A}$: $l_0(A) = \sum_{i=1}^n (b_i - a_i)$. In particular,

$$l_0((a, b]) = b - a = \text{the length of } (a, b].$$

We also note that $\mathcal{A}$ is not a σ -algebra because $A_n = (0, 1 - 1/n] \in \mathcal{A}$, but the union $\bigcup_{n=1}^{\infty} A_n = (0, 1) \notin \mathcal{A}$.

Even in the framework of this example, the following problem arises:

Determine a minimal σ -algebra $\mathcal{B}(R)$ containing $\mathcal{A}$ and a measure on the measurable space $(R, \mathcal{B}(R))$ which is called the Borel space.

Problem 1.2 Prove that:

1. $\mathcal{B}(R)$ is equivalent to a minimal σ -algebra containing all intervals $(a, b]$.
2. $\mathcal{B}(R)$ contains (a, b) , $[a, b]$, $\{a\}$.

Existence of $\mathcal{B}(R)$ is obvious, but an extension of l_0 from $\mathcal{A}$ to $\mathcal{B}(R)$ is a real mathematical problem, and its solution is the famous *Lebesgue measure* l on $(R, \mathcal{B}(R))$. To avoid real technical difficulties, we would like to get a solution of this problem from the fundamental *theorem of Caratheodory*:

Theorem 1.2 *If Ω is a space with an algebra $\mathcal{A}$ and with a σ -additive measure μ_0 on $(\Omega, \mathcal{A})$, then there exists a unique measure μ on $(\Omega, \sigma(\mathcal{A}))$ such that*

$$\mu(A) = \mu_0(A) \text{ for all } A \in \mathcal{A}.$$

Based on all these preparations and Theorems 1.1 and 1.2, one can arrive to the following natural *axioms proposed by Kolmogorov* for definition of the *probability space of the second level* (or simply, *probability space*):

An ordered triple $(\Omega, \mathcal{F}, P)$ with Ω as a set of elements ω , $\mathcal{F}$ as a σ -algebra on Ω , P as a probability measure (probability) on $\mathcal{F}$.

To go further, we give some other examples of measurable spaces playing an important role in many of our further constructions and statements.

Example 1.2 $(R^d, \mathcal{B}(R^d))$ is a multidimensional ($d \geq 1$) Borel space.

Take a set $I = \prod_{i=1}^d I_i$, $I_i = (a_i, b_i]$, $i = 1, \dots, d$, which is called a *rectangle with sides I_i* . The smallest σ -algebra containing all such sets is denoted $\mathcal{B}(R^d)$.

Instead of rectangles I , one can consider the sets $B = \prod_{i=1}^d B_i$, $B_i \in \mathcal{B}(R^1)$, $i = 1, 2, \dots, d$. The minimal σ -algebra containing all such sets is

$$\mathcal{B}(R^d) = \underbrace{\mathcal{B}(R^1) \otimes \dots \otimes \mathcal{B}(R^1)}_d.$$

Problem 1.3 Using an induction method, prove the above equality.

Example 1.3 Denote

$$R^\infty = \{x = (x_1, x_2, \dots), -\infty < x_i < +\infty\},$$

$$C(I) = C(I_1 \times \dots \times I_n) = \{x : x = (x_1, x_2, \dots), x_1 \in I_1, \dots, x_n \in I_n\}.$$

It is clear that the cylinders C satisfy the *consistency property*

$$C(I_1 \times \dots \times I_n) = C(I_1 \times \dots \times I_n \times R).$$

In this case, the system of the cylinders is an algebra. The smallest σ -algebra containing all cylinders is $\mathcal{B}(R^\infty)$, and we arrive to a measurable space $(R^\infty, \mathcal{B}(R^\infty))$.

Example 1.4 For $T = [0, \infty)$, we define R^T the space of all real-valued functions x , and $C_{t_1, \dots, t_n}(I_1 \times \dots \times I_n) = \{x : x_{t_1} \in I_1, \dots, x_{t_n} \in I_n\}$. The smallest σ -algebra containing all such cylinders is $\mathcal{B}(R^T)$. So, we arrive to the measurable space $(R^T, \mathcal{B}(R^T))$. Let us note a *structural property* of this space: For any $A \in \mathcal{B}(R^T)$, there exist $t_1 < t_2 < \dots$ and a Borel set $B \in \mathcal{B}(R^\infty)$ such that

$$A = \{x : (x_{t_1}, x_{t_2}, \dots) \in B\}.$$

1.2 Probabilistic Distributions and the Kolmogorov Consistency Theorem

After these theoretical considerations and the examples given above, we can put the problem:

How to construct a probability measure on given measurable space?

In the framework of Example 1.1, one can take a Borel set $A \in \mathcal{B}(R)$ in the form

$$A = (-\infty, x], \quad x \in R^1,$$

and define the function

$$F(x) = P((-\infty, x]),$$

where P is a probability measure on the Borel space.

We can note that the function $F = F(x)$ satisfies the following obvious properties:

1. F is nondecreasing.
2. $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$.
 $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$.
3. F is a right-continuous function with finite left limits.

Such a function, i.e., satisfying (1)–(3), is called a *distribution function*.

It is now clear that the problem of a construction of the probability can be solved if we have a distribution function F . In this case, we can define a finite-additive set function P_0 on algebra $\mathcal{A}$:

$$P_0(A) = \sum_{i=1}^n (F(b_i) - F(a_i)),$$

$$A = \sum_{i=1}^n (a_i, b_i] \in \mathcal{A}.$$

We can prove that P_0 is σ -additive, and according to the Caratheodory theorem P_0 admits a unique extension P to the whole space $(R^1, \mathcal{B}(R^1))$.

In particular, we can consider the measurable space $([0, 1], \mathcal{B}([0, 1]))$ and the distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ x, & x \in [0, 1], \\ 1, & x > 1. \end{cases}$$

Then the set function $l((a, b]) = F(b) - F(a) = b - a$ is the length of the interval $(a, b]$ presents the Lebesgue measure on this space.

Further, starting at the distribution function $F = F(x)$, which is a piecewise constant with change in values at points $x_1, x_2, \dots$, $\Delta F(x_i) > 0$, we define

$$p_i = P(\{x_i\}) = \Delta F(x_i) > 0, \quad \sum_{i=1}^{\infty} P(\{x_i\}) = 1.$$

This is a measure concentrated at $x_1, x_2, \dots$, and usually, the sequence $(p_1, p_2, \dots)$ is called a *discrete distribution*. Sometimes, the table

$$\begin{pmatrix} p_1 \dots p_n \dots \\ x_1 \dots x_n \dots \end{pmatrix}$$

is also called a *discrete distribution*. The corresponding distribution function F is also called *discrete*. Below we have the most well-known discrete distributions:

1. Uniform discrete distribution: $\begin{pmatrix} x_1 \dots x_N \\ 1/N \dots 1/N \end{pmatrix}$.
2. Bernoulli distribution: $\begin{pmatrix} x_1 & x_2 \\ p_1 & p_2 \end{pmatrix}$, $p_1 + p_2 = 1$.
3. Binomial distribution with parameter p : $\begin{pmatrix} 0 \dots i \dots n \\ \binom{n}{i} p^i (1-p)^{n-i} \end{pmatrix}$, $p \in (0, 1)$.
4. Poisson distribution with parameter λ : $\begin{pmatrix} 0 \dots i \dots n \dots \\ \frac{\lambda^i}{i!} e^{-\lambda} \end{pmatrix}$, $\lambda > 0$.

If the distribution function F admits the integral representation

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

where a nonnegative function (*density*) $f(y)$ satisfies the integral condition $\int_{-\infty}^{\infty} f(y) dy = 1$, then the distribution and its distribution function are *absolute continuous*.

Let us count the most important *absolute continuous distributions* as examples:

1. Uniform distribution on $[a, b]$:

$$f(y) = \frac{1}{b-a}, \quad y \in [a, b].$$

2. Normal distribution with parameters μ and σ^2 ($N \sim (\mu, \sigma^2)$) :

$$f(y) = (2\pi\sigma^2)^{-1/2} \exp \left\{ -\frac{(y-\mu)^2}{2\sigma^2} \right\}, \quad \mu \in \mathbb{R}^1, \sigma > 0, y \in \mathbb{R}^1.$$

3. Gamma distribution:

$$f(y) = \frac{y^{\alpha-1} e^{-y/\beta}}{\Gamma(\alpha)\beta^\alpha}, \quad y \geq 0, \alpha > 0, \beta > 0.$$

In particular, for $\beta = 1/\lambda$, and $\alpha = 1$, we get an *exponential distribution* with parameter

$$\lambda \geq 0: f(y) = \lambda e^{-\lambda y}, \quad y \geq 0;$$

for $\alpha = n/2$, and $\beta = 2$, we get the *Chi-squared distribution* with

$$f(y) = 2^{-n/2} y^{n/2-1} e^{-y/2} / \Gamma(n/2), \quad y \geq 0, n = 1, 2, \dots$$

4. Student distribution (t-distribution):

$$f(y) = \frac{\Gamma((n+1)/2)}{(n\pi)^{1/2} \Gamma(n/2)} (1 + y^2/n)^{-(n+1)/2}, \quad y \in \mathbb{R}^1, n = 1, 2, \dots$$

5. Cauchy distribution with parameter θ :

$$f(y) = \frac{\theta}{\pi(y^2 + \theta^2)}, \quad y \in \mathbb{R}^1, \theta > 0.$$

Besides discrete and absolute continuous distribution functions (measures, distributions), there are distribution functions which are continuous, but the *set of their points of increasing has the Lebesgue measure zero*. These types of distributions (measures) are called *singular*.

Let us explain it here with the help of the famous *Cantor function*. We define the functions

$$F_1(x) = \begin{cases} \text{linear function,} & \text{between } (0, 0) \text{ and } (1/3, 1/2), \\ 1/2, & x \in [1/3, 2/3], \\ \text{linear function,} & \text{between } (2/3, 1/2) \text{ and } (1, 1), \quad x \in [2/3, 1], \end{cases}$$

$$F_2(x) = \begin{cases} \text{linear function,} & \text{between } (0, 0) \text{ and } (1/9, 1/4), \ x \in [0, 1/9], \\ 1/4, & x \in [1/9, 2/9], \\ \text{linear function,} & \text{between } (2/9, 1/4) \text{ and } (1/3, 1/2), \ x \in (2/9, 1/3), \\ 1/2, & x \in [1/3, 2/3], \\ \text{On } (2/3, 1) & \text{is similar to the interval } (0, 1/3), \end{cases}$$

and so on. The sequence (F_n) converges to a nondecreasing continuous function $F_C(x)$, called the *Cantor function*. We can calculate the length of intervals on which $F_C(x)$ is constant and find

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \dots = \frac{1}{3} \sum_{n=0}^{\infty} \left(\frac{2}{3}\right)^n = 1.$$

So, the Lebesgue measure $l(\mathcal{N}) = 0$, where $\mathcal{N}$ is the set of points of increasing of F_C .

Denote μ the measure, generated by the Cantor function F_C , and find that $\mu(\mathcal{N}) = 1$ because $\mu(\mathcal{N}) = F_C(1) = 1$. It means that μ and l are *singular*, and this fact is denoted as $l \perp \mu$.

It is possible to give a *general description of arbitrary distribution functions* F :

$$F(x) = \alpha_1 F^1(x) + \alpha_2 F^2(x) + \alpha_3 F^3(x),$$

where $\alpha_i \geq 0$, $i = 1, 2$, and 3 , $\sum_{i=1}^3 \alpha_i = 1$, F^1 is discrete, F^2 is absolutely continuous, and F^3 is singular.

Let us pay our attention to a d -dimensional Borel space $(R^d, \mathcal{B}(R^d))$. If P is a probability measure on this space, we can define the d -dimensional distribution function:

$$F_d(x_1, \dots, x_d) = P((-\infty, x_1] \times \dots \times (-\infty, x_d]) = P((-\infty, x]).$$

Problem 1.4 Prove that:

1. $F_d(+\infty, \dots, +\infty) = 1$.
2. $F_d(x_1, \dots, x_d) \rightarrow 0$ if at least one of $x_1, \dots, x_d$ converges to $-\infty$.
3. Define operator (for $i = 1, \dots, d$, $a_i < b_i$)

$$\Delta_{a_i, b_i} F_d(x_1, \dots, x_d) = F_d(x_1, \dots, b_i, \dots, x_d) - F_d(x_1, \dots, a_i, \dots, x_d),$$

and then

$$\Delta_{a_1, b_1} \dots \Delta_{a_d, b_d} F_d(x_1, \dots, x_d) = P((a, b]),$$

where $(a, b] = (a_1, b_1] \times \dots \times (a_d, b_d]$.

In fact, there is a one-to-one correspondence between d -dimensional distribution functions and probabilities on the space $(R^d, \mathcal{B}(R^d))$.

Let us give *some examples of multidimensional distribution functions*:

1. If $F^1, \dots, F^d$ are the distribution functions on R^1 , then $F_d(x_1 \dots x_d) = \prod_{i=1}^d F^i(x_i)$ defines a d -dimensional distribution function.
2. If

$$F^i(x_i) = \begin{cases} 0, & x_i < 0, \\ x_i, & 0 \leq x_i \leq 1, \\ 1, & x_i > 1, \end{cases}$$

then $F_d(x_1, \dots, x_d) = x_1 \dots x_d$ corresponds to the d -dimensional Lebesgue measure.

3. A d -dimensional absolute continuous distribution function $F_d(x_1 \dots x_d)$ is defined by the same manner as in the real line case:

$$F_d(x_1, \dots, x_d) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_d} f_d(y_1 \dots y_d) dy_1 \dots dy_d,$$

where f_d is the density function, i.e., $f_d \geq 0$ and $\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_d(y_1 \dots y_d) dy_1 \dots dy_d = 1$.

The most important case is the *multidimensional Normal distribution function*. It is defined by the density

$$f_d(x_1 \dots x_d) = \frac{(\det A)^{1/2}}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2} \sum_{1 \leq i, j \leq d} a_{ij} (x_i - \mu_i)(x_j - \mu_j) \right\},$$

where $A = (a_{ij})$ is the inverse matrix of a symmetric positive definite matrix B .

In particular, in case $d = 2$, we have

$$f_2(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right] \right\},$$

where $\sigma_1, \sigma_2 > 0$, $\mu_1, \mu_2 \in R^1$, $|\rho| < 1$.

Now we need to consider the case of space $(R^\infty, \mathcal{B}(R^\infty))$. Let us take a set $B \in \mathcal{B}(R^n)$ and consider a cylinder

$$C_n(B) = \{x \in R^\infty : (x_1, \dots, x_n) \in B\}.$$

If P is a probability measure on $(R^\infty, \mathcal{B}(R^\infty))$, we can define $P_n(B) = P(C_n(B))$, $n = 1, 2, \dots$ a sequence of probability measures on space $(R^n, \mathcal{B}(R^n))$. By construction, we have

$$P_{n+1}(B \times R^1) = P_n(B),$$

which is called the *consistency condition*. Now we can formulate the *famous theorem of Kolmogorov* which is fundamental for foundations of probability theory.

Theorem 1.3 (Consistency Theorem of Kolmogorov) *Let $(P_n)_{n=1,2,\dots}$ be a system of probability measures on $(R^n, \mathcal{B}(R^n))$ correspondingly, $n = 1, 2, \dots$, satisfying the consistency property. Then there exists a unique probability measure P on $(R^\infty, \mathcal{B}(R^\infty))$ such that*

$$P(C_n(B)) = P_n(B).$$

We can give an example how to construct this sequence (P_n) . To do this, we start with the sequence of 1-dimensional distribution functions $(F^n(x))_{n=1,2,\dots}$, $x \in R^1$. Further, we construct another sequence of distribution functions as follows:

$$F_1(x_1) = F^1(x_1), \quad x_1 \in R^1; \quad F_2(x_1, x_2) = F^1(x_1) \cdot F^2(x_2), \quad x_1, x_2 \in R^1; \quad \text{etc.}$$

Denote $P_1, P_2, \dots$ probability measures on $(R^1, \mathcal{B}(R^1))$, $(R^2, \mathcal{B}(R^2))$, $\dots$, which correspond to the distribution functions $F_1, F_2, \dots$

We can also consider the space $(R^{[0,\infty)}, \mathcal{B}(R^{[0,\infty)}))$ for which one can state a *version of consistency theorem* in the same way as before. Using this theorem, one can construct an extremely important measure, called the *Wiener measure*.

Denote $(\phi_t(y|x))_{t \geq 0}$ a family of Normal (Gaussian) densities of y for a fixed x :

$$\phi_t(y|x) = \frac{1}{(2\pi t)^{1/2}} \exp \left\{ -\frac{(y-x)^2}{2t} \right\}, \quad y \in R^1.$$

For each $t_1 < t_2 < \dots < t_n$ and $B = \prod_{i=1}^n I_i$, $I_i = (a_i, b_i]$, $a_i < b_i$, we define the measure

$$P_{(t_1, t_2, \dots, t_n)}(B) = \int_{I_1} \dots \int_{I_n} \phi_{t_1}(x_1|0) \phi_{t_2-t_1}(x_2|x_1) \dots \phi_{t_n-t_{n-1}}(x_n|x_{n-1}) dx_1 \dots dx_n.$$

Furthermore, the measure P on cylinder sets can be defined as follows:

$$P(C_{t_1 \dots t_n}(I_1 \times \dots \times I_n)) = P_{(t_1 \dots t_n)}(I_1 \times \dots \times I_n).$$

The family of measures $(P_{(t_1 \dots t_n)})_{n=1,2,\dots}$ is consistent. Hence, according to Theorem 1.3, the measure P can be extended from cylinder sets to the whole space $(R^{[0,\infty)}, \mathcal{B}(R^{[0,\infty)}))$.

Chapter 2

Random Variables and Their Quantitative Characteristics



In the second chapter, random variables are introduced and investigated in the framework of axiomatic of Kolmogorov. It is shown a connection of probability distributions and distributions of random variables as well as their distribution functions. The notion of the Lebesgue integral is given in context of definition of moments of random variables (see [7, 13, 17, 28, 33, 37, 55], and [62]).

2.1 Distributions of Random Variables

P. L. Chebyshev was the first who introduced the notion of random variables as functions of elementary outcomes in the theory of probability. Here we develop this topic in the framework of probability spaces of the second level.

Definition 2.1 Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(R^1, \mathcal{B}(R^1))$ be a Borel space. Consider a mapping

$$X : (\Omega, \mathcal{F}) \rightarrow (R^1, \mathcal{B}(R^1)).$$

For a given set $B \in \mathcal{B}(R^1)$, the set $X^{-1}(B) = \{\omega : X(\omega) \in B\}$ is called the inverse image of B .

The mapping X is called a random variable (measurable function) if for any $B \in \mathcal{B}(R^1)$:

$$X^{-1}(B) \in \mathcal{F}. \quad (2.1)$$

Let us note that it is useful to allow to X can take values $\pm\infty$. In this case, we call a random variable as an *extended* random variable.

The notion of random variables is very productive as illustrated in the following example.

Example 2.1 Let us fix a set $A \in \mathcal{F}$ and define the *indicator* of A as follows:

$$I_A(\omega) = \begin{cases} 1, & \omega \in A, \\ 0, & \omega \in A^c. \end{cases}$$

Definitely, I_A is a mapping from the space $(\Omega, \mathcal{F})$ to $(R^1, \mathcal{B}(R^1))$. Then for any Borel set B , we have

$$\begin{cases} \{\omega : I_A(\omega) \in B\} = A \in \mathcal{F} \text{ if } B \text{ contains } 1 \text{ only,} \\ \{\omega : I_A(\omega) \in B\} = A^c \in \mathcal{F} \text{ if } B \text{ contains } 0 \text{ only;} \end{cases}$$

otherwise, we have $\emptyset$ or Ω . Hence, the condition (2.1) is fulfilled for I_A , and this mapping is a random variable taking two values 0 and 1. Therefore, the set of all random events is included to the set of random variables.

Problem 2.1 Assume $A_1, \dots, A_n$ are disjoint events from $\mathcal{F}$ such that $\bigcup_{i=1}^n A_i = \sum_{i=1}^n A_i = \Omega$. Define a mapping X with values $x_1, \dots, x_n$ on the sets $A_1, \dots, A_n$, correspondingly. Prove that $X = \sum_{i=1}^n x_i I_{A_i}$ is a random variable. Let us call X a *simple* random variable.

Further, if X takes countable numbers of values $x_1, \dots, x_n, \dots$ on disjoint sets $A_1, \dots, A_n, \dots$ with $\sum_1^\infty x_i I_{A_i}$, then $X = \sum_1^\infty x_i I_{A_i}$ is called a *discrete* random variable.

Condition (2.1) can be relaxed with the help of the next useful lemma.

Lemma 2.1 Assume $\mathcal{E}$ is a system of subsets such that $\sigma(\mathcal{E}) = \mathcal{B}(R^1)$. Then mapping X is a random variable if and only if

$$X^{-1}(E) \in \mathcal{F} \text{ for all } E \in \mathcal{E}.$$

Proof The direct implication is trivial. To prove the inverse implication, we define a system $\mathcal{Y}$ of those Borel sets Γ for which $X^{-1} \in \mathcal{F}$. Further, we have equalities

$$X^{-1}(\cup_\alpha \Gamma_\alpha) = \cup_\alpha (X^{-1}(\Gamma_\alpha)),$$

$$X^{-1}(\cap_\alpha \Gamma_\alpha) = \cap_\alpha (X^{-1}(\Gamma_\alpha)),$$

$$(X^{-1}(\Gamma_\alpha))^c = X^{-1}(\Gamma_\alpha^c).$$

It follows from here that $\mathcal{Y}$ is a σ -algebra and $\mathcal{E} \subseteq \mathcal{B}(R^1)$. Hence, $\mathcal{B}(R^1) = \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{Y}) = \mathcal{Y}$ and Lemma is proved. $\square$

Based on Lemma 2.1 and the previous description of the Borel space, we arrive to the corollary.

Corollary 2.1 *The function X is a random variable $\Leftrightarrow X^{-1}((-\infty, x]) \in \mathcal{F}$ for all $x \in \mathbb{R}^1$. This statement is true if we replace $(-\infty, x]$ with $(-\infty, x)$.*

Now we arrive to definition of the most important quantitative characteristic of random variable X .

Definition 2.2 Probability measure P_X on a Borel space $(\mathbb{R}^1, \mathcal{B}(\mathbb{R}^1))$ defined as

$$P_X(B) = P(X^{-1}(B)), \quad B \in \mathcal{B}(\mathbb{R}^1), \quad (2.2)$$

is called a *distribution (probability distribution)* of X .

Applying equality (2.2) to the set $B = (-\infty, x]$, $x \in \mathbb{R}^1$, we arrive to the *distribution function (cumulative distribution function)* of X :

$$F_X(x) = P\{\omega : X(\omega) \leq x\}. \quad (2.3)$$

In the previous sections, we already studied arbitrary distributions on $(\mathbb{R}^1, \mathcal{B}(\mathbb{R}^1))$. A very similar illustration can be found in the context of random variables.

Example 2.2 1. For *discrete* random variable X :

$$P_X(B) = \sum_{i: x_i \in B} p(x_i), \quad p(x_i) = P(\omega : X(\omega) = x_i) = \Delta F_X(x_i), \quad B \in \mathcal{B}(\mathbb{R}^1).$$

2. For *continuous* random variable X , its distribution function $F_X(x)$ is continuous.
3. For *absolute continuous* random variable X , its distribution function $F_X(x)$ has an integral representation with the density $f_X(x)$:

$$F_X(x) = \int_{-\infty}^x f_X(y) dy.$$

Usually, any measurable function

$$\varphi : (\mathbb{R}^1, \mathcal{B}(\mathbb{R}^1)) \rightarrow (\mathbb{R}^1, \mathcal{B}(\mathbb{R}^1))$$

is called Borelian. Using Borelian functions and a random variable, one can construct many other random variables. This is the essence of the next important lemma.

Lemma 2.2 *Let φ be a Borelian function and X be a random variable on a probability space $(\Omega, \mathcal{F}, P)$, and then the mapping*

$$Y = \varphi \circ X = \varphi(X),$$

is a random variable.

Proof For an arbitrary $B \in \mathcal{B}(R^1)$, we have

$$\begin{aligned}\{\omega : Y(\omega) \in B\} &= \{\omega : \varphi(X(\omega)) \in B\} = \\ &= \{\omega : X(\omega) \in \varphi^{-1}(B)\} \in \mathcal{F}.\end{aligned}$$

□

It means that Y is a random variable. In particular, the functions x^+ , x^- , $|x|$, etc., are random variables.

Let us note that there is a necessity to study *multidimensional* random variables or *random vectors*.

The corresponding definition is straightforward. This is a measurable mapping

$$X : (\Omega, \mathcal{F}) \rightarrow (R^d, \mathcal{B}(R^d)), \quad d \geq 1.$$

In this case, we have

$$X(\omega) = (X_1(\omega), \dots, X_d(\omega)),$$

where $X_i, i = 1, \dots, d$, are one-dimensional random variable. We can define a distribution function of X as follows:

$$F_X(x_1, \dots, x_d) = P(\omega : X_1(\omega) \leq x_1, \dots, X_d(\omega) \leq x_d).$$

Definition 2.3 We say that $X_1, \dots, X_d$ are independent if

$$F_X(x_1, \dots, x_d) = \prod_{i=1}^d F_{X_i}(x_i).$$

The following theorem is very important in providing many probabilistic constructions.

Theorem 2.1 *For any random variable (extended random variable) X , there exists a sequence of simple random variables $X_1, X_2, \dots, |X_i| \leq |X|$ such that $X_n(\omega) \rightarrow X(\omega)$ for all $\omega \in \Omega$. For a nonnegative X , such a sequence $(X_n)_{n=1,2,\dots}$ can be constructed as a nondecreasing sequence.*

Proof Assume $X \geq 0$ and define for $n = 1, 2, \dots$ simple random variables as follows:

$$X_n(\omega) = \sum_{i=1}^{n \cdot 2^n} \frac{i-1}{2^n} I_{i,n}(\omega) + n I_{\{X(\omega) \geq n\}}(\omega), \quad (2.4)$$

where $I_{i,n} = I_{\left\{\omega: \frac{i-1}{2^n} \leq X(\omega) < \frac{i}{2^n}\right\}}$.

The general case follows from here if we represent $X = X^+ - X^-$. $\square$

Problem 2.2 Prove that for (extended) random variables $X_1, X_2, \dots$ the mappings $\sup_n X_n, \inf_n X_n, \liminf_n X_n, \limsup_n X_n$ are (extended) random variables. As a *hint* to the corresponding solution, we note that for $\sup_n X_n$ we have

$$\{\omega : \sup_n X_n > x\} = \cup_n \{\omega : X_n > x\} \in \mathcal{F}, \quad x \in \mathbb{R}^1.$$

Theorem 2.2 Let $(X_n)_{n=1,2,\dots}$ be a sequence of (extended) random variables. Then $X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega)$ is a (extended) random variable.

Proof Follows from Problem 2.2 and the next equalities: For any $x \in \mathbb{R}^1$

$$\begin{aligned} \{\omega : X(\omega) < x\} &= \{\omega : \lim_{n \rightarrow \infty} X_n(\omega) < x\} \\ &= \{\omega : \limsup_n X_n(\omega) = \liminf_n X_n(\omega)\} \cap \{\omega : \limsup_n X_n(\omega) < x\} \\ &= \Omega \cap \{\omega : \limsup_n X_n(\omega) < x\} \\ &= \{\omega : \limsup_n X_n(\omega) < x\} \in \mathcal{F}. \end{aligned}$$

$\square$

Combining previous facts about limiting approximation of given (extended) random variables X and Y with the help of sequences of simple random variables $(X_n)_{n=1,2,\dots}$ and $(Y_n)_{n=1,2,\dots}$ correspondingly, we can prove the following convergence properties:

1. $\lim_{n \rightarrow \infty} (X_n(\omega) \pm Y_n(\omega)) = X(\omega) \pm Y(\omega), \omega \in \Omega.$
2. $\lim_{n \rightarrow \infty} (X_n(\omega) \cdot Y_n(\omega)) = X(\omega) \cdot Y(\omega), \omega \in \Omega.$
3. $\lim_{n \rightarrow \infty} (X_n(\omega) \cdot \tilde{Y}_n^{-1}(\omega)) = X(\omega) \cdot Y^{-1}(\omega), \omega \in \Omega,$ where $\tilde{Y}_n(\omega) = Y_n(\omega) + \frac{1}{n} I_{\{\omega: Y_n(\omega)=0\}}.$

Now, we focus on the question of how one random variable can be represented by another.

Definition 2.4 For a given random variable X , the family of events

$$((\omega : X(\omega) \in B))_{B \in \mathcal{B}(\mathbb{R}^1)}$$

is called a σ -algebra $\mathcal{F}_X$ generated by X .

Problem 2.3 Prove that $\mathcal{F}_X$ is a σ -algebra.

Theorem 2.3 Assume a random variable Y is measurable with respect to σ -algebra $\mathcal{F}_X$. Then there exists a Borelian function φ such that

$$Y = \varphi \circ X.$$

Proof Consider the set of $\mathcal{F}_X$ -measurable functions $Y = Y(\omega)$. Denote $\tilde{\mathcal{D}}_X$ the set of $\mathcal{F}_X$ -measurable functions of the form $\varphi \circ X$. It is clear that $\tilde{\mathcal{D}}_X \subseteq \mathcal{D}_X$. To prove the inverse inclusion, $\mathcal{D}_X \subseteq \tilde{\mathcal{D}}_X$, consider a set $A \in \mathcal{F}_X$ and $Y(\omega) = I_A(\omega)$.

Note that $A = X^{-1}(B)$ for some $B \in \mathcal{B}(R^1)$, and $Y = I_A(\omega) = I_B(X(\omega)) \in \tilde{\mathcal{D}}_X$.

Further, we consider functions Y of the form $\sum_{i=1}^n c_i I_{A_i}$, where $c_i \in R^1$, $A_i \in \mathcal{F}_X$, and find that $Y \in \tilde{\mathcal{D}}_X$.

For an arbitrary $\mathcal{F}_X$ -measurable function Y , we construct a sequence of simple $\mathcal{F}_X$ -measurable functions Y_n such that $Y = \lim_n Y_n$. We already know that $Y_n = \varphi_n(X)$ for some Borelian functions φ_n and $\varphi_n(X) \rightarrow Y(\omega)$ as $n \rightarrow \infty$. Now, we take the set $B = \{x : \lim_{n \rightarrow \infty} \varphi_n(x) \text{ exists}\} \in \mathcal{B}(R^1)$ and define a Borelian function

$$\varphi(x) = \begin{cases} \lim_n \varphi_n(x), & x \in B, \\ 0, & x \notin B. \end{cases}$$

Then $Y(\omega) = \lim_{n \rightarrow \infty} \varphi_n(X(\omega)) = \varphi(X(\omega))$ for all $\omega \in \Omega$, and we obtain $\tilde{\mathcal{D}}_X \subseteq \mathcal{D}_X$. $\square$

2.2 Expectations of Random Variables

In this section, we provide a construction of another important and convenient quantitative characteristic of a random variable X on a probability space $(\Omega, \mathcal{F}, P)$.

We start such a construction from the simplest case when

$$X(\omega) = \sum_{i=1}^n x_i \cdot I_{A_i}(\omega), \quad \sum_{i=1}^n A_i = \Omega, \quad x_i \in R^1. \quad (2.5)$$

Definition 2.5 For the simple random variable X with representation (2.5), we define its *expected value (expectation)* $\mathbf{E}X$ as follows:

$$\mathbf{E}X = \sum_{i=1}^n x_i P(A_i).$$

Problem 2.4

- For simple random variables $X_1, \dots, X_m$ and real numbers $a_1, \dots, a_m$, we have $\mathbf{E} \sum_{i=1}^m a_i X_i = \sum_{i=1}^m a_i \mathbf{E}X_i$ (linearity of expected values).
- For two simple random variables $X \leq Y$, we have $\mathbf{E}X \leq \mathbf{E}Y$ (monotonicity of expected values).

Let X be a nonnegative random variable. According to Theorem 2.1, one can construct a sequence of simple random variables $X_n^{(\omega)} \rightarrow X(\omega)$, $n \rightarrow \infty$, for all

$\omega \in \Omega$. Further, due to Problem 2.4, the sequence of expected values $(\mathbf{E}X_n)_{n=1,2,\dots}$ is nondecreasing, and therefore, $\lim_{n \rightarrow \infty} \mathbf{E}X_n$ does exist (finite number or $+\infty$). It gives a possibility to give the following definition.

Definition 2.6 The expected value (expectation) of a nonnegative random variable X is defined as

$$\mathbf{E}X = \lim_{n \rightarrow \infty} \mathbf{E}X_n.$$

Problem 2.5

1. If there exists another sequence of simple random variables $\tilde{X}_n(\omega) \uparrow X(\omega)$, $n \uparrow \infty$, $\omega \in \Omega$, then $\lim_{n \rightarrow \infty} \mathbf{E}\tilde{X}_n = \mathbf{E}X$.
2. If $0 \leq X(\omega) \leq Y(\omega)$ for all $\omega \in \Omega$, then $\mathbf{E}X \leq \mathbf{E}Y$.

There is a standard way of expected value extensions to random variables taking both positive and negative values. We can represent X as difference of its positive and negative parts: $X = X^+ - X^-$, where as usual $X^+ = \max(0, X)$ and $X^- = \max(-X, 0)$. In this case, we say that $\mathbf{E}X$ *does exist* if $\mathbf{E}X^+$ or $\mathbf{E}X^- < \infty$ and $\mathbf{E}X = \mathbf{E}X^+ - \mathbf{E}X^-$.

One can say that $\mathbf{E}X$ is *finite* if both $\mathbf{E}X^\pm < \infty$, and $\mathbf{E}|X| = \mathbf{E}(X^+ + X^-) = \mathbf{E}X^+ + \mathbf{E}X^- < \infty$. In this case, we also call X *integrable*.

There is another denotation for expected values:

$$\mathbf{E}X = \int_{\Omega} X dP.$$

This denotation came from functional analysis, where the notion of expected value of a measurable function X is called the *Lebesgue integral*. As a matter of fact, we note that the Lebesgue integral generalizes the Riemann integral at least in two directions:

1. It is constructed on the measurable space without any metric structure.
2. It is well-defined for any measurable bounded function X . But construction of the Riemann integral and its existence depends on the power of the set of discontinuity $\mathcal{D}_X$ of X , and the set $\mathcal{D}_X$ should have the Lebesgue measure zero. A standard example of function “integrable by Lebesgue and non-integrable by Riemann” is the Dirichlet function on $[0, 1]$:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega \in Q, \\ 0 & \text{if } \omega \in R \setminus Q, \end{cases}$$

where Q is the set of rational numbers.

Let us formulate several natural properties of expected values as a problem because their proofs are straightforward.

Problem 2.6

1. $\mathbf{E}c \cdot X = c\mathbf{E}X$ if $\mathbf{E}X$ does exist and $c \in \mathbb{R}$.
2. $\mathbf{E}X \leq \mathbf{E}Y$ if $X(\omega) \leq Y(\omega)$, $\omega \in \Omega$, $\mathbf{E}X > -\infty$ or $\mathbf{E}Y < \infty$.
3. $|\mathbf{E}X| \leq \mathbf{E}|X|$ if $\mathbf{E}X$ does exist.
4. If $\mathbf{E}X$ exists, then $\mathbf{E}X \cdot I_A$ exists for each $A \in \mathcal{F}$. Moreover, if $\mathbf{E}X$ is finite, then $\mathbf{E}XI_A$ is finite too.
5. If X and Y are nonnegative or integrable, then $\mathbf{E}(X + Y) = \mathbf{E}X + \mathbf{E}Y$.

To formulate other important properties of expectations, we need the next definition.

Definition 2.7 We say that a property holds *almost surely* (a.s.) if there exists a set $\mathcal{N} \in \mathcal{F}$ such that:

- (a) The property holds for every element $\omega \in \Omega \setminus \mathcal{N}$.
 - (b) $P(\mathcal{N}) = 0$.
6. If $X = 0$ (a.s.), then $\mathbf{E}X = 0$.
 7. If $X = Y$ (a.s.) and $\mathbf{E}|X| < \infty$, then $\mathbf{E}|Y| < \infty$ and $\mathbf{E}X = \mathbf{E}Y$.
 8. If $X \geq 0$ and $\mathbf{E}X = 0$, then $X = 0$ (a.s.).
 9. If X and Y are integrable and $\mathbf{E}X \cdot I_A \leq \mathbf{E}Y \cdot I_A$ for all $A \in \mathcal{F}$, then $X \leq Y$ (a.s.).

Let us give a *solution* of subproblem 8 in the above list of Problem 2.6:

Denote $A = \{\omega : X(\omega) > 0\}$ and $A_n = \{\omega : X(\omega) \geq \frac{1}{n}\} \uparrow A$ as $n \rightarrow \infty$.

We note that $0 \leq XI_{A_n} \leq XI_A$, and hence, $\mathbf{E}XI_{A_n} \leq \mathbf{E}X = 0$. Next, we have $0 = \mathbf{E}XI_{A_n} \geq \frac{1}{n}P(A_n)$. Hence, $P(A_n) = 0$, $n = 1, 2, \dots$, and $P(A) = \lim_{n \rightarrow \infty} P(A_n) = 0$.

Another useful property is the *formula of change of variables* in the Lebesgue integral. Let X be a random variable and ϕ be a Borelian function which is integrable with respect to the distribution P_X of X . Then for any $B \in \mathcal{B}(\mathbb{R})$, the formula of change of variables is true:

$$\int_B \phi(x) dP_X = \int_{X^{-1}(B)} \phi(X(\omega)) dP. \quad (2.6)$$

In particular, the formula (2.6) for $B = \mathbb{R}^1$ is reduced to the formula for calculation of expectation of X :

$$\mathbf{E}\phi(x) = \int_{\Omega} \phi(X) dP = \int_{-\infty}^{\infty} \phi(x) dP_X = \int_{-\infty}^{\infty} \phi(x) dF_X, \quad (2.7)$$

where F_X is a distribution function of X and the integral $\int_{-\infty}^{\infty} \phi(x) dF_X$ is the Lebesgue-Stieltjes integral.

Scheme of the Proof of the formula (2.6) is below.

Take $\phi(x) = I_C(x)$, $C \in \mathcal{B}(R^1)$, and (2.6) is reduced to $P_X(B \cap C) = P(X^{-1}(B) \cap X^{-1}(C))$, which follows from the definition of P_X and the equality:

$$X^{-1}(B) \cap X^{-1}(C) = X^{-1}(B \cap C).$$

Next steps are obvious: a nonnegative simple function ϕ , etc. □

If $\phi(x) = x^k$, $k = 1, 2, \dots$, the expectation $\mathbf{E}X^k$ is called the k -th moment of X with the help of its distribution and its distribution function. Suppose $\mathbf{E}X = \mu$ and $\phi(x) = (x - \mu)^k$, and then corresponding moments are called *centered moments*. The second centered moment is called the *variance* of X :

$$\text{Var}(X) = \mathbf{E}(X - \mu)^2,$$

and it is one of the key measures of the dispersion of values of X around the mean value μ .

The other common measures of this type are

$$\text{skewness} = \frac{\mathbf{E}(X - \mu)^3}{(\text{Var}(X))^{3/2}}$$

and

$$\text{kurtosis} = \frac{\mathbf{E}(X - \mu)^4}{(\text{Var}(X))^2}.$$

Problem 2.7 Let X be a normal random variable with parameters μ and σ^2 . Find $\mathbf{E}X$, $\text{Var}(X)$, *skewness* and *kurtosis* of X .

Chapter 3

Expectations and Convergence of Sequences of Random Variables



In the third chapter, asymptotic properties of sequences of random variables are studied. Lemma of Fatou and the Lebesgue dominated convergence theorem are presented as permanent technical tools of stochastic analysis. It is also emphasized the role of a uniform integrability condition of families of random variables. Classical probabilistic inequalities of Chebyshev, Jensen, and Cauchy-Schwartz are proved. It is shown how these inequalities work to investigate interconnections between different types of convergence of sequences of random variables. In particular, the large numbers law (LNL) is derived for the case of independent identically distributed random variables (see [7, 13, 17, 28, 33, 55], and [62]).

3.1 Limit Behavior of Sequences of Random Variables in Terms of Their Expected Values

One of the main questions here is how to take the limit under the sign of expectation. The first result exploits a monotonicity assumption. That is why the corresponding claim is called *monotonicity convergence theorem*.

Theorem 3.1 *Let $(X_n)_{n=1,2,\dots}$ be a sequence of random variables; then we have:*

1. *If there are random variables X and Y such that $X_n \geq Y$, $\mathbf{E}Y > -\infty$, $X_n \uparrow X$, $n \uparrow \infty$, then $\mathbf{E}X_n \uparrow \mathbf{E}X$, $n \uparrow \infty$.*
2. *If there are random variables X and Y such that $X_n \leq Y$, $\mathbf{E}Y < \infty$, $X_n \downarrow X$, $n \uparrow \infty$, then $\mathbf{E}X_n \downarrow \mathbf{E}X$, $n \uparrow \infty$.*

Proof Let us only prove the first part of the theorem because the second part can be derived in the same way. Consider only case $Y \geq 0$, and approximate X_i for

each $i = 1, 2, \dots$ by a sequence $(X_i^n)_{n=1,2,\dots}$ of simple random variables. Define $X^n = \max_{1 \leq i \leq n} X_i^n$ and note that

$$X^{n-1} \leq X^n = \max_{1 \leq i \leq n} X_i^n \leq \max_{1 \leq i \leq n} X_i = X_n.$$

Denote $Z = \lim_{n \rightarrow \infty} X^n$ and find that $X = Z$ because $X_i \leq Z \leq X$ for all $i = 1, 2, \dots$. This inequality follows from $X_i^n \leq X^n \leq X_n$, $i = 1, \dots, n$, by taking the limit as $n \rightarrow \infty$. The random variables $(X^n)_{n=1,2,\dots}$ are simple, and then

$$\mathbf{E}X = \mathbf{E}Z = \lim_{n \rightarrow \infty} \mathbf{E}X^n \leq \lim_{n \rightarrow \infty} \mathbf{E}X_n.$$

On the other hand, since $X_n \leq X_{n+1} \leq X$, we have that

$$\mathbf{E}X \geq \lim_{n \rightarrow \infty} \mathbf{E}X_n,$$

and hence,

$$\mathbf{E}X = \lim_{n \rightarrow \infty} \mathbf{E}X_n.$$

□

Problem 3.1 Give the proof in general case when $X_n \geq Y$ and $\mathbf{E}Y > -\infty$.

The next theorem is the most exploited result about taking the limit under the expectation sign. It is called the *Fatou Lemma*.

Theorem 3.2

1. Assume $X_n \geq Y$ and $\mathbf{E}Y > -\infty$, and then

$$\mathbf{E} \liminf_{n \rightarrow \infty} X_n \leq \liminf_{n \rightarrow \infty} \mathbf{E}X_n.$$

2. Assume $X_n \leq Y$ and $\mathbf{E}Y < \infty$, and then

$$\limsup_{n \rightarrow \infty} \mathbf{E}X_n \leq \mathbf{E} \limsup_{n \rightarrow \infty} X_n.$$

3. Assume $|X_n| \leq Y$ and $\mathbf{E}Y < \infty$, and then

$$\mathbf{E} \liminf_{n \rightarrow \infty} X_n \leq \liminf_{n \rightarrow \infty} \mathbf{E}X_n \leq \limsup_{n \rightarrow \infty} \mathbf{E}X_n \leq \mathbf{E} \limsup_{n \rightarrow \infty} X_n.$$

Proof In case (1), we define a sequence of random variables $Z_n = \inf_{m \geq n} X_m$ and find that $Z_n \uparrow \liminf_{n \rightarrow \infty} X_n = \sup \inf_{m \geq n} X_m$ and $Z_n \geq Y$. Further we have that

$$\liminf_{n \rightarrow \infty} X_n = \liminf_{m \geq n} X_m = \lim_{n \rightarrow \infty} Z_n,$$

and applying Theorem 3.1 we get the first claim of the theorem.

The case (2) is derived in a similar way. The case (3) is just a combination of (1) and (2). $\square$

Let us give the following definition of *convergence of almost surely (a.s.)*.

Definition 3.1 We say that sequence $(X_n)_{n=1,2,\dots}$ converges to a random variable X almost surely (a.s.), if

$$P(\omega : X_n(\omega) \xrightarrow[n \rightarrow \infty]{} X(\omega)) = 1.$$

We introduce the following denotations in this case: $X_n \xrightarrow[n \rightarrow \infty]{a.s.} X$ or $X_n \xrightarrow[n \rightarrow \infty]{} X$ (a.s.).

The following natural question arises here:

Under what conditions, one can provide convergence of expected values if $X_n \rightarrow X$ (a.s.)?

The classical result is the *dominated convergence theorem of Lebesgue*.

Theorem 3.3 Assume that $X_n \xrightarrow[n \rightarrow \infty]{} X$ (a.s.) and $|X_n| \leq Y$, $\mathbf{E}Y < \infty$. Then $\mathbf{E}|X| < \infty$, and $\mathbf{E}X_n \rightarrow \mathbf{E}X$, $n \rightarrow \infty$. Moreover, the sequence $(X_n)_{n=1,2,\dots}$ converges to X in space L^1 , i.e., $\mathbf{E}|X_n - X| \rightarrow 0$, $n \rightarrow \infty$.

Proof Using the Fatou lemma, we obtain $\mathbf{E} \liminf_{n \rightarrow \infty} X_n = \liminf_{n \rightarrow \infty} \mathbf{E}X_n = \limsup_{n \rightarrow \infty} \mathbf{E}X_n = \mathbf{E} \limsup_{n \rightarrow \infty} X_n = \mathbf{E}X$. It is clear that $|X| \leq Y$, and therefore $\mathbf{E}|X| < \infty$. Taking into account inequality $|X_n - X| \leq 2Y$ and applying the Fatou lemma again, we obtain the last statement of the theorem. $\square$

Now we introduce the weakest condition to provide the convergence of expected values.

Definition 3.2 A sequence of random variables $(X_n)_{n=1,2,\dots}$ is *uniformly integrable*, if

$$\sup_n \mathbf{E}|X_n| I_{\{\omega: |X_n| > c\}} \rightarrow 0, \quad c \rightarrow \infty. \quad (3.1)$$

Problem 3.2 If a sequence of $|X_n| \leq Y$, $n = 1, 2, \dots$, $\mathbf{E}Y < \infty$, then $(X_n)_{n=1,2,\dots}$ is uniformly integrable.

Theorem 3.4 If a sequence $(X_n)_{n=1,2,\dots}$ is uniformly integrable, then:

$$1. \quad \mathbf{E} \liminf_{n \rightarrow \infty} X_n \leq \liminf_{n \rightarrow \infty} \mathbf{E}X_n \leq \limsup_{n \rightarrow \infty} \mathbf{E}X_n \leq \mathbf{E} \limsup_{n \rightarrow \infty} X_n.$$

2. If also $X_n \rightarrow X$ (a.s.), $n \rightarrow \infty$, then $\mathbf{E}|X| < \infty$ and $\mathbf{E}X_n \rightarrow \mathbf{E}X$ and $\mathbf{E}|X_n - X| \rightarrow 0$, $n \rightarrow \infty$.

Proof To prove (1), we take $c > 0$ and find that

$$\mathbf{E}X_n = \mathbf{E}X_n I_{\{X_n < -c\}} + \mathbf{E}X_n I_{\{X_n \geq -c\}}. \quad (3.2)$$

Using (3.1) for $\epsilon > 0$, we obtain

$$\sup_n |\mathbf{E}X_n I_{\{X_n < -c\}}| < \epsilon.$$

Let us apply the Fatou lemma to the second term of (2) and get

$$\liminf_n \mathbf{E}X_n I_{\{X_n \geq -c\}} \geq \mathbf{E} \liminf_n X_n I_{\{X_n \geq -c\}}. \quad (3.3)$$

Taking into account that

$$X_n I_{\{X_n \geq -c\}} \geq X_n,$$

we obtain from (3.3) the following inequality

$$\liminf_n \mathbf{E}X_n \geq \mathbf{E} \liminf_n X_n - \epsilon.$$

The proof of part (2) is similar to the proof of Theorem 3.3. □

Theorem 3.5 Assume the sequence of integrable nonnegative random variables $(X_n)_{n=1,2,\dots}$ converges to a random variable X . Then $\mathbf{E}X_n \rightarrow \mathbf{E}X < \infty$, $n \rightarrow \infty$, if and only if $(X_n)_{n=1,2,\dots}$ is uniformly integrable.

Proof The inverse implication follows from Theorem 3.4. To prove the direct implication, we consider the set $A = \{a \in \mathbb{R}^1 : P(\omega : X(\omega) = a) > 0\}$ and find that

$$X_n I_{\{X_n < a\}} \rightarrow X I_{\{X < a\}}, \quad n \rightarrow \infty,$$

for each $a \in A$ and $(X_n I_{\{X_n < a\}})_{n=1,2,\dots}$ is uniformly integrable.

By Theorem 3.4, we obtain

$$\mathbf{E}X_n I_{\{X_n < a\}} \rightarrow \mathbf{E}X I_{\{X < a\}},$$

for $a \in A$, and hence,

$$\mathbf{E}X_n I_{\{X_n \geq a\}} \rightarrow \mathbf{E}X I_{\{X \geq a\}},$$

for $a \notin A$.

For each $\epsilon > 0$, we take $a_0 \notin A$ large enough that

$$\mathbf{E}X I_{\{X \geq a_0\}} < \epsilon/2.$$

Choose N_0 so large to provide inequalities (for all $n \geq N_0$):

$$\mathbf{E}X_n I_{\{X_n \geq a_0\}} \leq \mathbf{E}X I_{\{X \geq a_0\}} + \epsilon/2,$$

and $\mathbf{E}X_n I_{\{X_n \geq a_0\}} \leq \epsilon$. Now we choose $a_1 \geq a_0$ large enough to have the following inequality (for all $n \leq N_0$):

$$\mathbf{E}X_n I_{\{X_n \geq a_1\}} \leq \epsilon.$$

Finally, we obtain

$$\sup_n \mathbf{E}X_n I_{\{X_n \geq a_1\}} \leq \epsilon,$$

which means the condition (3.1). □

Let us note the following *necessary condition* of (3.1).

Problem 3.3 If the sequence $(X_n)_{n=1,2,\dots}$ is uniformly integrable, then

$$\sup_n \mathbf{E}|X_n| < \infty.$$

The following problem provides a convenient *sufficient condition* for (3.1).

Problem 3.4 Let $(X_n)_{n=1,2,\dots}$ be a sequence of integrable random variables and $G = G(u)$, $u \geq 0$, be a nonnegative increasing function such that

$$\lim_{u \rightarrow \infty} \frac{G(u)}{u} = \infty, \quad \sup_n \mathbf{E}G(|X_n|) < \infty.$$

Then $(X_n)_{n=1,2,\dots}$ is uniformly integrable.

Hint. Let $\epsilon > 0$ and $G = \sup_n \mathbf{E}G(X_n)$. For large enough c and $t \geq c$, $G(t)/t \geq G/\epsilon$. Then $\mathbf{E}|X_n| I_{\{|X_n| \geq c\}} \leq \frac{\epsilon}{G} \mathbf{E}G(|X_n|) I_{\{|X_n| \geq c\}} \leq \epsilon$.

To conclude our discussion of the notion of uniform integrability, we represent a necessary and sufficient condition in the following theorem.

Theorem 3.6 A sequence of random variables $(X_n)_{n=1,2,\dots}$ is uniformly integrable if and only if (3.3) is true and

$$\sup_n \mathbf{E}|X_n| I_A \rightarrow 0 \text{ as } P(A) \rightarrow 0. \quad (3.4)$$

Proof The direct implication follows from the next considerations. For $c > 0$, we have

$$\begin{aligned}\mathbf{E}|X_n|I_A &= \mathbf{E}|X_n|I_{A \cap \{|X_n| \geq c\}} + \mathbf{E}|X_n|I_{A \cap \{|X_n| < c\}} \\ &\leq \mathbf{E}|X_n|I_{\{|X_n| \geq c\}} + cP(A).\end{aligned}$$

Taking c large enough, we obtain for a fixed $\epsilon > 0$:

$$\sup_n \mathbf{E}|X_n|I_{A \cap \{|X_n| \geq c\}} \leq \epsilon/2. \quad (3.5)$$

From (3.5) with $P(A) \leq \epsilon/(2c)$, we get

$$\sup_n \mathbf{E}|X_n|I_A \leq \epsilon,$$

which means (3.4). To prove the inverse implication, we take $P(A) < \delta$ so that $\mathbf{E}|X_n|I_A \leq \epsilon$ uniformly over $n = 1, 2, \dots$. Let us note that

$$\mathbf{E}|X_n| \geq \mathbf{E}|X_n|I_{A \cap \{|X_n| \geq c\}} \geq cP(|X_n| \geq c).$$

Then

$$\sup_n P(|X_n| \geq c) \leq \frac{1}{c} \sup_n \mathbf{E}|X_n| \rightarrow 0, \quad c \rightarrow \infty.$$

So, we can take as $A = \{|X_n| \geq c\}$ for large enough c and get

$$\sup_n \mathbf{E}|X_n|I_{\{|X_n| \geq c\}} \leq \epsilon,$$

which is equivalent to (3.1). □

3.2 Probabilistic Inequalities and Interconnections Between Types of Convergence of Random Variables

We start with some important and well-known probabilistic inequalities.

Chebyshev inequality: For a nonnegative random variable X ,

$$P(\omega : X(\omega) \geq \epsilon) \leq \frac{\mathbf{E}X}{\epsilon}, \quad \epsilon > 0.$$

In particular, for arbitrary random variable X with finite $\mathbf{E}X$ and $\text{Var}(X)$:

$$P(\omega : X(\omega) \geq \epsilon) \leq \frac{\mathbf{E}X^2}{\epsilon^2} \quad \text{and}$$

$$P(\omega : |X(\omega) - \mathbf{E}X| \geq \epsilon) \leq \frac{\text{Var}(X)}{\epsilon^2}.$$

For the proof, we just note that

$$\mathbf{E}X \geq \mathbf{E}X I_{\{X \geq \epsilon\}} \geq \epsilon \mathbf{E}I_{\{X \geq \epsilon\}} = \epsilon P(\omega : X \geq \epsilon).$$

Cauchy-Schwartz Inequality Let X and Y be random variables which are integrable in square: $\mathbf{E}X^2 < \infty$, $\mathbf{E}Y^2 < \infty$, i.e., $X, Y \in L^2$. Then $\mathbf{E}|XY| < \infty$ and

$$\mathbf{E}|XY| \leq (\mathbf{E}X^2 \mathbf{E}Y^2)^{1/2}.$$

Proof Without loss of generality, we assume that $\mathbf{E}X^2 > 0$ and $\mathbf{E}Y^2 > 0$. Consider transformed random variables

$$\tilde{X} = X/(\mathbf{E}X^2)^{1/2}, \quad \tilde{Y} = Y/(\mathbf{E}Y^2)^{1/2}.$$

Note that

$$2|\tilde{X}\tilde{Y}| \leq \tilde{X}^2 + \tilde{Y}^2,$$

and obtain

$$2\mathbf{E}|\tilde{X}\tilde{Y}| \leq \mathbf{E}\tilde{X}^2 + \mathbf{E}\tilde{Y}^2 = 2,$$

or $\mathbf{E}|\tilde{X}\tilde{Y}| \leq 1$. □

Jensen Inequality Let a Borelian function $g = g(x)$ be convex downward and $\mathbf{E}g(|X|) < \infty$. Then

$$g(\mathbf{E}X) \leq \mathbf{E}g(X).$$

Proof Let us consider only the case of smooth function $g \in C^2$ with $g'(x) \geq 0$ for all $x \in R^1$. Using the Taylor decomposition at $\mu = \mathbf{E}X$, we have

$$g(x) = g(\mu) + g'(\mu)(x - \mu) + \frac{g''(\theta)(x - \mu)^2}{2}, \quad (3.6)$$

where θ is between x and μ .

Putting $x = X$ and taking expectation in (3.6), we get the desirable inequality. □

In context of the Cauchy-Schwartz inequality, we want to emphasize one important case when the inequality is transformed into equality.

Theorem 3.7 *Let X and Y be integrable independent random variables, i.e., $F_{XY}(x, y) = F_X(x) \cdot F_Y(y)$. Then $\mathbf{E}|XY| < \infty$ and*

$$\mathbf{E}XY = \mathbf{E}X\mathbf{E}Y. \quad (3.7)$$

Proof We start with nonnegative X and Y , constructing sequences $(X_n)_{n=1,2,\dots}$ and $(Y_n)_{n=1,2,\dots}$ of discrete random variables such that

$$X_n = \sum_{m=0}^{\infty} \frac{m}{n} I_{\{\frac{m}{n} \leq X(\omega) < \frac{m+1}{n}\}}, \quad n = 1, 2, \dots$$

(Y_n is similar)

$$X_n \leq X, \quad Y_n \leq Y, \quad |X_n - X| \leq \frac{1}{n}, \quad |Y_n - Y| \leq \frac{1}{n}.$$

Since X and Y are integrable, we get that $\mathbf{E}X_n \rightarrow \mathbf{E}X$, $\mathbf{E}Y_n \rightarrow \mathbf{E}Y$, $n \rightarrow \infty$, by the Lebesgue dominated convergence theorem. Due to the independence assumption, we have

$$\begin{aligned} \mathbf{E}X_n Y_n &= \sum_{m,l} \frac{ml}{n^2} \mathbf{E}I_{\{\frac{m}{n} \leq X(\omega) < \frac{m+1}{n}\}} I_{\{\frac{l}{n} \leq Y(\omega) < \frac{l+1}{n}\}} \\ &= \sum_{m,l} \frac{ml}{n^2} \mathbf{E}I_{\{\frac{m}{n} \leq X(\omega) < \frac{m+1}{n}\}} \mathbf{E}I_{\{\frac{l}{n} \leq Y(\omega) < \frac{l+1}{n}\}} \\ &= \mathbf{E}X_n \mathbf{E}Y_n. \end{aligned}$$

Let us note that for $n = 1, 2, \dots$

$$\begin{aligned} |\mathbf{E}XY - \mathbf{E}X_n Y_n| &\leq \mathbf{E}|XY - X_n Y_n| \\ &\leq \mathbf{E}|X||Y - Y_n| + \mathbf{E}|Y_n||X - X_n| \\ &\leq \frac{\mathbf{E}X}{n} + \frac{\mathbf{E}(Y + 1/n)}{n} \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Therefore,

$$\mathbf{E}XY = \lim_n \mathbf{E}X_n Y_n = \lim_n \mathbf{E}X_n \lim_n \mathbf{E}Y_n = \mathbf{E}X\mathbf{E}Y < \infty.$$

General case (3.7) can be treated in a similar way if we note equalities: $X = X^+ - X^-$, $Y = Y^+ - Y^-$, $XY = X^+Y^+ - X^-Y^+ - X^+Y^- + X^-Y^-$. $\square$

The brilliant Chebyshev inequality has a lot of applications, and one of the most important corollaries is the large numbers law (LNL).

Theorem 3.8 *Let $(Y_n)_{n=1,2,\dots}$ be a sequence of identically distributed independent random variables with mean μ and variance σ^2 , and $S_n = \sum_{i=1}^n Y_i$. Then for any $\epsilon > 0$*

$$P\left(\omega : \left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \rightarrow 0, \quad n \rightarrow \infty. \quad (3.8)$$

The proof of (3.8) follows from the Chebyshev inequality and Theorem 3.6. Denoting $\tilde{Y}_i = Y_i - \mu$, $i = 1, 2, \dots$, we calculate

$$\begin{aligned} \mathbf{E}\left(\frac{S_n}{n} - \mu\right)^2 &= \frac{1}{n^2} \mathbf{E}\left(\sum_{i=1}^n \tilde{Y}_i\right)^2 \\ &= \frac{1}{n^2} \left[\mathbf{E} \sum_{i=1}^n \tilde{Y}_i^2 + \sum_{i \neq j} \mathbf{E} \tilde{Y}_i \tilde{Y}_j \right] \\ &= \frac{1}{n^2} \left[n\sigma^2 + \sum_{i \neq j} \mathbf{E} \tilde{Y}_i \tilde{Y}_j \right] = \frac{\sigma^2}{n} \end{aligned}$$

and obtain that

$$P\left(\omega : \left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\mathbf{E}\left(\frac{S_n}{n} - \mu\right)^2}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0, \quad n \rightarrow \infty.$$

Let us give the following definition.

Definition 3.3 A sequence of random variables $(X_n)_{n=1,2,\dots}$ converges to a random variable X in probability, if for any $\epsilon > 0$

$$P(\omega : |X_n - X| \geq \epsilon) \rightarrow 0, \quad n \rightarrow \infty. \quad (3.9)$$

We denote it as follows $X_n \xrightarrow[n]{P} X$ or $X_n \xrightarrow{P} X$, $n \rightarrow \infty$.

Using this definition, one can reformulate Theorem 3.8 as convergence in probability of normed sums $X_n = S_n/n$ to the constant μ as $n \rightarrow \infty$.

We also introduce two types of convergence of random variables $(X_n)_{n=1,2,\dots}$ to a random variable X involving expected values.

Definition 3.4 We say that (X_n) converges to X *weakly* (*weak convergence*), if for any bounded continuous function f :

$$\mathbf{E}f(X_n) \rightarrow \mathbf{E}f(X), \quad n \rightarrow \infty. \quad (3.10)$$

Definition 3.5 Assume X_n , $n = 1, 2, \dots$ and X belongs to the space L^p with finite p -moment, i.e., $\mathbf{E}|X_n|^p < \infty$, $\mathbf{E}|X|^p < \infty$. We say that X_n converges to X in space L^p , if $\mathbf{E}|X_n - X|^p \rightarrow 0$, $n \rightarrow \infty$. This convergence is denoted as $X_n \xrightarrow[n]{L^p}$ or $X_n \xrightarrow{L^p}, n \rightarrow \infty$.

Now we are ready to discuss how the convergence (a.s.) (or convergence with probability 1), defined before and these convergences are related to each other.

It is not difficult to construct a sequence $(X_n)_{n=1,2,\dots}$ of random variables which converges to X for *all* $\omega \in \Omega$, and hence, almost surely.

For example, if X is a random variable and $(a_n)_{n=1,2,\dots}$ is a sequence of positive numbers converges to zero, then a sequence $X_n(\omega) = (1 - a_n)X(\omega)$ converges to X for all $\omega \in \Omega$.

Further, convergence with probability one provides both convergence in probability and the weak convergence.

If $X_n \rightarrow X$ (a.s.), $n \rightarrow \infty$, then $|X_n - X| \rightarrow 0$ (a.s.), $n \rightarrow \infty$, and for $\epsilon > 0$ $I_{\{\omega: |X_n - X| \geq \epsilon\}} \rightarrow 0$ (a.s.) as well. Taking expectation, we obtain $P(\omega : |X_n - X| \geq \epsilon) = \mathbf{E}I_{\{\omega: |X_n - X| \geq \epsilon\}} \rightarrow 0$, $n \rightarrow \infty$.

The weak convergence follows due to definition and the Lebesgue dominated convergence theorem.

Convergence in space L^p is failed when the moment of p -th order does not exist.

Convergence in probability implies the weak convergence.

To prove it, assume that $X_n \xrightarrow[n]{P} X$, f is a bounded continuous function, $|f(x)| \leq c$, $x \in \mathbb{R}^1$, $\epsilon > 0$. We can choose a big enough N such that $P(|X| > N) \leq \epsilon/(4c)$. Take $\delta > 0$ so that $|f(x) - f(y)| \leq \epsilon/(2c)$ for $|x| \leq N$ and $|x - y| < \delta$. Then

$$\begin{aligned} |\mathbf{E}f(X_n) - \mathbf{E}f(X)| &\leq \mathbf{E}|f(X_n) - f(X)|I_{\{\omega: |X_n - X| > \delta\}} \\ &\quad + \mathbf{E}|f(X_n) - f(X)|I_{\{\omega: |X_n - X| \leq \delta\}}. \end{aligned} \quad (3.11)$$

The first term in the right-hand side of (3.10) is dominated by $2cP(\omega : |X_n - X| > \delta) \rightarrow 0$, $n \rightarrow \infty$.

$$\begin{aligned} \mathbf{E}|f(X_n) - f(X)|I_{\{\omega: |X_n - X| \leq \delta\}} &= \mathbf{E}|f(X_n) - f(X)|I_{\{\omega: |X_n - X| \leq \delta\}}I_{\{\omega: |X| \leq N\}} \\ &\quad + \mathbf{E}|f(X_n) - f(X)|I_{\{\omega: |X_n - X| \leq \delta\}}I_{\{\omega: |X| > N\}} \\ &\leq \epsilon + 2cP(\omega : |X_n - X| > \delta) < 2\epsilon \end{aligned}$$

for a large enough n .

Chapter 4

Weak Convergence of Sequences of Random Variables



This chapter is devoted to a systematic study of a weak convergence of sequences of random variables. It is shown the equivalence between a weak convergence and convergence in distribution. It is shown that a weak compactness and tightness for families of probability distributions (Prokhorov's theorem) is given. It is discussed a connection between characteristic functions and distributions of random variables. The method of characteristic functions is applied to prove the central limit theorem (CLT) for sums of independent identically distributed random variables (see [12, 28, 37], and [55]).

4.1 Weak Convergence and Its Description in Terms of Distributions

In the previous section, we recognized that *weak convergence* corresponds to the word “*weak*” because all other types of convergence provide the weak one. It turns out this type of convergence plays the most significant role in the theory of probability. This is why we study it here in more details.

We start with the following considerations. Let $(X_n)_{n=1,2,\dots}$ be a sequence of independent Bernoulli's random variables taking values 1 and 0 with probabilities p and q , $p + q = 1$, correspondingly. In this case, the law of large numbers (Theorem 3.7) is reduced to the following

$$\bar{S}_n = \frac{S_n}{n} \xrightarrow{P} p, \quad n \rightarrow \infty. \quad (4.1)$$

Let us define distribution functions

$$F_n(x) = P(\omega : \bar{S}_n \leq x) \text{ and } F(x) = \begin{cases} 1, & x \geq p, \\ 0, & x < p, \end{cases} \quad x \in \mathbb{R}^1.$$

We also know from the previous section that

$$\mathbf{E}f(\bar{S}_n) \rightarrow \mathbf{E}f(p), n \rightarrow \infty, \quad (4.2)$$

for any bounded continuous function f .

Further, denote P_{F_n} and P_F the distributions on $(R^1, \mathcal{B}(R^1))$ which correspond to F_n and F , respectively. Using formula of change of variables in expectations (2.7), we can rewrite (4.2) in two equivalent forms:

$$\begin{aligned} \int_{R^1} f(x) dP_{F_n} &\rightarrow \int_{R^1} f(x) dP_F, \quad n \rightarrow \infty, \\ \int_{R^1} f(x) dF_n &\rightarrow \int_{R^1} f(x) dF, \quad n \rightarrow \infty, \end{aligned} \quad (4.3)$$

for any bounded continuous function f .

The limit relations (4.3) allow to speak about the weak convergence of distributions P_{F_n} and distribution functions F_n to P_F and F in the sense (4.3).

We also note that

$$F_n(x) \rightarrow F(x), \quad n \rightarrow \infty, \quad (4.4)$$

for all $x \in R^1 \setminus \{p\}$, and hence, the weak convergence can be characterized with the help of distribution functions as their convergence (4.4) generalized in certain way.

Namely, for an arbitrary sequence of random variables $(X_n)_{n=1,2,\dots}$ and a random variable X with distribution functions $(F_n)_{n=1,2,\dots}$ and F , we may consider a convergence

$$F_n(x) \rightarrow F(x), \quad n \rightarrow \infty,$$

for all points of continuity of F , $x \in R^1$.

Such type of convergence of X_n to X we will call a *convergence in distribution* or convergence in law and denote $X_n \xrightarrow{d} X, n \rightarrow \infty$.

Theorem 4.1

$$X_n \xrightarrow{w} X, \quad n \rightarrow \infty \Leftrightarrow X_n \xrightarrow{d} X, \quad n \rightarrow \infty.$$

Proof We show the *direct implication* only. The inverse implication is the *Helly's selection principle* application. That is why we omit it here.

For fixed $x \in R^1$ and for each integer $\alpha \geq 1$, we define two bounded continuous functions

$$f_\alpha = f_\alpha(y) = \begin{cases} 1, & y \leq x, \\ \alpha(x - y) + 1, & x < y < x + \alpha^{-1}, \\ 0, & x + \alpha^{-1} \leq y, \end{cases}$$

$$f_\alpha^- = f_\alpha^-(y) = \begin{cases} 1, & x - \alpha^{-1} \geq y, \\ \alpha(x - y), & x - \alpha^{-1} < y \leq x, \\ 0, & x \leq y. \end{cases}$$

Functions f_α and f_α^- admit the limits:

$$\lim_{\alpha \rightarrow \infty} f_\alpha(y) = I_{(-\infty, x]}(y) \text{ and } \lim_{\alpha \rightarrow \infty} f_\alpha^-(y) = I_{(-\infty, x)}(y). \quad (4.5)$$

Moreover, applying the Lebesgue dominated convergence theorem, we get

$$\lim_{\alpha \rightarrow \infty} \mathbf{E} f_\alpha(X) = \mathbf{E} I_{(-\infty, x]}(X) = P(\omega : X(\omega) \leq x) = F(x),$$

$$\lim_{\alpha \rightarrow \infty} \mathbf{E} f_\alpha^-(X) = \mathbf{E} I_{(-\infty, x)}(X) = P(\omega : X(\omega) < x) = F(x-). \quad (4.6)$$

Further, due to $X_n \xrightarrow{w} X$, $n \rightarrow \infty$, we have for f_α and f_α^- the limiting relations:

$$\lim_{n \rightarrow \infty} \mathbf{E} f_\alpha(X_n) = \mathbf{E} f_\alpha(X),$$

$$\lim_{n \rightarrow \infty} \mathbf{E} f_\alpha^-(X_n) = \mathbf{E} f_\alpha^-(X).$$

By construction of f_α and f_α^- , we obtain

$$\mathbf{E} f_\alpha^-(X_n) \leq F_n(x) \leq \mathbf{E} f_\alpha(X_n).$$

Hence, for any $\alpha \geq 1$, the following inequalities are true:

$$\mathbf{E} f_\alpha^-(X) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq \mathbf{E} f_\alpha(X). \quad (4.7)$$

Combining (4.5)–(4.7), we arrive to the inequalities

$$F(x-) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x),$$

which provide convergence $F_n(x) \rightarrow F(x)$, $n \rightarrow \infty$, if x is a point of continuity of F . $\square$

Let us also pay a brief attention to a characterization of weak convergence of *probability distributions* $(P_n)_{n=1,2,\dots}$ on space $(R^1, \mathcal{B}(R^1))$.

We start with a very simple observation. Let P and $\tilde{P}$ be two *different* probability measures on $(R^1, \mathcal{B}(R^1))$. Define a sequence of probability measures $(P_n)_{n=1,2,\dots}$ as follows:

$$P_{2n} = P \text{ and } P_{2n+1} = \tilde{P},$$

and we arrive to conclusion that such a sequence $(P_n)_{n=1,2,\dots}$ *does not converge weakly*.

Another observation: We define a sequence of probability measures such that $P_n(\{n\}) = 1$ and, hence, $P_n(R^1) = 1$. On the other hand, $\lim_{n \rightarrow \infty} P_n((a, b]) = 0$ for any $a < b \in R^1$.

What does it mean in terms of distribution functions?

Obviously, P_n has the distribution function

$$F_n(x) = \begin{cases} 1, & x \geq n, \\ 0, & x < n, \end{cases} \quad n = 1, 2, \dots,$$

and for every $x \in R^1$

$$\lim_{n \rightarrow \infty} F_n(x) = G(x) = 0.$$

Therefore, the limit above is *not a distribution function*. It means that the set of distribution functions is *not compact*. These observations lead us to the next definitions.

Definition 4.1 A collection of probability measures (P_α) is *relatively compact*, if every sequence of measures from this collection admits a subsequence which converges weakly to a probability measure.

Definition 4.2 A collection of probability measures (P_α) is *tight*, if for every $\epsilon > 0$ there exists a compact set $K \subseteq R^1$ such that

$$\sup_{\alpha} P_{\alpha}(R^1 \setminus K) \leq \epsilon.$$

Similar definitions can be reproduced word by word for a collection of distribution functions (F_α) .

The classical Prokhorov theorem below speaks us that both these notions are equivalent.

Theorem 4.2 Let (P_α) be a family of probability measures on $(R^1, \mathcal{B}(R^1))$. Then this family is *relatively compact* if and only if it is *tight*.

Proof We give the proof of the direct implication only. Assume that (P_α) is not tight. Hence, there exists $\epsilon > 0$ such that for any compact $K \subseteq R^1$:

$$\sup_{\alpha} P_{\alpha}(R^1 \setminus K) > \epsilon.$$

Taking as K compact intervals $I_n = [-n, n]$, $n = 1, 2, \dots$, we arrive to a sequence of measures (P_{α_n}) such that $P_{\alpha_n}(R^1 \setminus I_n) > \epsilon$, $n = 1, 2, \dots$.

Further, since (P_α) is relatively compact, we can select from (P_{α_n}) a subsequence $(P_{\alpha_{n_k}})$ such that $(P_{\alpha_{n_k}}) \xrightarrow{w} \tilde{P}$, $k \rightarrow \infty$, where $\tilde{P}$ is a probability measure on $(R^1, \mathcal{B}(R^1))$.

Now we have

$$\epsilon \leq \limsup_{k \rightarrow \infty} P_{\alpha_{n_k}}(R^1 \setminus I_n) \leq \tilde{P}(R^1 \setminus I_n),$$

but it is not possible as $n \rightarrow \infty$ and $\tilde{P}$ is a probability measure. $\square$

4.2 Weak Convergence and Central Limit Theorem

We already know that distribution function F is a key quantitative characteristics of a random variable on given probability space $(\Omega, \mathcal{F}, P)$. Moreover, it is clear from the Kolmogorov consistency theorem that there always exist such a probability space and a random variable with given distribution. To be more illustrative, we would like to describe a particular way how this procedure can be realized.

We assume for simplicity that $F(x)$ is strictly increasing. Then we choose the space $([0, 1], \mathcal{B}([0, 1]), l)$ as a probability space. Let us define the following random variable

$$X(\omega) = \begin{cases} F^{-1}(\omega) & \text{if } \omega \in [0, 1], \omega = F(x) \\ 0 & \text{if } \omega < 0 \text{ or } \omega > 1. \end{cases}$$

It is clear that X is $\mathcal{B}([0, 1])$ -measurable, and if $\omega' \leftrightarrow x'$ and $\omega \leftrightarrow x$, we obtain

$$\begin{aligned} P(\omega' : X(\omega') \leq x) &= P(\omega' : F^{-1}(\omega') \leq x) = \\ &= P(\omega' : \omega' \leq F(x)) = P(\omega' : \omega' \leq \omega) = F(x), \end{aligned}$$

i.e., X has the distribution function $F(x)$.

We also want to emphasize the following additional properties of expectations and variances.

Problem 4.1 Prove that:

1. $\mathbf{E} \sum_{i=1}^n a_i X_i = \sum_{i=1}^n a_i \mathbf{E} X_i$ for integrable random variables $X_1, \dots, X_n$ and real numbers $a_1, \dots, a_n$, $n = 1, 2, \dots$
2. $\mathbf{E} \prod_{i=1}^n a_i X_i = \prod_{i=1}^n a_i \mathbf{E} X_i$ for independent random variables $X_1, \dots, X_n$, $n = 1, 2, \dots$
3. $\mathbf{Var}(aX + b) = a^2 \mathbf{Var}(X)$ for a random variable X with well-defined variance and real numbers a, b .
4. $\mathbf{Var}(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i^2 \mathbf{Var}(X_i)$ for independent random variables $X_1, \dots, X_n$ and real numbers $a_1, \dots, a_n$, $n = 1, 2, \dots$

Let us pay more attention to normal (or Gaussian) random variables. Denote

$$p(x) = p_{\mu, \sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}$$

the density of a normal random variable with parameters μ and σ^2 , i.e., $N(\mu, \sigma^2)$. We note the following:

1. If $\mu = 0$ and $\sigma^2 = 1$, then the random variable $Z = N(0, 1)$ is called *standard*.
2. The integral $\int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}$ is known as the Poisson integral. Therefore, $\int_{-\infty}^{\infty} p_{0,1}(x) dx = \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = 1$, which means that $p_{0,1}$ is the density of some probability distribution.
3. $\mathbf{E} Z = \int_{-\infty}^{\infty} x p_{0,1}(x) dx = 0$ because $h(x) = x$ is the odd function, and $\int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy = \sqrt{2\pi}$ for the even function $h(x) = x^2$.

Hence,

$$\mathbf{Var}(Z) = \mathbf{E} Z^2 - (\mathbf{E} Z)^2 = \mathbf{E} Z^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2/2} dy = 1.$$

4. *Standardization procedure.*

A random variable $X = N(\mu, \sigma^2) = \sigma Z + \mu$ or, equivalently, $Z = \frac{X - \mu}{\sigma}$. Hence, $\mathbf{E} X = \mathbf{E}(\sigma Z + \mu) = \sigma \mathbf{E} Z + \mu = \mu$ and $\mathbf{Var}(X) = \mathbf{Var}(\sigma Z + \mu) = \sigma^2 \mathbf{Var}(Z) = \sigma^2$, and we get a nice *probabilistic interpretation* of parameters μ and σ^2 as *mean* and *variance* of X .

5. Let X be a normal random variable with parameters μ and σ^2 . Consider its exponential transformation $Y = e^X$. This equality can be rewritten in a logarithmic form $X = \ln Y$.

In this case, Y is called *log-normal*, and its density has the form

$$l(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(\ln y - \mu)^2}{2\sigma^2} \right) y^{-1}, y \in (0, \infty).$$

Problem 4.2 Denote μ_Y and σ_Y^2 the mean and variance of Y , respectively. Prove that

$$\begin{aligned}\mu_Y &= \mathbf{E}Y = \exp(\mu + \sigma^2/2), \\ \sigma_Y^2 &= \mathbf{Var}(Y) = \exp(2\mu + \sigma^2)(e^{\sigma^2} - 1).\end{aligned}$$

To give an alternative description of moments of a random variable and its distribution function, we introduce the notion of characteristic function.

Definition 4.3 Let X be a random variable. Then the function $\phi_X(t) = \mathbf{E}e^{itX}$, $i = \sqrt{-1}$, is called a *characteristic function* of X .

To simplify our further considerations, we assume that the distribution function $F_X(x) = F(x)$ admits a density $f_X(x)$. Now we note the following:

First, the characteristics function does exist because of

$$\mathbf{E}|e^{itx}| = \int_{-\infty}^{\infty} |e^{itx}| f_X(x) dx = \int_{-\infty}^{\infty} f_X(x) dx = 1 < \infty.$$

Second, if $\mathbf{E}|X|^m < \infty$, $m = 1, 2, \dots$, then the characteristic function admits m continuous derivatives such that

$$\mathbf{E}X = i\phi'_X(0), \quad \mathbf{E}X^2 = -\phi''_X(0), \quad \dots, \quad \mathbf{E}X^m = i^m \phi_X^{(m)}(0). \quad (4.8)$$

Relations (4.8) follow from direct calculations of the corresponding integrals:

$$\phi'_X(t) = \int_{-\infty}^{\infty} ix e^{itx} f_X(x) dx$$

and, hence,

$$\phi'_X(0) = i \int_{-\infty}^{\infty} x e^{i \cdot 0 \cdot x} f_X(x) dx = i\mathbf{E}X \text{ etc.}$$

Example 4.1

1. For the Bernoulli random variable X taking values 1 and 0 with probabilities p and q , respectively, $p + q = 1$, we have

$$\phi_X(t) = e^{it \cdot 0} q + e^{it \cdot 1} p = pe^{it} + q.$$

Hence, $\phi'_X(t) = pe^{it} \cdot i$ and $\phi'_X(0) = ip$.

2. For the Poisson random variable with parameter $\lambda > 0$, we have

$$\begin{aligned}\phi_X(t) &= \mathbf{E}e^{itX} = \sum_{m=0}^{\infty} e^{itm} P(\omega : X(\omega) = m) = \\ &= \sum_{m=0}^{\infty} e^{itm} \frac{\lambda^m}{m!} e^{-\lambda} = \sum_{m=0}^{\infty} \frac{(\lambda e^{it})^m}{m!} e^{-\lambda} = \\ &= e^{-\lambda} e^{\lambda e^{it}} = e^{\lambda(e^{it}-1)}\end{aligned}$$

and $\phi'_X(0) = \lambda i$.

3. For the normal random variable X with parameters μ and σ^2 , we have using standardized random variable Z :

$$\begin{aligned}\phi_Z(t) &= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \int_{-\infty}^{\infty} \frac{\cos(tx)}{\sqrt{2\pi}} e^{-x^2/2} dx + i \int_{-\infty}^{\infty} \frac{\sin(tx)}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos(tx) e^{-x^2/2} dx.\end{aligned}\tag{4.9}$$

Differentiating both sides of (4.9) and using the integration by parts, we obtain

$$\begin{aligned}\phi'_Z(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-x \sin(tx)) e^{-x^2/2} dx \\ &= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \cos(tx) e^{-x^2/2} dx = -t\phi_Z(t).\end{aligned}$$

Hence,

$$\phi'_Z(t)/\phi_Z(t) = -t.\tag{4.10}$$

Integrating Eq.(4.10), we get $\ln \phi_Z(t) = -t^2/2 + c$, $c = \text{const}$, and therefore, $\phi_Z(t) = e^c e^{-t^2/2}$ with $e^c = 1$ due to $\phi_Z(0) = 1$. Finally, we have

$$\phi_Z(t) = e^{-t^2/2}.$$

The case of arbitrary $X = N(\mu, \sigma^2)$ is considered with the help of standardization procedure $X = \mu + \sigma Z$. As a result, we arrive to the formula

$$\phi_X(t) = \exp\left(it\mu - \frac{t^2\sigma^2}{2}\right).$$

Now we want to discuss a correspondence between weak convergence (convergence in distribution, in law) and the convergence of characteristic functions.

There is a simple *sufficient condition* for convergence in distribution. Namely, assume $(f_n)_{n=1,2,\dots}$ is a sequence of densities of random variables $(X_n)_{n=1,2,\dots}$ such that $f_n(x) \rightarrow f(x)$, $n \rightarrow \infty$, implies the convergence of their distribution functions $F_n(x) \rightarrow F(x)$, $n \rightarrow \infty$. In fact, this result from Calculus is true for any bounded function $h(x)$:

$$\int_{-\infty}^{\infty} h(x) f_n(x) dx \rightarrow \int_{-\infty}^{\infty} h(x) f(x) dx, \quad n \rightarrow \infty.$$

Taking $h(x) = e^{itx}$ in (4.11), we get

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

The result (4.12) implies the convergence $F_n(x) \rightarrow F(x)$, $n \rightarrow \infty$, too. Summarizing all these findings, we arrive to the following *methodology* to prove the central limit theorem (CLT) of the theory of probability.

Theorem 4.3 *Let $(X_n)_{n=1,2,\dots}$ be a sequence of independent identically distributed (iid) random variables with $\mathbf{E}X_n = \mu$ and $\mathbf{Var}X_n = \sigma^2$. Denote*

$$S_n = \sum_{m=1}^n X_m, \quad Y_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Then

$$Y_n \xrightarrow{d} Y, \quad n \rightarrow \infty, \quad (4.11)$$

where Y is a standard normal random variable $N(0, 1)$.

This is a version of the CLT for iid random variables.

Proof First of all, we note that the characteristic function of the sum of independent random variables is equal to the product of characteristic functions of these random variables. Using this property, we have

$$\begin{aligned} \phi_{Y_n}(t) &= \phi_{\frac{1}{\sigma\sqrt{n}} \sum_{m=1}^n (X_m - \mu)}(t) = \phi_{\sum_{m=1}^n (X_m - \mu)}(t/\sigma\sqrt{n}) \\ &= \prod_{m=1}^n \phi_{X_m - \mu}(t/\sigma\sqrt{n}) \\ &= (\phi(t/\sigma\sqrt{n}))^n, \quad n = 1, 2, \dots \end{aligned} \quad (4.12)$$

Further, $\mathbf{E}(X_m - \mu) = 0$, $\mathbf{E}(X_m - \mu)^2 = \sigma^2$ and

$$\begin{aligned}\phi'(t) &= i\mathbf{E}(X_m - \mu)e^{it(X_m - \mu)}, \quad \phi'(0) = 0, \\ \phi''(t) &= -\mathbf{E}(X_m - \mu)^2 e^{it(X_m - \mu)}, \quad \phi''(0) = -\sigma^2.\end{aligned}$$

Hence, we can expand ϕ in a Taylor expansion at point $t = 0$ and find that

$$\phi(t) = 1 + 0 - \frac{\sigma^2 t^2}{2} + t^2 \Delta(t),$$

where $\Delta(t) \rightarrow 0$, $t \rightarrow 0$. Therefore, using (4.14), we get

$$\phi_{Y_n}(t) = \phi^n(t/\sigma\sqrt{n}) = \exp(n \log \phi(t/\sigma\sqrt{n})) = \exp(n \log(1 - \frac{t^2}{2n} + \frac{t^2}{n\sigma^2} \Delta(t))),$$

and hence $\lim_{n \rightarrow \infty} \phi_{Y_n}(t)e^{-t^2/2}$, which is the characteristics function of $N(0, 1)$. $\square$

A more *general version* of the CLT is usually formulated under the *Lindeberg Condition (L)*.

Let $(X_n)_{n=1,2,\dots}$ be a sequence of independent random variables with $\mathbf{E}X_n = \mu_n$, $\mathbf{Var}X_n = \sigma_n^2$. We say that the sequence $(X_n)_{n=1,2,\dots}$ satisfies the condition **(L)** if for any $\epsilon > 0$

$$\left(\sum_{m=1}^n \sigma_m^2 \right)^{-1} \sum_{m=1}^n \int_{\left(x: |x - \mu_m| \geq \epsilon \left(\sum_{m=1}^n \sigma_m^2 \right)^{1/2} \right)} (x - \mu_m)^2 dF_{X_m}(x) \rightarrow 0, \quad n \rightarrow \infty.$$

Under the conditions above,

$$\frac{S_n - \mathbf{E}S_n}{(\mathbf{Var}S_n)^{1/2}} \xrightarrow{d} N(0, 1), \quad n \rightarrow \infty.$$

The *proof* of this version of the CLT is provided by the same method as the proof of Theorem 4.3.

Chapter 5

Absolute Continuity of Probability Measures and Conditional Expectations



In this chapter, a special attention is devoted to the absolute continuity of measures. It is shown how this notion and the Radon-Nikodym theorem work to define conditional expectations. The list of properties of conditional expectations is given here. In particular, it is emphasized the optimality in the mean-square sense of conditional expectations (see [12, 28, 37], and [55]).

5.1 Absolute Continuity of Measures and the Radon-Nikodym Theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space and X be a random nonnegative variable. Define a set function

$$\tilde{P}(A) = \int_A X dP = \mathbf{E} X I_A, \quad A \in \mathcal{F}. \quad (5.1)$$

Let us take $A = \cup_{i=1}^{\infty} A_i$, $A_i \cap A_j = \emptyset, i \neq j$. Using a linearity of expected values, we find that $\tilde{P}$ from (5.1) is a *finite-additive measure*. Further, with the help of monotonic convergence theorem, we have

$$\begin{aligned} \tilde{P}(A) &= \mathbf{E} X I_A = \mathbf{E} X I_{\cup_{i=1}^{\infty} A_i} = \\ &= \mathbf{E} \sum_{n=1}^{\infty} X I_{A_n} = \sum_{n=1}^{\infty} \mathbf{E} X I_{A_n} = \sum_{n=1}^{\infty} \tilde{P}(A_n), \end{aligned}$$

and therefore, $\tilde{P}$ is *countable additive*.

In case of arbitrary random variable, we can use the standard representation $X = X^+ - X^-$. Assume that one of expected values $\mathbf{E}X^+$ or $\mathbf{E}X^-$ is finite. In such a case, one can define the following *signed measure*

$$\tilde{P}(A) = \int_A X^+ dP - \int_A X^- dP = \tilde{P}^+(A) - \tilde{P}^-(A), \quad A \in \mathcal{F}, \quad (5.2)$$

that satisfies the next property

$$P(A) = 0 \Rightarrow \tilde{P}(A) = 0. \quad (5.3)$$

The *proof* of (5.3) is standard. For simple random variable with values $x_1, \dots, x_n$, $X = \sum_{i=1}^n x_i I_{A_i}$, we have

$$\tilde{P}(A) = \mathbf{E}X I_A = \sum_{i=1}^n x_i P(A_i \cap A) = 0, \quad P(A) = 0.$$

For arbitrary random variable X , we monotonically approximate X by a sequence of simple random variables $X_n \uparrow X$, $n \rightarrow \infty$, and take a limit according to monotonic convergence theorem:

$$\tilde{P}(A) = \mathbf{E}X I_A = \lim_{n \rightarrow \infty} \mathbf{E}X_n I_A = 0.$$

Definition 5.1 Relation (5.3) between two measures P and $\tilde{P}$ (not necessarily probability measures) is called the absolute continuity of $\tilde{P}$ with respect to P , and denoted $\tilde{P} \ll P$. Moreover, if $P \ll \tilde{P}$, then measures P and $\tilde{P}$ are equivalent ($P \sim \tilde{P}$).

It turns out one can characterize the absolute continuity property with the help of relations (5.1)–(5.2). This fundamental fact is known as *Radon-Nikodym theorem*.

Theorem 5.1 Let $(\Omega, \mathcal{F})$ be a measurable space. Let μ be a σ -additive measure and λ be a signed measure which is absolute continuous with respect to μ , $\lambda = \lambda_1 - \lambda_2$, where one of λ_1 or λ_2 is finite. Then there exists $\mathcal{F}$ -measurable function $Z = Z(\omega)$, $\omega \in \Omega$, with values in $[-\infty, +\infty]$ such that

$$\lambda(A) = \int_A Z(\omega) d\mu, \quad A \in \mathcal{F}. \quad (5.4)$$

Moreover, the function Z is uniquely defined up to sets of μ -measure zero.

In representation (5.4), the function Z is called the *Radon-Nikodym density* or *derivative* and denoted $Z = \frac{d\lambda}{d\mu}$.

As the first consequence of Theorem 5.1, we get a convenient *rule of changing of measure in expected values*. Let X be a random variable on a probability space

$(\Omega, \mathcal{F}, P)$ and $\tilde{P} \ll P$ with the Radon-Nikodym density Z . Then we formally obtain

$$\tilde{\mathbf{E}}X = \int_{\Omega} X d\tilde{P} = \int_{\Omega} X \frac{d\tilde{P}}{dP} dP = \int_{\Omega} XZ dP = \mathbf{E}XZ.$$

5.2 Conditional Expectations and Their Properties

Now we are ready to introduce the conditional expected value of random variable X with respect to a σ -algebra $\mathcal{Y} \subseteq \mathcal{F}$. In this case, we also call $\mathcal{Y}$ sub- σ -algebra of $\mathcal{F}$.

Definition 5.2 Let X be a nonnegative random variable and $\mathcal{Y}$ be a sub- σ -algebra of $\mathcal{F}$. Then a random variable $\mathbf{E}(X|\mathcal{Y})$ is called the conditional expectation of X with respect to $\mathcal{Y}$ if $\mathbf{E}(X|\mathcal{Y})$ is $\mathcal{Y}$ -measurable and

$$\mathbf{E}X I_A = \mathbf{E}(\mathbf{E}(X|\mathcal{Y}) I_A) \quad (5.5)$$

for every $A \in \mathcal{Y}$.

In general case, we decompose $X = X^+ - X^-$ and assume that one of random variables $\mathbf{E}(X^+|\mathcal{Y})$ and $\mathbf{E}(X^-|\mathcal{Y})$ is finite, and define

$$\mathbf{E}(X|\mathcal{Y}) = \mathbf{E}(X^+|\mathcal{Y}) - \mathbf{E}(X^-|\mathcal{Y}). \quad (5.6)$$

Existence of $\mathbf{E}(X|\mathcal{Y})$ in (5.6) follows from the Radon-Nikodym theorem. To prove it, we define a (signed) measure $\tilde{P}(A) = \mathbf{E}X I_A$ on a measurable space $(\Omega, \mathcal{Y})$ such that $\tilde{P} \ll P$. According to Theorem 5.1, there exists a unique Radon-Nikodym density which can be denoted here $\mathbf{E}(X|\mathcal{Y})$:

$$\tilde{P}(A) = \int_A \mathbf{E}(X|\mathcal{Y}) dP.$$

Let us note that the above equality coincides with (5.5).

Using definition of $\mathbf{E}(X|\mathcal{Y})$, we can easily define:

Conditional probability w.r.t. $\mathcal{Y}$

$$P(B|\mathcal{Y}) = \mathbf{E}(I_B|\mathcal{Y}), \quad B \in \mathcal{F},$$

conditional variance w.r. to $\mathcal{Y}$

$$\mathbf{Var}(X|\mathcal{Y}) = \mathbf{E}\left((X - \mathbf{E}(X|\mathcal{Y}))^2|\mathcal{Y}\right).$$

It is often the σ -algebra $\mathcal{Y}$ is generated by another random variable Y . Therefore, we can define conditional expected value and conditional probability X w.r. to Y as follows:

$$\mathbf{E}(X|Y) = \mathbf{E}(X|\mathcal{Y}), \mathbf{P}(B|Y) = P(B|\mathcal{Y}), B \in \mathcal{F},$$

where $\mathcal{Y} = \mathcal{Y}^Y$.

Let us count the list of properties of conditional expectations:

1. If $X = c$ (a.s.), then $\mathbf{E}(X|\mathcal{Y}) = c$ (a.s.).
2. If $X \leq Y$ (a.s.), then $\mathbf{E}(X|\mathcal{Y}) \leq \mathbf{E}(Y|\mathcal{Y})$ (a.s.).
3. $|\mathbf{E}(X|\mathcal{Y})| \leq \mathbf{E}(|X| |\mathcal{Y})$ (a.s.).
4. For random variables X and Y and constants a and b , we have

$$\mathbf{E}(aX + bY|\mathcal{Y}) = a\mathbf{E}(X|\mathcal{Y}) + b\mathbf{E}(Y|\mathcal{Y}) \text{ (a.s.)}.$$

5. $\mathbf{E}(X|\mathcal{F}_*) = \mathbf{E}X$ (a.s.).
6. $\mathbf{E}(X|\mathcal{F}) = X$ (a.s.).
7. $\mathbf{E}\mathbf{E}(X|\mathcal{Y}) = \mathbf{E}X$.
8. If sub- σ -algebras $\mathcal{Y}_1 \subseteq \mathcal{Y}_2$, then

$$\mathbf{E}(\mathbf{E}(X|\mathcal{Y}_2)|\mathcal{Y}_1) = \mathbf{E}(X|\mathcal{Y}_1) \text{ (a.s.)}.$$

9. If sub- σ -algebras $\mathcal{Y}_2 \subseteq \mathcal{Y}_1$, then

$$\mathbf{E}(\mathbf{E}(X|\mathcal{Y}_2)|\mathcal{Y}_1) = \mathbf{E}(X|\mathcal{Y}_2) \text{ (a.s.)}.$$

10. We say that two sub- σ -algebras $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are *independent*, if for any $B_1 \in \mathcal{Y}_1$ and $B_2 \in \mathcal{Y}_2$ we have $P(B_1 \cap B_2) = P(B_1)P(B_2)$.

For a random variable X and a sub- σ -algebra $\mathcal{Y}$, we say that X *does not depend on* $\mathcal{Y}$ if $\mathcal{Y}^X$ and $\mathcal{Y}$ are independent. In this case, we have

$$\mathbf{E}(X|\mathcal{Y}) = \mathbf{E}X \text{ (a.s.)},$$

if $\mathbf{E}X$ is well-defined.

11. For random variables X and Y such that Y is $\mathcal{Y}$ -measurable, $\mathbf{E}|X| < \infty$, $\mathbf{E}|XY| < \infty$, we have

$$\mathbf{E}(XY|\mathcal{Y}) = Y\mathbf{E}(X|\mathcal{Y}) \text{ (a.s.)}.$$

12. For a (generalized) sequence of random variables $(X_n)_{n=1,2,\dots}$ such that $|X_n| \leq Y$, $n = 1, 2, \dots$, $\mathbf{E}Y < \infty$ and $X_n \rightarrow X$, $n \rightarrow \infty$, a.s. Then $\mathbf{E}(X_n|\mathcal{Y}) \rightarrow \mathbf{E}(X|\mathcal{Y})$, (a.s.) $n \rightarrow \infty$.
13. If $X_n \geq Y$, $\mathbf{E}Y > -\infty$ and $X_n \uparrow X$, $n \rightarrow \infty$, a.s., then $\mathbf{E}(X_n|\mathcal{Y}) \uparrow \mathbf{E}(X|\mathcal{Y})$, (a.s.) $n \rightarrow \infty$.

14. If $X_n \leq Y$, $\mathbf{E}Y < \infty$ and $X_n \downarrow X$, $n \rightarrow \infty$, a.s., then $\mathbf{E}(X_n|\mathcal{Y}) \downarrow \mathbf{E}(X|\mathcal{Y})$, (a.s.) $n \rightarrow \infty$.
15. If $X_n \geq Y$, $\mathbf{E}Y > -\infty$, then

$$\mathbf{E}(\liminf_{n \rightarrow \infty} X_n|\mathcal{Y}) \leq \liminf_{n \rightarrow \infty} \mathbf{E}(X_n|\mathcal{Y}) \text{ (a.s.)}.$$

16. For nonnegative random variables $(X_n)_{n=1,2,\dots}$, we have

$$\mathbf{E}\left(\sum_{n=1}^{\infty} X_n|\mathcal{Y}\right) = \sum_{n=1}^{\infty} \mathbf{E}(X_n|\mathcal{Y}) \text{ (a.s.)}.$$

Let us show ways of proof of such properties. In case 2), we have for any $A \in \mathcal{Y}$ that

$$\int_A X dP \leq \int_A Y dP.$$

Therefore, we get from (5.5) that

$$\int_A \mathbf{E}(X|\mathcal{Y}) dP \leq \int_A \mathbf{E}(Y|\mathcal{Y}) dP,$$

which means that $\mathbf{E}(X|\mathcal{Y}) \leq \mathbf{E}(Y|\mathcal{Y})$ (a.s.).

In case (8), we have for any $A \in \mathcal{Y}_1 \subseteq \mathcal{Y}_2$ that

$$\int_A \mathbf{E}(X|\mathcal{Y}_1) dP = \int_A X dP = \int_A \mathbf{E}(X|\mathcal{Y}_2) = \int_A \mathbf{E}(\mathbf{E}(X|\mathcal{Y}_2)|\mathcal{Y}_1) dP,$$

which certifies the statement in (8).

Below we provide some comments and detailed calculations about conditional expectations.

Consider $\mathbf{E}(X|Y)$ for two (integrable) random variables X and Y . According to our definition, $\mathbf{E}(X|Y)$ is $\mathcal{Y}^Y$ -measurable, and by the representation theorem, there exists a Borelian function $\phi(\cdot)$ such that

$$\phi(Y(\omega)) = \mathbf{E}(X|Y)(\omega), \quad \omega \in \Omega. \quad (5.7)$$

For $A \in \mathcal{Y}^Y$, we get from (5.7):

$$\int_A X dP = \int_A \mathbf{E}(X|Y) dP = \int_A \phi(Y) dP. \quad (5.8)$$

Further, taking $A = Y^{-1}(B)$, $B \in \mathcal{B}(R^1)$, we have

$$\int_{Y^{-1}(B)} \phi(Y) dP = \int_B \phi(y) dP_Y,$$

and hence,

$$\int_{Y^{-1}(B)} X dP = \int_B \phi(y) dP_Y. \quad (5.9)$$

Having equalities (5.7)–(5.9), we can take $\phi(y) = \mathbf{E}(X|Y = y)$, $y \in R^1$.

To provide more details, one can consider a special case when the pair (X, Y) admits a joint density $f_{XY}(x, y)$. In such a case, we can put

$$f_{X|Y}(x|y) = \frac{f_{xy}(x, y)}{f_Y(y)}$$

and find that for $C \in \mathcal{B}(R^1)$

$$P(X \in C | Y = y) = \int_C f_{X|Y}(x|y) dx$$

and

$$\mathbf{E}(X|Y = y) = \int_{R^1} x f_{X|Y}(x|y) dx.$$

This is a reason that the function $f_{X|Y}(x|y)$ is called a density of a conditional distribution (conditional density).

In case of discrete random variables X and Y with values $x_1, x_2, \dots$ and $y_1, y_2, \dots$ correspondingly, denote $\mathcal{Y} = \sigma(Y) = \mathcal{Y}^Y$ and $(\omega : Y(\omega) = y_i) = D_i$, $i = 1, 2, \dots$. Sets D_i , $i = 1, 2, \dots$ are *atoms* for P in the sense that $P(D_i) > 0$ and any subset $A \subseteq D_i$ has a probability zero or its completion to D_i (i.e., $D_i \setminus A$) has a probability zero. We can define

$$\mathbf{E}(X|D_i) = \frac{\mathbf{E}X I_{D_i}}{P(D_i)}, \quad i = 1, 2, \dots$$

Then the calculation of $\mathbf{E}(X|Y)$ is reduced to the claim:

$$\mathbf{E}(X|Y) = \mathbf{E}(X|\mathcal{Y}^Y) = \mathbf{E}(X|D_i) \quad (a.s.)$$

on the atom D_i , $i = 1, 2, \dots$

To connect these definitions and results to the general definition, we need to check (5.5). In particular, we have

$$\begin{aligned}
 \mathbf{E}\mathbf{E}(X|\mathcal{Y}) &= \sum_{i=1}^{\infty} \mathbf{E}(X|D_i)P(D_i) \\
 &= \sum_{i=1}^{\infty} \mathbf{E}(XI_{D_i}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_j P(X = x_j, Y = y_i) \\
 &= \sum_{j=1}^{\infty} x_j \sum_{i=1}^{\infty} P(X = x_j, Y = y_i) \\
 &= \sum_{j=1}^{\infty} x_j \sum_{i=1}^{\infty} P(X = x_j, Y = y_i) = \sum_{j=1}^{\infty} x_j P(X = x_j) \\
 &= \mathbf{E}X.
 \end{aligned}$$

Let us take note of that calculations of $\mathbf{E}(X|Y_1, \dots, Y_n)$ can be given in the same way.

Finally, we demonstrate an important application of $\mathbf{E}(X|Y)$. We will interpret Y as an *observable* variable and X as a non-observable. This is a typical situation in many areas, including option pricing theory, where X is a pay-off of option and Y is a stock price.

How can we estimate X based on observations of Y to provide the optimality of the estimate?

A satisfactory solution of the problem can be given as follows.

Let $\phi = \phi(x)$, $x \in R^1$, be a Borelian function. Taking $\phi(Y)$, we get an estimate for X , and we should choose from a variety of such estimates an *optimal* estimate. As criteria, we can take $\mathbf{E}(X - \phi(Y))^2$ and define the *optimal* estimate $\phi^*(Y)$ as follows:

$$\mathbf{E}(X - \phi^*(Y))^2 = \inf_{\phi} \mathbf{E}(X - \phi(Y))^2,$$

where we assume $\mathbf{E}X^2 < \infty$ and $\mathbf{E}\phi^2(Y) < \infty$.

Theorem 5.2 *The optimal estimate has the following representation $\phi^*(x) = \mathbf{E}(X|Y = x)$.*

Proof Taking $\phi^*(Y) = \mathbf{E}(X|Y)$, we have for any other estimate ϕ :

$$\begin{aligned}
 \mathbf{E}(X - \phi(Y))^2 &= \mathbf{E}[(X - \phi^*(Y)) + (\phi^*(Y) - \phi(Y))]^2 \\
 &= \mathbf{E}(X - \phi^*(Y))^2 + 2\mathbf{E}(X - \phi^*(Y))(\phi^*(Y) - \phi(Y)) \\
 &\quad + \mathbf{E}(\phi^*(Y) - \phi(Y))^2
 \end{aligned}$$

$$\begin{aligned} &= \mathbf{E}(X - \phi^*(Y))^2 + \mathbf{E}(\phi^*(Y) - \phi(Y))^2 \\ &\quad + 2\mathbf{E}[\mathbf{E}((X - \phi^*(Y))(\phi^*(Y) - \phi(Y)) | Y)] \\ &= \mathbf{E}(X - \phi^*(Y))^2 + \mathbf{E}(\phi^*(Y) - \phi(Y))^2 \\ &\geq \mathbf{E}(X - \phi^*(Y))^2. \end{aligned}$$

□

Chapter 6

Discrete Time Stochastic Analysis: Basic Results



This chapter is completely devoted to a discrete time stochastic analysis. It contains the key notions adapted to discrete time like stochastic basis, filtration, predictability, stopping times, martingales, sub- and supermartingales, local densities of probability measures, discrete stochastic integrals, and stochastic exponents. It is stated the Doob decomposition for stochastic sequences, maximal inequalities, and other Doob's theorems. The developed martingale technique is further applied to prove several asymptotical properties for martingales and submartingales (see [7, 13, 15, 17, 28, 55], and [62]).

6.1 Basic Notions: Stochastic Basis, Predictability, and Martingales

We have seen already that a sequence of random variables $X_1, X_2, \dots, X_n$ can be interpreted as a *n-times repetition* (realization) of the underlying random experiment. One can note that the order of appearance of new information is important. For example, such order and such information are valuable in stock exchange trading. One can also observe that a numeration of X_i may not be directly connected to *real (physical) time*, because an *operation time* is often exploited instead of real time in finance. As we know, to provide an accurate work with random variables, we need to assume that these random variables are defined on some probability space $(\Omega, \mathcal{F}, P)$. In this setting, for *each fixed outcome* $\omega \in \Omega$, one can observe a sequence of numbers $X_1(\omega), X_2(\omega), \dots, X_n(\omega), \dots$, called a *trajectory* or *sample path*. It may present a behavior of stock prices or indexes over given time interval. So, thinking in this way, we must emphasize the difference between probability theory and stochastic/random processes. It can be roughly explained as follows. If theory of probability studies probabilities of occurrence of random events, including the events connected to random variables, theory of stochastic processes

considers probabilities of occurrence of trajectories and families of trajectories of stochastic processes. Such a difference calls for a “supplementary equipment” of given probability space with an *information flow* or *filtration*. It is a nondecreasing family $(\mathcal{F}_n)_{n=0,1,\dots}$ together with $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0,1,\dots}, P)$ called a *stochastic basis*. Usually, the main σ -algebra $\mathcal{F}$ is determined by the filtration $(\mathcal{F}_n)_{n=0,1,\dots}$ in the sense that $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n) = \mathcal{F}$. We sometimes will count that $\mathcal{F}_0$ is trivial, i.e., $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Filtration creates a new class of random variables, called *stopping times*. A random variable $\tau : \Omega \rightarrow \{0, 1, \dots, n, \dots, \infty\}$ is called a stopping time if for each $n = 0, 1, \dots$

$$\{\omega : \tau(\omega) \leq n\} \in \mathcal{F}_n,$$

i.e., each value n is taken based on information until time n without having information from the future time.

We can interpret a stopping time τ as a random time, and therefore we can speak about information $\mathcal{F}_\tau$ before this time τ . It is formally realized as follows:

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\omega : \tau(\omega) \leq n\} \in \mathcal{F}_n\},$$

and $\mathcal{F}_\tau$ is a σ -algebra.

If two stopping times τ and σ are connected to each other through inequality $\tau \leq \sigma$ (a.s.), then obviously $\mathcal{F}_\tau \subseteq \mathcal{F}_\sigma$.

Now if $(X_n)_{n=0,1,\dots}$ is a sequence of random variables on $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0,1,\dots}, P)$ with information flow $(\mathcal{F}_n)$ and $X_n - \mathcal{F}_n$ -measurable for each $n = 0, 1, \dots$. In this case, we will call (X_n) *adapted* to $(\mathcal{F}_n)_{n=0,1,\dots}$, or *stochastic*.

For a stochastic sequence $(X_n)_{n=0,1,\dots}$ and a stopping time τ , we define

$$X_\tau = \sum_{n=0}^{\infty} X_n I_{\{\tau=n\}} + X_\infty I_{\{\tau=\infty\}},$$

where X_∞ may be a fixed constant or $X_\infty = \lim_{n \rightarrow \infty} X_n$ (a.s.), if the limit does exist.

The random variable X_τ is a “superposition” of a stochastic sequence and τ , and hence, $X_\tau - \mathcal{F}_\tau$ -measurable. We can go even further if we take a stochastic sequence $(X_n)_{n=0,1,\dots}$ and a nondecreasing sequence of stopping times $(\tau_n)_{n=0,1,\dots}$ and define a new sequence of random variables

$$Y_n = X_{\tau_n}, \quad \tau_n \leq \tau_{n+1}.$$

It is quite natural to call (τ_n) as a *time change* and (Y_n) as a time-changed sequence. The idea of “time change” is very productive for stochastic analysis. For example, in mathematical finance, this type of time change is often based on market volatility and is called the “operation time.”

Problem 6.1 Prove that:

1. τ is a stopping time if and only if $\{\tau = n\} \in \mathcal{F}_n$ for all n .
2. $\mathcal{F}_\tau$ is a σ -algebra, if τ is a stopping time.
3. $\mathcal{F}_\tau \subseteq \mathcal{F}_\sigma$ if stopping times $\tau \leq \sigma(\omega)$ (a.s.).
4. X_τ is $\mathcal{F}_\tau$ -measurable if τ is a stopping time.
5. $\tau \wedge \sigma$ is a stopping time.

We can also consider $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$ as a system of probability spaces $(\Omega, \mathcal{F}_0, P_0), (\Omega, \mathcal{F}, P), \dots$, where P_0 on $\mathcal{F}_0$ is just a restriction of P to $\mathcal{F}_0$ and so on. In this framework, we can consider another probability measure $\tilde{P}$ and $\tilde{P}_n$ as its restriction to $\mathcal{F}_n$. Assume that $\tilde{P}_n \ll P_n$, $n = 0, 1, \dots$, and denote $Z_n = \frac{d\tilde{P}_n}{dP_n}$ the density of $\tilde{P}_n$ with respect to P_n , $n = 0, 1, \dots$. The sequence $(Z_n)_{n=0,1,\dots}$ is called a *local density* of $\tilde{P}$ with respect to P , and for each $n = 0, 1, \dots$, we have

$$\mathbf{E}Z_n = \int_{\Omega} Z_n dP = \int_{\Omega} Z_n dP_n = \tilde{P}_n(\Omega) = 1.$$

Further, for any $A \in \mathcal{F}_{n-1}$, we formally have that

$$\int_A Z_n dP = \int_A Z_n dP_n = \int_A \frac{d\tilde{P}_n}{dP_n} dP_n = \tilde{P}_n(A) = \tilde{P}_{n-1}(A) = \int_A Z_{n-1} dP.$$

It means that (a.s.)

$$\mathbf{E}(Z_n | \mathcal{F}_{n-1}) = Z_{n-1}, \quad n = 1, 2, \dots \quad (6.1)$$

We can also ask the question:

Is it possible to calculate for some integrable random variable Y $\tilde{\mathbf{E}}(Y | \mathcal{F}_{n-1})$ via $\mathbf{E}(Y | \mathcal{F}_{n-1})$?

Again, take $A \in \mathcal{F}_{n-1}$ and find that

$$\begin{aligned} \int_A \mathbf{E}(Y Z_n | \mathcal{F}_{n-1}) dP &= \int_A Y Z_n dP = \int_A Y d\tilde{P}_n \\ &= \int_A Y d\tilde{P}_{n-1} = \int_A \tilde{\mathbf{E}}(Y | \mathcal{F}_{n-1}) d\tilde{P}_{n-1} \\ &= \int_A \tilde{\mathbf{E}}(Y | \mathcal{F}_{n-1}) dP_{n-1} = \int_A \tilde{\mathbf{E}}(Y | \mathcal{F}_{n-1}) Z_{n-1} dP, \end{aligned}$$

and hence,

$$\tilde{\mathbf{E}}(Y | \mathcal{F}_{n-1}) = Z_{n-1}^{-1} \mathbf{E}(Y Z_n | \mathcal{F}_{n-1}) \quad (a.s.). \quad (6.2)$$

Relation (6.2) is called a *rule of change of probability in conditional expectations*.

Besides adapted sequences of random variables, we can introduce sequences of random variables which are in between deterministic and stochastic sequences. We say that an adapted sequence $(A_n)_{n=0,1,\dots}$ is *predictable*, if A_n - $\mathcal{F}_{n-1}$ -measurable for $n = 1, 2, \dots$. We can also take the property (6.1) to introduce the whole class of stochastic sequences called *martingales*. We say that an integrable adapted sequence $(M_n)_{n=0,1,\dots}$ is a martingale, if

$$\mathbf{E}(M_n | \mathcal{F}_{n-1}) = M_{n-1} \text{ (a.s.) for } n = 1, 2, \dots \quad (6.3)$$

Relation (6.3) can be rewritten as follows $\mathbf{E}(M_n - M_{n-1} | \mathcal{F}_{n-1}) = 0$ (a.s.). It means the sequence $Y_n = M_n - M_{n-1}$ presents the so-called *martingale difference*.

We can also see that the local density $(Z_n)_{n=0,1,\dots}$ in (6.1) is a martingale with respect to P . These two types of stochastic sequences is a natural basis for many others. To demonstrate this claim, we consider an arbitrary integrable stochastic sequence $(X_n)_{n=0,1,\dots}$ with $X_0 = 0$ (a.s.) for simplicity. Define for each $n = 1, 2, \dots$, $\Delta X_n = X_n - X_{n-1}$ and write an obvious equality

$$\Delta X_n = \Delta X_n - \mathbf{E}(\Delta X_n | \mathcal{F}_{n-1}) + \mathbf{E}(\Delta X_n | \mathcal{F}_{n-1}) = \Delta M_n + \Delta A_n, \quad (6.4)$$

where

$$\Delta M_n = \Delta X_n - \mathbf{E}(\Delta X_n | \mathcal{F}_{n-1}),$$

$$\Delta A_n = \mathbf{E}(\Delta X_n | \mathcal{F}_{n-1}).$$

It is clear from (6.4) that for $n = 0, 1, \dots$,

$$X_n = M_n + A_n, \quad (6.5)$$

where $M_n = \sum_{i \leq n} \Delta M_i$ and $A_n = \sum_{i \leq n} \Delta A_i$, and therefore, $(M_n)_{n=0,1,\dots}$ is a martingale and $(A_n)_{n=0,1,\dots}$ is predictable.

Such a unique decomposition (6.5) is called the *Doob decomposition* of $(X_n)_{n=0,1,\dots}$. In particular, it is true for sequences $(X_n)_{n=0,1,\dots}$ satisfying conditions

$$\mathbf{E}(X_n | \mathcal{F}_{n-1}) \geq X_{n-1} \text{ (a.s.)}$$

and

$$\mathbf{E}(X_n | \mathcal{F}_{n-1}) \leq X_{n-1} \text{ (a.s.)}.$$

Such sequences are called *submartingales* and *supermartingales*, and in their decompositions (6.5) $\Delta A_n \geq 0$ and $\Delta A_n \leq 0$ (a.s.), $n = 1, 2, \dots$, correspondingly.

Remark 6.1 Given a martingale $(X_n)_{n=0,1,\dots}$, there is a simple way of constructing submartingales (supermartingales). Suppose ϕ is a convex downward measurable

function such that $\mathbf{E}|\phi(X_n)| < \infty$, $n = 0, 1, \dots$. Then Jensen's inequality implies that $(\phi(X_n))_{n=0,1,\dots}$ is a submartingale.

Let us consider a martingale $(M_n)_{n=0,1,\dots}$ such that $\mathbf{E}M_n^2 < \infty$, $n = 0, 1, \dots$, and then it is called *square integrable*. Further, due to the Jensen's inequality, $X_n = M_n^2$ is a submartingale. Using the Doob decomposition (6.5), one can conclude that

$$M_n^2 = m_n + \langle M, M \rangle_n, \quad n = 0, 1, \dots,$$

where $(m_n)_{n=0,1,\dots}$ is a martingale and $(\langle M, M \rangle_n)_{n=0,1,\dots}$ is a nondecreasing predictable sequence called the *quadratic characteristic (compensator)* of M . Moreover,

$$\langle M, M \rangle_n = \sum_{i=1}^n \mathbf{E}((\Delta M_i)^2 | \mathcal{F}_{i-1}), \quad \langle M, M \rangle_0 = 0,$$

and

$$\begin{aligned} \mathbf{E}((M_k - M_l)^2 | \mathcal{F}_l) &= \mathbf{E}(M_k^2 - M_l^2 | \mathcal{F}_l) \\ &= \mathbf{E}(\langle M, M \rangle_k - \langle M, M \rangle_l | \mathcal{F}_l), \quad l \leq k, \end{aligned}$$

and $\mathbf{E}M_n^2 = \mathbf{E}\langle M, M \rangle_n$, $n = 0, 1, \dots$. It is possible to define a measure of correlation between two square-integrable martingales $(M_n)_{n=0,1,\dots}$ and $(N_n)_{n=0,1,\dots}$:

$$\langle M, N \rangle_n = \frac{1}{4} \{ \langle M + N, M + N \rangle_n - \langle M - N, M - N \rangle_n \},$$

$n = 0, 1, \dots$, which is called the *joint quadratic characteristic* of $(M_n)_{n=0,1,\dots}$ and $(N_n)_{n=0,1,\dots}$.

It is almost obvious to show that the sequence $(M_n N_n - \langle M, N \rangle_n)_{n=0,1,\dots}$ is a martingale, and if $\langle M, N \rangle_n = 0$ for all $n = 0, 1, \dots$, then such martingales are called *orthogonal*.

Further, the following property that connects the martingale property and absolute continuity of probability measures is related to formula (6.2) and can be referred to as a discrete time version of the *Girsanov theorem*.

Let $(M_n)_{n=0,1,\dots}$, $M_0 = 0$, be a martingale with respect to the original measure P and assume $\mathbf{E}|Z_n Z_{n-1}^{-1} \Delta M_n| < \infty$, $n = 1, 2, \dots$. Define $(\tilde{M}_n)_{n=0,1,\dots}$, $\tilde{M}_0 = 0$, by relations

$$\Delta \tilde{M}_n = \Delta M_n - \mathbf{E}(Z_n Z_{n-1}^{-1} \Delta M_n | \mathcal{F}_{n-1}), \quad n = 1, 2, \dots$$

Using the rule of change of measure (6.2), we obtain

$$\tilde{\mathbf{E}}(\Delta \tilde{M}_n | \mathcal{F}_{n-1}) = \tilde{\mathbf{E}}(\Delta M_n - \mathbf{E}(Z_n Z_{n-1}^{-1} \Delta M_n | \mathcal{F}_{n-1}) | \mathcal{F}_{n-1})$$

$$= \tilde{\mathbf{E}}(\Delta M_n | \mathcal{F}_{n-1}) - \tilde{\mathbf{E}}(\Delta M_n | \mathcal{F}_{n-1}) = 0 \quad (a.s.),$$

which implies that $(\tilde{M}_n)_{n=0,1,\dots}$ is a martingale with respect to $\tilde{P} \ll P$ with the local density $(Z_n)_{n=0,1,\dots}$.

Let us use predictable and martingale stochastic sequences to construct more complicated objects. For a predictable $(H_n)_{n=0,1,\dots}$ and a martingale $(m_n)_{n=0,1,\dots}$, define a *discrete stochastic integral*

$$H * m_n = \sum_{i=0}^n H_i \Delta m_i. \quad (6.6)$$

If martingale $(m_n)_{n=0,1,\dots}$ is square integrable, sequence $(H_n)_{n=0,1,\dots}$ is predictable, and $\mathbf{E} H_n^2 \Delta \langle m, m \rangle_n < \infty$, $n = 0, 1, \dots$, then $(H * M_n)_{n=0,1,\dots}$ is a square-integrable martingale with quadratic characteristic

$$\langle H * m, H * m \rangle_n = \sum_{i=0}^n H_i^2 \Delta \langle m, m \rangle_i.$$

Further, let $(M_n)_{n=0,1,\dots}$ be a fixed square-integrable martingale, then one can consider all square-integrable martingales $(N_n)_{n=0,1,\dots}$ that are orthogonal to $(M_n)_{n=0,1,\dots}$. Introduce a family of square-integrable martingale of the form

$$X_n = M_n + N_n. \quad (6.7)$$

On the other hand, any square-integrable martingale $(X_n)_{n=0,1,\dots}$ can be written in the form (6.7), where the orthogonal term $(N_n)_{n=0,1,\dots}$ satisfies (6.6) with the martingale $(m_n)_{n=0,1,\dots}$ that is orthogonal to the given martingale $(M_n)_{n=0,1,\dots}$. Such a version of decomposition (6.7) is referred to as the *Kunita-Watanabe decomposition*.

Discrete stochastic integrals are related to *discrete stochastic differential equations*. Solutions of such equations are used in modeling the dynamics of asset prices in financial markets.

Consider a stochastic sequence $(U_n)_{n=0,1,\dots}$ with $U_0 = 0$ and define a new stochastic sequence $(X_n)_{n=0,1,\dots}$ with $X_0 = 1$ by

$$\Delta X_n = X_{n-1} \Delta U_n, \quad n = 1, 2, \dots \quad (6.8)$$

Solution of (6.8) has the form

$$X_n = \prod_{i=1}^n (1 + \Delta U_i) = \mathcal{E}_n(U), \quad n = 1, 2, \dots,$$

and is called a *stochastic exponential*.

One can consider a nonhomogeneous version of Eq. (6.8)

$$\Delta X_n = \Delta N_n + X_{n-1} \Delta U_n, \quad X_0 = N_0, \quad (6.9)$$

where $(N_n)_{n=0,1,\dots}$ is a given sequence.

Solution of (6.9) is determined with the help of stochastic exponential as follows:

$$X_n = \mathcal{E}_n(U) \left[N_0 + \sum_{i=1}^n \mathcal{E}_i^{-1}(U) \Delta U_i \right].$$

Let us list very helpful properties of stochastic exponentials:

1. $\mathcal{E}_n^{-1}(U) = \mathcal{E}_n(-U^*)$, where $\Delta U_n^* = \frac{\Delta U_n}{1 + \Delta U_n}$, $\Delta U_n \neq -1$.
2. $(\mathcal{E}_n(U))_{n=0,1,\dots}$ is a martingale if and only if $(U_n)_{n=0,1,\dots}$ is a martingale.
3. $\mathcal{E}_n(U) = 0$ (a.s.) for $n \geq \tau_0 = \inf(i : \mathcal{E}_i(U) = 0)$.
4. For two stochastic sequences (U_n) and (V_n) , the next multiplication rule of stochastic exponentials is true:

$$\mathcal{E}_n(U) \mathcal{E}_n(V) = \mathcal{E}_n(U + V + [U, V]),$$

where $\Delta[U, V]_n = \Delta U_n \Delta V_n$.

6.2 Martingales on Finite Time Interval

Here we study stochastic sequences on the interval $[0, N] = \{0, 1, \dots, N\}$. The first important theorem is devoted to an interesting characterization of the notion of a martingale.

Theorem 6.1 *Let $(X_n)_{n=0,1,\dots,N}$ be an integrable stochastic sequence on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0,1,\dots,N}, P)$. Then the following statements are true:*

1. The sequence $(X_n)_{n=0,1,\dots,N}$ is a martingale if and only if $X_n = \mathbf{E}(X_N | \mathcal{F}_n)$ (a.s.) for all $n = 0, 1, \dots, N$.
2. If for all stopping times τ the equality $\mathbf{E}X_\tau = \mathbf{E}X_0$ is fulfilled, then $(X_n)_{n=0,1,\dots,N}$ is a martingale.

Proof

- (a) For a direct implication, using the definition of conditional expectations and their telescopic property, we have that (a.s.)

$$\mathbf{E}(X_N | \mathcal{F}_{N-2}) = \mathbf{E}(\mathbf{E}(X_N | \mathcal{F}_{N-1}) | \mathcal{F}_{N-2}) = \mathbf{E}(X_{N-1} | \mathcal{F}_{N-2}) = X_{N-2}, \text{ etc.}$$

For an inverse implication, we also use the definition and the telescopic property and find that for all n (a.s.)

$$\mathbf{E}(X_N|\mathcal{F}_{N-1}) = \mathbf{E}(\mathbf{E}(X_N|\mathcal{F}_N)|\mathcal{F}_{N-1}) = \mathbf{E}(X_N|\mathcal{F}_{N-1}) = X_{N-1}, \text{ etc.}$$

(b) For a fixed $n \leq N$ and $A \in \mathcal{F}_n$, we define the following stopping time:

$$\tau_A(\omega) = \begin{cases} n & \text{if } \omega \in A, \\ N & \text{if } \omega \notin A. \end{cases}$$

According to the assumption, we have

$$\begin{aligned} \mathbf{E}X_0 &= \mathbf{E}X_{\tau_A} = \mathbf{E}X_{\tau_A}I_A + \mathbf{E}X_{\tau_A}I_{A^c} \\ &= \mathbf{E}X_nI_A + \mathbf{E}X_NI_{A^c}. \end{aligned}$$

It means that $\mathbf{E}(X_N|\mathcal{F}_n) = X_n$ (a.s.), and by (a), we conclude that $(X_n)_{n=0,1,\dots,N}$ is a martingale. □

Theorem 6.1 gives us an understanding about a strong connection between martingales and stopping times. The next statement which has a special name as “optional sampling theorem” of Doob tells us more about these connections.

Theorem 6.2 *Let $(X_n)_{n=0,1,\dots,N}$ be a martingale (submartingale, supermartingale) and $\tau_1 \leq \tau_2$ (a.s.) are stopping times, and then*

$$\mathbf{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) = X_{\tau_1} \text{ (a.s.)}$$

($\mathbf{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) \geq X_{\tau_1}$ and $\mathbf{E}(X_{\tau_2}|\mathcal{F}_{\tau_1}) \leq X_{\tau_1}$ (a.s.), respectively). In particular, $\mathbf{E}X_{\tau_1} = \mathbf{E}X_{\tau_2}$.

Proof of this theorem, which generalizes the obvious property for deterministic times, readily follows from next considerations.

For $A \in \mathcal{F}_{\tau_1}$, $n \leq N$ and $B = A \cap \{\tau_1 = n\}$ in order to prove $\int_A X_{\tau_2} dP = \int_A X_{\tau_1} dP$, we need to prove that

$$\int_{B \cap \{\tau_2 \geq n\}} X_{\tau_2} dP = \int_{B \cap \{\tau_2 \geq n\}} X_n dP.$$

It follows from the next equalities

$$\int_{B \cap \{\tau_2 \geq n\}} X_n dP = \int_{B \cap \{\tau_2 = n\}} X_n dP + \int_{B \cap \{\tau_2 > n\}} X_n dP$$

$$\begin{aligned}
&= \int_{B \cap \{\tau_2 = n\}} X_n dP + \int_{B \cap \{\tau_2 > n\}} \mathbf{E}(X_{n+1} | \mathcal{F}_n) dP \\
&= \int_{B \cap \{\tau_2 = n\}} X_n dP + \int_{B \cap \{\tau_2 > n\}} X_{n+1} dP \\
&= \int_{B \cap \{n \leq \tau_2 \leq n+1\}} X_n dP + \int_{B \cap \{\tau_2 \geq n+2\}} X_{n+2} dP = \dots = \\
&= \int_{B \cap \{\tau_2 \geq n\}} X_{\tau_2} dP.
\end{aligned}$$

It turns out for martingales (submartingales, supermartingales) one can get inequalities called *maximal* (or the *Kolmogorov-Doob inequalities*) which are stronger than the Chebyshev inequality.

Theorem 6.3

1. If $(X_n)_{n=0,1,\dots,N}$ is a submartingale, then for any $\lambda > 0$:

$$P(\omega : \max_{n \leq N} X_n \geq \lambda) \leq \frac{\mathbf{E}X_N^+}{\lambda}.$$

2. If $(X_n)_{n=0,1,\dots,N}$ is a supermartingale, then for any $\lambda > 0$:

$$P(\omega : \max_{n \leq N} X_n \geq \lambda) \leq \frac{\mathbf{E}X_0 + \mathbf{E}X_N^-}{\lambda}.$$

3. If $(X_n)_{n=0,1,\dots,N}$ is a martingale, then for any $\lambda > 0$:

$$P(\omega : \max_{n \leq N} |X_n| \geq \lambda) \leq \frac{\mathbf{E}|X_N|}{\lambda}.$$

Proof Let us define the following stopping time

$$\tau = \inf\{n \leq N : X_n \geq \lambda\},$$

where we put $\tau = N$ if the set in brackets above is $\emptyset$.

Denote $A = \{\omega : \max_{n \leq N} X_n \geq \lambda\}$ and find that

$$A \cap \{\tau = n\} = \{\tau = n\} \in \mathcal{F}_n,$$

if $n < N$ and

$$A \cap \{\tau = N\} \in \mathcal{F}_N,$$

due to $A \in \mathcal{F}_N$.

Hence, $A \in \mathcal{F}_\tau$.

To prove the statement (1) of the theorem, we derive from Theorem 6.2 that

$$\mathbf{E}X_\tau I_A \leq \mathbf{E}X_N I_A$$

because stopping time $\tau \leq N$.

If the event A occurs, then $X_\tau \geq \lambda$ and $X_\tau I_A \geq \lambda I_A$. Therefore,

$$\lambda P(A) \leq \mathbf{E}X_N I_A \leq \mathbf{E}X_N^+ I_A \leq \mathbf{E}X_N^+,$$

and we obtain (1).

Further, due to $0 \leq \tau \leq N$, we have from Theorem 6.2

$$\mathbf{E}X_0 \geq \mathbf{E}X_\tau = \mathbf{E}X_\tau I_A + \mathbf{E}X_\tau I_{A^c} \geq \mathbf{E}X_\tau I_A + \mathbf{E}X_N I_{A^c}.$$

Hence, we get statement (2) after next calculations

$$\begin{aligned} \lambda P(A) &\leq \mathbf{E}X_\tau I_A \leq \mathbf{E}X_0 - \mathbf{E}X_N I_{A^c} \\ &= \mathbf{E}X_0 + \mathbf{E}(-X_N) I_{A^c} \leq \mathbf{E}X_0 + \mathbf{E}(-X_N)^+ I_{A^c} \\ &\leq \mathbf{E}X^0 + \mathbf{E}X_N^-. \end{aligned}$$

The last statement (3) is a combination of (1) and (2). □

We finish this section by the well-known and helpful results of Doob about estimation of the number of upcrossings (downcrossings) of given interval by submartingales and supermartingales.

Let $a < b$ be real numbers and $(X_n)_{n=0,1,\dots,N}$ be a stochastic sequence. Define the following sequence of stopping times:

$$\begin{aligned} \tau_0 &= 0, \quad \tau_1 = \inf\{n > 0 : X_n \leq a\}, \\ \tau_2 &= \inf\{n > \tau_1 : X_n \geq b\}, \dots, \\ \tau_{2m-1} &= \inf\{n > \tau_{2m-2} : X_n \leq a\}, \\ \tau_{2m} &= \inf\{n > \tau_{2m-1} : X_n \geq b\}, \end{aligned}$$

where $\tau_k = N$ if the set in brackets above is $\emptyset$.

It is clear that between τ_{2m-1} and τ_{2m} that is an *upcrossing* of the interval (a, b) by the sequence (X_n) occurs.

Define the following random variable

$$\beta_+^X(N, a, b) = \begin{cases} 0, & \text{if } \tau_2 > N, \\ \max\{m : \tau_{2m} \leq N\}, & \text{if } \tau_2 \leq N, \end{cases}$$

which is called the *number of upcrossings* of (a, b) by the sequence (X_n) during the time interval $[0, N]$.

Definition of notion of downcrossings is given in a similar way. Denote the corresponding number of downcrossings by $\beta_-^X(N, a, b)$.

Theorem 6.4 *Let $(X_n)_{n=0,1,\dots,N}$ be a submartingale and $\beta_{\pm}^X(N, a, b)$ be the number of upcrossings (downcrossings) of the interval (a, b) . Then*

$$\mathbf{E}\beta_+^X(N, a, b) \leq \frac{\mathbf{E}(X_N - a)^+}{b - a}, \quad (6.10)$$

$$\mathbf{E}\beta_-^X(N, a, b) \leq \frac{\mathbf{E}(X_N - b)^+}{b - a}. \quad (6.11)$$

Proof We prove only (6.10) in view a symmetry of formulas (6.10) and (6.11). Let us reduce the initial problem to estimation of the number of upcrossings by a nonnegative submartingale $((X_n - a)^+)_{n=0,1,\dots,N}$ of the interval $(0, b - a)$. Moreover, we can put $a = 0$ and prove that

$$\mathbf{E}\beta_+^X(N, 0, b) \leq \frac{\mathbf{E}X_N}{b}, \quad (6.12)$$

for a nonnegative submartingale $(X_n)_{n=0,1,2,\dots,N}$, $X_0 = 0$.

Define for $i = 1, 2, \dots$ a sequence of random variables:

$$\phi_i = \begin{cases} 1, & \text{if } \tau_m < i \leq \tau_{m+1} \text{ for some non-even } m, \\ 0, & \text{if } \tau_m < i \leq \tau_{m+1} \text{ for some even } m. \end{cases}$$

It follows from definitions of $\beta_+^X(N, a, b)$ and ϕ_i that

$$b\beta_+^X(N, 0, b) \leq \sum_{i=1}^N \phi_i (X_i - X_{i-1}).$$

Let us note that

$$\{\phi_i = 1\} = \cup_{\text{non-even } m} (\{\tau_m < i\} \notin \{\tau_{m+1} < i\}) \in \mathcal{F}_{i-1},$$

and therefore by properties of conditional expected values and submartingales

$$\begin{aligned} b\mathbf{E}\beta_+^X(N, 0, b) &\leq \mathbf{E} \sum_{i=1}^N \phi_i (X_i - X_{i-1}) \\ &= \sum_{i=1}^N \int_{\{\phi_i\}} (X_i - X_{i-1}) dP \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^N \int_{\{\phi_i\}} \mathbf{E}(X_i - X_{i-1} | \mathcal{F}_{i-1}) dP \\
&\leq \sum_{i=1}^N \mathbf{E} \mathbf{E}(X_i - X_{i-1} | \mathcal{F}_{i-1}) \\
&= (\mathbf{E}X_1 - \mathbf{E}X_0) + (\mathbf{E}X_2 - \mathbf{E}X_1) + \dots \\
&\quad + (\mathbf{E}X_N - \mathbf{E}X_{N-1}) = \mathbf{E}X_N,
\end{aligned}$$

and we arrive to (6.12), and further to (6.10). $\square$

Corollary 6.1 *Let $(X_n)_{n=0,1,\dots,N}$ be a supermartingale and $\alpha_+^X(N, a, b)$ be the number of upcrossings of the interval (a, b) . Then*

$$\mathbf{E}\alpha_+^X(N, a, b) \leq \frac{E(a - X_N)^+}{b - a}. \quad (6.13)$$

Proof The inequality (6.13) follows from (6.11) because $\alpha_+^X(N, a, b)$ can be interpreted as the number of downcrossings of the interval $(-b, -a)$ by the submartingale $(-X_n)_{n=0,1,\dots,N}$:

$$\mathbf{E}\alpha_+^X(N, a, b) = \mathbf{E}\beta_-^{-X}(N, -b, -a) \leq \frac{\mathbf{E}(-X_N - (-a))^+}{(-a) - (-b)} = \frac{\mathbf{E}(a - X_N)^+}{b - a}.$$

$\square$

6.3 Martingales on Infinite Time Interval

Here martingales (submartingales, supermartingales) are studied for $\mathbb{Z}_+ = \{0, 1, \dots, n, \dots\}$. Statements of Theorems 6.3 and 6.4 are transformed by taking the limits as $N \rightarrow \infty$, and we get the following list of inequalities (6.14)–(6.19):

$$P(\omega : \sup_n X_n \geq \lambda) \leq \frac{\sup_n \mathbf{E}X_n^+}{\lambda}, \quad \lambda > 0, \quad (6.14)$$

for a submartingale $(X_n)_{n=0,1,\dots}$ with $\sup_n \mathbf{E}X_n^+ < \infty$;

$$P(\omega : \sup_n X_n \geq \lambda) \leq \frac{\mathbf{E}X_0 + \sup_n \mathbf{E}X_n^-}{\lambda}, \quad \lambda > 0, \quad (6.15)$$

for a supermartingale $(X_n)_{n=0,1,\dots}$ with $\sup_n \mathbf{E}X_n^- < \infty$;

$$P(\omega : \sup_n |X_n| \geq \lambda) \leq \frac{\sup_n \mathbf{E}|X_n|}{\lambda}, \quad \lambda > 0, \quad (6.16)$$

for a martingale $(X_n)_{n=0,1,\dots,N}$ with $\sup_n \mathbf{E}|X_n| < \infty$;

$$\mathbf{E}\beta_+^X(\infty, a, b) \leq \frac{\sup_n \mathbf{E}(X_N - a)^+}{b - a}, \quad (6.17)$$

for a submartingale $(X_n)_{n=0,1,\dots}$ with $\sup_n \mathbf{E}(X_n - a)^+ < \infty$;

$$\mathbf{E}\beta_-^X(\infty, a, b) \leq \frac{\sup_n \mathbf{E}(X_N - b)^+}{b - a}, \quad (6.18)$$

for a submartingale $(X_n)_{n=0,1,\dots}$ with $\sup_n \mathbf{E}(X_n - b)^+ < \infty$;

$$\mathbf{E}\alpha_+^X(\infty, a, b) \leq \frac{\sup_n \mathbf{E}(a - X_N)^+}{b - a}, \quad (6.19)$$

for a supermartingale $(X_n)_{n=0,1,\dots}$ with $\sup_n \mathbf{E}(a - X_n)^+ < \infty$.

The *Doob optional stopping (sampling) theorem* admits a natural extension to $\mathbb{Z}_+$ too.

Theorem 6.5 *Let $(X_n)_{n=0,1,\dots}$ be a uniformly integrable martingale (submartingale, supermartingale) and τ be a finite stopping time. Then $\mathbf{E}|X_\tau| < \infty$ and*

$$\mathbf{E}X_\tau = \mathbf{E}X_0 \quad (\mathbf{E}X_\tau \geq \mathbf{E}X_0, \mathbf{E}X_\tau \leq \mathbf{E}X_0). \quad (6.20)$$

Moreover, for finite stopping times $\tau \leq \sigma$ (a.s.):

$$\mathbf{E}(X_\sigma | \mathcal{F}_\tau) = X_\tau \quad (a.s.), \quad (\mathbf{E}(X_\sigma | \mathcal{F}_\tau) \geq X_\tau \quad (a.s.), \quad \mathbf{E}(X_\sigma | \mathcal{F}_\tau) \leq X_\tau \quad (a.s.)). \quad (6.21)$$

Proof Both formulas (6.20) and (6.21) are proved with the help of limit arguments. Let us only provide the proof of (6.20).

For a fixed N , denote $\tau_N = \tau \wedge N$, which is a bounded stopping time, and by Theorem 6.2, we have

$$\mathbf{E}X_0 = \mathbf{E}X_{\tau_N}. \quad (6.22)$$

Let us note from (6.22) that

$$\mathbf{E}X_{\tau_N} = 2\mathbf{E}X_{\tau_N}^+ - \mathbf{E}X_{\tau_N} \leq 2\mathbf{E}X_{\tau_N}^+ - \mathbf{E}X_0. \quad (6.23)$$

Further, using (6.23) and a submartingale property of $(X_n^+)_{n=0,1,\dots}$, we obtain that

$$\begin{aligned} \mathbf{E}X_{\tau_N}^+ &= \sum_{j=0}^N \mathbf{E}X_j^+ I_{\{\tau_N=j\}} + \int_{\{\tau>N\}} X_N^+ dP \\ &\leq \sum_{j=0}^N \mathbf{E}X_j^+ I_{\{\tau_N=j\}} + \mathbf{E}X_N^+ I_{\{\tau>N\}} \\ &= \mathbf{E}X_N^+ \leq \mathbf{E}|X_N| \leq \sup_n \mathbf{E}|X_n|, \end{aligned}$$

and, hence,

$$\mathbf{E}|X_{\tau_N}| \leq 3 \sup_n \mathbf{E}|X_n| < \infty. \quad (6.24)$$

Using (6.24) and the Fatou lemma, we get

$$\mathbf{E}|X_\tau| \leq \limsup_N \mathbf{E}|X_{\tau_N}| \leq 4 \sup_n \mathbf{E}|X_n| < \infty.$$

Let us note that for a finite stopping time τ we have

$$\lim_{n \rightarrow \infty} \mathbf{E}|X_n| I_{\{\tau>n\}} = 0, \quad (6.25)$$

because $(X_n)_{n=0,1,\dots}$ is uniformly integrable and $P(\omega : \tau(\omega) > n) \rightarrow 0, n \rightarrow \infty$.

We can write a decomposition

$$X_\tau = X_{\tau \wedge n} + (X_\tau - X_n) I_{\{\tau>n\}}$$

and its average

$$\mathbf{E}X_\tau = \mathbf{E}X_{\tau \wedge n} + \mathbf{E}X_\tau I_{\{\tau>n\}} - \mathbf{E}X_n I_{\{\tau>n\}}. \quad (6.26)$$

In Eq. (6.26), $\mathbf{E}X_{\tau \wedge n} = \mathbf{E}X_0$, because $(X_{\tau \wedge n})$ is a martingale.

The second term of (6.26)

$$\mathbf{E}(X_\tau) I_{\{\tau>n\}} = \sum_{i=n+1}^{\infty} \mathbf{E}X_i I_{\{\tau=i\}} \rightarrow 0, n \rightarrow \infty,$$

because the series

$$\mathbf{E}X_\tau = \sum_{i=0}^{\infty} \mathbf{E}X_i I_{\{\tau \geq i\}} \text{ converges.}$$

The third term of (6.26) converges to zero because (6.25). $\square$

Let us prove a key theorem of Doob about (a.s.)-convergence of submartingales.

Theorem 6.6 *Let $(X_n)_{n=0,1,\dots}$ be a submartingale satisfying condition*

$$\sup_n \mathbf{E}|X_n| < \infty. \quad (6.27)$$

Then there exists an integrable random variable $X_\infty = \lim_{n \rightarrow \infty} X_n$ (a.s.).

Proof Assume that this limit does not exist. It means that

$$P(\omega : \limsup_{n \rightarrow \infty} X_n(\omega) > \liminf_{n \rightarrow \infty} X_n(\omega)) > 0. \quad (6.28)$$

The set in (6.28) can be written as

$$\begin{aligned} & \{\omega : \limsup_n X_n > \liminf_n X_n\} \\ &= \cup_{a < b, a, b \in \mathbb{Q}} \{\omega : \limsup_n X_n > b > a > \liminf_n X_n\}, \end{aligned} \quad (6.29)$$

where $\mathbb{Q} \subseteq \mathbb{R}^1$ is the set of rational numbers.

It follows from (6.28)–(6.29) that for some rational numbers $a < b$ we have

$$P(\omega : \limsup_n X_n > b > a > \liminf_n X_n) > 0. \quad (6.30)$$

We also note that for each $n \in \mathbb{Z}$

$$\mathbf{E}X_n^+ \leq \mathbf{E}|X_n| = 2\mathbf{E}X_n^+ - \mathbf{E}X_n \leq 2\mathbf{E}X_n^+ - \mathbf{E}X_0$$

and find that for submartingale $(X_n)_{n \in \mathbb{Z}}$ the following conditions are equivalent:

$$\sup_n \mathbf{E}|X_n| < \infty \Leftrightarrow \sup_n \mathbf{E}X_n^+ < \infty. \quad (6.31)$$

Applying (6.17) together with (6.31), we obtain

$$\mathbf{E}\beta_+^X(\infty, a, b) \leq \frac{\sup_n \mathbf{E}X_n^+ + |a|}{b - a} < \infty. \quad (6.32)$$

Condition (6.32) is a contradiction with the assumption (6.30) and (6.28). Hence, there exists $X_\infty = \lim_{n \rightarrow \infty} X_n$ (a.s.) which is integrable by the Fatou lemma and $\mathbf{E}|X_\infty| \leq \sup_n \mathbf{E}|X_n| < \infty$. $\square$

Corollary 6.2

- (a) If $(X_n)_{n=0,1,\dots}$ is a nonnegative martingale, then $\sup_n \mathbf{E}|X_n| = \sup_n \mathbf{E}X_n = \mathbf{E}X_0 < \infty$, and therefore, there exists integrable $X_\infty = \lim_{n \rightarrow \infty} X_n$ (a.s.).
 (b) If $(X_n)_{n=0,1,\dots}$ is a nonpositive submartingale, then there exists integrable $X_\infty = \lim_{n \rightarrow \infty} X_n$ (a.s.) and $(X_n, \mathcal{F}_n)_{n=0,1,\dots,\infty}$ is also submartingale, where $\mathcal{F}_\infty = \sigma(\cup_{n=1}^\infty \mathcal{F}_n)$.

Proof Statement (a) is obvious. In case (b), one can apply the Fatou lemma and get

$$\mathbf{E}X_\infty = \mathbf{E} \lim_n X_n \geq \limsup_n \mathbf{E}X_n \geq \mathbf{E}X_0 > -\infty.$$

The submartingale property follows from relations:

$$\mathbf{E}(X_\infty | \mathcal{F}_m) = \mathbf{E}(\lim_n X_n | \mathcal{F}_m) \geq \limsup_n \mathbf{E}(X_n | \mathcal{F}_m) \geq X_m \text{ (a.s.)}, \quad m = 0, 1, \dots$$

$\square$

To study another type of convergence of martingales (submartingales, supermartingales) which is L^1 -convergence, we give the following example.

Example 6.1 Let $(Y_n)_{n=0,1,\dots}$ be a sequence of independent random variables such that

$$Y_n = \begin{cases} 2 & \text{with probability } 1/2, \\ 0 & \text{with probability } 1/2. \end{cases}$$

Define $X_n = \prod_{i=0}^n Y_i$ and $\mathcal{F}_n = \sigma(Y_0, \dots, Y_n)$. It is clear that $(X_n)_{n=0,1,\dots}$ is a martingale with $\mathbf{E}X_n = 1$, and hence, $X_n \rightarrow X_\infty = 0$ (a.s.).

At the same time, we have $\mathbf{E}|X_n - X_\infty| = \mathbf{E}X_n = 1$, and therefore, this martingale does not converge in space L^1 .

It means that the condition $\sup_n \mathbf{E}|X_n| < \infty$, appeared in Doob's theorem about (a.s.)-convergence, is not enough to achieve L^1 -convergence.

Theorem 6.7 If $(X_n)_{n=0,1,\dots}$ is a uniformly integrable submartingale, then there exists an integrable random variable X_∞ such that $X_n \xrightarrow[n \rightarrow \infty]{} X_\infty$ (a.s.) in space L^1 .

Moreover, $(X_n)_{n=0,1,\dots,\infty}$ will be a submartingale with respect to $(\mathcal{F}_n)_{n=0,1,\dots,\infty}$, where $\mathcal{F}_\infty = \sigma(\cup_{n=1}^\infty \mathcal{F}_n)$.

Proof Existence of X_∞ and convergence (a.s.) follow from Theorem 6.6. Further, take $c > 0$ and represent

$$\mathbf{E}X_n = \mathbf{E}X_n I_{\{X_n < -c\}} + \mathbf{E}X_n I_{\{X_n \geq -c\}}. \quad (6.33)$$

For any $\epsilon > 0$ due to uniform integrability (X_n), we can choose $c > 0$ large enough to provide inequality

$$\sup_n |\mathbf{E}X_n I_{\{X_n < -c\}}| < \epsilon, \quad (6.34)$$

for the first term in (6.33).

We also note that

$$X_n I_{\{X_n \geq -c\}} \geq X_n,$$

and by the Fatou lemma, we obtain that

$$\liminf_n \mathbf{E}X_n I_{\{X_n \geq -c\}} \geq \mathbf{E} \liminf_n X_n I_{\{X_n \geq -c\}} \geq \mathbf{E} \liminf_n X_n.$$

Hence, due to (6.34), we get from the above inequalities

$$\liminf_n \mathbf{E}X_n \geq \mathbf{E} \liminf_n X_n - \epsilon.$$

We can prove in a similar way

$$\limsup_n \mathbf{E}X_n \leq \mathbf{E} \limsup_n X_n$$

and further just repeat the same steps of the proof of the Lebesgue dominated convergence theorem and find that $\mathbf{E}|X_n - X_\infty| \rightarrow 0$, $n \rightarrow \infty$.

To finish the proof, we take $A \in \mathcal{F}_n$, and for $m \geq n$, we have

$$\mathbf{E}I_A |X_m - X_\infty| \rightarrow 0, \quad m \rightarrow \infty,$$

$$\int_A X_m dP \rightarrow \int_A X_\infty dP, \quad m \rightarrow \infty,$$

$\{\int_A X_m dP\}_{m \geq n}$ is nondecreasing.

Hence, for any $A \in \mathcal{F}_n$, we have

$$\int_A X_n dP \leq \int_A X_m dP \leq \int_A X_\infty dP = \int_A \mathbf{E}(X_\infty | \mathcal{F}_n) dP,$$

which certifies a submartingale property. □

Corollary 6.3 *Let $(X_n)_{n=0,1,\dots}$ be a submartingale with $\sup_n \mathbf{E}|X_n|^p < \infty$, $p > 1$. Then there exists an integrable random variable X_∞ such that $X_n \rightarrow X_\infty$, $n \rightarrow \infty$, (a.s.) and in L^1 .*

Theorem 6.8 (Levy's Theorem) *Consider an integrable random variable X on stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F})_{n=0,1,\dots}, P)$. Then*

$$\mathbf{E}(X|\mathcal{F}_n) \xrightarrow{n \rightarrow \infty} \mathbf{E}(X|\mathcal{F}_\infty) \text{ (a.s.) and in } L^1.$$

Proof Denote $X_n = \mathbf{E}(X|\mathcal{F}_n)$, $n = 0, 1, \dots$ and find for $a > 0$ and $b > 0$ that

$$\begin{aligned} \int_{\{|X_n| \geq a\}} |X_n| dP &\leq \int_{\{|X_n| \geq a\}} \mathbf{E}(|X_n||\mathcal{F}_n) dP = \int_{\{|X_n| \geq a\}} |X| dP \\ &= \int_{\{|X_n| \geq a\} \cap \{|X_n| \leq b\}} |X| dP + \int_{\{|X_n| \geq a\} \cap \{|X_n| > b\}} |X| dP \\ &\leq bP(|X_n| \geq a) + \int_{\{|X_n| > b\}} |X| dP \\ &\leq \frac{b}{a} \mathbf{E}|X| + \int_{\{|X_n| > b\}} |X| dP. \end{aligned}$$

So, taking $b \rightarrow \infty$ and after $a \rightarrow \infty$, we get

$$\lim_{a \rightarrow \infty} \sup_n \mathbf{E}|X_n| I_{\{|X_n| \geq a\}} = 0,$$

which means that $(X_n)_{n=0,1,\dots}$ is uniformly integrable. Applying Theorem 6.7, we get the statement of this theorem. $\square$

Corollary 6.4 *A uniformly integrable stochastic sequence $(X_n)_{n=0,1,\dots}$ on stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F})_{n=0,1,\dots}, P)$ is a martingale $\Leftrightarrow$ there exists an integrable random variable X such that $X_n = \mathbf{E}(X|\mathcal{F}_n)$, $n = 0, 1, \dots$*

Proof The inverse implication $\Leftarrow$ is just the Levy theorem (Theorem 6.8). The direct implication $\Rightarrow$ is just a consequence of Theorem 6.7 if we take $X = X_\infty$. $\square$

Problem 6.2 In Levy's theorem, prove that

$$X_\infty = \mathbf{E}(X|\mathcal{F}_\infty).$$

Chapter 7

Discrete Time Stochastic Analysis: Further Results and Applications



In this chapter, a characterization of sets of convergence of martingale is given in predictable terms. As a consequence, the strong LNL for square-integrable martingales is proved. This result is applied for derivation of strong consistency of the least-squares estimates in the framework of regression model with martingale errors. Moreover, the CLT for martingales is stated, and further this theorem together with the martingale LNL is applied to derive the asymptotic normality and strong consistency of martingale stochastic approximation procedures. A discrete version of the Girsanov theorem is given here with its further application for derivation of a discrete time Bachelier option pricing formula. In the last section, the notion of a martingale is extended in several directions: from asymptotic martingales and local martingales to martingale transforms and generalized martingales (see [10, 15, 19, 20, 28, 43, 47, 49], and [55]).

7.1 Limiting Behavior of Martingales with Statistical Applications

Let us investigate limiting behavior of martingales. We start with some facts about sets of convergence of martingales and submartingales.

Definition 7.1 Let $(X_n)_{n=0,1,\dots}$ be a stochastic sequence. Then we denote $\{\omega : X_n \rightarrow \}$ the set of those $\omega \in \Omega$ such that $X_n(\omega)$ converges to a finite limit as $n \rightarrow \infty$.

We also say that $A \subseteq B \in \mathcal{F}$ (a.s.), if $P(A \cap B^c) = 0$, and $A = B$ (a.s.) if $A \subseteq B$ (a.s.) and $B \subseteq A$ (a.s.).

Lemma 7.1 *Let $(X_n)_{n=0,1,\dots}$ be a square-integrable martingale. Then (a.s.)*

$$\{\omega : \langle X, X \rangle_\infty < \infty\} \subseteq \{\omega : X_n \rightarrow\}. \quad (7.1)$$

Proof For a positive $a \in \mathbb{R}$, we define a stopping time

$$\tau_a = \inf\{n : \langle X, X \rangle_{n+1} \geq a\},$$

assuming that $\inf\{\emptyset\} = \infty$. One can observe that $\langle X, X \rangle_{\tau_a} \leq a$ and the stopped sequence $(X_n^{\tau_a})_{n=0,1,\dots} = (X_{\tau_a \wedge n})_{n=0,1,\dots}$ is a square-integrable martingale such that

$$\mathbf{E}(X_{\tau_a \wedge n})^2 = \mathbf{E}(X_n^{\tau_a})^2 = \mathbf{E}\langle X^{\tau_n}, X^{\tau_n} \rangle_n = \mathbf{E}\langle X, X \rangle_{\tau_a \wedge n} \leq a.$$

According to Theorem 6.6, $\mathbf{P}\{\omega : X_n^{\tau_n} \rightarrow\} = 1$. We also note that $X_n^{\tau_n} = X_{\tau_a \wedge n} = X_n$ on the set $\{\tau_a = \infty\} = \{\omega : \langle X, X \rangle_\infty \leq a\}$.

Further, (a.s.) $\{\omega : \tau_a = \infty\} \subseteq \{\omega : X_n \rightarrow\}$ and $\cup_{a>0}\{\tau_a = \infty\} \subseteq \{\omega : X_n \rightarrow\}$. Hence, (a.s.)

$$\begin{aligned} \{\omega : \langle X, X \rangle_\infty < \infty\} &= \cup_{a>0}\{\omega : \langle X, X \rangle_\infty \leq a\} \\ &= \cup_{a>0}\{\omega : \tau_a = \infty\} \subseteq \{\omega : X_n \rightarrow\}, \end{aligned}$$

and we get (7.1). □

Lemma 7.2 (Stochastic Kronecker's Lemma) *Let $(A_n)_{n=0,1,\dots}$, $A_0 > 0$, be a predictable nondecreasing sequence, $(M_n)_{n=0,1,\dots}$ be a square-integrable martingale and*

$$N_n = \sum_{i=0}^n A_i^{-1} \Delta M_i.$$

Then (a.s.)

$$\Omega' = \{\omega : A_\infty = \infty\} \cap \{\omega : N_n \rightarrow\} \subseteq \{\omega : A_n^{-1} M_n \rightarrow\}. \quad (7.2)$$

Proof First of all, we note the following formula of “summation by parts”:

$$A_n X_n = A_0 X_0 + \sum_{i=1}^n A_i \Delta X_i + \sum_{i=1}^n X_{i-1} \Delta A_i, \quad (7.3)$$

for two stochastic sequences $(A_n)_{n=0,1,\dots}$ and $(X_n)_{n=0,1,\dots}$.

Applying the formula (7.3) to $(A_n)_{n=0,1,\dots}$ and $(N_n)_{n=0,1,\dots}$, we represent

$$\sum_{i=1}^n A_i \Delta N_i = A_n N_n - \sum_{i=1}^n N_{i-1} \Delta A_i,$$

and using $\sum_{i=1}^n A_i \Delta N_i = M_n$, we obtain

$$\begin{aligned} A_n^{-1} M_n &= A_n^{-1} \sum_{i=1}^n A_i \Delta N_i = A_n^{-1} \left[A_n N_n - \sum_{i=1}^n N_{i-1} \Delta A_i \right] \\ &= A_n^{-1} \sum_{i=1}^n (N_n - N_{i-1}) \Delta A_i. \end{aligned} \quad (7.4)$$

Further, for $\epsilon > 0$, define

$$n_\epsilon = \begin{cases} \sup\{i : |N_\infty - N_{i-1}| \leq \epsilon\} & \text{on the set } \{N_n \rightarrow N_\infty\}, \\ 0, & \text{on } \Omega \setminus \{N_n \rightarrow N_\infty\}. \end{cases}$$

By (7.4), on the set $\{\omega : N_n \rightarrow N_\infty\}$, we have

$$\begin{aligned} |A_n^{-1} M_n| &= |A_n^{-1} \sum_{i=1}^n (N_n - N_{i-1}) \Delta A_i| \\ &\leq A_n^{-1} \left| \sum_{i=1}^n (N_n - N_{i-1}) \Delta A_i \right| \\ &\leq A_n^{-1} \left[\sum_{i=1}^{n \wedge n_\epsilon} |N_n - N_{i-1}| \Delta A_i + \sum_{i=n \wedge n_\epsilon}^n [|N_\infty - N_n| \right. \\ &\quad \left. + |N_\infty - N_{i-1}|] \Delta A_i \right] \\ &\leq A_n^{-1} \left[\text{const.} A_{n_\epsilon} \sup_i |N_i| + \text{const.} A_n |N_\infty - N_n| + \text{const.} \epsilon A_n \right]. \end{aligned} \quad (7.5)$$

Hence, for almost all $\omega \in \Omega'$, one can find $n'_\epsilon(\omega)$ such that in (7.5) for $n \geq n'_\epsilon(\omega)$: $|A_n^{-1} M_n(\omega)| \leq \text{const.}(\omega) \cdot \epsilon$, and therefore (7.2). $\square$

Theorem 7.1 (Strong LNL for Martingales) *Let $(A_n)_{n=0,1,\dots}$ be nonnegative predictable nondecreasing sequence and $(M_n)_{n=0,1,\dots}$, $M_0 = 0$, be a square-*

integrable martingale such that (a.s.) $A_\infty = \infty$ and (a.s.)

$$\sum_{n=1}^{\infty} \frac{\Delta \langle M, M \rangle_n}{A_n^2} < \infty. \quad (7.6)$$

Then $A_n^{-1} M_n \rightarrow 0$ (a.s.), $n \rightarrow \infty$.

Proof Due to assumptions (7.6) and Lemma 7.1, we observe in Lemma 7.2 $P(\Omega') = 1$, and hence, we get the statement of the theorem. $\square$

Corollary 7.1 (Strong LNL of Kolmogorov) Let $(Y_n)_{n=1,2,\dots}$ be a sequence of independent random variables with $\mathbf{E}Y_n = 0$ and $\sigma_n^2 = \mathbf{E}Y_n^2$ such that

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty. \quad (7.7)$$

Then for a sequence $X_n = \sum_{i=1}^n Y_i$ the following strong large numbers law is true: (a.s.) $\frac{X_n}{n} \rightarrow 0$, $n \rightarrow \infty$.

Proof We put $A_n = n$, $M_n = X_n$ and find that the Kolmogorov variance condition (7.7) is transformed into the condition (7.6) of Theorem 7.1 and we get the claim. $\square$

Let us give an interesting and valuable application of strong LNL for martingales to regression analysis.

Example 7.1 Suppose the observations are performed at times $n = 0, 1, \dots$ and obey the formula

$$x_n = f_n \theta + e_n, \quad (7.8)$$

where $\theta \in \mathbb{R}$ is an unknown parameter, $(e_n)_{n=0,1,\dots}$ is a martingale difference, and $(f_n)_{n=0,1,\dots}$ is a predictable regressor.

One can consider a structural least-squares estimate (LS estimate) in the framework of regression model (7.8):

$$\theta_n = \left(\sum_{i=0}^n f_i^2 \right)^{-1} \sum_{i=0}^n f_i x_i. \quad (7.9)$$

Denote $D_n = \mathbf{E}(e_n^2 | \mathcal{F}_{n-1})$ and $F_n = \sum_{i=0}^n f_i^2$ and assume that $F_n \rightarrow \infty$ (a.s.), $n \rightarrow \infty$, and $\sum_{i=0}^{\infty} F_i^{-2} D_i < \infty$ (a.s.). Then according to Theorem 7.1, LS estimate (7.9) is strongly consistent in the sense that $\theta_n \rightarrow \theta$ (a.s.), $n \rightarrow \infty$.

Remark 7.1 If in the model (7.8) we take $f_n = x_{n-1}$, then we arrive to the first order autoregression model for which the condition $\sum_{i=1}^{\infty} x_{i-1}^2 = \infty$ is well-known in regression analysis as a standard guarantee for consistency of LS estimates.

Below we give some additional facts about asymptotic behavior of martingales and submartingales.

Definition 7.2 A stochastic sequence $(X_n)_{n=0,1,\dots} \in C^+$, if for each $\tau_a = \inf\{n \geq 0 : X_n > a\}$, $a > 0$:

$$\mathbf{E}(\Delta X_{\tau_a})^+ I_{\{\omega: \tau_a < \infty\}} < \infty.$$

Problem 7.1 Prove that the condition C^+ is true for any stochastic sequence with $\mathbf{E} \sup_n |\Delta X_n| < \infty$.

Theorem 7.2

(a) Assume a submartingale with its Doob decomposition $X_n = X_0 + A_n + M_n$, $n = 0, 1, \dots$, belongs to a class C^+ , and then (a.s.)

$$\{X_n \rightarrow\} = \{\sup_n X_n < \infty\} \subseteq \{A_\infty < \infty\}.$$

(b) Assume a submartingale $X_n = X_0 + A_n + M_n$, $n = 0, 1, \dots$, is nonnegative, and then (a.s.)

$$\{A_\infty < \infty\} \subseteq \{X_n \rightarrow\} \subseteq \{\sup_n X_n < \infty\}.$$

(c) If a nonnegative submartingale belongs to C^+ , then (a.s.)

$$\{A_\infty < \infty\} = \{X_n \rightarrow\} = \{\sup_n X_n < \infty\}.$$

(d) Let $(Y_n)_{n=0,1,\dots}$ be a nonnegative stochastic sequence with the structure

$$Y_n = Y_0 + A'_n + M_n - A_n^2, \quad n = 0, 1, \dots,$$

where $(A'_n)_{n=0,1,\dots}$ and $(A_n^2)_{n=0,1,\dots}$ are nondecreasing predictable sequences and $(M_n)_{n=0,1,\dots}$ is a martingale. Then (a.s.)

$$\{A'_\infty < \infty\} \subseteq \{Y_n \rightarrow\} = \{\sup_n A_n^2 < \infty\}.$$

Proof

(a) It is clear that (a.s.) $\{X_n \rightarrow\} \subseteq \{\sup_n X_n < \infty\}$. To prove the inverse inclusion, we have

$$\sup_n \mathbf{E} X_{\tau_a}^+ \leq a + \mathbf{E} X_{\tau_a}^+ I_{\tau_a < \infty} \leq 2a + \mathbf{E}(\Delta X_{\tau_a})^+ I_{\{\tau_a < \infty\}} < \infty.$$

Hence, the stopped submartingale $(X_n^{\tau_a})_{n=0,1,\dots} = (X_{\tau_a \wedge n})_{n=0,1,\dots}$ converges (a.s.) as $n \rightarrow \infty$, and $\{\tau_a = \infty\} \subseteq \{X_n \rightarrow\}$ (a.s.).

Due to $\cup_{a>0}\{\tau_a = \infty\} = \{\sup_n X_n < \infty\}$, we get

$$\{\sup_n X_n < \infty\} = \{X_n \rightarrow\} \text{ (a.s.)}.$$

(b) Define $\sigma_a = \inf(n : A_{n+1} > a)$, $a > 0$, and get

$$\mathbf{E}X_{n \wedge \sigma_a} = \mathbf{E}A_{n \wedge \sigma_a} \leq a.$$

Hence, $\sup_n \mathbf{E}X_{n \wedge \sigma_a} \leq a < \infty$, and due to positivity of $(X_n)_{n=0,1,\dots}$, we can apply the Doob convergence theorem and find that

$$\{A_\infty \leq a\} \subseteq \{\sigma_a = \infty\} = \{X_n \rightarrow\} \text{ (a.s.)},$$

and therefore,

$$\{A_\infty < \infty\} = \cup_{a>0}\{A_\infty \leq a\} \subseteq \{X_n \rightarrow\} \text{ (a.s.)}.$$

(c) is a combination of (a) and (b).

(d) Let us rewrite $Y_n = X_n - A_n^2$, $n = 0, 1, \dots$, where $X_n = X_0 + A'_n + M_n$. $X_0 = Y_0$, is a nonnegative submartingale due to $0 \leq Y_n = X_n - A_n^2$, $0 \leq A_n^2 \leq X_n$ (a.s.), and (a.s.) $\{A'_\infty < \infty\} \subseteq \{X_n \rightarrow\} \subseteq \{A_\infty^2 < \infty\}$ by (b). Hence, (a.s.) $\{A'_\infty < \infty\} \subseteq \{Y_n \rightarrow\} \cap \{A_\infty^2 < \infty\}$.

□

Corollary 7.2 Assume a martingale $(X_n)_{n=0,1,\dots}$ satisfies to condition $\mathbf{E} \sup_n |\Delta X_n| < \infty$. Then (a.s.)

$$\Omega = \{X_n \rightarrow\} \cup \{\liminf_n X_n = -\infty, \limsup_n X_n = +\infty\}. \quad (7.10)$$

Proof We apply Theorem 7.2 to $\pm(X_n)_{n=0,1,\dots}$ and get (a.s.)

$$\{\limsup_n X_n < \infty\} = \{\sup_n X_n < \infty\} = \{X_n \rightarrow\},$$

$$\{\liminf_n X_n > -\infty\} = \{\inf_n X_n > -\infty\} = \{X_n \rightarrow\},$$

which leads to (7.10). □

We already studied the large numbers law for martingales and a very nice generalization of the corresponding results for sums of independent random variables observed.

A natural question arises.

Is it possible to get a similar version of the *central limit theorem (CLT)* for *martingales*?

The next theorem gives a positive answer to this question.

Theorem 7.3 *Let $(X_n)_{n=0,1,\dots}$ be a martingale difference such that (a.s.) $\mathbf{E}(X_n^2|\mathcal{F}_{n-1}) = 1$ and $\mathbf{E}(|X_n|^3|\mathcal{F}_{n-1}) \leq C < \infty$. Then for a martingale $M_n = \sum_{i=1}^n X_i$, the CLT is true:*

$$Y_n = \frac{M_n}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} Z \sim N(0, 1).$$

Proof Here, as in Chap. 4 for a random variable X , we define its characteristic function $\phi_X(t) = \mathbf{E}e^{itx}$, $t \in \mathbb{R}$, and denote

$$\phi_{n,j}(t) = \mathbf{E} \left(e^{it \frac{X_j}{\sqrt{n}}} | \mathcal{F}_{j-1} \right).$$

Let us write the Taylor decomposition

$$e^{it \frac{X_j}{\sqrt{n}}} = 1 + it \frac{X_j}{\sqrt{n}} - \frac{t^2}{2n} X_j^2 - \frac{it^3}{6n^{3/2}} \bar{X}_j^3, \quad (7.11)$$

with a random reminder bounded as follows $0 \leq \bar{X}_j \leq |X_j|$.

Taking the conditional expectation in (7.11) and exploiting a martingale difference assumption, we obtain the corresponding representation for $\phi_{n,j}(t)$:

$$\phi_{n,j}(t) = 1 + it \frac{\mathbf{E}(X_j|\mathcal{F}_{j-1})}{\sqrt{n}} - \frac{t^2}{2n} \mathbf{E}(X_j^2|\mathcal{F}_{j-1}) - \frac{it^3}{6n^{3/2}} \mathbf{E}(\bar{X}_j^3|\mathcal{F}_{j-1}). \quad (7.12)$$

It follows from (7.12) that

$$\phi_{n,j}(t) - 1 - \frac{t^2}{2n} = \frac{it^3}{6n^{3/2}} \mathbf{E}(\bar{X}_j^3|\mathcal{F}_{j-1}).$$

Hence, for $m \leq n$, we obtain from (7.10) that

$$\begin{aligned} \mathbf{E}e^{it \frac{M_m}{\sqrt{n}}} &= \mathbf{E}e^{it \frac{M_{m-1}}{\sqrt{n}}} e^{it \frac{X_m}{\sqrt{n}}} = \mathbf{E} \left(e^{it \frac{M_{m-1}}{\sqrt{n}}} \mathbf{E} \left(e^{it \frac{X_m}{\sqrt{n}}} | \mathcal{F}_{m-1} \right) \right) \\ &= \mathbf{E}e^{it \frac{M_{m-1}}{\sqrt{n}}} \phi_{n,m}(t) \\ &= \mathbf{E}e^{it \frac{M_{m-1}}{\sqrt{n}}} \left(1 - \frac{t^2}{2n} - \frac{it^3}{6n^{3/2}} \bar{X}_j^3 \right). \end{aligned} \quad (7.13)$$

Let us rewrite (7.11) as follows:

$$\mathbf{E} \left(e^{it \frac{M_m}{\sqrt{n}}} - \left(1 - \frac{t^2}{2n}\right) e^{it \frac{M_{m-1}}{\sqrt{n}}} \right) = \mathbf{E} e^{it \frac{M_{m-1}}{\sqrt{n}}} \frac{it^3}{6n^{3/2}} \bar{X}_j^3. \quad (7.14)$$

Using the boundedness of the third moment of $(X_n)_{n=0,1,\dots}$, we derive from (7.13) that

$$\begin{aligned} & \left| \mathbf{E} \left(e^{it \frac{M_m}{\sqrt{n}}} - \left(1 - \frac{t^2}{2n}\right) e^{it \frac{M_{m-1}}{\sqrt{n}}} \right) \right| \\ & \leq \mathbf{E} \left(\left| e^{it \frac{M_{m-1}}{\sqrt{n}}} \right| \frac{|t|^3}{6n^{3/2}} \mathbf{E} (|X_m|^3 | \mathcal{F}_{m-1}) \right) \leq c \frac{|t|^3}{6n^{3/2}}. \end{aligned} \quad (7.15)$$

Let us fix $t \in \mathbb{R}$ and choose large enough $n \geq \frac{t^2}{2}$ to provide the following inequality $0 \leq 1 - t^2/2n \leq 1$.

For such t and n , we obtain from (7.13)–(7.15) that

$$\left| \left(1 - \frac{t^2}{2n}\right)^{n-m} \mathbf{E} e^{it \frac{M_{m-1}}{\sqrt{n}}} - \left(1 - \frac{t^2}{2n}\right)^{n-m+1} \mathbf{E} e^{it \frac{M_{m-1}}{\sqrt{n}}} \right| \leq c \frac{|t|^3}{6n^{3/2}}. \quad (7.16)$$

Let us note that

$$\begin{aligned} & \mathbf{E} e^{it \frac{M_m}{\sqrt{n}}} - \left(1 - \frac{t^2}{2n}\right)^n \\ & = \sum_{m=1}^n \left[\left(1 - \frac{t^2}{2n}\right)^{n-m} \mathbf{E} e^{it \frac{M_{m-1}}{\sqrt{n}}} - \left(1 - \frac{t^2}{2n}\right)^{n-m+1} \mathbf{E} e^{it \frac{M_{m-1}}{\sqrt{n}}} \right]. \end{aligned} \quad (7.17)$$

Relations (7.16)–(7.17) for $n \geq t^2/2$ lead to the following inequality

$$\left| \mathbf{E} e^{it \frac{M_m}{\sqrt{n}}} - \left(1 - \frac{t^2}{2n}\right)^n \right| \leq nC \frac{|t|^3}{6n^{3/2}} = C \frac{|t|^3}{6\sqrt{n}}. \quad (7.18)$$

Due to the right-hand side of (7.18) tends to zero, now one can conclude that

$$\lim_{n \rightarrow \infty} \mathbf{E} e^{it \frac{M_m}{\sqrt{n}}} = e^{-t^2/2}, \quad (7.19)$$

where we used the well-known fact that

$$\lim_{n \rightarrow \infty} \left(1 - \frac{t^2}{2n}\right)^n = e^{-t^2/2}.$$

Relation (7.19) shows that characteristic functions of Y_n converge to the characteristic function of $N(0, 1)$. Hence, $Y_n \xrightarrow[n \rightarrow \infty]{d} Z \sim N(0, 1)$. $\square$

Another area of applications of a martingale LNL and a martingale CLT is stochastic approximation algorithms. The classical theory of stochastic approximation is concerned with the problem to construct a stochastic sequence $(\theta_n)_{n=0,1,\dots}$ that converges in some probabilistic sense to a unique root θ of the regression equation

$$R(\theta) = 0, \quad \theta \in \mathbb{R}, \quad (7.20)$$

where R is a regression function.

An approximate solution of (7.20) is given by the Robbins-Monro procedure

$$\theta_n = \theta_{n-1} - \gamma_n y_n, \quad (7.21)$$

with the sequence (y_n) satisfying to equation

$$y_n = R(\theta_{n-1}) + e_n, \quad (7.22)$$

where $(e_n)_{n=0,1,\dots}$ are errors of observations usually modeled by a sequence of independent random variables with mean zero and bounded variance, and $(\gamma_n)_{n=0,1,\dots}$ is a sequence of positive real numbers converging to 0.

The convergence of the procedure (7.21)–(7.22) for a continuous linearly bounded regression function $R(x)$ with $R(x)(x - \theta) > 0$ for all $x \in \mathbb{R}$ is guaranteed by the following classical conditions:

$$\sum_1^\infty \gamma_k = \infty, \quad (7.23)$$

$$\sum_1^\infty \gamma_k^2 < \infty. \quad (7.24)$$

Let us demonstrate how the martingale technique and convergence work here and how classical stochastic approximation results are extended to models (7.21)–(7.22) with martingale errors $(e_n)_{n=0,1,\dots}$ and predictable sequences $(\gamma_n)_{n=0,1,\dots}$.

Example 7.2 We consider for simplicity the case of linear function $R(x) = \beta(x - \theta)$, $\beta > 0$, assuming in (7.22) that $(e_n)_{n=0,1,\dots}$ is a martingale difference with $\mathbf{E}e_n = 0$ and $\mathbf{E}(e_n^2 | \mathcal{F}_{n-1}) \leq \xi < \infty$ (a.s.). Regarding (γ_n) , we assume that (γ_n) is predictable, $0 < \gamma_n \leq \beta^{-1}$ (a.s.), and conditions (7.23)–(7.24) are fulfilled almost surely. One can rewrite the algorithm (7.21)–(7.22) as follows: ($\theta = 0$ for simplicity)

$$\Delta\theta_n = -\gamma_n \beta \theta_{n-1} - \gamma_n e_n. \quad (7.25)$$

Moreover, (7.25) can be rewritten in the form of an inhomogeneous linear stochastic differential equation (6.9) with $X_n = \theta_n$, $\Delta N_n = -\gamma_n e_n$, $\Delta U_n = -\beta \gamma_n$, and $\mathcal{E}_n(-\beta \gamma)$ is a stochastic exponential of sequence (U_n) .

In these denotations, the solution of (7.25) is expressed as follows:

$$\begin{aligned} \theta_n &= X_n = \mathcal{E}_n(U) X_0 + \mathcal{E}_n(U) \sum_1^n \mathcal{E}_i^{-1}(U) \Delta N_i \\ &= \mathcal{E}_n(-\beta \gamma) \theta_0 + \mathcal{E}_n(-\beta \gamma) \sum_1^n \mathcal{E}_i^{-1}(-\beta \gamma) \gamma_i e_i. \end{aligned} \quad (7.26)$$

According to the assumption $\gamma_n < \beta^{-1}$ (a.s.), the stochastic exponential $\mathcal{E}_n(-\beta \gamma) = \prod_1^n (1 - \beta \gamma_i)$ is positive (a.s.) if (7.23) is fulfilled (a.s.).

Further, the first term of the right-hand side of (7.26) tends to zero (a.s.), $n \rightarrow \infty$, because $\mathcal{E}_n(-\beta \gamma) \rightarrow 0$ (a.s.), $n \rightarrow \infty$.

The structure of the second term of the right-hand side of (7.26) is exactly (see Theorem 7.1) as in the strong LNL for martingales with $A_n^{-1} = \mathcal{E}_n(-\beta \gamma)$ and $M_n = \sum_1^n \mathcal{E}_i^{-1}(-\beta \gamma) \gamma_i e_i$.

So, we have to check the condition (7.6) in this case:

$$\begin{aligned} \sum_1^\infty A_n^{-2} \Delta \langle M, M \rangle_n &= \sum_1^n \mathcal{E}_n^2(-\beta \gamma) \mathcal{E}_n^{-2}(-\beta \gamma) \gamma_n^2 \mathbf{E}(e_n^2 | \mathcal{F}_{n-1}) \\ &\leq \xi \sum_1^\infty \gamma_n^2 < \infty \quad (\text{a.s.}). \end{aligned}$$

Applying Theorem 7.1, we get the convergence $\theta_n \rightarrow \theta$ (a.s.), $n \rightarrow \infty$.

Let us show how the martingale CLT (see Theorem 7.3) works to study asymptotic normality properties of algorithms (7.21)–(7.22). Under some reasonable simplifications, we will show it to avoid technical difficulties.

Example 7.3 We consider the linear model (7.21)–(7.22) assuming $\gamma_n = \frac{\alpha}{n}$, $\alpha > 0$, $e_n \sim N(0, \sigma^2)$ and independent, and n must be greater than $\alpha \beta$ to provide a positivity of $\mathcal{E}_n(-\beta \gamma)$.

Problem 7.2 Prove that under the assumptions above

$$\mathcal{E}_n(-\beta \gamma) \equiv \mathcal{E}_n(-\beta \alpha) \sim n^{-\beta \alpha}, \quad n \rightarrow \infty. \quad (7.27)$$

To derive an asymptotic normality of the procedure (θ_n) , we multiply (7.26) by $n^{1/2}$ and obtain

$$\sqrt{n} \theta_n = \sqrt{n} \mathcal{E}_n(-\beta \alpha) \theta_0 - \sqrt{n} \mathcal{E}_n(-\beta \alpha) \sum_1^n \frac{\alpha}{k} \mathcal{E}_k^{-1}(-\beta \alpha) e_k. \quad (7.28)$$

It follows from (7.27) that $\sqrt{n}\mathcal{E}_n(-\beta\alpha) \sim n^{1/2-\beta\alpha}$, and, hence, under additional assumption $2\beta\alpha > 1$, we find the first term of the right side of (7.28) tends to zero (a.s.), $n \rightarrow \infty$. The second term of (7.28) has a normal distribution. Therefore, we need to calculate its limiting variance.

As $n \rightarrow \infty$, we have

$$n\mathcal{E}_n^2(-\beta\alpha) \sum_1^n \mathcal{E}_k^{-2}(-\beta\alpha) \frac{\alpha^2 \sigma^2}{k^2} \rightarrow \frac{\alpha^2 \sigma^2}{(2\beta\alpha - 1)}. \quad (7.29)$$

Thus we obtain from (7.29) that

$$\sqrt{n}\theta_n \xrightarrow[n \rightarrow \infty]{d} N\left(0, \frac{\alpha^2 \sigma^2}{2\beta\alpha - 1}\right).$$

Remark 7.2 One can develop this approach to a nonlinear regression function $R(x) = \beta(x - \theta) + U(x)$, where $U(x) = O((x - \theta)^2)$, and prove both strong consistency and asymptotic normality of stochastic approximation procedures (7.21)–(7.22). We also note that Lemma 7.2 works for such extension.

7.2 Martingales and Absolute Continuity of Measures. Discrete Time Girsanov Theorem and Its Financial Application

We start this section with a martingale characterization of absolute continuity of a probability measure $\tilde{P}$ with respect to measure P on given stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0,1,\dots}, P)$. It was already shown in Sect. 6.1 that under assumption of “local absolute continuity” $\tilde{P} \ll_{loc} P$ the corresponding local density $Z_n = \frac{d\tilde{P}_n}{dP_n}$ is a martingale w.r.t. P . The next theorem states conditions under which a “local absolute continuity” is transformed into “absolute continuity” of $\tilde{P}$ w.r.t. P .

Theorem 7.4 Assume that $\tilde{P} \ll_{loc} P$ and $(Z_n)_{n=0,1,\dots}$, $Z_0 = 1$, is a local density $\frac{d\tilde{P}_n}{dP_n}$. Then the following statements are equivalent:

- (1) $\tilde{P} \ll P$.
- (2) $(Z_n)_{n=0,1,\dots}$ is uniformly integrable.
- (3) $\tilde{P}(\omega : \sup_n Z_n < \infty) = 1$.

Proof (1) $\Rightarrow$ (3): According to Theorem 6.6, there exists $Z_\infty = \lim_{n \rightarrow \infty} Z_n$ P -a.s. Due to $\tilde{P} \ll P$, such a limit does exist $\tilde{P}$ -a.s.. Hence, $\tilde{P}(\omega : \sup_n Z_n < \infty) = 1$.

(3) $\Rightarrow$ (2): For a constant $c > 0$, we have

$$\mathbb{E} Z_n I_{\{\omega : Z_n > c\}} = \tilde{P}(\omega : Z_n > c) \leq \tilde{P}(\omega : \sup_i Z_i > c) \rightarrow 0, \quad c \rightarrow \infty.$$

(2) $\Rightarrow$ (1): Due to the uniform integrability of $(Z_n)_{n=0,1,\dots}$, we can apply Theorem 6.7 and find the existence $Z_\infty = \lim_{n \rightarrow \infty} Z_n$ P -a.s., and $\mathbf{E}|Z_n - Z_\infty| \rightarrow 0$, $n \rightarrow \infty$.

Further, for any $A \in \mathcal{F}_m$ and for $n \geq m$, we can write a martingale property of $(Z_n)_{n=0,1,\dots}$ as follows:

$$\tilde{P}(A) = \mathbf{E}I_A Z_m = \mathbf{E}I_A Z_n. \quad (7.30)$$

It follows from L^1 -convergence that one can take a limit as $n \rightarrow \infty$ in the equality (7.30) and obtain

$$\tilde{P}(A) = \mathbf{E}I_A Z_\infty,$$

$$\text{i.e., } Z_\infty = \frac{d\tilde{P}}{dP}.$$

□

Now we show a method of construction of probability measure $\tilde{P}$ and its (local) density w.r.t. P . Aiming it, we formulate one of the simplest versions of the *Girsanov theorem*.

Suppose $(\epsilon_n)_{n=0,1,\dots}$ is a sequence of independent random variables $\epsilon_n \sim N(0, 1)$. Define $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(\epsilon_1, \dots, \epsilon_n)$, $n = 1, 2, \dots$, a bounded predictable sequence $(\alpha_n)_{n=0,1,\dots}$ and $Z_0 = 1$,

$$Z_n = \exp \left\{ - \sum_1^n \alpha_k \epsilon_k - \frac{1}{2} \sum_1^n \alpha_k^2 \right\}, \quad n = 1, 2, \dots \quad (7.31)$$

Problem 7.3 Prove that $(Z_n)_{n=0,1,\dots}$ is a martingale with $\mathbf{E}Z_n = 1$.

Theorem 7.5 Define a new probability measure $\tilde{P}_N(A) = \mathbf{E}I_A Z_N$, $A \in \mathcal{F}_N$. Then $\tilde{\epsilon}_n = \alpha_n + \epsilon_n$, $n = 1, \dots, N$, is a sequence of independent standard normal random variables w.r.t. $\tilde{P}_N$.

Proof For real numbers $(\lambda_n)_{n=1,\dots,N}$, we are using the rule of changing of measures, properties of conditional expectations, and the form of characteristic function for standard normal random variables:

$$\begin{aligned} \tilde{E}_N e^{i \sum_1^N \lambda_n \tilde{\epsilon}_n} &= \mathbf{E} Z_N e^{i \sum_1^N \lambda_n \tilde{\epsilon}_n} \\ &= \mathbf{E} \left[e^{i \sum_1^{N-1} \lambda_n \tilde{\epsilon}_n} Z_{N-1} \mathbf{E} \left(e^{i \lambda_N (\alpha_N + \epsilon_N) + \alpha_N \epsilon_N - \alpha_N^2/2} \middle| \mathcal{F}_{N-1} \right) \right] \\ &= \mathbf{E} \left[e^{i \sum_1^{N-1} \lambda_n \tilde{\epsilon}_n} Z_{N-1} \right] e^{-\lambda_N^2/2} = \dots = e^{-1/2 \sum_1^N \lambda_n^2}. \end{aligned}$$

It means that characteristic function of $\sum_1^N \lambda_n \tilde{\epsilon}_n$ is a product of characteristic functions of standard normal random variables, and we get the statement of the theorem. □

Remark 7.3 One can extend Theorem 7.5 to infinite the time interval. According to Theorem 7.4, we need to provide a condition to guarantee that $(Z_n)_{n=0,1,\dots}$ is uniformly integrable. Usually, it is achieved with the help of the Novikov type condition $\mathbf{E} \exp(1/2 \sum_1^\infty \alpha_n^2) < \infty$.

Let us give an example of application of this theorem to Mathematical Finance.

Example 7.4 We assume that financial market consists of two assets $(B_n)_{n=0,1,\dots,N}$ and $(S_n)_{n=0,1,\dots,N}$ (bank account and stock prices relatively). We put for simplicity that $B_n \equiv 1$ (interest rate is zero) and

$$\Delta S_n = S_n - S_{n-1} = \alpha_n + \epsilon_n, S_0 > 0, n = 1, \dots, N. \quad (7.32)$$

We consider a standard financial contract “call option” with a strike K . The holder of the contract has the right to buy a stock share by the price K at the maturity date N . In terms of having the right, he/she must pay a premium $\mathbb{C}_N$ at time 0. The basic problem here is to determine $\mathbb{C}_N$.

The market (7.32) can be considered as the *discrete time Bachelier model* with unit volatility. It is well established that the premium must be calculated to avoid *arbitrage*. According to this financial no-arbitrage principle, $\mathbb{C}_N$ is determined as expected value of payoff $(S_N - K)^+$ w.r.t. a risk-neutral measure $\tilde{P}_N$ given in the Girsanov theorem.

To provide concrete calculations, we formulate an auxiliary fact for a Normal distribution as a problem.

Problem 7.4 Prove that

$$\mathbf{E}(a + b\epsilon)^+ = a\Phi(a/b) - b\phi(a/b), \quad (7.33)$$

where $\epsilon \sim N(0, 1)$, $a \in \mathbb{R}$, $b > 0$, $\phi = \phi(x)$ is a standard normal density and $\Phi(x) = \int_{-\infty}^x \phi(y)dy$.

To calculate $\mathbb{C}_N$, we apply (7.33) and obtain the following discrete time *formula of Bachelier*

$$\begin{aligned} \mathbb{C}_N &= \tilde{\mathbf{E}}_N(S_N - K)^+ = \tilde{\mathbf{E}}_N \left(S_0 + \sum_1^N \alpha_i + \sum_1^n \epsilon_i - K \right)^+ \\ &= \mathbf{E} \left(S_0 - K + \sum_1^N \epsilon_i \right)^+ = \mathbf{E}(S_0 - K + \sqrt{N}\epsilon)^+ \\ &= (S_0 - K)\Phi \left(\frac{S_0 - K}{\sqrt{N}} \right) + \sqrt{N}\phi \left(\frac{S_0 - K}{\sqrt{N}} \right). \end{aligned}$$

7.3 Asymptotic Martingales and Other Extensions of Martingales

We start with an idea to avoid conditional expected values to extend the notion of martingale. Let us assume that $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n=0,1,\dots}, P)$ be a stochastic basis on which all stochastic sequences are considered. Denote T_b the set of all bounded stopping times on this stochastic basis. We defined before that two stopping times $\tau \leq \sigma$ if $\tau(\omega) \leq \sigma(\omega)$ (a.s.). With this definition, the set T_b is a *directed set* filtering to the right.

Definition 7.3 A stochastic sequence of integrable random variables $(X_n)_{n=0,1,\dots}$ is called an *asymptotic martingale (amart)*, if the family $(\text{net}) (EX_\tau)_{\tau \in T_b}$ converges.

Remark 7.4 It follows from the definition of amart $(X_n)_{n=0,1,\dots}$ that the set $(EX_\tau)_{\tau \in T_b}$ is bounded. It is also clear that a linear combination of amarts is an amart.

According to these definitions and Theorems 6.6 and 6.7, we can conclude that martingales belong to the class of asymptotic martingales. In this case, a natural question arises:

“What elements of previously developed martingale theory can be extended to a wider class of asymptotic martingales?”

If another filtration $(G_n)_{n=0,1,\dots}$ is included to filtration $(\mathcal{F}_n)_{n=0,1,\dots}$, i.e., $G_n \subseteq \mathcal{F}_n$, $n = 0, 1, \dots$, then each stopping time w.r.t. (G_n) will be a stopping time w.r.t. $(\mathcal{F}_n)$. Therefore, every amart $(X_n)_{n=0,1,\dots}$ w.r.t. $(\mathcal{F}_n)$ will be amart w.r.t. (G_n) if it is adapted to (G_n) . Moreover, from every $(\mathcal{F}_n)$ -amart (X_n) , one can construct a (G_n) -amart (Y_n) if we put $Y_n = \mathbf{E}(X_n | G_n)$, $n = 0, 1, \dots$, because of $\mathbf{E}X_\tau = \mathbf{E}Y_\tau$ for all $\tau \in T_b$.

Further, we have seen the role of “maximal inequalities” in martingale theory. It turns out such inequality can be stated in more general setting.

Lemma 7.3 Assume that a stochastic sequence $(X_n)_{n=0,1,\dots}$ such that $\sup_{\tau \in T_b} \mathbf{E}|X_\tau| < \infty$. Then for $\lambda > 0$:

$$P \left\{ \sup_n |X_n| > \lambda \right\} \leq \lambda^{-1} \sup_{\tau \in T_b} \mathbf{E}|X_\tau|. \quad (7.34)$$

Proof For a fixed number N , we define the set $A = \{\omega : \sup_{0 \leq n \leq N} |X_n| > \lambda\}$ and a bounded stopping time

$$\sigma(\omega) = \begin{cases} \inf(n \leq N : |X_n(\omega)| > \lambda) & \text{if } \omega \in A, \\ N & \text{if } \omega \notin A. \end{cases}$$

Then

$$\sup_{\tau \in T_b} \mathbf{E}|X_\tau| \geq \mathbf{E}|X_\sigma| \geq \lambda P(A). \quad (7.35)$$

Taking the limit in (7.35) as $N \rightarrow \infty$, we get (7.34). $\square$

Lemma 7.4 *If $(X_n)_{n=0,1,\dots}$ and $(Y_n)_{n=0,1,\dots}$ are L^1 -bounded amarts, then $(\max(X_n, Y_n))_{n=0,1,\dots}$ and $(\min(X_n, Y_n))_{n=0,1,\dots}$ are amarts.*

Proof We consider only the case of $Z_n = \max(X_n, Y_n)$ due to a symmetry. First of all, we prove that the net $(\mathbf{E}Z_\tau)_{\tau \in T_b}$ is bounded. To do it, take an arbitrary $\tau \in T_b$ and choose $n \geq \tau$. Define

$$\tau_X = \begin{cases} \tau, & \text{if } X_\tau \geq 0, \\ n, & \text{if } X_\tau < 0. \end{cases}$$

$$\tau_Y = \begin{cases} \tau, & \text{if } Y_\tau \geq 0, \\ n, & \text{if } Y_\tau < 0. \end{cases}$$

Find that

$$\begin{aligned} \mathbf{E}Z_\tau &\leq \mathbf{E}X_\tau I_{\{X_\tau \geq 0\}} + \mathbf{E}Y_\tau I_{\{Y_\tau \geq 0\}} \\ &= \mathbf{E}X_{\tau_X} - \mathbf{E}X_n I_{\{X_\tau < 0\}} + \mathbf{E}Y_{\tau_Y} - \mathbf{E}Y_n I_{\{Y_\tau < 0\}} \\ &\leq \sup_{\sigma \in T_b} \mathbf{E}X_\sigma + \sup_m \mathbf{E}|X_m| + \sup_{\sigma \in T_b} \mathbf{E}Y_\sigma + \sup_m \mathbf{E}|Y_m|, \end{aligned}$$

and we arrive to the conclusion.

Let us prove now that $(Z_n)_{n=0,1,\dots}$ is an amart. For $\epsilon > 0$, we can choose $\tau_0 \in T_b$ such that

$$|\mathbf{E}X_\sigma - \mathbf{E}X_\tau| < \epsilon, \quad |\mathbf{E}Y_\sigma - \mathbf{E}Y_\tau| < \epsilon \quad (7.36)$$

for $\sigma, \tau \geq \tau_0$. Further, due to the boundedness of $(\mathbf{E}Z_\tau)_{\tau \in T_b}$, one can choose $\tau_1 \geq \tau_0$ such that if $\sigma \geq \tau_1$, then

$$\mathbf{E}Z_\sigma \leq \mathbf{E}Z_{\tau_1} + \epsilon. \quad (7.37)$$

For a bounded stopping time $\sigma \geq \tau_1$, we define $A = \{X_{\tau_1} < Y_{\tau_1}\}$ and $\sigma_1 \in T_b$ as follows:

$$\sigma_1 = \begin{cases} \tau_1 & \text{on } A, \\ \sigma & \text{on } A^c. \end{cases}$$

Then we obtain

$$\mathbf{E}X_{\tau_1} = \mathbf{E}Z_{\tau_1} I_{A^c} + \mathbf{E}X_{\tau_1} I_A, \quad (7.38)$$

$$\mathbf{E}X_{\sigma_1} = \mathbf{E}X_{\sigma} I_{A^c} + \mathbf{E}X_{\tau_1} I_A. \quad (7.39)$$

Subtracting (7.38) from (7.37), we get with the help of (7.35) that

$$\mathbf{E}Z_{\tau_1} I_{A^c} = \mathbf{E}X_{\sigma} I_{A^c} + \mathbf{E}X_{\tau_1} - \mathbf{E}X_{\sigma_1} \leq \mathbf{E}Z_{\sigma} I_{A^c} + \epsilon. \quad (7.40)$$

We can write the same relations as (7.38) and (7.39) for (Y_n) :

$$\mathbf{E}Y_{\tau_1} = \mathbf{E}Z_{\tau_1} I_{A^c} + \mathbf{E}Y_{\tau_1} I_A,$$

$$\mathbf{E}Y_{\sigma_1} = \mathbf{E}Y_{\sigma} I_{A^c} + \mathbf{E}Y_{\tau_1} I_A,$$

and find

$$\mathbf{E}Z_{\tau_1} I_A = \mathbf{E}Y_{\sigma} I_A + \mathbf{E}Y_{\tau_1} - \mathbf{E}Y_{\sigma_1} \leq \mathbf{E}Z_{\sigma} I_A + \epsilon. \quad (7.41)$$

Combining (7.40) and (7.41), we obtain inequality

$$\mathbf{E}Z_{\tau_1} \leq \mathbf{E}Z_{\sigma} + 2\epsilon,$$

which leads together with (7.36) that

$$|\mathbf{E}Z_{\sigma} - \mathbf{E}Z_{\tau_1}| \leq 2\epsilon,$$

and, hence, the net $(\mathbf{E}Z_{\tau})_{\tau \in T_b}$ is the Cauchy net. $\square$

Applying Lemmas 7.3 and 7.4, we arrive to the following theorem.

Theorem 7.6 *Let $(X_n)_{n=0,1,\dots}$ be amart with $\sup_n \mathbf{E}|X_n| < \infty$. Then:*

- (1) *Its positive (negative) parts X_n^+ (X_n^-) and its absolute value $|X_n|$ are L^1 -bounded amarts.*
- (2) *For each $\lambda \geq 0$, the sequence*

$$X_n^\lambda = \text{sign}(X_n) \lambda I_{|X_n| > \lambda} + X_n I_{|X_n| \leq \lambda}.$$

- (3) *$\sup_{\tau \in T_b} \mathbf{E}|X_{\tau}| < \infty$ and $\sup_n |X_n| < \infty$ (a.s.).*

The next statement can be considered as a version of the optional sampling theorem.

Theorem 7.7 *Let $(X_n)_{n=0,1,\dots}$ be an amart for $(\mathcal{F}_n)_{n=0,1,\dots}$ and let $(\tau_m)_{m=0,1,\dots}$ be a nondecreasing sequence of bounded stopping times for $(\mathcal{F}_n)_{n=0,1,\dots}$. Then $(X_{\tau_m})_{m=0,1,\dots}$ is an amart for $(\mathcal{F}_{\tau_m})_{m=0,1,\dots}$.*

Proof For fixed $\epsilon > 0$, we can choose N such that $|\mathbf{E}X_\tau - \mathbf{E}X_{\tau'}| < \epsilon$ for all bounded stopping times τ and $\tau' \geq N$. Denote $\tau_\infty = \lim_{m \rightarrow \infty} \tau_m$ (a.s.) and find that $X_{\tau_m \wedge N} \rightarrow X_{\tau_\infty \wedge N}$ (a.s.), $m \rightarrow \infty$, and

$$\mathbf{E} \sup_m |X_{\tau_m \wedge N}| \leq \mathbf{E} \max(|X_1|, \dots, |X_N|) < \infty.$$

Therefore, by the dominated convergence theorem, the sequence $(X_{\tau_m \wedge N})_{m=0,1,\dots}$ is an amart. Taking a big enough M so that $|\mathbf{E}X_{\tau_\sigma \wedge N} - \mathbf{E}X_{\tau_{\sigma'} \wedge N}|$ for all bounded stopping times σ and $\sigma' \geq M$ for $(\mathcal{F}_{\tau_m})_{m=0,1,\dots}$, we have

$$|\mathbf{E}X_{\tau_\sigma} - \mathbf{E}X_{\tau_{\sigma'}}| \leq |\mathbf{E}X_{\tau_\sigma \vee N} - \mathbf{E}X_{\tau_{\sigma'} \vee N}| + |\mathbf{E}X_{\tau_\sigma \wedge N} - \mathbf{E}X_{\tau_{\sigma'} \wedge N}| \leq \epsilon + \epsilon = 2\epsilon.$$

□

Example 7.5 An integrable stochastic sequence $(X_n)_{n=0,1,\dots}$ is called a *quasi-martingale* if

$$\sum_{n=0}^{\infty} \mathbf{E}|X_n - \mathbf{E}(X_{n+1}|\mathcal{F}_n)| < \infty. \quad (7.42)$$

Condition (7.42) is trivial in case of a martingale. In view (7.42) for given $\epsilon > 0$, one can find a big enough N such that

$$\sum_{n=N}^{\infty} \mathbf{E}|X_n - \mathbf{E}(X_{n+1}|\mathcal{F}_n)| \leq \epsilon. \quad (7.43)$$

Taking $\tau \in T_b$ such that $N \leq \tau \leq M$, we have

$$\begin{aligned} |\mathbf{E}X_\tau - \mathbf{E}X_M| &= \left| \sum_{m=N}^M \mathbf{E}(X_m - X_M)I_{\{\tau=m\}} \right| \\ &= \left| \sum_{m=N}^M \sum_{n=m}^{M-1} \mathbf{E}(X_n - X_{n+1})I_{\{\tau=m\}} \right| \\ &= \sum_{n=N}^{M-1} \sum_{m=N}^n \mathbf{E}(|X_n - \mathbf{E}(X_{n+1}|\mathcal{F}_n)| I_{\{\tau=m\}}) \\ &\leq \sum_{n=N}^{\infty} \mathbf{E}|X_n - \mathbf{E}(X_{n+1}|\mathcal{F}_n)| \leq \epsilon. \end{aligned}$$

If $\tau_1, \tau_2 \geq N$, then choose $M \geq \tau_1 \vee \tau_2$ and get

$$|\mathbf{E}X_{\tau_1} - \mathbf{E}X_{\tau_2}| \leq |\mathbf{E}X_{\tau_1} - \mathbf{E}X_M| + |\mathbf{E}X_{\tau_2} - \mathbf{E}X_M| \leq 2\epsilon.$$

Hence, the net $(\mathbf{E}X_\tau)_{\tau \in T_b}$ is Cauchy and converges.

So, any quasimartingale is an amart.

The definition of amarts is certainly directed to provide their asymptotic behavior similar to martingales. Below we confirm these expectations.

Lemma 7.5 *Let $(X_n)_{n=0,1,\dots}$ be a stochastic L^1 -bounded sequence. Then the following are equivalent:*

- (1) $(X_n)_{n=0,1,\dots}$ converges (a.s.).
- (2) $(X_n)_{n=0,1,\dots}$ is an amart.

Proof The implication (1) $\Rightarrow$ (2) follows from the dominated convergence theorem.

To prove that (2) $\Rightarrow$ (1), we put $X^* = \limsup X_n$ and $X_* = \liminf X_n$. Then we can find sequences $\tau_n, \sigma_n \in T_b$ with $\tau_n \uparrow \infty$, $\sigma_n \uparrow \infty$ such that $X_{\tau_n} \rightarrow X^*$ (a.s.) and $X_{\sigma_n} \rightarrow X_*$ (a.s.). Applying again the dominated convergence theorem, we obtain that

$$\mathbf{E}(X^* - X_*) = \lim_n \mathbf{E}(X_{\tau_n} - X_{\sigma_n}) = 0.$$

It means $X^* = X_*$ (a.s.). □

Theorem 7.8 *Let $(X_n)_{n=0,1,\dots}$ be an amart with $\sup_n \mathbf{E}|X_n| < \infty$. Then $(X_n)_{n=0,1,\dots}$ converges (a.s.).*

Proof By Theorem 7.6, $\sup_n |X_n| < \infty$ (a.s.), and therefore $P\{\omega : \sup_n |X_n| > \lambda\}$ is arbitrary small if λ is big enough. By Theorem 7.7, X_n^λ is an amart, and by Lemma 7.5, it converges (a.s.). Taking $\lambda \uparrow \infty$, we get (a.s.)-convergence of (X_n) . □

In the end of this section, we would like to mention some other extensions of the notion of martingales.

Definition 7.4 A stochastic sequence $(X_n)_{n=0,1,\dots}$, $X_0 = 0$, is a *local martingale* if there exists a sequence of stopping times $(\tau_m)_{m=1,2,\dots}$ increasing to $+\infty$ such that $(X_{\tau_m \wedge n})_{n=0,1,\dots}$ is a martingale for each $m = 1, 2, \dots$. The sequence (τ_m) is called a *localizing sequence*.

Definition 7.5 A stochastic sequence $(X_n)_{n=0,1,\dots}$, $X_0 = 0$, is a *generalized martingale* if (a.s.)

$$\{\omega : \mathbf{E}(X_n^+ | \mathcal{F}_{n-1}) < \infty\} \cup \{\omega : \mathbf{E}(X_n^- | \mathcal{F}_{n-1}) < \infty\} = \Omega,$$

and for $n = 1, 2, \dots$

$$\mathbf{E}(X_n | \mathcal{F}_{n-1}) = X_{n-1} \quad (a.s.).$$

Definition 7.6 A stochastic sequence $(X_n)_{n=0,1,\dots}$, $X_0 = 0$, is a *martingale transform* (stochastic integral) if it admits the following representation

$$X_n = \sum_{m=0}^n H_m \Delta M_m,$$

where $(M_n)_{n=0,1,\dots}$, $M_0 = 0$, is a martingale, $(H_n)_{n=0,1,\dots}$ is predictable, $H_0 = \Delta M_0 = M_0 = 0$.

All these generalizations are directed to relax the condition of integrability of stochastic sequences. All these classes of stochastic sequences coincide. We omit the proof of this statement here.

Chapter 8

Elements of Classical Theory of Stochastic Processes



This chapter contains a general notion of random processes with continuous time. It is given in context of the Kolmogorov consistency theorem. The notion of a Wiener process with a variety of its properties is also presented here. Its existence is stated by two ways: with the help of the Kolmogorov theorem as well as with the help of orthogonal functional systems. Besides the Wiener process as a basic process for many others, the Poisson process is also considered here. Stochastic integration with respect to Wiener process is developed for a class of progressively measurable functions. It leads to the Ito processes, the Ito formula, the Girsanov theorem, and representation of martingales (see [11, 12, 27, 31, 37, 50, 57], and [61]).

8.1 Stochastic Processes: Definitions, Properties, and Classical Examples

In classical theory, the notion of a stochastic process is associated with a family of random variables (X_t) on a probability space $(\Omega, \mathcal{F}, P)$ with values in space $\mathbb{R}^d, d \geq 1$, where parameter $t \in [0, \infty) = \mathbb{R}_+$. Usually, for simplicity, we put $d = 1$. It is supposed that for time parameter $t \in \mathbb{R}_+$, we have a random variable $X_t(\omega), \omega \in \Omega$. On the other hand, one can fix $\omega \in \Omega$ and get a function $X_\cdot(\omega) : \mathbb{R}_+ \rightarrow \mathbb{R}^d$, which is called a *trajectory*. This point of view opens up the possibility of studying stochastic processes that have been exhaustively determined by their distributions as probability measures on the functional space $(\mathbb{R}^{[0, \infty)}, \mathcal{B}^{[0, \infty)})$. Usually, a family of finite dimensional distributions $P_{t_1, \dots, t_n}(B_1, \dots, B_n), 0 \leq t < t_1 < t_2 < \dots < t_n < \infty, B_i \in \mathcal{B}(\mathbb{R}^d), i = 1, \dots, n$, is exploited as the basic probabilistic characteristic of a stochastic process. A natural question arises in this

regard. Is there a stochastic process $(X_t(\omega))$ such that $P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) = P_{t_1, \dots, t_n}(B_1, \dots, B_n)$. As it was noted in Chap. 1, this question is solved with the help of the Kolmogorov theorem if for the system $(P_{t_1, \dots, t_n})$ the following *consistency conditions* are fulfilled:

1. $P_{t_1, \dots, t_n}(B_1, \dots, B_{i-1}, \cdot, B_{i+1}, \dots, B_n)$, $i = 1, 2, \dots, n$, is a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
2. $P_{t_1, \dots, t_n}(B_1, \dots, B_n) = P_{t_{i_1}, \dots, t_{i_n}}(B_{i_1}, \dots, B_{i_n})$, where $(i_1, \dots, i_n)$ is arbitrary permutation of numbers $(1, 2, \dots, n)$.
3. $P_{t_1, \dots, t_{n-1}, t_n}(B_1, \dots, B_{n-1}, \mathbb{R}^d) = P_{t_1, \dots, t_{n-1}}(B_1, \dots, B_{n-1})$. In such a case, there exists a probability measure P_X in the functional space $(\mathbb{R}^{[0, \infty)}, \mathcal{B}^{[0, \infty)})$, and the process $(X_t(\omega))$ is constructed as $X_t(\omega) = \omega_t$, where ω_t is the value of function $\omega \in \mathbb{R}^{[0, \infty)}$ at time t . Measure P_X will be a distribution of $(X_t(\omega))$ which has $P_{t_1, \dots, t_n}$ as a system of finite dimensional distributions of this process.

After these necessary explanations, we introduce the notion of a *Wiener process* or *Brownian motion*.

Definition 8.1 A stochastic process $(W_t)_{t \geq 0}$ such that:

- (1) $W_0 = 0$ (a.s.),
- (2) $W_t(\omega) \in C[0, \infty)$ for almost all $\omega \in \Omega$,
- (3) $P(W_{t_1} \in B_1, \dots, W_{t_n} \in B_n) = \int_{B_1} \dots \int_{B_n} p(t, 0, x_1) p(t_2 - t_1, x_1, x_2) \dots p(t_n - t_{n-1}, x_{n-1}, x_n) dx_1 \dots dx_n$,
for arbitrary $0 < t_1 < t_2 < \dots < t_n$, $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$, where $p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right)$ is called a Wiener process.

It follows directly from the definition that $W_t \sim N(0, t)$ for every fixed $t \in \mathbb{R}_+$.

Let us formulate the first *properties* of the process $(W_t)_{t \geq 0}$:

- (1) $\mathbf{E}W_t W_s = s \wedge t$ for arbitrary $s, t \in \mathbb{R}_+$.
- (2) $\mathbf{E}(W_t - W_s)^2 = |t - s|$.
- (3) $\mathbf{E}(W_t - W_s)^4 = 3(t - s)^2$.

We only prove (2), leaving (1) and (3) as problems.

Problem 8.1 Prove properties (1) and (3) of $(W_t)_{t \geq 0}$.

Assume that $s < t$ and find that

$$\begin{aligned}
 \mathbf{E}W_s W_t &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp(s, 0, x)p(t-s, x, y)dx dy \\
 &= \int_{-\infty}^{\infty} xp(s, 0, x) \left(\int_{-\infty}^{\infty} yp(t-s, x, y)dy \right) dx \\
 &= \int_{-\infty}^{\infty} xp(s, 0, x) \left(\int_{-\infty}^{\infty} (x+u)p(t-s, x, x+u)du \right) dx
 \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} xp(s, 0, x) \left(\int_{-\infty}^{\infty} (x+u)p(t-s, 0, u)du \right) dx \\
&= \int_{-\infty}^{\infty} xp(s, 0, x) \left(x \int_{-\infty}^{\infty} p(t-s, 0, u)du \right. \\
&\quad \left. + u \int_{-\infty}^{\infty} p(t-s, 0, u)du \right) dx \\
&= \int_{-\infty}^{\infty} xp(s, 0, x) (x+0) dx \\
&= \int_{-\infty}^{\infty} x^2 p(s, 0, x) dx = s.
\end{aligned}$$

Now we calculate for $s < t$ and $B \in \mathcal{B}(\mathbb{R})$.

$$\begin{aligned}
P(W_t - W_s \in B) &= \int_{\{(x,y): y-x \in B\}} p(s, 0, x) p(t-s, x, y) dx dy \\
&= \int_{-\infty}^{\infty} p(s, 0, x) \left(\int_{\{y: y-x \in B\}} p(t-s, x, y) dy \right) dx \\
&= \int_{-\infty}^{\infty} p(s, 0, x) \left(\int_B p(t-s, x, x+u) du \right) dx \\
&= \int_{-\infty}^{\infty} p(s, 0, x) \left(\int_B p(t-s, 0, u) du \right) dx \\
&= \int_{-\infty}^{\infty} p(s, 0, x) dx \int_B p(t-s, 0, u) du = \int_B p(t-s, 0, u) du,
\end{aligned}$$

and we obtain that $W_t - W_s \sim N(0, t-s)$.

As a result, we arrive at the fourth property of W_t :

- (4) For $0 = t_0 < t_1 < \dots < t_n$, the increments $W_{t_1} - W_{t_0}, \dots, W_{t_n} - W_{t_{n-1}}$ of a Wiener process are *independent* random variables.

Proof By “normality” of increments, we need to prove that they are *uncorrelated*. Let us take $r < s < t < u$ and find that

$$\mathbf{E}(W_u - W_t) = \mathbf{E}W_u W_s - \mathbf{E}W_t W_s - \mathbf{E}W_u W_r + \mathbf{E}W_t W_r = s - s - r + r = 0.$$

□

To go further, we introduce filtration, generated by a Wiener process $\mathcal{F}_t = \sigma(W_s, s \leq t)$. We assume that $\mathcal{F}_t$ is *complete*, i.e., it contains all sets of P -measure zero.

Definition 8.2 A process $(M_t)_{t \geq 0}$ satisfying conditions

- (1) $M_t - \mathcal{F}_t$ -measurable,
- (2) M_t is integrable,
- (3) $\mathbf{E}(M_t | \mathcal{F}_s) = M_s$ (a.s.) for all $s < t$,
is called a *martingale*.

After this definition, we formulate the *martingale* property of a Wiener process. If in (3) the equality is related by the inequality $\geq$ ($\leq$), then (M_t) is a submartingale (supermartingale).

- (5) Processes $(W_t)_{t \geq 0}$ and $(W_t^2 - t)_{t \geq 0}$ are martingales w.r.to $(\mathcal{F})_{t \geq 0}$.

Remark 8.1 In fact, the inverse statement to (5) is well-known as *Levy's characterization of a Wiener process*.

- (6) The Doob maximal inequality for a Wiener process $(W_t)_{t \geq 0}$ has the following form: For $t > 0$,

$$\mathbf{E} \max_{s \leq t} |W_s|^2 \leq 4 \mathbf{E} |W_t|^2. \quad (8.1)$$

To prove it, we consider a subdivision of the interval $[0, t]$ as of the interval $[0, t]$ as $t_k^{(n)} = \frac{kt}{2^n}$, $0 \leq k \leq 2^n$ and define a square integrable martingale with discrete time

$$M_k = M_k^n = W_{\frac{kt}{2^n}}, \mathcal{F}_k^n = \mathcal{F}_{\frac{kt}{2^n}}, 0 \leq k \leq 2^n.$$

For $(|M_k|)$, we have, using its submartingale property that,

$$\begin{aligned} \mathbf{E}(\max_k |M_k|)^2 &= 2 \int_0^\infty y P(\max_k |M_k| > y) dy \\ &\leq 2 \int_0^\infty \mathbf{E} |M_{2^n}| I_{(\max_k |M_k| \geq y)} dy \\ &= 2 \int_0^\infty \left(\int_{(\max_k |M_k| \geq y)} |M_{2^n}| dP \right) dy \\ &= 2 \int_\Omega |M_{2^n}|^k \left(\int_0^{\max_k |M_k|} dy \right) dP \\ &= 2 \int_\Omega |M_{2^n}| \max_k |M_k| dP = 2 \mathbf{E} |M_{2^n}| \max_k |M_k| \\ &\leq 2 (\mathbf{E} (M_{2^n})^2)^{1/2} \left(\mathbf{E} (\max_k |M_k|)^2 \right)^{1/2}. \end{aligned}$$

As a result, we obtain that

$$(\mathbf{E}(\max_k |M_k|)^2)^{1/2} \leq 2(\mathbf{E}(M_{2^n})^2)^{1/2},$$

and, hence,

$$\mathbf{E} \max_k |M_k^n|^2 \leq \mathbf{E}(\max_k |M_k^n|)^2 \leq 4\mathbf{E}(M_{2^n})^2 = 4\mathbf{E}|W_t|^2. \quad (8.2)$$

Taking the limit as $n \rightarrow \infty$ in (8.2) and using the continuity of trajectories of a Wiener process, we arrive to inequality (8.1).

- (7) For any positive constant c the process $W_t^c = c^{-1}W_{c^2t}$ is a Wiener process. This property of a Wiener process is called *self-similarity*.

To prove the self-similarity of a Wiener process, one can use the Levy characterization of this process.

- (8) *Existence* of a Wiener process follows from the consistency theorem of Kolmogorov (see Chap. 1). However, such a reference is incomplete because this theorem guarantees existence of this process only in space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}^{[0,\infty)})$. To make the proof complete, one can use another *theorem of Kolmogorov*:

Suppose $(X_t)_{t \geq 0}$ is a stochastic process satisfying the following condition

$$\mathbf{E}|X_t - X_s|^\alpha \leq C|t - s|^{1+\beta} \text{ for all } t, s \geq 0,$$

and for some $\alpha > 0, \beta > 0, 0 < C < \infty$. Then the process (X_t) admits a continuous *modification*, i.e., there exists a continuous process (Y_t) such that $P(X_t = Y_t) = 1$ for all $t \in \mathbb{R}_+$.

Property (3) of a Wiener process (W_t) means that this theorem can be applied, and, therefore, the process (W_t) can be identified with its continuous modification.

- (9) Now we give some other characterizations of trajectories of $(W_t)_{t \geq 0}$. The continuity property brings an element of a regularity of trajectories. But this description of sample paths properties is clearly incomplete as it is shown below.

Let us analyze the *first* question of *differentiability* of trajectories of $(W_t)_{t \geq 0}$. We take $t \geq 0$ and consider the ratio $R_\Delta^W(t) = \frac{W_{t+\Delta t} - W_t}{\Delta t}$, where $\Delta t \rightarrow 0$. One can easily observe that

$$\mathbf{E}R_\Delta^W(t) = \frac{\mathbf{E}W_{t+\Delta t} - \mathbf{E}W_t}{\Delta t} = 0,$$

and

$$\mathbf{Var}R_\Delta^W(t) = (\Delta t)^{-2}\mathbf{Var}(W_{t+\Delta t} - W_t) = (\Delta t)^{-2}\Delta t = (\Delta t)^{-1} \rightarrow \infty \text{ as } \Delta t \rightarrow 0.$$

It shows that W is *not differentiable* in the L^2 -sense. Moreover, one can prove that almost all trajectories of W are non-differentiable.

Let us analyze the second question of how *vary* the trajectories of $(W_t)_{t \geq 0}$ are. This property is characterized with the help of the notion of p -variation of given function $f : [0, T] \rightarrow \mathbb{R}$, $p \geq 1$:

We say that

$$\limsup_{\Delta t \rightarrow 0} \sum_{i=1}^n |f(t_i) - f(t_{i-1})|^p, 0 = t_0 < t_1 < \dots < t_n = T$$

is a p -variation of f on $[0, T]$.

Usually, the first and second variations are most important in this regard.

For the second variation of $(W_t)_{t \geq 0}$, we have the following observations. In view of independent increments, we obtain for $0 = t_0 < t_1 < \dots < t_n = T$ that $\mathbf{E} \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 = \sum_{i=1}^n \mathbf{E}(W_{t_i} - W_{t_{i-1}})^2 = \sum_{i=1}^n \mathbf{Var}(W_{t_i} - W_{t_{i-1}}) = \sum_{i=1}^n (t_i - t_{i-1}) = T$.

Further, with the help of (3),

$$\begin{aligned} \mathbf{Var} \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 &= \sum_{i=1}^n \mathbf{Var}(W_{t_i} - W_{t_{i-1}})^2 \\ &= \sum_{i=1}^n \left[\mathbf{E}(W_{t_i} - W_{t_{i-1}})^4 - \left(\mathbf{E}(W_{t_i} - W_{t_{i-1}})^2 \right)^2 \right] \\ &= \sum_{i=1}^n \left[3(t_i - t_{i-1})^2 - (t_i - t_{i-1})^2 \right] \\ &= 2 \sum_{i=1}^n (t_i - t_{i-1})^2. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{E} \left[\sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 - T \right]^2 &= \mathbf{Var} \left(\sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 \right) \\ &= 2 \sum_{i=1}^n (t_i - t_{i-1})^2 \leq 2T \max_{0 < i \leq n} (t_i - t_{i-1}) \rightarrow 0, \end{aligned}$$

and we conclude that the second variation of (W_t) on $[0, T]$ converges to the *length* of this interval in L^2 sense, and hence in the *sense of convergence in probability*. To investigate the *first variation* of (W_t) , we consider the probability $P(\omega : \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| > N)$. First of all, we note that $\mathbf{E} \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| =$

$\sqrt{\frac{2}{\pi}} \sum_{i=1}^n \sqrt{t_i - t_{i-1}} \rightarrow \infty$ as $\max(t_i - t_{i-1}) \rightarrow 0$, and

$$\mathbf{Var} \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| = \sum_{i=1}^n \left(1 - \frac{2}{\pi}\right) (t_i - t_{i-1}) = \left(1 - \frac{2}{\pi}\right) T.$$

Further, for $\mathbf{E} \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| > n$ with the help of the Chebyshev inequality, we get

$$\begin{aligned} & P\left(\omega : \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| \leq N\right) \\ & \leq P\left(\omega : \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| - \mathbf{E} \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| \geq \mathbf{E} \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| - N\right) \\ & \leq \mathbf{Var} \left(\sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| \right) / \left(\mathbf{E} \sum_{i=1}^n |W_{t_i} - W_{t_{i-1}}| - N \right)^2 \rightarrow 0, \end{aligned}$$

as $\max(t_i - t_{i-1}) \rightarrow 0$.

It means that the first variation of (W_t) converges in probability to 0.

Remark 8.2 It is possible to prove that all mentioned results are true in the sense of (a.s.)-convergence.

Another process that plays an important role in both theory and applications is a Poisson process.

Definition 8.3 The process $(N_t)_{t \geq 0}$ is called a *Poisson process* with parameter $\lambda > 0$, if the following conditions are satisfied:

- (1) $N_0 = 0$. (a.s.).
- (2) Its increments $N_{t_1} - N_{t_0}, \dots, N_{t_n} - N_{t_{n-1}}$ are independent random variables for any subdivision $0 \leq t_0 < t_1 < \dots < t_n < \infty$.
- (3) $N_t - N_s$ is a random variable that has a Poisson distribution with parameter $\lambda(t - s)$:

$$P(\omega : N_t - N_s = i) = e^{-\lambda(t-s)} \frac{(\lambda(t-s))^i}{i!}, 0 \leq s \leq t < \infty, i = 0, 1, \dots$$

Let us present some properties of sample paths of this process.

First of all, we note that its almost all trajectories are *nondecreasing*, because for $s \leq t$

$$\begin{aligned} P(\omega : N_t - N_s \geq 0) &= \sum_{i=0}^{\infty} P(\omega : N_t - N_s = i) = \sum_{i=0}^{\infty} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^i}{i!} \\ &= e^{-\lambda(t-s)} \sum_{i=0}^{\infty} \frac{(\lambda(t-s))^i}{i!} = e^{-\lambda(t-s)} e^{\lambda(t-s)} = 1. \end{aligned}$$

The process $(N_t)_{t \geq 0}$ is *stochastically continuous*: in the sense $N_t \rightarrow N_s$ in probability as $t \rightarrow s$. This property is an obvious consequence of the application of the Chebyshev inequality. It is an interesting effect when a jumping process satisfies a continuity property. Moreover, one can say something about a *differentiability* of trajectories of $(N_t)_{t \geq 0}$ in the sense of convergence in probability, i.e.,

$$\frac{N_{t+\Delta t} - N_t}{\Delta t} \xrightarrow[\Delta t \rightarrow 0]{p} 0. \quad (8.3)$$

Problem 8.2 Prove relation (8.3).

Further, one can derive from $P(\omega : N_s \leq N_t) = 1$ for all $s \leq t$ that there is a *right-continuous modification* of (N_t) , almost all trajectories of which are nondecreasing integer-valued functions with unit jumps.

Let us note a martingale property of $(N_t)_{t \geq 0}$. To do this, we introduce a natural filtration $(\mathcal{F}_t)_{t \geq 0}$, generated by $(N_t)_{t \geq 0}$ and completed by sets of P -measure zero. We note that

$$\begin{aligned} \mathbf{E}N_t &= \sum_{i=0}^{\infty} i P(\omega : N_t = i) = \sum_{i=0}^{\infty} i e^{-\lambda t} \frac{(\lambda t)^i}{i!} \\ &= (\lambda t) e^{-\lambda t} \sum_{i=1}^{\infty} \frac{(\lambda t)^{i-1}}{(i-1)!} = (\lambda t) e^{-\lambda t} e^{\lambda t} = \lambda t. \end{aligned}$$

Let us define a new process $M_t = N_t - \lambda t$ and find that for $s \leq t$ by independence $(N_t - N_s)$ of $\mathcal{F}_s$:

$$\begin{aligned} \mathbf{E}(M_t | \mathcal{F}_s) &= \mathbf{E}(N_t - \lambda t | \mathcal{F}_s) = \mathbf{E}(N_t - N_s + N_s - \lambda t | \mathcal{F}_s) \\ &\quad \times \mathbf{E}(N_t - N_s | \mathcal{F}_s) + N_s - \lambda t \\ &= \mathbf{E}(N_t - N_s) + N_s - \lambda t = \lambda(t-s) + N_s - \lambda t = N_s - \lambda s = M_s. \end{aligned}$$

Hence, $(M_t)_{t \geq 0}$ is a *martingale* w.r.to $(\mathcal{F}_t)_{t \geq 0}$.

We have seen that both processes $(W_t)_{t \geq 0}$ and (N_t) are stochastically continuous with independent increments. The following example shows why a consideration of *stochastic processes with independent values is not productive*.

Example 8.1 Let $(X_t)_{t \geq 0}$ be a family of independent random variables with the same density $f = f(x) \geq 0$, $\int_{-\infty}^{\infty} f(x)dx = 1$. Then for a fixed $s \geq 0$, and $t \neq s$, $\epsilon > 0$ we have that

$$P(\omega : |X_t - X_s| \geq \epsilon) = \int \int_{|x-y| \geq \epsilon} f(x)f(y)dx dy. \quad (8.4)$$

The integral on the right-hand side of equality (8.4) converges (as $\epsilon \rightarrow 0$) to

$$\int \int_{x \neq y} f(x)f(y)dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)f(y)dx dy. \quad (8.5)$$

The relation (8.5) shows that for some $\epsilon > 0$ the probability on the left-hand side of (8.4) does not converge to zero as $t \rightarrow s$. It means the process with independent values is not stochastically continuous even.

The existence of a Wiener process was derived from the Kolmogorov consistency theorem. Such a derivation does not look constructive. That is why we want to demonstrate a *direct way of construction* of a Wiener process, based on orthogonal systems of functions of Haar and Schauder, and sequences of independent normal random variables.

Let us define a system of functions $(H_k(t))_{k=1,2,\dots}$, $t \in [0, 1]$, called the *Haar system*:

$$H_1(t), H_2(t) = I_{[0,2^{-1}]}(t) - I_{[2^{-1},1]}(t), \dots,$$

$$H_k(t) = 2^{n/2} \left(I_{[a_{n,k}, a_{n,k} + 2^{-n-1}]}(t) - I_{[a_{n,k} + 2^{-n-1}, a_{n,k} + 2^{-n}]}(t) \right),$$

where $w^n < k \leq 2^{n+1}$, $a_{n,k} = 2^{-n}(k - 2^n - 1)$, $n = 1, 2, \dots$

In the space $L^2([0, 1], \mathcal{B}(0, 1), dt)$ with the scalar product $\langle f, g \rangle = \int_0^1 f_s g_s ds$, $f, g \in L^2$, the system $(H_k(t))_{k=1,2,\dots}$ is complete and orthogonal. Hence,

$$f = \sum_{k=1}^{\infty} \langle f, H_k \rangle H_k, \quad g = \sum_{k=1}^{\infty} \langle g, H_k \rangle H_k, \quad \langle f, g \rangle = \sum_{k=1}^{\infty} \langle f, H_k \rangle \langle g, H_k \rangle.$$

Using the system $(H_k(t))_{k=1,2,\dots}$, one can construct the *Schauder system* $(S_k(t))_{k=1,2,\dots}$ as follows:

$$S_k(t) = \int_0^t H_k(y)dy = \langle I_{[0,t]}, H_k \rangle, \quad t \in [0, 1].$$

Lemma 8.1 *If a sequence of real numbers $a_k = O(k^\epsilon)$, $k \rightarrow \infty$, for some $\epsilon \in (0, 1/2)$, then the series $\sum_{k=1}^{\infty} a_k S_k(t)$ converges uniformly on $[0, 1]$, and, hence, it is a continuous function.*

Proof Denote $R_m = \sup_t \sum_{k > 2^m} |a_k| S_k(t)$, $m = 1, 2, \dots$, and note that $|a_k| \leq ck^\epsilon$ for all $k \geq 1$ and some constant $c > 0$. Therefore, for all t and $n = 1, 2, \dots$, we obtain that

$$\begin{aligned} \sum_{2^n < k < 2^{n+1}} |a_k| S_k(t) &\leq c 2^{(n+1)\epsilon} \sum_{2^n < k < 2^{n+1}} S_k(t) \\ &\leq c 2^{(n+1)\epsilon} 2^{-n/2-1} \\ &\leq c 2^{\epsilon-n(1/2-\epsilon)}. \end{aligned} \quad (8.6)$$

It follows from here that $R_m \leq c 2^\epsilon \sum_{n \geq m} 2^{-n(1/2-\epsilon)} \rightarrow 0$ as $m \rightarrow \infty$. $\square$

Lemma 8.2 *Let $(\xi_k)_{k=1,2,\dots}$ be a sequence of standard normal random variable on a complete probability space $(\Omega, \mathcal{F}, P)$. Then for every constant $c > \sqrt{2}$ and almost all $\omega \in \Omega$, there exists a number $N_0 = N_0(\omega, c)$ such that $|\xi_k(\omega)| < c\sqrt{\ln k}$ for all $k \geq N_0$.*

Proof For a standard normal random variable $\xi \sim N(0, 1)$ and $x > 0$, we have that

$$\begin{aligned} P(\omega : \xi \geq x) &= (2\pi)^{-1/2} \int_x^\infty e^{-y^2/2} dy \\ &= (2\pi)^{-1/2} \int_x^\infty \left(-\frac{1}{y}\right) de^{-y^2/2} \\ &= (2\pi)^{-1/2} \left(x^{-1} e^{-x^2/2} - \int_x^\infty y^{-2} e^{-y^2/2} dy \right) \\ &\leq x^{-1} (2\pi)^{-1/2} (x^{-1} e^{-x^2/2}) \end{aligned} \quad (8.7)$$

and, hence,

$$P(\omega : |\xi| \geq x) \leq x^{-1} (2\pi)^{1/2} e^{-x^2/2}.$$

Applying the above inequality, we can estimate the series as follows:

$$\sum_{k \geq 2} P\left(\omega : |\xi_k| \geq c(\ln k)^{1/2}\right) \leq c^{-1} (2/\pi)^{1/2} \sum_{k \geq 2} k^{-c^2/2} (\ln k)^{-1/2} < \infty,$$

and the statement of Lemma 8.2 follows from the Borel-Cantelli lemma. $\square$

Theorem 8.1 *Let $(\xi_k)_{k=1,2,\dots}$ be a sequence of independent standard normal random variables $(\Omega, \mathcal{F}, P)$. Define a stochastic process $W_t = W(t, \omega) = \sum_{k=1}^{\infty} \xi_k(\omega) S_k(t)$, $t \in [0, 1]$, $\omega \in \Omega$. Then $(W_t)_{t \in [0,1]}$ is a Wiener process.*

Proof First of all, we note that (W_t) has continuous trajectories by Lemmas 8.1 and 8.2.

Denote $Z_n(t) = \sum_{k=1}^n \xi_k S_k(t)$, $t \in [0, 1]$, and find that

$$\begin{aligned} \mathbf{E}|Z_{n+m}(t) - Z_n(t)|^2 &= \mathbf{E} \left| \sum_{k=n+1}^{n+m} \xi_k S_k(t) \right|^2 \\ &= \sum_{k=n+1}^{n+m} \sum_{l=n+1}^{n+m} S_k(t) S_l(t) \mathbf{E} \xi_k \xi_l \\ &= \sum_{k=n+1}^{n+m} S_k^2(t). \end{aligned} \quad (8.8)$$

Further, $\sum_{k=1}^{\infty} S_k^2(t) \leq \sum_{k=1}^{\infty} (2^{-k/2-1})^2 < \infty$, and, hence, by the completeness of $L^2(\Omega, \mathcal{F}, P)$, we arrive at the conclusion that there exists a limit $Z(t)$ of $Z_n(t)$ in this space such that

$$\mathbf{E}|Z_n(t) - Z(t)|^2 \rightarrow 0, \quad n \rightarrow \infty, \quad t \in [0, 1].$$

This limit $Z(t)$ coincides with W_t up to equivalence and $Z_n(t) \xrightarrow{t} W_t$ in L^2 -sense.

Now by $\mathbf{E}Z_n(t) = 0$, we obtain that

$$|\mathbf{E}W_t| = |\mathbf{E}W_t - \mathbf{E}Z_n(t)| \leq \left(\mathbf{E}|W_t - Z_n(t)|^2 \right)^{1/2} \xrightarrow{n \rightarrow \infty} 0,$$

which means that $\mathbf{E}W_t = 0$.

Using a continuity property of the scalar product, we find that

$$\langle Z_n(s), Z_n(t) \rangle_{L^2(\Omega)} \rightarrow \langle W_s, W_t \rangle_{L^2(\Omega)} = \mathbf{E}W_s W_t = \mathbf{Cov}(W_s, W_t), \quad s, t \in [0, 1].$$

On the other hand, we can use the independence of $(\xi_k)_{k=1,2,\dots}$ and $\mathbf{E}\xi_k = 0$, $\mathbf{E}\xi_k^2 = 1$ and observe that

$$\begin{aligned} \langle Z_n(s), Z_n(t) \rangle_{L^2(\Omega)} &= \mathbf{E}Z_n(s)Z_n(t) = \sum_{k=1}^n S_k(s)S_k(t) \mathbf{E}\xi_k^2 \\ &\rightarrow \sum_{k=1}^{\infty} S_k(s)S_k(t), \quad n \rightarrow \infty. \end{aligned}$$

Applying the Parseval equality for scalar products, we finally derive that

$$\begin{aligned} \sum_{k=1}^{\infty} S_k(s) S_k(t) &= \sum_{k=1}^{\infty} \langle H_k, I_{[0,s]} \rangle \langle H_k, I_{[0,t]} \rangle = \langle I_{[0,s]}, I_{[0,t]} \rangle \\ &= \min(s, t) = \mathbf{Cov}(W_s, W_t). \end{aligned}$$

Let us prove that $(W_t)_{t \in [0,1]}$ is a Gaussian process. Take $\mu = (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$, $t_1, \dots, t_n \in [0, 1]$ and define $Y = \sum_{m=1}^n \mu_m W_{t_m}$. We have now that

$$Y = \sum_{m=1}^n \mu_m \sum_{k=1}^{\infty} \xi_k S_k(t_m) = \sum_{k=1}^{\infty} b_k \xi_k,$$

where $b_k = b_k(\mu_1 \dots \mu_n, t_1, \dots, t_n) = \sum_{m=1}^n \mu_m S_k(t_m)$. Consider $Y_N = \sum_{k=1}^N b_k \xi_k \sim N(0, \sigma_N^2)$, $\sigma_N^2 = \sum_{k=1}^N b_k^2$ and find that

$$\mathbf{E} Y_N^2 = \sigma_N^2 \rightarrow \mathbf{E} Y^2 = \sigma^2, \quad N \rightarrow \infty,$$

in the sense of $L^2(\Omega)$. Therefore, we also have the convergence in distribution $Y_N \xrightarrow{d} Y$, $N \rightarrow \infty$. Hence, we have that the convergence of

$$\phi_Y(\lambda) = \exp(-\sigma_N^2 \lambda^2 / 2) \rightarrow \exp(-\sigma^2 \lambda^2 / 2) = \phi_Y(\lambda), \quad \lambda \in \mathbb{R},$$

i.e., $Y \sim N(0, \sigma^2)$.

Finally, we can extend this construction from the unit time interval to the whole $[0, \infty)$ as follows:

$$W(t, \omega) = \begin{cases} W_t = W(t, \omega) = W_1(t, \omega), & t \in [0, 1), \\ \sum_{j=1}^k W_j(1, \omega) + W_{k+1}(t - k, \omega), & t \in [k, k+1), \quad k = 1, 2, \dots, \end{cases}$$

where $W_n = (W_n(t))_{t \in [0,1]}$ is a sequence of independent Wiener processes. □

8.2 Stochastic Integrals with Respect to a Wiener Process

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, $(W_t)_{t \geq 0}$ be a Wiener process, and $(\mathcal{F}_t)_{t \geq 0}$ be a complete filtration generated by this process.

We describe here a scheme of construction of a stochastic integral

$$I(f) = \int_0^\infty f_s(\omega) dW_s \tag{8.9}$$

for any $\mathcal{F}_t$ -adapted and $\mathcal{B}(0, \infty) \times \mathcal{F}$ -measurable random function/process $f_t(\omega) = f(t, \omega)$ integrable in square with respect to $dt \times dP$:

$$\mathbf{E} \int_0^\infty f_s^2 ds < \infty. \quad (8.10)$$

The class of these random functions is denoted by S^2 , and the stochastic integral (8.9) must be a *linear and isometric operator* from $L^2(\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \times \mathcal{F}, dt \times dP)$ to $L^2(\Omega, \mathcal{F}, dP)$. We start its construction from the class of *step functions* $f \in S_0^2 \subseteq S^2$. For each function f , there exists a subdivision $0 = t_0 \leq t_1 \leq \dots \leq t_n < \infty$ such that f_{t_i} - $\mathcal{F}_{t_i}$ -measurable and square-integrable, and $f_t = f_{t_i}$ for $t \in [t_i, t_{i+1})$, $i \leq n$, whereas $f_t = 0$ for $t \geq t_n$. Hence, the weighted sum

$$\sum_{i=0}^{n-1} f_{t_i} I_{[t_i, t_{i+1})}(t) = f_t(\omega). \quad (8.11)$$

For function $f_t(\omega)$, defined by (8.11), we introduce a stochastic integral $I(f)$ as follows:

$$I(f) = \int_0^\infty f_s dW_s = \sum_{i=0}^{n-1} f_{t_i} (W_{t_{i+1}} - W_{t_i}). \quad (8.12)$$

It is almost obvious that the definition of $I(f)$ in (8.12) does not depend on changes of subdivisions (t_i) , and it is a linear mapping from S_0^2 to $L^2(\Omega, \mathcal{F}, P)$, i.e.,

$$I(\alpha f + \beta g) = \alpha I(f) + \beta I(g), \quad (8.13)$$

where $\alpha, \beta \in \mathbb{R}$, $f, g \in S_0^2$.

Using properties (1), (2), and (4) of a Wiener process $(W_t)_{t \geq 0}$ and properties of conditional expected values, we obtain that

$$\begin{aligned} \mathbf{E} I^2(f) &= \mathbf{E} \left[\sum_{i=0}^{n-1} f_{t_i}^2 (W_{t_{i+1}} - W_{t_i})^2 + 2 \sum_{j < i} f_{t_j} (W_{t_{j+1}} - W_{t_j}) f_{t_i} (W_{t_{i+1}} - W_{t_i}) \right] \\ &= \mathbf{E} \left[\sum_{i=0}^{n-1} f_{t_i}^2 \mathbf{E} \left((W_{t_{i+1}} - W_{t_i})^2 | \mathcal{F}_{t_i} \right) \right. \\ &\quad \left. + 2 \sum_{j < i} f_{t_j} (W_{t_{j+1}} - W_{t_j}) f_{t_i} \mathbf{E} \left((W_{t_{i+1}} - W_{t_i}) | \mathcal{F}_{t_i} \right) \right] \\ &= \sum_{i=0}^{n-1} \mathbf{E} f_{t_i}^2 (t_{i+1} - t_i) = \mathbf{E} \int_0^\infty f_s^2 ds. \end{aligned} \quad (8.14)$$

The *isometry* relation (8.14) means that the norms $\|I(f)\|_{L^2(\Omega, \mathcal{F}, P)}$ and $\|f\|_{L^2(\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \times \mathcal{F}, dt dP)}$ coincide. Using similar reasoning as in proving (8.14), one can derive for $f, g \in S_0^2$ that

$$\mathbf{E} I(f) I(g) = \mathbf{E} \int_0^\infty f_s g_s ds. \quad (8.15)$$

Equalities (8.13)–(8.15) show that the stochastic integral defined in (8.12) is a *linear* and *isometric* operator.

For function $f \in S_0^2$, we define a stochastic integral over interval $[0, t)$, $t > 0$, as follows:

$$I_t(f) = I(I_{[0,t)} f) = \int_0^t f_s dW_s. \quad (8.16)$$

It turns out that $(I_t(f))_{t \geq 0}$ is a *martingale* with respect to $(\mathcal{F}_t)_{t \geq 0}$.

To prove it, we fix $t > 0$, and for simplicity, assume that $t = t_k$.

Then we have

$I_{[0,t)} f_s = \sum_{i=0}^{n-1} f_{t_i} I_{[t_i, t_{i+1})}(t)$ and $I_t(f) = \sum_{i=0}^{n-1} f_{t_i} (W_{t_{i+1}} - W_{t_i})$, which are $\mathcal{F}_t$ –measurable. Further, if $s \leq t$ and $s = t_u$ and $t = t_k$, then

$$\mathbf{E}(I_t(f) - I_s(f) | \mathcal{F}_s) = \sum_{i=u}^{k-1} \mathbf{E}(f_{t_i} \mathbf{E}(W_{t_{i+1}} - W_{t_i} | \mathcal{F}_{t_i}) | \mathcal{F}_s) = 0.$$

For function $f \in S_0$, one can rewrite

$$I_t(f) = \sum_{i=0}^{n-1} f_{t_i} (W_{t_{i+1} \wedge t} - W_{t_i \wedge t}),$$

and notice that the process $(I_t(f))_{t \geq 0}$ is *continuous*.

Now we take a monotonic sequence (s_j) such that $0 \leq s_0 \leq s_1 \leq \dots \leq s_m < \infty$. Then the sequence $(I_{s_j}(f), \mathcal{F}_j)_{j=0, \dots, m}$ is a martingale for which the Doob inequality is true

$$\mathbf{E} \sup_{0 \leq j \leq m} I_{s_j}^2(f) \leq 4 \mathbf{E} I_{s_m}^2(f) \leq 4 \mathbf{E} \int_0^{s_m} f_t^2 dt \leq 4 \mathbf{E} \int_0^\infty f_t^2 dt. \quad (8.17)$$

Considering (s_j) rational and using the continuity of the trajectories of $I_t(f)$, we arrive at the *Doob inequality* for $f \in S_0^2$ taking *sup* over $t \in \mathbb{R}_+$ in (8.17):

$$\mathbf{E} \sup_t I_t^2(f) \leq 4 \mathbf{E} \int_0^\infty f_t^2 dt. \quad (8.18)$$

One can extend the operator $I_t(f)$ from S_0^2 to its closure S^2 . Any function $f \in S^2$ can be approximated by a sequence $f_n \in S_0^2$.

A concrete construction of such a sequence $(f_n)_{n=1,2,\dots} \in S_0^2$ can be provided as follows. For any function $f \in S^2$, we define

$$f_n(t) = \begin{cases} n \int_{(k-1)/n}^{k/n} f(s) ds, & \text{for } t \in [k/n, (k+1)/n), \quad k = 1, 2, \dots, n^2 - 1, \\ n = 1, 2, \dots \\ 0, & \text{otherwise.} \end{cases}$$

Using the Jensen inequality, we can find that

$$\int_{k/n}^{(k+1)/n} |f_n(t)|^2 dt = n \left| \int_{(k-1)/n}^{k/n} f(t) dt \right|^2 \leq \int_{(k-1)/n}^{k/n} |f(t)|^2 dt, \quad k = 1, 2, \dots$$

Hence, (a.s.) $\int_0^\infty |f_n(t)|^2 dt \leq \int_0^\infty |f(t)|^2 dt < \infty$,

$$\int_0^\infty |f(t) - f_n(t)|^2 dt \leq 2 \int_0^\infty (|f(t)|^2 + |f_n(t)|^2) dt \leq 4 \int_0^\infty |f(t)|^2 dt < \infty.$$

Now, for arbitrary $N = 1, 2, \dots$, we easily observe that $\mathbf{E} \int_0^N |f - f_n(t)|^2 dt \rightarrow 0$, $n \rightarrow \infty$, by construction of $(f_n)_{n=1,2,\dots}$. Further, we have

$$\begin{aligned} \int_0^\infty |f(t) - f_n(t)|^2 dt &= \int_0^N |f(t) - f_n(t)|^2 dt + \int_N^\infty |f(t) - f_n(t)|^2 dt \\ &\leq \int_0^N |f(t) - f_n(t)|^2 dt + 2 \int_N^\infty (|f(t)|^2 + |f_n(t)|^2) dt \\ &\leq \int_0^N |f(t) - f_n(t)|^2 dt \\ &\quad + 2 \int_N^\infty (|f(t)|^2 + \int_{N-1/n}^\infty |f_n(t)|^2) dt \\ &\leq \int_0^N |f(t) - f_n(t)|^2 dt + 2 \int_{N-1}^\infty |f(t)|^2 dt. \end{aligned} \quad (8.19)$$

The term $\int_{N-1}^\infty |f(t)|^2 dt \rightarrow 0$ as $N \rightarrow \infty$, and by the dominated convergence theorem, $\mathbf{E} \int_0^\infty |f(t) - f_n(t)|^2 dt \rightarrow 0$ as $n \rightarrow \infty$.

Then for m and $n \in \mathbb{Z}_+$, we have

$$\|I(f_m) - I(f_n)\|_{L^2(\Omega)} = \|I(f_m - f_n)\|_{L^2(\Omega)} = \|f_m - f_n\|_{S^2}.$$

It means that $(I(f_n))_{n \geq 1}$ is a fundamental sequence in a complete space $L^2(\Omega, \mathcal{F}, P)$. Hence, it converges to some element of this space:

$$I(f) = \lim_{n \rightarrow \infty} I(f_n), \quad (8.20)$$

and we call $I(f)$ a *stochastic integral* of $f \in S^2$. We also define for $t \geq 0$

$$\int_0^t f_s dW_s = \int_0^\infty I_{[0,t)}(s) f_s dW_s, \quad (8.21)$$

and we prove that the stochastic integral (8.21) admits a *continuous modification*.

To prove it, we take a sequence $f_n = f_n(t, \omega) \in S_0^2$ such that

$$\mathbf{E} \int_0^\infty |f(s) - f_n(s)|^2 ds \leq 2^{-n}, \quad n = 1, 2, \dots$$

For each $t \geq 0$, we have in space $L^2(\mathbb{R}_+ \times \Omega, \mathcal{B}(0, \infty) \times \mathcal{F}, dt dP)$ that

$$I_{[0,t)}(s)(f) = I_{[0,t)}(s)f_1 + \dots + I_{[0,t)}(s)(f_{n+1} - f_n) + \dots \quad (8.22)$$

The equality (8.22) by continuity in the sense of this space is transformed to the next one

$$\int_0^t f(s) dW_s = \int_0^t f_1(s) dW_s + \dots + \int_0^t |f_{n+1}(s) - f_n(s)| dW_s + \dots \quad (8.23)$$

On the right-hand side of (8.23), each term is continuous as functions $f_n \in S_0^2$. Therefore, to prove the continuity of stochastic integral (8.21), it is enough to provide a uniform convergence (a.s.) of series (8.21). First, by the Doob inequality, we obtain that

$$\mathbf{E} \sup_{t \geq 0} \left| \int_0^t (f_{n+1} - f_n(s)) dW_s \right|^2 \leq 4 \mathbf{E} \int_0^\infty (f_{n+1}(s) - f_n(s))^2 ds \quad (8.24)$$

$$\leq 16 \cdot 2^{-n}. \quad (8.25)$$

Second, combining (8.24) with the Chebyshev inequality, we get

$$P \left(\sup_t \left| \int_0^t (f_{n+1}(s) - f_n(s)) dW_s \right| \geq n^{-2} \right) \leq 16n^4 2^{-n}. \quad (8.26)$$

Application of (8.25) together with the Borel-Cantelli lemma certifies that the series (8.23) converges (a.s.).

Finally, we can conclude that the extension of $I_t(f)$ from space S_0^2 to S^2 leaves its *basic properties*:

- (1) Linearity
- (2) Isometry
- (3) Martingality
- (4) Continuity of trajectories

Before starting to state other properties of stochastic integrals, we introduce a special class of random variables needed here.

Let τ be a random variable taking values in $[0, \infty]$. We call it a *stopping time* on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$, generated by a Wiener process $(W_t)_{t \geq 0}$, if $\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$ for every $t \in \mathbb{R}_+$.

A natural *example* is the *first hitting time* of the point a by W_t :

$$\begin{aligned}\tau &= \tau_a = \inf\{t : W_t \geq a\}, \\ \tau &= \infty, \text{ if the set in brackets } \{.\} \text{ is empty.}\end{aligned}$$

Indeed, one can observe that

$$\{\omega : \tau(\omega) \leq t\} = \Omega \setminus \{\omega : \tau(\omega) > t\}. \quad (8.27)$$

Further, we have a relation

$$\{\omega : \tau(\omega) > t\} = \{\omega : \max_{s \leq t} W_s(\omega) < a\}. \quad (8.28)$$

Let us note that

$$\max_{s \leq t} W_s = \sup_{r \leq t, r \in \mathbb{Q}} W_r, \quad (8.29)$$

where $\mathbb{Q}$ is the set of rational numbers, and hence $\max_{s \leq t} W_s$ is $\mathcal{F}_t$ -measurable. Using continuity of trajectories of W_t and relations (8.27)–(8.29), we arrive at the conclusion that $\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$.

Let us connect a stochastic integral $I_t(f) = X_t = \int_0^t f_s dW_s$, $f \in S^2$, and a stopping time τ . We have the two following equalities which are very useful:

$$X_\tau = \int_0^\infty I_{s < \tau} f_s dW_s = \int_0^\infty I_{s \leq \tau} f_s dW_s \quad (a.s.), \quad (8.30)$$

$$\mathbf{E} \left(\int_0^\tau f_s dW_s \right)^2 = \mathbf{E} \int_0^\tau f_s^2 ds \quad (\text{Wald's identity}). \quad (8.31)$$

To prove (8.30), we start with a stopping time τ taking discrete values $t_1, t_2, \dots$ and find that $I_{s < t_k} f_s = I_{s < \tau} f_s$ on the set $\{\omega : \tau = t_k\}$.

On each such set (a.s.),

$$X_\tau = X_{t_k} = \int_0^\infty I_{(s < t_k)} f_s dW_s = \int_0^\infty I_{(s < \tau)} f_s dW_s.$$

Hence, it is true (a.s.) on $\Omega = \cup_k \{\tau = t_k\}$. To prove (8.30) for an arbitrary stopping time τ , we approximate τ by a sequence of *discrete* stopping times

$$\tau_n = 2^{-2} [2^n \tau] + 2^{-n},$$

such that $\tau \leq \tau_n$ and $\tau_n - \tau \leq 2^{-n}$. Further, by a continuity of X_t , we get convergence $X_{\tau_n} \rightarrow X_\tau$ (a.s.) and by the dominated convergence theorem

$$\begin{aligned} & \mathbf{E} \left(\int_0^\infty I_{\{s < \tau\}} f_s dW_s - \int_0^\infty I_{s < \tau_n} f_s dW_s \right)^2 \\ &= \mathbf{E} \int_0^\infty I_{\{\tau \leq s < \tau_n\}} f_s^2 ds \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

Therefore, we obtain (8.30), and using isometry property, we get (8.31):

$$\mathbf{E} X_\tau^2 = \mathbf{E} \int_0^\infty I_{s < \tau} f_s^2 ds = \mathbf{E} \int_0^\tau f_s^2 ds.$$

It turns out that a certain extension of the notion of a stochastic integral with respect to a Wiener process can be done using the following *localization procedure*. It is given with the help of stopping times.

Consider the set of $\mathcal{B}(0, \infty) \times \mathcal{F}$ -measurable adapted random functions $f = f_t(\omega)$ such that

$$\int_0^T f_s^2 ds < \infty \quad (\text{a.s.}) \text{ for all } T \in \mathbb{R}_+. \quad (8.32)$$

The family of such functions satisfying (8.32) is denoted by S .

We define a sequence of stopping time $(\tau_n)_{n=1,2,\dots}$ such that

$$\tau_n = \inf \left(t : \int_0^t f_s^2 ds \geq n \right).$$

One can easily observe that (a.s.)

$$\int_0^{\tau_n} f_s^2 ds \leq n \text{ and } \tau_n \uparrow \infty, \quad n \uparrow \infty, \text{ and } I_{\{t < \tau_n\}} f \in S^2 \text{ for } n = 1, 2, \dots$$

It follows from here that

$$X_t^n = \int_0^t I_{\{s < \tau_n\}} f_s dW_s$$

is well defined for $n = 1, 2, \dots$. One can also note that

$$X_t^n(\omega) = X_t^m(\omega),$$

for $m \geq n$, $0 \leq t \leq \tau_n(\omega)$ and $\omega \in \Omega'$, $P(\Omega') = 1$.

Moreover, the sequence $(X_t^n)_{n=1,2,\dots}$ converges uniformly on finite time intervals.

All arguments above give a possibility to define a stochastic integral for each $f \in S$ as follows:

$$\int_0^t f_s dW_s = \lim_{n \rightarrow \infty} \int_0^t I_{s < \tau_n} f_s dW_s. \quad (8.33)$$

Stochastic integral (8.33) for $f \in S$ admits a continuous modification and saves a linearity property. But it is not a martingale anymore. Nevertheless, the martingale property is fulfilled *locally*:

$$\left(\int_0^{t \wedge \tau_n} f_s dW_s \right)_{t \geq 0} \text{ is a square-integrable martingale for } n = 1, 2, \dots$$

Remark 8.3 Let us make a remark regarding *finite-dimensional processes*.

We say that (W_t) is a d -dimensional Wiener process if $W_t = (W_t^1, \dots, W_t^d)$, where (W_t^i) , $i = 1, \dots, d$, is a one-dimensional independent Wiener process.

For d -dimensional random function $f_t = (f_t^1, \dots, f_t^d)$ with components from S , we define

$$\int_0^t f_s dW_s = \sum_{i=1}^d \int_0^t f_s^i dW_s^i.$$

As far as a stochastic integration of matrix-valued random functions is concerned, $\sigma_t = (\sigma_t^{ij})$, $i = 1, \dots, d$, $j = 1, \dots, d$, and $\sigma^{ij} \in S$, we define $\int_0^t \sigma_s dW_s$ as a d -dimensional process with the i -th coordinate $\sum_{j=1}^d \int_0^t \sigma_s^{ij} dW_s^j$.

Remark 8.4 The construction of a stochastic integral developed in this section was given by K. Ito. That is why this integral is often called the *Ito integral*. A similar construction was given by R. Stratonovich who proposed to use a different class of integral sums:

$$\sum_{i=0}^{N-1} f(t_i^*) (W_{t_{i+1}} - W_{t_i}),$$

where $t_i^* = \frac{1}{2}(t_{i+1} + t_i)$.

The corresponding stochastic integral is called the *Stratonovich integral*.

Remark 8.5 It is well-known that the construction of the Riemann integral as a limit of the integral sums does not depend on the choice of intermediate points inside of subdivision intervals. It was shown in Remark 8.4 that the construction of a stochastic integral is not invariant with respect to such a choice. Let us give another example of this type. Take $f(t) = W(t) = W_t$ and consider a subdivision of $[0, T]$:

$$0 = t_0 < t_1 < \dots < t_n = T, \quad t_j = \frac{Tj}{n}, \quad j = 1, \dots, n.$$

For this subdivision, we define two integral sums as follows:

$$I_1^{(n)} = \sum_{i=1}^n W_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}) \text{ and } I_2^{(n)} = \sum_{i=1}^n W_{t_i} (W_{t_i} - W_{t_{i-1}}).$$

For two real numbers a and b , we note that $a(b - a) = \frac{1}{2}(b^2 - a^2) - \frac{1}{2}(a - b)^2$ and $b(b - a) = \frac{1}{2}(b^2 - a^2) + \frac{1}{2}(a - b)^2$.

Applying these elementary equalities to the integral sums, we obtain that

$$I_1^{(n)} = \frac{1}{2} \sum_{i=1}^n (W_{t_i}^2 - W_{t_{i-1}}^2) - \frac{1}{2} \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 = \frac{1}{2} W_T^2 - \frac{1}{2} \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2,$$

$$I_2^{(n)} = \frac{1}{2} \sum_{i=1}^n (W_{t_i}^2 - W_{t_{i-1}}^2) + \frac{1}{2} \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 = \frac{1}{2} W_T^2 + \frac{1}{2} \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2.$$

Hence,

$$\mathbf{E} I_1^{(n)} = \frac{1}{2} T - \frac{1}{2} \mathbf{E} \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 \rightarrow \frac{T}{2} - \frac{T}{2} = 0, \quad n \rightarrow \infty,$$

$$\mathbf{E} I_2^{(n)} = \frac{1}{2} T + \frac{1}{2} \mathbf{E} \sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 \rightarrow \frac{T}{2} + \frac{T}{2} = T, \quad n \rightarrow \infty,$$

and we observe that L^2 -limits of integral sums are different.

Remark 8.6 It is very often another type of measurability of f is exploited in the definition of stochastic integrals, i.e., f is *progressively measurable* in the sense: for each t ,

$$f : ([0, t] \times \Omega, \mathcal{B}(0, t) \times \mathcal{F}_t) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})).$$

Without loss of generality, we can count S_0^2, S^2, S as classes of progressively measurable random functions.

8.3 The Ito Processes: Formula of Changing of Variables, Theorem of Girsanov, and Representation of Martingales

Let us take two progressively measurable (or $\mathcal{F}_t$ -adapted and $\mathcal{B}(0, \infty) \times \mathcal{F}$ -measurable) $(b_t(\omega))_{t \geq 0}$ and $(\sigma_t(\omega))_{t \geq 0}$ satisfying the next integrability conditions: for each $T > 0$ (a.s.),

$$\int_0^T |b_s(\omega)| ds < \infty, \quad \int_0^T |\sigma_s(\omega)|^2 ds < \infty. \quad (8.34)$$

Conditions (8.34) imply a possibility to define a new stochastic process

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s, \quad (8.35)$$

where X_0 is a finite (a.s.) random variable, and $(W_t)_{t \geq 0}$ is a Wiener process generating filtration $(\mathcal{F}_t)_{t \geq 0}$.

The process $(X_t)_{t \geq 0}$ is called the *Ito process*. It is very convenient to write equality (8.35) in the following differential form:

$$dX_t = b_t dt + \sigma_t dW_t. \quad (8.36)$$

In the above representation (8.36), we can speak about dX_t as a *stochastic differential* of $(X_t)_{t \geq 0}$.

We will study the class of the Ito processes using both form (8.35) and (8.36). As the first result playing a very important role in Stochastic Analysis is the *formula of changing of variables* (the *Ito formula*) for the class of smooth functions $F(t, x)$, having one continuous derivative $\frac{\partial F}{\partial t}$ and two continuous derivatives $\frac{\partial F}{\partial x}$ and $\frac{\partial^2 F}{\partial x^2}$, i.e., $F \in C^{1,2}$.

We start with the next lemma.

Lemma 8.3 *Let $(W_t)_{t \geq 0}$ be a Wiener process, then for each $t \geq 0$ (a.s.),*

$$W_t^2 = 2 \int_0^t W_s dW_s + t. \quad (8.37)$$

Proof Without loss of generality, we consider only $t = 1$ and find that in the sense of convergence in probability (P -lim)

$$\begin{aligned} \int_0^1 W_s dW_s &= \lim_{n \rightarrow \infty} \sum_{k < n} W_{\frac{k}{n}} \left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}} \right) \\ &= \lim_{n \rightarrow \infty} \left[\sum_{k < n} \frac{1}{2} \left(W_{\frac{k+1}{n}} + W_{\frac{k}{n}} \right) \left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}} \right) \right. \\ &\quad \left. - \sum_{k < n} \frac{1}{2} \left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}} \right)^2 \right]. \end{aligned} \quad (8.38)$$

It is clear that as $n \rightarrow \infty$

$$\begin{aligned} \sum_{k < n} \frac{1}{2} \left(W_{\frac{k+1}{n}} + W_{\frac{k}{n}} \right) \left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}} \right) &= \frac{1}{2} \sum_{k < n} \left(W_{\frac{k+1}{n}}^2 - W_{\frac{k}{n}}^2 \right) \rightarrow \frac{1}{2} W_1^2, \\ \sum_{k < n} \left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}} \right)^2 &\rightarrow \frac{1}{2}. \end{aligned}$$

Hence, we can get from (8.38) the equality (8.37). $\square$

The next step is the formula for the product of two Ito's processes $(X_t^i)_{t \geq 0}$, $i = 1, 2$, with coefficients b^i , σ^i correspondingly.

Lemma 8.4 *For each $t \geq 0$ (a.s.),*

$$dX_t^1 X_t^2 = X_t^1 dX_t^2 + X_t^2 dX_t^1 + \sigma_t^1 \sigma_t^2 dt. \quad (8.39)$$

Proof Dividing $[0, t]$ with the help of a subdivision $(t_k)_{k \leq n}$ with the diameter $\max_{k \leq n} |t_k - t_{k-1}| \rightarrow 0$, $n \rightarrow \infty$, we can reduce the proof to the “small” intervals (t_k, t_{k+1}) of this subdivision.

For such small enough interval, we have a discrete analog of (8.39):

$$\begin{aligned} X_{t_{k+1}}^1 X_{t_{k+1}}^2 - X_{t_k}^1 X_{t_k}^2 &= \int_{t_k}^{t_{k+1}} X_s^1 \left[b_s^2 ds + \sigma_s^2 dW_s \right] \\ &\quad + \int_{t_k}^{t_{k+1}} X_s^2 \left[b_s^1 ds + \sigma_s^1 dW_s \right] + \int_{t_k}^{t_{k+1}} \sigma_s^1 \sigma_s^2 ds, \end{aligned}$$

where we can think about b^i, σ^i , $i = 1, 2$ as constants. Therefore, the problem is reduced to the following four cases:

- (1) $X_t^1 = t, X_t^2 = t$.
- (2) $X_t^1 = t, X_t^2 = W_t$.
- (3) $X_t^1 = W_t, X_t^2 = t$.
- (4) $X_t^1 = W_t, X_t^2 = W_t$.

Case (1) is the usual differentiation case.

Case (4) is Lemma 8.3.

Cases (2) and (3) are equivalent, and we need to show that for each $t \geq 0$ (a.s.),

$$t W_t = \int_0^t W_s ds + \int_0^t s dW_s,$$

which is true according to the following obvious relation: for $t = 1$,

$$\begin{aligned} 1 \cdot W_1 &= P - \lim_{n \rightarrow \infty} \sum_{k < n} \left[\frac{k+1}{n} W_{\frac{k+1}{n}} - \frac{k}{n} W_{\frac{k}{n}} \right] \\ &= P - \lim_{n \rightarrow \infty} \left[\sum_{k < n} \frac{k}{n} \left(W_{\frac{k+1}{n}} - W_{\frac{k}{n}} \right) + \sum_{k < n} \frac{1}{n} W_{\frac{k+1}{n}} \right]. \end{aligned}$$

So, we arrive at the equality (8.39). □

Lemma 8.5 For any polynomial function $P_n(x)$ of degree $n = 1, 2, \dots$, we have

$$dP_n(W_t) = P'_n(W_t)dW_t + \frac{1}{2}P''_n(W_t)dt. \quad (8.40)$$

Proof It is sufficient to prove (8.40) for $P_n(x) = x^n$. We can use the induction method to prove that

$$dW_t^n = nW_t^{n-1}dW_t + \frac{n(n-1)}{2}W_t^{n-2}dt. \quad (8.41)$$

For $n = 1$, the formula (8.41) is trivial. Further, we take $X_t^1 = W_t^n$ and $X_t^2 = W_t$; use Lemma 8.4 with

$$b_t^1 = \frac{n(n-1)}{2}W_t^{n-1}, \sigma_t^1 = nW_t^{n-2}, b_t^2 = 0, \sigma_t^2 = 1,$$

and obtain (8.41):

$$\begin{aligned} dW_t^{n+1} &= W_t^n dW_t + W_t \left[n W_t^{n-1} dW_t + \frac{n(n-1)}{2} W_t^{n-2} dt \right] \\ &\quad + n W_t^{n-1} dt \\ &= (n+1) W_t^n dW_t + \frac{n(n+1)}{2} W_t^{n-1} dt. \end{aligned}$$

□

Corollary 8.1 *If $F = F(x)$ is a continuously differentiable function with continuous derivatives $F'(x)$ and $F''(x)$, then for $t \geq 0$ (a.s.),*

$$dF(W_t) = F'(W_t)dW_t + \frac{1}{2}F''(W_t)dt. \quad (8.42)$$

To prove (8.42), we uniformly (on compact intervals) approximate $F(x)$ by a sequence of polynomials P_n such that

$$F(x) = \lim_{n \rightarrow \infty} P_n(x), \quad F'(x) = \lim_{n \rightarrow \infty} P'_n(x), \quad F''(x) = \lim_{n \rightarrow \infty} P''_n(x)$$

and take the limit in the formula (8.40).

Corollary 8.2 *Assume $F : \mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$ is a function of class $C^{1,2}$, then for $t \geq 0$ (a.s.),*

$$dF(t, W_t) = \left[\frac{\partial F}{\partial t}(t, W_t) + \frac{1}{2} \frac{\partial^2 F}{\partial x^2}(t, W_t) \right] dt + \frac{\partial F}{\partial x}(t, W_t) dW_t. \quad (8.43)$$

The first step of the proof is the case $F(t, x) = F_1(t)F_2(x)$. Then (8.43) follows from Lemma 8.4 and Corollary 8.1. The general case of F can be derived by approximation of $F(t, x)$ by a sequence of functions $F_n(t, x) = \sum_{k \leq n} F_{1,n}(t)F_{2,n}(x)$ with $\frac{\partial F_n}{\partial t}, \frac{\partial F_n}{\partial x}, \frac{\partial^2 F_n}{\partial x^2}$ converging uniformly on compact sets.

Remark 8.7 Let us define the local time of W at the level a during $[0, t]$:

$$l_W(t, a) = \lim_{\epsilon \rightarrow 0} \frac{1}{2\epsilon} \int_0^t I_{(a-\epsilon, a+\epsilon)}(W_s) ds,$$

which exists (a.s.) and in the sense of L^2 -space. Using this notion, one can get a version of the Ito formula for non-differentiable functions, say, $F(x) = |x|$:

$$|W_t - a| - |W_0 - a| = \int_0^t \text{sgn}(W_s - a) dW_s + l_W(t, a),$$

where

$$\operatorname{sgn}(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$$

Such a formula is called as the Tanaka formula.

The final formula (for Ito's processes) is given in the next theorem.

Theorem 8.2 Assume $F \in C^{1,2}$ and

$$dX_t = b_t(\omega)dt + \sigma_t(\omega)dW_t.$$

Then for $t \geq 0$ (a.s.),

$$\begin{aligned} dF(t, X_t) &= \left[\frac{\partial F}{\partial t}(t, X_t) + b_t \frac{\partial F}{\partial x}(t, X_t) + \frac{\sigma_t^2}{2} \frac{\partial^2 F}{\partial x^2}(t, X_t) \right] dt \\ &\quad + \sigma_t \frac{\partial F}{\partial x}(t, X_t) dW_t. \end{aligned} \quad (8.44)$$

The formula (8.44) admits a generalization to a multidimensional case. Namely, assume $F : \mathbb{R}_+ \times \mathbb{R}^k \rightarrow \mathbb{R}$ of class $C^{1,2}$; the process (X_t) in (8.35) is defined by a vector function $b_t(\omega) = b(t, \omega) = (b_t^1(\omega), \dots, b_t^k(\omega))$ and a matrix-value function $\sigma_t(\omega) = \sigma(t, \omega) = (\sigma_t^{ij}(\omega))_{i \leq k, j \leq d}$:

$$X_t^i = X_0^i + \int_0^t b_s^i ds + \int_0^t \sum_{j=1}^d \sigma_s^{ij} dW_s^j, \quad i = 1, \dots, k,$$

and $W_t = (W_t^1, \dots, W_t^d)$ is a d -dimensional Wiener process. Then for $t \geq 0$ (a.s.),

$$dF(t, X_t) = \left[\frac{\partial F}{\partial t}(t, X_t) + \sum_{i=1}^k \frac{\partial F}{\partial x^i}(t, X_t) b_t^i + \frac{1}{2} \sum_{i,j=1}^k \frac{\partial^2 F}{\partial x_i \partial x_j}(t, X_t) \right] \quad (8.45)$$

$$\times \sum_{l=1}^d \sigma_t^{il} \sigma_t^{lj} \Big] dt + \sum_{j=1}^d \left[\sum_{i=1}^k \frac{\partial F}{\partial x_i}(t, X_t) \sigma_t^{ij} \right] dW_t^j. \quad (8.46)$$

Example 8.2 (Stochastic Exponential) Consider the Ito process (8.35). Define the process

$$\mathcal{E}_t(X) = \exp \left\{ X_t - \frac{1}{2} \int_0^t \sigma_s^2 ds \right\}. \quad (8.47)$$

Applying the Ito formula to function $F(t, x) = e^{-\frac{\sigma^2}{2}t} e^x$, $(t, x) \in \mathbb{R}_+ \times \mathbb{R}$, we find that

$$d\mathcal{E}_t(X) = \mathcal{E}_t(X)dX_t, \quad \mathcal{E}_0(X) = 1. \quad (8.48)$$

We will call $(\mathcal{E}_t(X))_{t \geq 0}$ a *stochastic exponent* of process (X_t) .

Let us count the following properties of stochastic exponentials (8.47) and (8.48).

Problem 8.3

1. $\mathcal{E}_t(X) > 0$ (a.s.), $t \in \mathbb{R}_+$.
2. $\frac{1}{\mathcal{E}_t(X)} = \mathcal{E}_t(\tilde{X})$, (a.s.) $t \in \mathbb{R}_+$.
3. If $b_t = 0$ a.s., $t \in \mathbb{R}_+$, then $(\mathcal{E}_t(X))_{t \geq 0}$ is a (local) martingale.
4. $\mathcal{E}_t(X^1)\mathcal{E}_t(X^2) = \mathcal{E}_t(X^1 + X^2 + [X^1, X^2])$ (a.s.), $t \in \mathbb{R}_+$, where $dX_t^i = b_t^i dt + \sigma^i dW_t$, $d[X^1, X^2]_t = \sigma_t^1 \sigma_t^2 dt$, $i = 1, 2$, and this property is called a *multiplication rule* of stochastic exponentials.

We already demonstrated how important and useful the Ito formula is. The next consideration has an incredible meaning for stochastic analysis and its applications. Let us consider a d -dimensional random function $b_t = (b_t^1, \dots, b_t^d)$ belonging to the space S (or S^2) and define a new process

$$\begin{aligned} Z_t &= \exp \left\{ \int_0^t b_s dW_s - \frac{1}{2} \int_0^t |b_s|^2 ds \right\} \\ &= \exp \left\{ \sum_{i=1}^d \int_0^t b_s^i dW_s^i - \frac{1}{2} \sum_{i=1}^d \int_0^t (b_s^i)^2 ds \right\}, \end{aligned} \quad (8.49)$$

$$Z_0 = 1.$$

The process (8.49) satisfies the following properties:

- (1) $dZ_t = Z_t b_t dW_t$, $Z_0 = 1$.
- (2) $(Z_t)_{t \geq 0}$ is a supermartingale, which is a martingale, if $(b_t)_{t \geq 0}$ is bounded or $\mathbf{E}Z_t = 1$, $t \in \mathbb{R}_+$.
- (3) if $\mathbf{E}Z_T = 1$ for some $T > 0$, then $(Z_t)_{t \leq T}$ is a martingale, and for any sequence $b_n \in S(S^2)$ with the property $\int_0^T |b_n - b|^2 ds \rightarrow 0$, $n \rightarrow \infty$, (a.s.) the corresponding sequence $Z_T(b_n)$ converges to $Z_T(b)$ in the space L^1 .

To prove (1), we just apply the Ito formula. For the proof of (2), we consider a sequence of stopping times

$$\tau_n = \inf \left(t : \int_0^t |b_s|^2 Z_s^2 ds \geq n \right)$$

and find that the process

$$\left(\int_0^t I_{(s < \tau_n)} b_s Z_s dW_s \right)_{t \geq 0}$$

is a martingale, as well as $(Z_{t \wedge \tau_n})_{t \geq 0}$.

For $t_1 \leq t_2$, we have (a.s.)

$$\mathbf{E}(Z_{t_2 \wedge \tau_n} | \mathcal{F}_{t_1}) = Z_{t_1 \wedge \tau_n}, \quad n=1,2,\dots,$$

and as $n \rightarrow \infty$, we get by Fatou's Lemma that

$$\mathbf{E}(Z_{t_2} | \mathcal{F}_{t_1}) \leq Z_{t_1} \quad (a.s.),$$

which certifies that $(Z_t)_{t \geq 0}$ is a supermartingale, and therefore

$$\mathbf{E} \exp \left(\int_0^t b_s dW_s - \frac{1}{2} \int_0^t |b_s|^2 ds \right) \leq 1. \quad (8.50)$$

To prove (3), we first note that for $|b_s| \leq K < \infty$, we derive from (8.50) that

$$\begin{aligned} \mathbf{E} \int_0^t |b_s|^2 Z_s^2 ds &\leq K^2 \mathbf{E} \int_0^t Z_s^2 ds \\ &= K^2 \int_0^t \mathbf{E} Z_s^2(2b) \exp \left(\int_0^t |b_u|^2 du \right) ds \\ &\leq K^2 \int_0^t e^{K^2 s} ds < \infty \end{aligned}$$

and conclude that $\left(\int_0^t b_s Z_s ds \right)_{t \geq 0}$ and $(Z_t)_{t \geq 0}$ are martingales.

The claim (3) follows from claim (2) above, where $\mathbf{E} Z_T(b) = 1$ and $Z_T(b^n) \xrightarrow{P} Z_T(b)$, $n \rightarrow \infty$. Also, $\mathbf{E}|Z_T(b^n) - Z_T(b)| \rightarrow 0$, $n \rightarrow \infty$, and hence $Z_t(b^n) = \mathbf{E}(Z_T(b^n) | \mathcal{F}_t)$, $t \leq T$, a.s. Taking the limit in the above equality as $n \rightarrow \infty$, we get the martingale property of $(Z_t(b))_{t \leq T}$.

Now we are ready to formulate a *beautiful Girsanov Theorem*.

Theorem 8.3 Assume $(W_t)_{t \leq T}$ is a Wiener process on a probability space $(\Omega, \mathcal{F}, P)$. Let b be a random function from space S (or S^2) such that $\mathbf{E} Z_T = 1$. Define the process $X_t = W_t - \int_0^t b_s ds$. Then

- (1) A random variable Z_T defines a new probability measure $\tilde{P}$, equivalent to P , i.e., $\tilde{P} \ll P$ and $P \ll \tilde{P}$ with the density $Z_T = \frac{d\tilde{P}}{dP}$.
- (2) The process (X_t) is a Wiener process with respect to $\tilde{P}$.

Proof For simplicity, consider a bounded function (b_t) . For $0 \leq t_0 \leq t_1 \leq \dots \leq t_n = T$ and fixed $(\lambda_j)_{j=0, \dots, n-1}$ from $\mathbb{R}^d$, we define the process λ_s as $i\lambda_j$ on $[t_j, t_{j+1})$. Further, we obtain that

$$\begin{aligned} \tilde{\mathbf{E}} \exp \left\{ i \sum_{j=0}^{n-1} \lambda_j (X_{t_{j+1}} - X_{t_j}) \right\} &= \mathbf{E} \exp \left\{ \int_0^T \lambda_s dW_s - \int_0^T \lambda_s b_s ds \right\} Z_T(b) \\ &= \mathbf{E} Z_T(\lambda + b) \exp \left(\frac{1}{2} \int_0^T \lambda_s^2 ds \right) \\ &= e^{1/2 \int_0^T \lambda_s^2 ds}, \end{aligned}$$

which corresponds to a Gaussian distribution. Taking into account $X_0 = 0$, we arrive at the claim (2) of the theorem. $\square$

To finish this section, we study the structure of any martingale $(M_t)_{t \geq 0}$ on stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t = \mathcal{F}_t^W)_{t \geq 0}, P)$. It turns out that the set of all martingales on the stochastic basis, generated by a Wiener process $(W_t)_{t \geq 0}$, is determined by stochastic integrals. Namely, any square integrable martingale $(M_t)_{t \geq 0}$, $M_0 = 0$, admits

$$M_t = \int_0^t \phi_s dW_s \quad (a.s.), \quad t \geq 0, \quad (8.51)$$

where $(\phi_t)_{t \geq 0}$ is uniquely defined progressively measurable function from space S^2 .

The relation (8.51) is called a martingale representation. Let us prove it for a finite interval $[0, T]$. For a deterministic function $\beta \in L^2(0, T)$, we define the Girsanov exponent

$$Z_t = Z_t(\beta) = \exp \left\{ \int_0^t \beta_u du - \frac{1}{2} \int_0^t \beta_u^2 du \right\}, \quad Z_0 = 1$$

Let us note that

$$Z_t^2(\beta) = Z_t(2\beta) \exp \left\{ \int_0^t \beta_u^2 du \right\},$$

and $\mathbf{E} Z_t(2\beta) \leq 1$ for all $t \leq T$.

Hence, we obtain

$$\mathbf{E} Z_t^2(\beta) \leq \exp \left\{ \int_0^t \beta_u^2 du \right\} < \infty.$$

Let $\mathfrak{X}^2$ be a class of random variables X from the space $L^2 = L^2(\Omega, \mathcal{F} = \mathcal{F}_T^W, P) = L^2$ such that $X = \int_0^T \phi_s dW_s$ with a progressively measurable ϕ

and $\int_0^T \mathbf{E} \phi_s^2 ds < \infty$. Further, if Y is an arbitrary random variable from L^2 such that $\mathbf{E}Y = 0$ and $Y \perp \mathfrak{X}^2$ in the L^2 -sense, then $\mathbf{E}Y \cdot Z_T(\beta) = 0$ for all $\beta \in L^2(0, T)$. Therefore, taking an arbitrary subdivision $(t_i), i = 1, \dots, n$ of the interval $[0, T]$ and constructing an arbitrary step function with (t_i) , we obtain that $\mathbf{E}(Y|W_{t_1}, \dots, W_{t_n}) = 0$. Hence, $Y = 0$ (a.s.) as $\mathcal{F}_T = \mathcal{F}_T^W$. The last step of the proof is clear. We take M_T as $X \in \mathfrak{X}^2$ and using the martingale and stochastic integral properties, we obtain

$$M_t = \mathbf{E}(M_T|\mathcal{F}_t) = \int_0^t \phi_s dW_s.$$

Chapter 9

Stochastic Differential Equations, Diffusion Processes, and Their Applications



The chapter presents stochastic differential equations (SDEs) and their connections with diffusion processes and partial differential equations (PDEs). The existence and uniqueness of solutions of SDEs are proved under Lipschitz's conditions. Two important processes geometric Brownian motion (GBM) and the Ornstein-Uhlenbeck process are constructed on this theoretical base. The difference between ordinary differential equations and SDEs is discussed. As a part of this discussion, the existence of a solution (weak solution) of any SDE with measurable bounded drift coefficient and unit diffusion is proved with the help of the miracle Girsanov theorem. Moreover, it is shown by mean of the method of monotonic approximations that such a solution will be strong if the drift coefficient is a bounded piece-wise smooth function. Diffusion processes are defined as Markov processes for which their transition densities satisfy the asymptotic properties of Kolmogorov. The backward and forward equations of Kolmogorov are derived. A connection between SDEs and PDEs are stated with the help of the Feynman-Kac theorem. Absolute continuity of distributions of diffusion processes is studied with the help of the Girsanov theorem. A special attention is paid to the class of controlled diffusion processes for which the Hamilton-Jacobi-Bellman optimality equation is derived. It is shown how the theory of diffusion processes and SDEs are helpful in mathematical finance (Bachelier and Black-Scholes models) and in statistics of random processes (see [9, 11, 27, 31, 36, 37, 39, 40, 44, 47, 50, 56, 57, 59], and [61]).

9.1 Stochastic Differential Equations

Let $b(t, x)$ and $\sigma(t, x)$ be (t, x) —measurable locally bounded functions from $\mathbb{R}_+ \times \mathbb{R} \rightarrow \mathbb{R}$, i.e., these functions are bounded on each compact set of $\mathbb{R}_+ \times \mathbb{R}$.

If a continuous $\mathcal{F}_t$ -adapted process $X(t, \omega) = X_t(\omega)$ such that for all $t \geq 0$ (a.s.),

$$X_t(\omega) = X_0(\omega) + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dW_s, \quad (9.1)$$

where $X_0(\omega)$ is a finite random variable, measurable w.r. to $\mathcal{F}_0$, $(W_t)_{t \geq 0}$ is a Wiener process, then it is called a *solution* of (9.1). We rewrite (9.1) in an equivalent differential form

$$dX_t = b_t(X_t)dt + \sigma_t(X_t)dW_t. \quad (9.2)$$

We say that such a solution is *unique* if for any other solution $(\tilde{X}_t)$, we have

$$P(\omega : \sup_t |X_t(\omega) - \tilde{X}_t(\omega)| > 0) = 0. \quad (9.3)$$

Theorem 9.1 Assume that coefficients $b(t, x)$ and $\sigma(t, x)$ of stochastic equations (9.1) and (9.2) satisfy the conditions:

1) For each $T \geq 0$, there exists a constant k_T such that

$$|b(t, x)| + |\sigma(t, x)| \leq k_T(1 + |x|)$$

for $(t, x) \in [0, T] \times \mathbb{R}$.

2) For all $c > 0$, there exists a constant $l_c > 0$ such that

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq l_c|x - y|$$

for all $(t, x) \in [0, c] \times [-c, c]$.

Then the Eqs. (9.1) and (9.2) admits one and only one solution.

Proof First, we note that conditions 1) and 2) of this theorem are called a *linear growth* and a *local Lipschitz* conditions, correspondingly. Second, we give the proof for a standard Lipschitz condition ($l_c = l$) and for a finite time interval $[0, T]$, and for $X_0(\omega) = x \in \mathbb{R}$.

Let us prove the uniqueness of solutions. Assume $X(t, \omega)$ and $\tilde{X}(t, \omega)$ are two solutions with the initial condition $X(0, \omega) = \tilde{X}(0, \omega) = x$. Then we have

$$\begin{aligned} \tilde{X}(t) - X(t) &= \int_0^t [b(s, \tilde{X}(s)) - b(s, X(s))] ds \\ &\quad + \int_0^t [\sigma(s, \tilde{X}(s)) - \sigma(s, X(s))] dW_s. \end{aligned}$$

Further, using the Cauchy-Schwartz inequality and the isometry property, we obtain for each $t \in [0, T]$ that

$$\mathbf{E}|\tilde{X}(t) - X(t)|^2 \leq 2\mathbf{E} \left(\int_0^t [b(s, \tilde{X}(s)) - b(s, X(s))] ds \right)^2 \quad (9.4)$$

$$\begin{aligned} &+ 2\mathbf{E} \left(\int_0^t [\sigma(s, \tilde{X}(s)) - \sigma(s, X(s))] dW_s \right)^2 \\ &\leq 2t\mathbf{E} \int_0^t [b(s, \tilde{X}(s)) - b(s, X(s))]^2 ds \\ &\quad + 2\mathbf{E} \int_0^t [\sigma(s, \tilde{X}(s)) - \sigma(s, X(s))]^2 ds \\ &\leq 2(t+1)t^2 \int_0^t \mathbf{E}|\tilde{X}(s) - X(s)|^2 ds. \end{aligned} \quad (9.5)$$

To continue, we need the following *Gronwall lemma*:

If $\phi(t) \geq 0$ is an integrable function satisfying inequality for all $t \geq 0$ and fixed constants A and B ,

$$\phi(t) \leq A + B \int_0^t \phi(s) ds. \quad (9.6)$$

Then $\phi(t) \leq Ae^{Bt}$ for all $t \geq 0$.

To prove Gronwall lemma, we just differentiate the logarithm of the right-hand side of (9.6):

$$\begin{aligned} \left(\ln \left(A + B \int_0^t \phi(s) ds \right) \right)' &= \frac{B\phi(t)}{A + B \int_0^t \phi(s) ds} \leq B, \\ \ln \left(A + B \int_0^t \phi(s) ds \right) &\leq \ln A + Bt, \\ A + B \int_0^t \phi(s) ds &\leq Ae^{Bt}. \end{aligned}$$

In relations (9.5), the function $\phi(t) = \mathbf{E}|\tilde{X}(t) - X(t)|^2$ and $A = 0$, $B = 2(T+1)l^2$. Hence, by the Gronwall lemma, we obtain $\mathbf{E}|\tilde{X}(t) - X(t)|^2$. We derive from here that for the set $\mathbb{Q}_T$ of rational numbers of $[0, T]$ that

$$P \left(\sup_{t \in \mathbb{Q}_T} |\tilde{X}(t) - X(t)| = 0 \right) = 1.$$

By the continuity of processes $X(t)$ and $\tilde{X}(t)$, we have

$$1 = P(\sup_{\mathbb{Q}_T} |\tilde{X}(t) - X(t)| = 0) = P(\sup_{t \in [0, T]} |\tilde{X}(t) - X(t)| = 0),$$

which means that X and $\tilde{X}$ are equivalent.

To prove the *existence* of a solution of (9.1) and (9.2), we apply the Picard method of successive approximations:

$$\begin{aligned} X^0(t) &= x, \\ X^n(t) &= x + \int_0^t b(s, X^{n-1}(s)) ds + \int_0^t \sigma(s, X^{n-1}(s)) dW_s. \end{aligned} \quad (9.7)$$

For the sequence of approximations (9.7), we get for $t \in [0, T]$ that

$$|X^n(t)|^2 \leq 3|x|^2 + 3 \left| \int_0^t b(s, X^{n-1}(s)) ds \right|^2 + 3 \left(\int_0^t b(s, X^{n-1}(s)) dW_s \right)^2. \quad (9.8)$$

Inequality (9.8) implies for $t \leq T$ that

$$\mathbf{E}|X^n(t)|^2 \leq 3|x|^2 + L_T \int_0^t \mathbf{E}|X^{n-1}(s)|^2 ds, \quad (9.9)$$

where L_T depends on k_T .

Taking $\sup_{k \leq n}$ in (9.9), we derive from Gronwall's Lemma that

$$\sup_{k \leq n} \mathbf{E}|X^k(t)|^2 \leq 3|x|^2 e^{TL_T}, \quad t \in [0, T]. \quad (9.10)$$

Further, we have that

$$\begin{aligned} X^{n+1}(t) - X^n(t) &= x + \int_0^t \left[b(s, X^n(s)) - b(s, X^{n-1}(s)) \right] ds \\ &\quad + \int_0^t \left[\sigma(s, X^n(s)) - \sigma(s, X^{n-1}(s)) \right] dW_s, \end{aligned}$$

and applying a similar estimation as the proof of uniqueness, we get for $t \leq T$,

$$\mathbf{E}|X^{n+1}(t) - X^n(t)|^2 \leq K_T \int_0^t \mathbf{E}|X^n(s) - X^{n-1}(s)|^2 ds, \quad (9.11)$$

$$\mathbf{E}|X^1(t) - X^0(t)|^2 \leq \mathbf{E} \left(\int_0^t b(s, x) ds + \int_0^t \sigma(s, x) dW_s \right)^2 \leq C_T t, \quad (9.12)$$

with some constants C_T and K_T .

Using the induction, we derive from (9.11) and (9.12) that

$$\mathbf{E}|X^{n+1}(t) - X^n(t)|^2 \leq C_T K_T^n \frac{t^{n+1}}{(n+1)!}. \quad (9.13)$$

From (9.13), we conclude that the series

$$X^0(t) + \sum_{n=0}^{\infty} (X^{n+1}(t) - X^n(t))$$

converges uniformly to continuous process $X(t)$ in the L^2 -sense.

To prove that this process is a solution of (9.1) and (9.2), we note for $t \in [0, T]$ that

$$\begin{aligned} \mathbf{E} \left(\int_0^t [b(s, X(s)) - b(s, X^n(s))] ds + \int_0^t [\sigma(s, X(s)) - \sigma(s, X^n(s))] dW_s \right)^2 \\ \leq K_T \int_0^T \mathbf{E}|X(s) - X^n(s)|^2 ds. \end{aligned}$$

Using this inequality, we obtain that

$$\begin{aligned} & \mathbf{E} \left(X(t) - x - \int_0^t b(s, X(s)) ds + \int_0^t \sigma(s, X(s)) dW_s \right)^2 \\ & \leq 2\mathbf{E}|X(t) - X^n(t)|^2 \\ & \quad + 2\mathbf{E} \left(\int_0^t b(s, X^{n-1}(s)) ds + \int_0^t \sigma(s, X^{n-1}(s)) dW_s \right. \\ & \quad \left. - \int_0^t b(s, X(s)) ds - \int_0^t \sigma(s, X(s)) dW_s \right)^2 \\ & \leq w\mathbf{E}|X(t) - X^n(t)|^2 + 2K_T\mathbf{E}|X(t) - X^{n-1}(t)|^2 \rightarrow 0, \end{aligned}$$

as $n \rightarrow \infty$. □

Remark 9.1 To emphasize a dependence of the limit of successive approximations from an initial point x (correspondently, y), denote it by $X_x(t, \omega)$ (correspondently, $X_y(t, \omega)$). Similar to inequality (9.10), one can derive for $t \in [0, T]$:

$$\begin{aligned} \mathbf{E}|X_x(t) - X_y(t)|^2 & \leq \text{const}(T)|x - y|^2 e^{t \cdot \text{const}(T)}, \\ \mathbf{E}|X_x(t) - X_x(t + \Delta t)|^2 & \leq \text{const}(T)|1 + |x|^2|^2 e^{t \cdot \text{const}(T)}. \end{aligned}$$

Due to these inequalities, the function $X_x(t)$ can be chosen measurable w.r. to x .

Example 9.1 Consider a linear SDE

$$dX_t = \mu X_t dt + \sigma X_t dW_t, \quad (9.14)$$

where X_0 is $\mathcal{F}_0$ -measurable finite random variable. According to Theorem 8.1, the Eq. (9.14) has a unique solution. To find a concrete formula for X_t , we rewrite (9.1) in the form

$$dX_t = (c + \sigma^2/2)X_t dt + \sigma X_t dW_t, \quad (9.15)$$

where $c = \mu - \sigma^2/2$. Applying the Ito formula to $X_0 \exp\{ct + \sigma W_t\} = \tilde{X}_t$.

Applying the Ito formula to $X_0 \exp\{ct + \sigma W_t\} = \tilde{X}_t$, we find that

$$\begin{aligned} d\tilde{X}_t &= d\left(X_0 e^{ct + \sigma W_t}\right) \\ &= \left(c X_0 e^{ct + \sigma W_t} + \frac{\sigma^2}{2} X_0 e^{ct + \sigma W_t}\right) dt + \sigma X_0 e^{ct + \sigma W_t} dW_t \\ &= \left(c + \frac{\sigma^2}{2}\right) \tilde{X}_t dt + \sigma \tilde{X}_t dW_t. \end{aligned}$$

So, using the uniqueness of solution of (9.14), we arrive at the conclusion that $X_t = \tilde{X}_t$ (a.s.) for all $t \geq 0$. Hence,

$$X_t = X_0 e^{ct + \sigma W_t} = X_0 e^{(\mu - \sigma^2/2)t + \sigma W_t}, \quad (9.16)$$

which is called a *geometric Brownian motion* (GBM). The GBM (9.16) is exploited as a stock price model in mathematical finance.

Example 9.2 Consider another linear stochastic differential equation

$$dX_t = -\alpha X_t dt + \sigma dW_t; \quad (9.17)$$

X_0 is $\mathcal{F}_0$ -measurable finite random variable, $\alpha \in \mathbb{R}$.

Let us transform (X_t) with the help of function $f(t, x) = e^{\alpha t} x$, $x \in \mathbb{R}$. By the Ito formula, we can find a stochastic differential of $Y_t = f(t, X_t)$:

$$\begin{aligned} dY_t &= (f'_t - \alpha X_t f'_x + \frac{1}{2} \sigma^2 f''_{xx}) dt + \sigma f'_x dW_t \\ &= (\alpha e^{\alpha t} X_t - \alpha e^{\alpha t} X_t) dt + \sigma e^{\alpha t} dW_t = \sigma e^{\alpha t} dW_t. \end{aligned}$$

It leads to $Y_t = X_0 + \sigma \int_0^t e^{\alpha s} dW_s$ and hence

$$X_t = e^{-\alpha t} Y_t = e^{-\alpha t} X_0 + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dW_s.$$

This is a famous *Ornstein-Uhlenbeck process*, which is called a *Vasicek model* for interest rate modeling in mathematical finance.

We observed here that for SDEs with Lipschitz-type coefficients, existence and uniqueness of solutions are stated in a similar way as for ordinary differential equations. A natural question arises: Are there new results in this theory, which are different from the deterministic theory?

We start our answer with the following *nice result of Girsanov*.

Theorem 9.2 *Let $b(t, x)$, $(t, x) \in [0, T] \times \mathbb{R}^d$, $d \geq 1$, be a bounded measurable function. Then there exists a probability space $(\Omega, \mathcal{F}, P)$, a continuous process (X_t) , and a Wiener process on this space such that $(X_t)_{t \leq T}$ is a solution of the SDE:*

$$dX_t = b(t, X_t)dt + dW_t, \quad X_0 = 0, \quad t \leq T. \quad (9.18)$$

Proof Take any probability space $(\Omega, \mathcal{F}, \tilde{P})$ with a Wiener process $(X_t)_{t \leq T}$. Define another process

$$W_t = X_t - \int_0^t b(s, X_s)ds \quad (9.19)$$

and a new probability measure P with the density w.r.to $\tilde{P}$:

$$\frac{dP}{d\tilde{P}} = \exp \left(\int_0^T b(s, X_s)dX_s - \frac{1}{2} \int_0^T |b(s, X_s)|^2 ds. \right)$$

By the Girsanov theorem, the process $(W_t)_{t \leq T}$ is a Wiener process on the probability space $(\Omega, \mathcal{F}, P)$. Hence, if we rewrite the equality (9.19) as

$$X_t = \int_0^t b(s, X_s)ds + W_t, \quad t \leq T,$$

we arrive at the conclusion that $(X_t)_{t \leq T}$ is a solution of (9.18). $\square$

Remark 9.2 The result of Theorem 9.2 guarantees the existence of a solution of (9.18) for any bounded measurable function $b(t, x)$. It is well-known that the ordinary differential equation $dx_t = b(t, x_t)dt$ does not admit a solution for any bounded measurable function $b(t, x)$.

Remark 9.3 Comparing solutions in Theorems 9.1 and 9.2, we can observe the following difference. The solution in Theorem 9.1 was constructed on *given probability space*. On the other hand, the solution in Theorem 9.2 was built on *some probability space*. They are naturally called *strong* and *weak* solutions, respectively, because any strong solution will be also a weak solution, but not vice versa.

Example 9.3 Let (β_t) be a Brownian motion (BM) with $\beta_0 = y \in \mathbb{R}$. Define a process $B_t = \int_0^t \text{sgn}(\beta_s) d\beta_s$, where

$$\text{sgn}(x) = \begin{cases} 1, & x > 0, \\ -1, & x \leq 0. \end{cases}$$

The process (B_t) is a BM by the Levy characterization. Further, we note that by a property of stochastic integral

$$\begin{aligned} y + \int_0^t \text{sgn}(\beta_s) d\beta_s &= y + \int_0^t (\text{sgn}(\beta_s))^2 d\beta_s \\ &= y + \int_0^t d\beta_s = y + \beta_t - y = \beta_t. \end{aligned}$$

Hence, (β_t) solves the SDE

$$dX_t = \text{sgn}(X_t) dB_t, \quad X_0 = y, \quad (9.20)$$

in the "weak" sense. Moreover, any other solution will be again a BM, and in the sense of its distribution, such a solution is unique (weak uniqueness). But taking $y = 0$, we find that (β_t) and $(-\beta_t)$ solve the Eq. (9.20) with some BM (B_t) and $X_0 = 0$. It means there is no uniqueness as (9.5).

Let us develop the *method of monotonic approximations*, which is powerful enough to prove the existence of strong solutions without the Lipschitz conditions. We start with the simplest version of the *comparison theorem* of strong solutions of SDEs.

Lemma 9.1 Assume that $b^i = b^i(x)$, $x \in \mathbb{R}$, $i = 1, 2$, are bounded continuous functions, and X^i , $i = 1, 2$, are strong (continuous) solutions of the equations

$$dX_t^i = b^i(X_t^i)dt + dW_t, \quad X_0^i = 0, \quad t \geq 0.$$

Then the inequality $b^1(x) < b^2(x)$ for all $x \in \mathbb{R}$ implies that $X_t^1(\omega) \leq X_t^2(\omega)$ for almost all $\omega \in \Omega$ and all $t \geq 0$.

Proof For each $\omega \in \tilde{\Omega}$, $P(\tilde{\Omega}) = 1$, we define

$$\theta = \begin{cases} \inf\{t > 0 : X_t^1(\omega) = X_t^2(\omega)\}, \\ \infty \text{ otherwise.} \end{cases}$$

Note that $X_t^2(\omega) - X_t^1(\omega) = \int_0^t [b^2(X_s^2(\omega)) - b^1(X_s^1(\omega))] ds = \Delta_t(\omega)$. By the continuity of $b^i(x)$, $X_t^i(\omega)$, $i = 1, 2$, and assumption $b^1(x) < b^2(x)$, $x \in \mathbb{R}$, we find that $\Delta_0(\omega) > 0$, $\Delta_t(\omega)$ is continuous and $\Delta_t(\omega) > 0$ for all $t < \theta(\omega)$. If $\theta(\omega) = \infty$, then $X_t^2(\omega) = X_t^1(\omega)$, and Lemma is proved. Otherwise, we have

$\theta(\omega) < \infty$ and $\Delta_\theta(\omega) = X_\theta^2 - X_\theta^1 = 0$. It gives a possibility to reproduce for θ the same consideration as for $t = 0$ and find that $X_t^2(\omega) - X_t^1(\omega) = \Delta_t(\omega) > 0$ for a neighborhood of $\theta(\omega)$, etc. Therefore, $X_t^1(\omega) - X_t^1(\omega) = \Delta_t(\omega) > 0$ for all $t \geq 0$. $\square$

Theorem 9.3 Consider a homogeneous SDE (9.18) and suppose that the drift coefficient $b_t(x) = b(x)$ satisfies the following conditions:

- (1) $b(x)$ has a finite number of points of discontinuity $\{x_1, \dots, x_N\}$.
- (2) $b(x)$ is piece-wise smooth, i.e., for any $x \in \mathbb{R} \setminus \{x_1, \dots, x_N\}$ the derivative $b'(x)$ exists.
- (3) $|b(x)|$ and $|b'(x)| \leq K < \infty$.

Then the Eq. (9.18) has a strong solution.

Proof The first step is to show that $b(x)$ can be approximated by a sequence of functions $b_n(x)$ such that a strong solution of the equations

$$dX_n(t, \omega) = b_n(X_n(t, \omega))dt + dW(t, \omega), \quad (9.21)$$

$$X_n(0, \omega) = 0, t \in [0, T], n = 1, 2, \dots$$

does exist and is unique. Then the properties of these solutions are studied with the aim to show that the sequence $(X_n(t, \omega))_{n=1,2,\dots}$ converges to a continuous $\mathcal{F}_t$ -adapted process $X_\infty(t, \omega)$, which is the desired solution of the limit equation (9.18).

To construct the auxiliary sequence of functions $(b_n(x))_{n=1,2,\dots}$, we take neighborhoods $V_r(x_i)$ of the points $x_1 < \dots < x_N$ of radius $r \leq \frac{1}{4} \min_{1 \leq i \leq N-1} |x_{i+1} - x_i|$ and set

$$\tilde{b}_n(x) = b(x),$$

$$x \in \mathbb{R} \setminus \bigcup_{i=1}^N V_r(x_i).$$

It suffices now to build $\tilde{b}_n(x)$ in the neighborhood $V_r(x_i)$, $i = 1, 2, \dots, N$. Consider $V_r(x_i)$ and let $b(x_i-) > b(x_i+)$ for limits from the left and right side of x_i correspondingly.

It follows from the smoothness of $b(x)$ that there exists a deterministic sequence $(\alpha_n^i)_{n=1,2,\dots}$ such that $\alpha_n^i \in V_r(x_i)$, $\alpha_n^i > x_i$, $\alpha_n^i \downarrow x_i$ as $n \rightarrow \infty$, while the continuous function $\tilde{b}_n(x)$ is determined as follows:

$$\tilde{b}_n(x) = b(x) \text{ for } x \in V_r(x_i) \cap \left\{ (x_{i-1}, x_i) \cup [\alpha_n^i, x_{i+1}) \right\},$$

$$\tilde{b}_n(x_i) = b(x_i-);$$

$\tilde{b}_n(x)$ is linear on $[x_i, \alpha_n^i]$ satisfying the condition $\tilde{b}_n(x) \geq b(x)$, $x \in (x_i, \alpha_n^i)$.

If $b(x_i-) < b(x_i+)$, the construction of $\tilde{b}_n(x)$ in $V_r(x_i)$ is analogous. Finally, we define $b_n(x) = \tilde{b}_n(x) + 1/n$, $n = 1, 2, \dots$, with desirable properties for us:

$$\begin{aligned} b_n(x) &> b_{n+1}(x), \quad x \in \mathbb{R}, \\ b_n(x) &\downarrow b(x) \text{ as } n \rightarrow \infty, \quad x \in \mathbb{R} \setminus \{x_1, \dots, x_N\}; \end{aligned}$$

$b_n(x)$ satisfies a Lipschitz condition and simultaneously for $n = 1, 2, \dots$,

$$|b_n(x)| \leq K + 1 < \infty.$$

Applying Theorems 9.1 to (9.21) with $b_n(x)$, constructed above, we get the existence (and uniqueness) of strong solutions $X_n(t)$, which are continuous and $\mathcal{F}_t$ -adapted.

Further, according to Lemma 9.1, we obtain the system of inequalities

$$X_1(t, \omega) \geq X_2(t, \omega) \geq \dots \geq X_n(t, \omega) \geq \dots \text{ (a.s.)}, \quad t \leq T.$$

Therefore, there exists a $\mathcal{F}_t$ -adapted process $X_\infty(t, \omega)$ such that $X_\infty(t, \omega) = \lim_{n \rightarrow \infty} X_n(t, \omega)$ (a.s.), $t \in [0, T]$. $\square$

Lemma 9.2 *The family of continuous functions $(X_n(\cdot, \omega))_{n=1,2,\dots}$, $\omega \in \tilde{\Omega}$, $P(\tilde{\Omega}) = 1$, is uniformly bounded and equicontinuous.*

Proof For each ω from $\tilde{\Omega}$, $P(\tilde{\Omega}) = 1$, we have for all $n = 1, 2, \dots$ that

$$|X_n(t, \omega)| \leq \left| \int_0^t b_n(X_n(s, \omega)) ds \right| + |W(t, \omega)| \quad (9.22)$$

$$\leq \int_0^t |b_n(X_n(s, \omega))| ds + |W(t, \omega)| \quad (9.23)$$

$$\leq (K + 1)T + \max_{0 \leq t \leq T} |W(t, \omega)|, \quad (9.24)$$

which means that $(X_n(\cdot, \omega))_{n=1,2,\dots}$ is uniformly bounded.

Further, for $\omega \in \tilde{\Omega}$, $P(\tilde{\Omega}) = 1$, $\delta > 0$, we obtain

$$\begin{aligned} \omega_{X_n(\cdot, \omega)}(\delta) &= \sup_{s, t, |t-s| < \delta} |X_n(t, \omega) - X_n(s, \omega)| \\ &\leq \sup_{s, t, |t-s| < \delta} \left| \int_s^t b_n(X_n(u, \omega)) du \right| + |W_n(t, \omega) - W(s, \omega)| \\ &\leq (K + 1)\delta + \omega_{W(\cdot, \omega)}(\delta). \end{aligned}$$

Hence, for almost all ω , we have as $\delta \rightarrow 0$ that $\sup_n \omega_{X_n(\cdot, \omega)}(\delta) \leq (K + 1)\delta + \omega_{W(\cdot, \omega)}(\delta) \rightarrow 0$.

In view of Lemma 9.2 and the Ascoli-Arzelà test, there exists (for $\omega \in \tilde{\Omega}$, $P(\tilde{\Omega}) = 1$) a subsequence $n'(\omega)$ such that

$$\sup_{0 \leq t \leq T} |X_{n'(\omega)}(t, \omega) - X_\infty(t, \omega)| \rightarrow 0, \quad n'(\omega) \rightarrow \infty,$$

and therefore almost all paths of $X_\infty(t, \omega)$ are continuous on $[0, T]$.

Now we put $\phi_n(t, \omega) = b_n(W(t, \omega))$, $n = 1, 2, \dots$, $\phi_\infty(t, \omega) = b(W(t, \omega))$ and define probability measures $(P_n)_{n=1,2,\dots}$ and P_∞ with the help of the Girsanov exponents:

$$\begin{aligned} \frac{dP_n}{dP} &= \exp \left\{ \int_0^T \phi_n(s, \omega) dW(s, \omega) - \frac{1}{2} \int_0^T \phi_n^2(s, \omega) ds \right\}, \\ \frac{dP_\infty}{dP} &= \exp \left\{ \int_0^T \phi_\infty(s, \omega) dW(s, \omega) - \frac{1}{2} \int_0^T \phi_\infty^2(s, \omega) ds \right\}. \end{aligned}$$

□

Problem 9.1 Prove that for $y \in \mathbb{R}$ and $t \in [0, T]$,

$$P_n(\omega : W(t, \omega) \leq y) = P(\omega : X_n(t, \omega) \leq y), \quad (9.25)$$

$$P_n(\omega : W(t, \omega) \leq y) \xrightarrow{n \rightarrow \infty} P(\omega : W(t, \omega) \leq y), \quad (9.26)$$

$$P_n(\omega : X_n(t, \omega) \leq y) \rightarrow P(\omega : X_\infty(t, \omega) \leq y). \quad (9.27)$$

Relations (9.25)–(9.27) yield

$$P(\omega : X_\infty(t, \omega) \leq y) \rightarrow P_\infty(\omega : W(t, \omega) \leq y). \quad (9.28)$$

Lemma 9.3 If A is a Borel set in $\mathbb{R}$ with the Lebesgue measure zero ($l(A) = 0$), then

$$P(\omega : l(t \in [0, T] : X_\infty(t, \omega) \in A) = 0) = 1. \quad (9.29)$$

Proof For fixed $\epsilon > 0$, we first apply Chebyshev's inequality and Fubini's theorem, and then we use (9.28) to derive that

$$\begin{aligned} P(\omega : l(t \in [0, T] : X_\infty(t, \omega) \in A) > \epsilon) &\leq \epsilon^{-1} \mathbb{E} \int_0^T I_{(t: X_\infty(t, \omega) \in A)} dt \\ &= \epsilon^{-1} \int_0^T \mathbb{E} I_{(t: X_\infty(t, \omega) \in A)} dt \\ &= \epsilon^{-1} \int_0^T \left[\int_\Omega I_{(t: X_\infty(t, \omega) \in A)} dP \right] dt \end{aligned}$$

$$\begin{aligned}
&= \epsilon^{-1} \int_0^T \left[\int_{\Omega} I_{(t: X_{\infty}(t, \omega) \in A)} dP_{\infty} \right] dt \\
&= \epsilon^{-1} \int_0^T P_{\infty}(\omega : W(t, \omega) \in A) dt = 0.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary, we obtain (9.29). $\square$

Let us now define the following process

$$X(t, \omega) = \int_0^t b(X_{\infty}(s, \omega)) ds + W(t, \omega)$$

and consider the difference

$$X_n(t, \omega) - X(t, \omega) = \int_0^t b_n(X_n(s, \omega)) ds - \int_0^t b(X_{\infty}(s, \omega)) ds.$$

Applying Lemma 9.3 with $A = \{x_1, \dots, x_N\}$, we find for almost all $t \in [0, T]$ w.r.to the Lebesgue measure and almost all $\omega \in \Omega$ that

$$\lim_{n \rightarrow \infty} b_n(X_n(t, \omega)) = b(X_{\infty}(t, \omega)).$$

Further, we derive from here by using the Lebesgue dominated convergence theorem that

$$\int_0^t b_n(X_n(s, \omega)) ds \xrightarrow{n \rightarrow \infty} \int_0^t b(X_{\infty}(s, \omega)) ds,$$

and, hence, for $t \in [0, T]$ and P -a.s.

$$X(t, \omega) = X_{\infty}(t, \omega),$$

which completes the proof of Theorem 9.3.

9.2 Diffusion Processes and Their Connection with SDEs and PDEs

A general conception of a *Markov process* $(X_t)_{t \geq 0}$ on the given stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ can be described as follows. The characteristic (Markov) property of such adapted process means that

$$P(\omega : X(t, \omega) \in B | \mathcal{F}_s) = P(\omega : X(t, \omega) \in B | X(s, \omega)) \quad (9.30)$$

for every $B \in \mathcal{B}(\mathbb{R})$ and all $s < t$.

It is convenient to treat (9.30) with the help of a *transition probability function* $P(s, x, t, B)$ satisfying conditions:

- (1) It is a measurable function of $x \in \mathbb{R}$.
- (2) It is a probability measure on $\mathbb{R}$, $\mathcal{B}(\mathbb{R})$ such that $P(s, x, s, B) = I_B(x)$.
- (3)

$$P(s, x, u, B) = \int P(s, x, t, dy) P(t, y, u, B) \quad (9.31)$$

for $x \in \mathbb{R}$, $B \in \mathcal{B}(\mathbb{R})$, $0 \leq s < t < u$.

Equation (9.31) is called the *Kolmogorov-Chapman* equation, and the process $(X_t)_{t \geq 0}$ is a *Markov random function* (in a restricted sense). Substituting x by $X(s, \omega)$ into (9.30) and using properties of conditional expectations, we get

$$\begin{aligned} P(s, X(s, \omega), u, B) &= \mathbf{E}(P(t, X(t, \omega), u, B) | \mathcal{F}_s) \\ &= \int P(t, y, u, B) P(s, X(s, \omega), t, dy), \end{aligned}$$

which coincides with (9.31) under $x = X(s, \omega)$.

We know that many properties of stochastic processes can be described using their finite-dimensional distributions. Let us derive the formula of a finite-dimensional distribution of $(X(t))$ in terms of its transition probability function. Putting $I_k = I_{B_k}(X(t_k, \omega))$, $0 \leq t_1 < t_2 < \dots < t_k < \dots < t_n$, $B_k \in \mathcal{B}(\mathbb{R})$, we obtain by backward integration:

$$\begin{aligned} P(\omega : X(t_1, \omega) \in B_1, \dots, X(t_n, \omega) \in B_n) &= \mathbf{E} I_1 \dots I_{n-1} \mathbf{E}(I_n | \mathcal{F}_{t_{n-1}}) \\ &= \mathbf{E} I_1 \dots I_{n-1} \int_{B_n} P(t_{n-1}, X(t_{n-1}, \omega), t_n, dy_n) \\ &= \mathbf{E} I_1 \dots I_{n-1} \int_{B_{n-1}} P(t_{n-2}, X(t_{n-2}, \omega), t_{n-1}, dy_{n-1}) \\ &\quad \times \int_{B_n} P(t_{n-1}, X(t_{n-1}, \omega), t_n, dy_n) \\ &= \mathbf{E} I_1 \int_{B_2} P(t_1, X(t_1, \omega), t_2, dy_2) \times \dots \\ &\quad \times \int_{B_n} P(t_{n-1}, X(t_{n-1}, \omega), t_n, dy_n). \end{aligned} \quad (9.32)$$

In the expression above, we can treat

$$\int_{B_2} P(t_1, X(t_1, \omega), t_2, dy_2) \dots \int_{B_n} P(t_{n-1}, X(t_{n-1}, \omega), t_n, dy_n) \quad (9.33)$$

as a conditional expectation

$$P(\omega : X(t_2, \omega)) \in B_2, \dots, X(t_n, \omega) \in B_n | X(t_1, \omega) = x_1).$$

The expressions (9.32) and (9.33) present a consistent system of finite-dimensional distributions of the process $X(t, \omega)$ for fixed t_1 and x_1 . So, to provide a full description of this system, it is reasonable to have a family of probability spaces $(\Omega, \mathcal{F}, P_{t,x})$. So, in the case of (9.32) and (9.33), we arrive at the probabilities $P_{t,x}(\omega : X(t_2, \omega) \in B_2, \dots, X(t_n, \omega) \in B_n)$ expressed with the help of $P(s, x, t, B)$. We can also see that the process $X(t, \omega)$ can be started at each $t \in \mathbb{R}_+$ from each point $x \in \mathbb{R}$. The process $(X(t, \omega))$ in such a framework is called a *Markov family*, but we will also call it a *Markov process*. All these arguments show that in the theory of Markov processes, it is necessary to describe the class of transition probability functions.

Before further steps in developments of this theory, we give some examples and problems.

First of all, $(W_t)_{t \geq 0}$ presents an example of a Markov process, because its finite-dimensional distributions were given in its definition via transition probability function.

Problem 9.2 Prove that $(|W_t|)_{t \geq 0}$ is a Markov process.

Hint: Its transition function is

$$P(s, x, t, B) = \frac{1}{\sqrt{2\pi(t-s)}} \int_B \left[e^{-\frac{(y-x)^2}{2(t-s)}} + e^{\frac{(y+x)^2}{2(t-s)}} \right] dy.$$

Example 9.4 Let $(Y_n)_{n=1,2,\dots}$ be a sequence of independent random variables with a positive density $p(x)$. Define $X_0 = 0, X_1 = Y_1^+, X_2 = (Y_1 + Y_2)^+, \dots$. The process $(X_n)_{n=0,1,\dots}$ is non-Markov. To show it, we take $x > 0$ and consider the next conditional probability

$$P(\omega : X_3 > 0 | X_2 = 0, X_1 = x) = \int_{-\infty}^0 \int_0^\infty p(y-x)p(z-y)dydz.$$

This probability depends on x , and therefore the process (X_n) cannot be Markov.

Problem 9.3 Let $(Y_n)_{n=1,2,\dots}$ be a sequence of independent r.v.'s with the density $p(x) > 0, x \in \mathbb{R}$. Define the process $(Y_0 = 0)$

$$X_t = (Y_1 + \dots + Y_k)(t-k) + (Y_1 + \dots + Y_{k+1})(k+1-t)$$

for $k \leq t \leq k+1, t \geq 0$. Prove that such a process is not Markov.

Remark 9.4 The transition probability function $P(s, x, t, B)$ is *homogeneous*, if $P(s + h, x, t + h, B) = P(s, x, t, B)$ for any positive h . Hence, $P(\cdot, \cdot, \cdot, \cdot)$ depends on the difference $(t - s)$ only, and one can use another transition probability function $P(t, x, B)$ for which the Kolmogorov-Chapman equation has the form

$$P(t + s, x, B) = \int_{\mathbb{R}} P(t, x, dy) P(s, y, B).$$

Definition 9.1 Markov process $X(t, \omega)$ with its transition probability function $P(s, x, t, B)$ satisfying the properties (1)–(3) is called a *diffusion process*, if there exist measurable functions $b(t, x)$ and $\sigma^2(t, x)$, $(t, x) \in \mathbb{R}_+ \times \mathbb{R}$ such that for $\epsilon > 0$ and $V_\epsilon = \{y \in \mathbb{R} : |y - x| > \epsilon\}$

$$\lim_{h \rightarrow 0} \frac{1}{h} P(t, x, t + h, V_\epsilon) = 0, \quad (9.34)$$

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_{|y-x| \leq \epsilon} (y - x) P(t, x, t + h, dy) = b(t, x), \quad (9.35)$$

$$\lim_{h \rightarrow 0} \frac{1}{h} \int_{|y-x| \leq \epsilon} (y - x)^2 P(t, x, t + h, dy) = \sigma^2(t, x). \quad (9.36)$$

Functions $b(t, x)$ and $\sigma^2(t, x)$ are called *drift* and *diffusion coefficients*.

It turns out relations (9.34)–(9.36) admit an adequate characterization with the help of the following differential operator (generator of $X(t, \omega)$):

$$L_t \phi(x) = a(t, x) \frac{\partial \phi(x)}{\partial x} + \frac{1}{2} \sigma(t, x) \frac{\partial^2 \phi(x)}{\partial x^2}, \quad (9.37)$$

in class of twice continuously differentiable functions ϕ .

Lemma 9.4 For all bounded functions ϕ described above

$$\int_{\mathbb{R}} |\phi(y) - \phi(x)| P(t, x, t + h, dy) = h L_t \phi(x) + o(h) \quad (9.38)$$

is fulfilled $\iff$ (9.34)–(9.36).

Proof To prove (9.38), we decompose function ϕ at point x : for $|y - x| < \epsilon$

$$\int_{\mathbb{R}} |\phi(y) - \phi(x)| P(t, x, t + h, dy) = \alpha_\epsilon(h) + h L_t \phi(x) + h \gamma_\epsilon,$$

where L_t is defined by (9.37), $\alpha_\epsilon(h) h^{-1} \rightarrow 0$, $\epsilon > 0$, and $\gamma_\epsilon \rightarrow 0$ as $\epsilon \rightarrow 0$.

Hence, we get (9.38). To prove the direct implication, we take the function $\phi_{x_0}(x) = 1 - \exp(-|x - x_0|^4)$ and find that $\frac{\partial}{\partial x} \phi_{x_0}(x)|_{x=x_0} = 0$ and $\frac{\partial^2}{\partial x^2} \phi_{x_0}(x)|_{x=x_0} = 0$. It implies (9.34).

Now we take the function $\phi_{x_0}(y) = y - x_0$ for $|y - x_0| < 1$, extending this function to be a bounded twice continuously differentiable function. For such a function and $0 < \epsilon < 1$, we get

$$\int_{|y-x_0| \leq \epsilon} (y - x_0) P(t, x_0, t+h, dy) = a(t, x)h + o(h) \quad (9.39)$$

and hence (9.35). To prove (9.36), we take a square of function in the previous construction (9.39). $\square$

Let us find differential equations for transition probabilities of diffusion processes, known as the *Kolmogorov backward and forward equations*.

For some bounded continuous function $\phi(x)$ and $T > 0$, we define the following function:

$$u_T(t, x) = \int_{\mathbb{R}} \phi(y) P(t, x, T, dy).$$

Theorem 9.4 Assume that (9.34)–(9.36) are uniformly satisfied on bounded time intervals with continuous functions $b(t, x)$ and $\sigma^2(t, x)$. Let function $u_T(t, x)$ be (t, x) -continuous together with derivatives $\frac{\partial u_T}{\partial x}$ and $\frac{\partial^2 u_T}{\partial x^2}$. Then $u_T(t, x)$ has the derivative w.r.to t and satisfies the equation

$$\frac{\partial u_T(t, x)}{\partial t} + L_t u_T(t, x) = 0, \quad (9.40)$$

with the boundary condition $u_T(T, x) = \phi(x)$.

Proof The boundary condition is fulfilled by (9.34). Further, for $h > 0$, we have by Lemma 8.4 that

$$\begin{aligned} u_T(t-h, x) &= \int_{\mathbb{R}} u_T(t, y) P(t-h, x, t, dy) \\ &= u_T(t, x) + \int_{\mathbb{R}} [u_T(t, y) - u_T(t, x)] P(t-h, x, t, dy) \\ &= hL_{t-h}u_T(t, x) + u_T(t, x) + o(h), \end{aligned}$$

whence

$$\frac{1}{h} [u_T(t-h, x) - u_T(t, x)] = L_{t-h}u_T(t, x) + \frac{o(h)}{h}.$$

This implies (9.40). $\square$

In fact, Eq. (9.40) can be called as the Kolmogorov backward equation. But usually, it is formulated as the equation for the *density* $p(s, z, t, y)$ of the transition

probability function $p(s, x, t, B)$:

$$p(s, x, t, y)dy = P(s, x, t, dy).$$

So, we arrive at the *backward Kolmogorov equation* for $p(s, x, t, y)$:

$$\frac{\partial p(s, x, t, y)}{\partial s} + b(s, x, y) \frac{\partial p(s, x, t, y)}{\partial x} + \frac{\sigma^2(s, x)}{2} \frac{\partial^2 p(s, x, t, y)}{\partial x^2} = 0. \quad (9.41)$$

On the other hand, one can also write the *forward Kolmogorov equation* or the *Fokker-Plack equation*:

$$\frac{\partial p(s, x, t, y)}{\partial t} + \frac{\partial}{\partial y} (b(t, y)p(s, x, t, y)) - \frac{\sigma^2}{2\partial y^2} (\sigma^2(t, y)p(s, x, t, y)) = 0. \quad (9.42)$$

So, the same function $p(s, x, t, y)$ plays the role of a *fundamental solution* for two differential equations:

Equation (9.41) as the function of (s, x) and

Equation (9.42) as the function of (t, y) .

Let us come back to the stochastic differential equation (9.1) and its solution $X(t)$ derived under the Lipschitz conditions in Theorem 9.1. Applying the same reasoning as in Theorem 9.1 we can prove that there exists a unique solution $X_{s,x}(t)$ of the equation ($t \geq s$) :

$$X_{s,x}(t) = x + \int_s^t b(u, X_{s,x}(u))du + \int_s^t \sigma(u, X_{s,x}(u))dW_u. \quad (9.43)$$

Putting for $B \in \mathcal{B}(\mathbb{R})$

$$P(s, x, t, B) = P(X_{s,x}(t) \in B), \quad (9.44)$$

we arrive at the conclusion that the solution $X(t)$ of (9.1) is a *Markov process with the transition probability function* (9.44).

To prove this claim, we note that $X_{s,X(s)}(t)$ is a solution of (9.43). We just note from (9.1) and Remark 9.1 that

$$X(t, \omega) - X(s, \omega) = \int_s^t b(u, X(u, \omega))du + \int_s^t \sigma(u, X(u, \omega))dW_u.$$

Thus, $X(t, \omega) = X_{s,X(s,\omega)}(t)$. Further, the σ -algebra $\mathcal{F}_s$ is generated by $W_u, u \leq s$. Let us take arbitrary bounded $\mathcal{F}_s$ -measurable r.v. θ . Then for any bounded function $g(x)$, we obtain that

$$\begin{aligned} \mathbf{E}\theta g(X(t, \omega)) &= \mathbf{E}\theta g(X_{s,X(s,\omega)}(t)) \\ &= \mathbf{E}\theta \mathbf{E}(g(X_{s,x}(t)))_{x=X(s,\omega)} \end{aligned}$$

$$\begin{aligned}
&= \mathbf{E}\theta \left(\int_{\mathbb{R}} g(y) P(s, x, t, dy) \right)_{x=X(s, \omega)} \\
&= \mathbf{E}\theta \int_{\mathbb{R}} g(y) P(s, X(s, \omega), t, dy),
\end{aligned}$$

where we used the independence of $X_{s,x}(t)$ and $\mathcal{F}_s$ because $X_{s,x}(t)$ depends on $W(u) - W(s)$ for $u \geq s$ only. Hence,

$$\mathbf{E}(g(X(t, \omega)) | \mathcal{F}_s) = \int_{\mathbb{R}} g(y) P(s, X(s, \omega), t, dy),$$

and we get the claim.

Remark 9.5 As a methodological result, one can conclude that diffusion processes are described in two ways: as a class of Markov processes with transition probabilities determined by the drift and diffusion coefficients and as solutions of SDEs. Let us discuss other interesting connections. We consider a stochastic differential equation with coefficients $b = b(t, x)$ and $\sigma = \sigma(t, x)$:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t.$$

As before, we define the following differential operator based on functions b and σ :

$$L\phi = \frac{1}{2}\sigma^2 \frac{\partial^2 \phi}{\partial x^2} + b \frac{\partial \phi}{\partial x}.$$

Having a smooth function $F \in C^{1,2}$, we can transform (X_t) with the help of the Ito formula as follows:

$$F(X_t) = F(X_0) + \int_0^t \left[\frac{\partial F}{\partial s} + LF(X_s) \right] ds + \int_0^t \sigma(s, X_s) \frac{\partial F}{\partial x}(X_s) dW_s. \quad (9.45)$$

Equality (9.45) is a smooth transformation of the diffusion process (X_t) . Such a problem was first formulated and solved by Kolmogorov in 1931: a new process $Y_t = F(X_t)$ will be again a diffusion process with the drift $b_Y = \frac{\partial F}{\partial t} + LF(X_t)$ and the diffusion coefficient $\sigma_Y = \sigma \frac{\partial F}{\partial x}(X_t)$. Putting together both these formulas, we arrive at the Ito formula in the form (9.45). That is why the formula (9.45) can also be called the *Kolmogorov-Ito formula*.

Remark 9.6 We can also put the Cauchy problem for the parabolic differential equation: Find a smooth function $v = v(t, x) : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\frac{\partial v}{\partial t} + Lv = 0, \quad (t, x) \in [0, T] \times \mathbb{R}$$

with $v(T, x) = f(x)$, $x \in \mathbb{R}$.

This boundary value problem is solved in the theory of partial differential equations (PDEs) under wide conditions. It turns out that one can write a probabilistic form of this solution using the theory of diffusion processes. We take a diffusion process (X_t) with a generator L and

$$v(t, x) = \mathbf{E}_{t,x} f(X_T), \quad X_t = x. \quad (9.46)$$

To get formula (9.46), called the *Feynman-Kac representation*, we apply the Ito formula to $v(t, X_t)$ and obtain that

$$dv(t, X_t) = \left[\frac{\partial v}{\partial t} + Lv \right] dt + \sigma \frac{\partial v}{\partial x} dW_t.$$

It is clear that $v(t, X_t)$ is a martingale, and therefore

$$v(t, X_t) = \mathbf{E}(f(X_T) | \mathcal{F}_t) = \mathbf{E}_{t,x} f(X_T) \Big|_{x=X_t}.$$

Let us investigate absolute continuity of distributions of diffusion processes. We shall do it in the form of solutions of stochastic differential equations.

Assume $(X(t))_{t \leq T}$ is a continuous random process on a probabilistic space $(\Omega, \mathcal{F}, P)$. Denote $C[0, T]$ as the space of continuous functions on $[0, T]$, and $\mathcal{B}^{[0, T]}$ is a σ -algebra in this space, generated by cylinders. We define a distribution of $(X(t))_{t \leq T}$ as the measure on $(C[0, T], \mathcal{B}^{[0, T]})$

$$\mu_X(B) = P\{\omega : X(\cdot, \omega) \in B\}.$$

It means that the measure μ_X is just an image of P under mapping

$$X(\cdot, \omega) : \Omega \rightarrow C[0, T].$$

Assume, there is a probability measure $\tilde{P} \ll P$, i.e., for $A \in \mathcal{F}$,

$$\tilde{P}(A) = \int_A Z(\omega) dP,$$

where Z is a measurable nonnegative random variable such that

$$\int_{\Omega} Z(\omega) dP = 1.$$

We can consider the process $(X_t)_{t \leq T}$ under a new measure $\tilde{P}$ and denote this process by $(\tilde{X}_t)_{t \leq T}$. Then

$$\frac{d\mu_{\tilde{X}}}{d\mu_X}(X(\cdot, \omega)) = \mathbf{E}(Z(\omega) | \mathcal{F}^X), \quad (9.47)$$

where $\mathcal{F}^X$ is a σ -algebra, generated by the process $(X_t)_{t \leq T}$. The equality (9.47) is almost obvious because for $A \in \mathcal{B}^{[0, T]}$, we have

$$\begin{aligned} \mu_{\tilde{X}}(A) &= \tilde{P}(\omega : X(\cdot, \omega) \in A) \\ &= \int_A Z(\omega) dP \\ &= \mathbb{E}Z(\omega) I_A(X(\cdot, \omega)) \\ &= \mathbb{E}I_A(X(\cdot, \omega)) \mathbb{E}(Z(\omega) | \mathcal{F}^X) \\ &= \int_A \mathbb{E}(Z(\omega) | \mathcal{F}^X) \mu_X(dx). \end{aligned}$$

Assume that two diffusion processes $(X_i(t))_{t \leq T}$, $i = 1, 2$, satisfy the following stochastic differential equations:

$$\begin{aligned} dX_t(t) &= b_i(t, X_i(t))dt + \sigma(t, X_i(t))dW_t, \\ X_i(0) &= x \in \mathbb{R}. \end{aligned} \tag{9.48}$$

Theorem 9.5 *Assume the coefficients of Eqs. (9.48) are bounded and satisfy the Lipschitz conditions, and there exists a bounded continuous function $\theta = \theta(t, x)$ such that $\sigma(t, x) > 0$ and*

$$b_2(t, x) - b_1(t, x) = \theta(t, x)\sigma(t, x),$$

$(t, x) \in [0, T] \times \mathbb{R}$. Then $\mu_{X_2} < \mu_{X_1}$ and

$$\frac{d\mu_{X_2}}{d\mu_{X_1}}(X_1(\cdot, \omega)) = \exp \left\{ \int_0^T \theta(s, X_1(s))dW_s - \frac{1}{2} \int_0^T \theta^2(s, X_1(s))ds \right\}. \tag{9.49}$$

Proof Obviously, we can apply the Girsanov theorem here with the measure $\tilde{P}$ and

$$Z(\omega) = \frac{d\tilde{P}}{dP} = \exp \left\{ \int_0^T \theta(s, X_1(s))dW_s - \frac{1}{2} \int_0^T \theta^2(s, X_1(s))ds \right\}$$

and by boundedness of coefficients $\mathbb{E}Z(\omega) = 1$.

Hence, $\tilde{P}(\Omega) = 1$ and the process

$$\tilde{W}(t) = W(t) - \int_0^t \theta(s, X_1(s))ds$$

is a Wiener process w.r. to $\tilde{P}$.

Let us rewrite the process $(X_1(t))_{t \leq T}$ as follows, using (9.5):

$$\begin{aligned}
 X_1(t) - x &= \int_0^t b_1(s, X_1)ds + \int_0^t \sigma(s, X_1)dW_s \\
 &= \int_0^t b_1(s, X_1)ds + \int_0^t \sigma(s, X_1) \left[d\tilde{W}_s + \theta(s, X_1)ds \right] \\
 &= \int_0^t [b_1(s, X_1) + \theta(s, X_1)\sigma(s, X_1)]ds + \int_0^t \sigma(s, X_1)d\tilde{W}_s \\
 &= \int_0^t b_2(s, X_1)ds + \int_0^t \sigma(s, X_1)d\tilde{W}_s.
 \end{aligned} \tag{9.50}$$

Representation (9.50) certifies that $(X_1(t))_{t \leq T}$ coincides with the solution of the equation

$$d\tilde{X}_2(t) = b_2(t, \tilde{X}_2(t))dt + \sigma(t, \tilde{X}_2(t))d\tilde{W}_t$$

on the space $(\Omega, \mathcal{F}, \tilde{P})$. Hence, distributions $\mu_{\tilde{X}_2}$ and μ_{X_2} are the same, and applying (9.47), we obtain

$$\frac{d\mu_{X_2}}{d\mu_{X_1}}(X_1(\cdot, \omega)) = \mathbb{E} \left(Z(\omega) | \mathcal{F}^{X_1} \right). \tag{9.51}$$

Let us prove that $Z(\omega)$ is $\mathcal{F}^{X_1}$ -measurable and therefore the formula (9.49) is fulfilled. To establish it, we represent $\theta(t, x)$ as follows

$$\theta(t, x) = \frac{b_2(t, x) - b_1(t, x)}{\sigma(t, x)}$$

by the assumption $\sigma(t, x) > 0$ for $(t, x) \in [0, T] \times \mathbb{R}$. Further, for a subdivision $0 = s_0 < s_1 < \dots < s_n = t$, we have (in the sense of convergence in probability) that

$$W_t = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \sigma^{-1}(s_i, X_1(s_i)) \left[X_1(s_{i+1}) - X_1(s_i) - \int_{s_i}^{s_{i+1}} b_1(s, X_1(s))ds \right]. \tag{9.52}$$

The relation (9.52) from observation that

$$\begin{aligned}
 \sum_{i=0}^{n-1} \sigma^{-1}(s_i, X_1(s_i)) \left[X_1(s_{i+1}) - X_1(s_i) - \int_{s_i}^{s_{i+1}} b_1(s, X_1(s))ds \right] \\
 = \sum_{i=0}^{n-1} \sigma^{-1}(s_i, X_1(s_i)) \sigma(s, X_1(s)) dW_s,
 \end{aligned}$$

which by the continuity of coefficients tends to W_t in probability as $\max_i \Delta s_i \rightarrow 0$ by properties of stochastic integrals. $\square$

Example 9.5 Consider processes W_t , $X_t = 10 + W_t$ and $Y_t = 3W_t$, $t \in [0, 1]$. Let us define a functional on $C[0, 1]$, $f(x(\cdot)) = x(0)$. For the process X_t , this functional takes value 10 with probability 1, but for other processes W_t and Y_t , this functional takes value 0 with probability one. Hence, μ_X is singular w.r.to μ_W and μ_Y . We also note that μ_W and μ_Y are singular too. It follows from the fact that in probability $\sum_{i=1}^n (W_{t_i} - W_{t_{i-1}})^2 \rightarrow 1$ as the diameter of subdivision $0 = t_0 < t_1 < \dots < t_n = 1$ tends to 1. A similar limit for (Y_t) will be equal to 9.

Remark 9.7 As it was noted in Remark 8.3 for Ito's processes, a similar theory of stochastic differential equations with a multidimensional Wiener process, a vector-valued coefficient b and a matrix-valued coefficient σ can be developed. Respectively, it is connected with diffusion processes with a vector-valued drift b , a matrix-valued diffusion $a = \sigma\sigma^*$, and a generator $Lu = \sum_i b^i \frac{\partial u}{\partial x^i} + \frac{1}{2} \sum_{i,j} a^{ij} \frac{\partial^2 u}{\partial x^i \partial x^j}$.

9.3 Applications to Mathematical Finance and Statistics of Random Processes

In financial context, a Wiener process was mathematically introduced and developed by L. Bachelier. His model of price evolution of stocks has the following simple form:

$$S_t = S_0 + \mu t + \sigma W_t, \quad t \leq T, \quad (9.53)$$

where $\mu \in \mathbb{R}$, $\sigma > 0$, and (W_t) is a Wiener process on the given stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t^W), P)$.

Besides price dynamics of a risky asset (9.53), we assume for simplicity that evolution of non-risky asset (bank account) is trivial, i.e., $B_t = 1$.

One of the main subjects of mathematical finance is option pricing. We consider only standard and the most exploited contracts which are called call and put options. These derivative securities (or simply, derivatives) give the right to the holder to buy (call option) and to sell (put option) a stock at maturity time T . To get such a derivative, it is necessary to buy it by some price at time $t \leq T$. Denote such prices for call and put options for $t = 0$ by $\mathbb{C}_T$ and $\mathbb{P}_T$, correspondently. It is convenient to identify these options with their pay-off functions at maturity time as $(S_T - K)^+$ and $(K - S_T)^+$. The problem is to find the so-called *fair price* $\mathbb{C}_T(\mathbb{P}_T)$ in the beginning ($t = 0$) of the contract period. According to the theory of option pricing, such a price $\mathbb{C}_T(\mathbb{P}_T)$ is calculated as $\mathbf{E}^*(S_T - K)^+$ (correspondingly, $\mathbf{E}^*(K - S_T)^+$), where $\mathbf{E}^*$ is the expectation with respect to a measure P^* such that the process

$(S_t)_{t \leq T}$ is a P^* -martingale. In the case of model (9.53), such *martingale* measure P^* is determined by the Girsanov exponent

$$Z_T^* = \exp \left\{ -\frac{\mu}{\sigma} W_T - \frac{1}{2} \left(\frac{\mu}{\sigma} \right)^2 T \right\},$$

and according to the Girsanov theorem, the process

$$W_t^* = W_t + \frac{\mu}{\sigma} t$$

is a Wiener process w.r. to P^* . Hence, for distributions Law_{P^*} and Law_P w.r. to measures P^* and P relatively, we have equality

$$Law_{P^*}(S_0 + \mu t + \sigma W_t; t \leq T) = Law_P(S_0 + \sigma W_t; t \leq T). \quad (9.54)$$

Theorem 9.6 *In the framework of the Bachelier model (9.53), the initial price of a call option is determined by the formula*

$$\mathbb{C}_T = (S_0 - K) \Phi \left(\frac{S_0 - K}{\sigma \sqrt{T}} \right) + \sigma \sqrt{T} \phi \left(\frac{S_0 - K}{\sigma \sqrt{T}} \right), \quad (9.55)$$

where $\Phi(x) = \int_{-\infty}^x \phi(y) dy$, $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

In particular, for $S_0 = K$, we have $\mathbb{C}_T = \sigma \sqrt{\frac{T}{2\pi}}$.

To prove (9.55), we use (9.54) and self-similarity property of (W_t) and rewrite $\mathbb{C}_T$ as follows:

$$\begin{aligned} \mathbb{C}_T &= \mathbf{E}^*(S_T - K)^+ = \mathbf{E}^*(S_0 + \mu T + \sigma W_T - K)^+ = \\ &= \mathbf{E}^*(S_0 + \sigma W_T - K)^+ = \mathbf{E}^*(S_0 + \sigma \sqrt{T} W_1 - K)^+ \\ &= \mathbf{E}^*(S_0 - K + \sigma \sqrt{T} \xi)^+, \end{aligned} \quad (9.56)$$

where $W_1 = \xi \sim N(0, 1)$.

Denote $a = S_0 - K$ and $b = \sigma \sqrt{T}$ and obtain from (9.56) that

$$\begin{aligned} \mathbb{C}_T &= \mathbf{E}(a + b\xi)^+ = \int_{-a/b}^{\infty} (a + bx) \phi(x) dx \\ &= a \Phi(a/b) + b \int_{-a/b}^{\infty} x \phi(x) dx \\ &= a \Phi(a/b) - b \int_{-a/b}^{\infty} d\phi(x) \end{aligned}$$

$$\begin{aligned}
&= a\Phi(a/b) + b\phi(a/b) \\
&= (S_0 - K)\Phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right) + \sigma\sqrt{T}\phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right).
\end{aligned}$$

Moreover, using the Markov property of (S_t) , we can find the price of call option $\mathbb{C}(t, S_t)$ at any time $t \leq T$ taking conditional expected value of $(S_T - K)^+$ with respect to $\mathcal{F}_t$:

$$\begin{aligned}
\mathbb{C}(t, S_t) &= \mathbf{E}^* \left((S_T - K)^+ | \mathcal{F}_t \right) \\
&= \mathbf{E}^* \left((S_T - K)^+ | S_t \right) \\
&= \mathbf{E}_{t,x} (S_T - K)^+ \Big|_{S_t=x} \\
&= \mathbf{E}(a_t + b_t \xi)^+ \\
&= a_t \Phi(a_t/b_t) + b_t \phi(a_t/b_t),
\end{aligned}$$

where $a_t = S_t - K$, $b = \sigma\sqrt{T-t}$.

Therefore,

$$\mathbb{C}(t, S_t) = (S_t - K)\Phi\left(\frac{S_t - K}{\sigma\sqrt{T-t}}\right) + \sigma\sqrt{T-t}\phi\left(\frac{S_t - K}{\sigma\sqrt{T-t}}\right). \quad (9.57)$$

Applying the Ito formula to (9.57) and using a martingale property of $\mathbf{E}^*((S_T - K)^+ | \mathcal{F}_t)$, we arrive at

$$d\mathbb{C}(t, S_t) = \frac{\partial \mathbb{C}}{\partial S_t} dS_t + \left(\frac{\partial \mathbb{C}}{\partial t} + \frac{1}{2} \sigma^2 \frac{\partial^2 \mathbb{C}}{\partial S_t^2} \right) dt$$

and a PDE

$$\frac{\partial \mathbb{C}}{\partial t}(t, x) + \frac{1}{2} \sigma^2 \frac{\partial^2 \mathbb{C}(t, x)}{\partial x^2} = 0, \quad (9.58)$$

with the boundary condition

$$\mathbb{C}(T, x) = (x - K)^+.$$

Equation (9.58) gives the opportunity to apply methods of PDEs in pricing of option.

Remark 9.8 As far as the put price $\mathbb{P}_T$ is often considered, it follows from the *put-call parity*: for $r = 0$,

$$\mathbb{C}_T = \mathbb{P}_T + S_0 - K. \quad (9.59)$$

The parity (9.59) follows from the next elementary equality

$$(x - K)^+ = (K - x) + x - K, \quad x, K \geq 0. \quad (9.60)$$

Putting S_T instead of x into (9.60) and taking expected value w.r.to P^* , we arrive at (9.59).

The Bachelier model (9.53) has at least one disadvantage that prices can take negative values that contradicts their financial sense. To make a reasonable improvement of the model, P. Samuelson (1965) proposed to transform (9.53) with the help of an exponential function. The resulting model

$$S_t = S_0 \exp\left((\mu - \sigma^2/2)t + \sigma W_t\right) \quad (9.61)$$

became the name of geometric Brownian motion (GBM).

Applying the Ito formula to (9.61), we get

$$dS_t = S_t(\mu dt + \sigma dW_t), \quad S_0 > 0. \quad (9.62)$$

The model of financial market in the form (9.62) is called the Black-Scholes model. As in the case of the Bachelier model, we assume here $B_t = 1$ for simplicity. As in the case of the Bachelier model, we use the Girsanov theorem with the same Girsanov exponent Z_T^* , the martingale measure P^* , and a Wiener process W_t^* w.r. to P^* . Thus, we can conclude that

$$Law_{P^*}(\sigma W_t, t \leq T) = Law_{P^*}(\sigma W_t^*, t \leq T) = Law_P(\sigma W_t, t \leq T)$$

and, hence,

$$Law_{P^*}(S_t; t \leq T) = Law_P\left(S_0 e^{-\sigma^2/2t + \sigma W_t}, t \leq T\right).$$

It gives a possibility to calculate the price $\mathbb{C}_T$ of a call option $(S_T - K)^+$ in the model (9.62), taking the expected value w.r. to P^* :

$$\begin{aligned} \mathbb{C}_T &= \mathbf{E}^*(S_T - K)^+ = \mathbf{E}(S_0 e^{-\sigma^2/2T + \sigma W_T} - K)^+ \\ &= \mathbf{E}(S_0 e^{-\sigma^2/2T + \sigma \sqrt{T} W_1} - K)^+ \\ &= \mathbf{E}(a e^{b\xi - b^2/2} - K)^+ \\ &= a \Phi\left(\frac{\ln(\frac{a}{K}) + \frac{1}{2}b^2}{b}\right) - K \Phi\left(\frac{\ln(\frac{a}{K}) - \frac{1}{2}b^2}{b}\right), \end{aligned} \quad (9.63)$$

where $\xi \sim N(0, 1)$, $a = S_0$, and $b = \sigma \sqrt{T}$.

The formula (9.63) is the famous formula of Black and Scholes for call option. A similar formula for put option is derived with the help of put-call parity.

Example 9.6 To recognize how close the prices of call options in the model of Bachelier and the model of Black and Scholes are, we consider the simplest equations for them:

$$\begin{aligned} dS_t^B &= S_0 \sigma dW_t, \\ dS_t^{BS} &= S_t^{BS} \sigma dW_t, \quad t \leq T, \sigma > 0. \end{aligned}$$

We put $S_0 = K$, and in this case, we compare prices $\mathbb{C}_T^B$ and $\mathbb{C}_T^{BS}$, for which

$$0 \leq \mathbb{C}_T^B - \mathbb{C}_T^{BS} \leq \frac{S_0}{12\sqrt{2\pi}} \sigma^3 T^{3/2} = O((\sigma\sqrt{T})^3). \quad (9.64)$$

To prove (9.64), we note the following inequalities: $e^y \geq 1 + y$ for all y , and, hence, $y^2/2 \geq 1 - e^{-y^2/2}$ for all y . Using these inequalities, we obtain that

$$\begin{aligned} 0 \leq \mathbb{C}_T^B - \mathbb{C}_T^{BS} &= \left(\frac{S_0}{\sqrt{2\pi}} x - S_0 \left[\Phi\left(\frac{x}{2}\right) - \Phi\left(-\frac{x}{2}\right) \right] \right)_{x=\sigma\sqrt{T}} \\ &= \frac{S_0}{\sqrt{2\pi}} \left[\int_{-x/2}^{x/2} (1 - e^{-y^2/2}) dy \right]_{x=\sigma\sqrt{T}} \\ &\leq \frac{S_0}{\sqrt{2\pi}} \int_{-x/2}^{x/2} y^2/2 dy \Big|_{x=\sigma\sqrt{T}} \\ &= \frac{S_0}{\sqrt{2\pi}} x^3/12 \Big|_{x=\sigma\sqrt{T}} \\ &= \frac{S_0}{12\sqrt{2\pi}} \sigma^3 T^{3/2} = O((\sigma\sqrt{T})^3). \end{aligned}$$

Assuming $\sigma\sqrt{T} \ll 1$ in (9.64), we can observe that call prices in both models are pretty close to each other. In particular, for $T = 1/2$ and $\sigma = 2.4\%$, we obtain $(\sigma\sqrt{T})^3 \simeq 5 \cdot 10^{-7}$ and, therefore, $\frac{S_0}{12\sqrt{2\pi}} - 5 \cdot 10^{-7} \simeq 1.6 \cdot 10^{-8} S_0$.

Let us show how methods of stochastic processes work in solving statistical problems.

Assume that the observation process has the following structure:

$$X_t = \int_0^t f_s ds + W_t, \quad (9.65)$$

where $(f_t)_{t \geq 0}$ is a progressively measurable process for which $P(\omega : \int_0^t f_s^2 ds < \infty) = 1$, $t \geq 0$, $\theta \in \mathbb{R}$ is an unknown parameter to be estimated based on observations of the process $(X_t)_{t \geq 0}$, and $(W_t)_{t \geq 0}$ is a Wiener process.

The model (9.65) is a continuous time version of a regression model in discrete time. So, it is quite natural to use the *least-squares estimates* for the estimation of θ . In the case of model (9.65), such estimate has the following structure:

$$\theta_t = \left(\int_0^t f_s^2 ds \right)^{-1} \int_0^t f_s dX_s.$$

We need to investigate some properties of estimates (9.3). We demonstrate it in the form of the following example.

Example 9.7 Assume that (a.s.)

$$0 < c^2 \leq f_t^2(\omega) \leq C^2 < \infty, \quad (9.66)$$

for every $t > 0$. We can provide a stochastic representation of θ_t using (9.65) as follows:

$$\begin{aligned} \theta_t &= \left(\int_0^t f_s^2 ds \right)^{-1} \int_0^t f_s dX_s \\ &= \left(\int_0^t f_s^2 ds \right)^{-1} \int_0^t f_s^2 ds \theta + \left(\int_0^t f_s^2 ds \right)^{-1} \int_0^t f_s dW_s \end{aligned}$$

and find that

$$\theta_t - \theta = \left(\int_0^t f_s^2 ds \right)^{-1} \int_0^t f_s dW_s. \quad (9.67)$$

Using representation (9.67), condition (9.66), and the Chebyshev inequality, we obtain for $\epsilon > 0$ that

$$\begin{aligned} P(\omega : |\theta_t - \theta| \geq \epsilon) &\leq \epsilon^{-2} \mathbb{E} \left(\left(\int_0^t f_s^2 ds \right)^{-1} \int_0^t f_s dW_s \right)^2 \\ &\leq \epsilon^{-2} c^{-4} t^{-2} \mathbf{E} \left(\int_0^t f_s dW_s \right)^2 \\ &= \epsilon^{-2} c^{-4} t^{-2} \mathbf{E} \int_0^t f_s^2 ds \\ &\leq \epsilon^{-2} c^{-4} t^{-2} C^2 t \\ &= \epsilon^{-2} \frac{C^2}{c^4} t^{-1} \rightarrow 0, \quad t \rightarrow \infty. \end{aligned}$$

It means that θ_t is a consistent estimate of parameter θ .

Example 9.8 The least-squares estimate (9.3) is consistent, as was shown in the previous example, but it does not explain the accuracy of these estimates. One can modify them with the help of a specially chosen stopping times to get an estimate with fixed accuracy. To do this, we fix a positive number H and define

$$\tau_H = \inf \left(t : \int_0^t f_s^2 ds \geq H \right).$$

Assume $\int_0^t f_s^2 ds \rightarrow \infty$ (a.s.) as $t \rightarrow \infty$ to prove that

$$P(\omega : \tau_H(\omega) < \infty) = 1.$$

Further, define a *sequential least-squares* estimate

$$\hat{\theta}_H = H^{-1} \int_0^{\tau_H} f_s dX_s. \quad (9.68)$$

It follows from (9.68) that

$$\begin{aligned} \mathbf{E}\hat{\theta}_H &= H^{-1} \mathbf{E} \int_0^{\tau_H} f_s dX_s \\ &= H^{-1} \mathbf{E} \int_0^{\tau_H} f_s^2 ds \theta + H^{-1} \mathbf{E} \int_0^{\tau_H} f_s dW_s \\ &= H^{-1} H \theta + H \cdot 0 = \theta. \end{aligned}$$

So, the estimate $\hat{\theta}_H$ is unbiased.

To estimate the accuracy of $\hat{\theta}_H$, we use its variance:

$$\begin{aligned} \mathbf{Var}\hat{\theta}_H &= \mathbf{E}(\hat{\theta}_H - \theta)^2 \\ &= \mathbf{E} \left(H^{-1} \int_0^{\tau_H} f_s dX_s - \theta \right)^2 \\ &= H^{-2} \mathbf{E} \left(\int_0^{\tau_H} f_s dW_s \right)^2 \\ &= H^{-2} \mathbf{E} \int_0^{\tau_H} f_s^2 ds \\ &= H^{-2} \cdot H = H^{-1}. \end{aligned} \quad (9.69)$$

The relation (9.69) shows that the accuracy of the estimate $\hat{\theta}_H$ can be controlled with the help of the level H .

9.4 Controlled Diffusion Processes and Applications to Option Pricing

Suppose there is a family of diffusion processes $X_t = X_t^\alpha$ satisfying a stochastic differential equation

$$dX_t := dX_t^\alpha = b_\alpha(t, X_t^\alpha)dt + \sigma_\alpha(t, X_t^\alpha)dW_t, \quad X_0 = x. \quad (9.70)$$

Here, b_α and σ_α are functions satisfying some reasonable conditions for existence and uniqueness of solutions of (9.70). Parameter α is a *control process* adapted to filtration $(\mathcal{F}_t)_{t \geq 0}$. Equation (9.70) is also called a *stochastic control system*. We will call the process (X_t) a *controlled diffusion process*.

For estimating the quality of control α , we introduce a function $f^\alpha(t, x)$, $(t, x) \in \mathbb{R}_+ \times \mathbb{R}$. The process α takes values in a domain $D \subseteq \mathbb{R}$. Function f^α can be interpreted as the density of the value flow. Then the total value on the interval $[0, t]$ will be identified with $\int_0^t f^\alpha(X_s^\alpha)ds$, which is assumed to be well-defined.

Let us put the problem and provide some heuristic explanations of its solution for a time *homogeneous* stochastic control system (9.70). Denoting

$$v^\alpha(x) = \mathbf{E}_{0,x} \left[\int_0^\infty f^\alpha(X_s^\alpha)ds \right] = \mathbf{E}_x \left[\int_0^\infty f^\alpha(X_s^\alpha)ds \right],$$

we define the *optimal control* α^* , if

$$v(x) = \sup_{\alpha} v^\alpha(x) = v^{\alpha^*}(x), \quad x \in \mathbb{R}. \quad (9.71)$$

In the theory of controlled diffusion processes, the following *Hamilton-Jacobi-Bellman principle of optimality* is exploited to determine $v(x)$:

$$v(x) = \sup_{\alpha} \mathbf{E}_x \left[\int_0^t f^\alpha(X_s^\alpha)ds + v(X_t^\alpha) \right]. \quad (9.72)$$

Let us explain a motivation for using (9.72). We rewrite the total value using strategy (control) α on $[0, \infty)$ as follows:

$$\int_0^\infty f^\alpha(X_s^\alpha)ds = \int_0^t f^\alpha(X_s^\alpha)ds + \int_t^\infty f^\alpha(X_s^\alpha)ds. \quad (9.73)$$

If this strategy was used only up to moment t , then the first term on the right-hand side of (9.73) represents the value on the interval $[0, t]$. Suppose the controlled process $X_t^\alpha = X_t$ has the value $y = X_t$ at time t . If we wish to continue the control process after time t with the goal of maximization of the value over the whole time interval $[0, \infty)$, then we have to maximize $\mathbf{E}_y \left[\int_0^t f^\alpha(X_s^\alpha)ds \right]$, where

α also denotes the continuation of the control process to $[t, \infty)$. Changing variable $s = t + u$, $u \geq 0$, and using independence and stationarity of increments of the Wiener process W , we obtain

$$\mathbf{E}_{X_t} \int_0^\infty f^\alpha(X_s^\alpha) ds = v^\alpha(X_t) \leq v(X_t).$$

Thus, a strategy that is optimal after time t gives the average value such that

$$\mathbf{E}_x \int_0^t f^\alpha(X_s^\alpha) ds \geq v^\alpha(x). \quad (9.74)$$

One can choose α_s , $s \leq t$, so that the corresponding value is close enough to the average value. Therefore, taking supremum of both sides (9.74), we arrive at the Hamilton-Jacobi-Bellman (HJB)-principle of optimality (9.72), and we will call $v(x)$ the *value function*.

Moreover, one can rewrite (9.72) in a differential form if the value function is smooth enough. Applying the Ito formula, we obtain

$$v(X_t) = v(x) + \int_0^t \left[\frac{\partial v}{\partial x} b_\alpha(X_s) + \frac{1}{2} \sigma_\alpha^2 \frac{\partial^2 v}{\partial x^2}(X_s) \right] ds + \int_0^t \sigma_\alpha \frac{\partial v}{\partial x}(X_s) dW_s. \quad (9.75)$$

Since the last term on the right-hand side of (9.75) is a martingale, we get from (9.72) that

$$v(x) = \sup_\alpha \mathbf{E}_x \left\{ \int_0^t \left[\frac{\partial v}{\partial x} b_\alpha(X_s) + \frac{\sigma_\alpha^2}{2} \frac{\partial^2 v}{\partial x^2}(X_s) + f^\alpha(X_s) \right] ds + v(x) \right\},$$

and hence

$$\sup_\alpha [L_\alpha v(x) + f^\alpha(x)] = 0, \quad (9.76)$$

where $L_\alpha v = b_\alpha \frac{\partial v}{\partial x} + \frac{1}{2} \sigma_\alpha^2 \frac{\partial^2 v}{\partial x^2}$.

Equation (9.76) is usually referred to as *HJB-differential equation*.

Keeping in mind an adequate application of stochastic control theory in option pricing, we would like to reformulate it for the inhomogeneous case and for a finite interval time $[0, T]$. We consider the following stochastic control system

$$\begin{aligned} dX_s &= b_\alpha(s, X_s) ds + \sigma_\alpha(s, X_s) dW_s, \\ X_t &= x, \quad s \in [t, T], \end{aligned} \quad (9.77)$$

where α is a D -valued adapted process. Then the *optimal control problem* is to maximize (minimize) the value function

$$\begin{aligned} v^\alpha(t, x) &= \mathbf{E}_{t,x} \left[\int_t^T f^\alpha(s, X_s) ds + g(X_T) \right], \\ v(t, x) &= \sup_{\alpha} v^\alpha(t, x), \end{aligned} \quad (9.78)$$

where $X_s = X_s^\alpha$ satisfies (9.77) and function g determines the terminal value.

For system (9.77) with the value function (9.78), the corresponding HJB-equation is derived in the form

$$\begin{aligned} \frac{\partial v(t, x)}{\partial x} + \sup_{\alpha} L_{\alpha} v(t, x) &= 0, \\ v(T, x) &= g(x), \quad x \in \mathbb{R}. \end{aligned} \quad (9.79)$$

Let us show how this mathematical technique works for option pricing. We start with the *Bachelier model with stochastic volatility* (with interest rate $r = 0$):

$$S_t = S_0 + \mu t + \sigma_t W_t, \quad t \in [0, T], \quad (9.80)$$

where $S_0, \sigma_t^2 = \sigma^2 + (-1)^{N_t} \delta \sigma^2$, $\delta \sigma^2 < \sigma^2$, $(N_t)_{t \leq T}$ is a Poisson process with intensity $\lambda > 0$.

It is well-known that pricing of option with pay-off function g leads to the *interval of non-arbitrage prices* of option with end points

$$\mathbb{C}_* = \inf_{P^*} \mathbf{E}^* g(S_T) \text{ and } \mathbb{C}^* = \sup_{P^*} \mathbf{E}^* g(S_T), \quad (9.81)$$

where P^* runs a family of martingale measures for the model (9.80).

Number $\mathbb{C}_*$ and $\mathbb{C}^*$ are called (initial) *lower* and *upper* price of option, respectively. For arbitrary time t before maturity date T , such prices are determined as follows:

$$\begin{aligned} v(t, S_t) &= \sup_{P^*} \mathbf{E}^*(g(S_T) | \mathcal{F}_t) \text{ and} \\ u(t, S_t) &= \inf_{P^*} \mathbf{E}^*(g(S_T) | \mathcal{F}_t). \end{aligned} \quad (9.82)$$

By the Markov property of the process (X_t) from (9.80) and a parametrization of martingale measures $P^* = P^*(\alpha)$, $\alpha_t^2 = \sigma^2 + (-1)^{N_t} \delta \sigma^2$, formulas (9.81) and (9.82) can be rewritten as

$$\begin{aligned} v(t, S_t) &= \sup_{\alpha} \mathbf{E}^*(g(S_T) | S_t) \text{ and} \\ u(t, S_t) &= \inf_{\alpha} \mathbf{E}^*(g(S_T) | S_t). \end{aligned} \quad (9.83)$$

Let us consider $v(t, S_t)$ only because the case of the lower price $u(t, S_t)$ (9.83) can be treated in the same way. Now we rewrite model (9.80) via $P^*(\alpha)$:

$$dS_t = \alpha(t, S_t)dW_t^*, \quad (9.84)$$

where W^* is a Wiener process w.r. to P^* .

According the Ito formula, we obtain

$$dv(t, S_t) = \left[\frac{\partial v(t, S_t)}{\partial t} + \frac{1}{2} \frac{\partial^2 v(t, S_t)}{\partial x^2} \alpha^2 \right] dt + \alpha \frac{\partial v(t, S_t)}{\partial x} dW_t^*.$$

Hence, the equality (9.83) can be rewritten as

$$v(t, S_t) = \sup_{\alpha} \mathbf{E}_{t, S_t} \left[v(t, S_t) + \int_t^T \left[\frac{\partial v}{\partial s} + \frac{\alpha^2}{2} \frac{\partial^2 v}{\partial x^2} \right] ds + \int_t^T \alpha \frac{\partial v}{\partial x} dW_s^* \right]. \quad (9.85)$$

Using the martingale property of the last term of (9.85), we arrive at the equation

$$0 = \sup_{\alpha} \mathbf{E}_{t, S_t}^* \left[\int_t^T \left[\frac{\partial v}{\partial s} + \frac{\alpha^2}{2} \frac{\partial^2 v}{\partial x^2} \right] ds \right]. \quad (9.86)$$

Dividing both sides of (9.86) by $T - t$ and letting $T \rightarrow t$, we get

$$0 = \sup_{\alpha} \left[\frac{\partial v(t, S_t)}{\partial t} + \frac{\alpha^2}{2} \frac{\partial^2 v(t, S_t)}{\partial x^2} \right]. \quad (9.87)$$

Similarly, we can obtain that

$$0 = \inf_{\alpha} \left[\frac{\partial u(t, S_t)}{\partial t} + \frac{\alpha^2}{2} \frac{\partial^2 u(t, S_t)}{\partial x^2} \right]. \quad (9.88)$$

Putting (9.87) and (9.88) into $\alpha^2 = \sigma^2 + (-1)^{N_t} \delta \sigma^2$, we determine $D = (\sigma^2 - \delta \sigma^2, \sigma^2 + \delta \sigma^2)$ and

$$\begin{aligned} 0 &= \frac{\partial v(t, S_t)}{\partial t} + \frac{1}{2} \frac{\partial^2 v(t, S_t)}{\partial x^2} \left(\sigma^2 + \operatorname{sgn} \left(\frac{\partial^2 v}{\partial x^2}(t, S_t) \right) \delta \sigma^2 \right) \\ 0 &= \frac{\partial u(t, S_t)}{\partial t} + \frac{1}{2} \frac{\partial^2 u(t, S_t)}{\partial x^2} \left(\sigma^2 - \operatorname{sgn} \left(\frac{\partial^2 u}{\partial x^2}(t, S_t) \right) \delta \sigma^2 \right). \end{aligned}$$

So, we arrive at the following theorem of pricing of a call option in the model (9.80).

Theorem 9.7 *The bounds of non-arbitrage prices of a call option for any $t \leq T$ are calculated as solutions of the HJB-equations:*

$$\frac{\partial v(t, S_t)}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 v(t, S_t)}{\partial x^2} + \frac{1}{2} \left| \frac{\partial^2 v}{\partial x^2} \right| \delta\sigma^2 = 0, \quad (9.89)$$

$$v(T, x) = (x - K)^+;$$

$$\frac{\partial u(t, S_t)}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 u(t, S_t)}{\partial x^2} - \frac{1}{2} \left| \frac{\partial^2 u}{\partial x^2} \right| \delta\sigma^2 = 0, \quad (9.90)$$

$$u(T, x) = (x - K)^+.$$

The next theorem shows how Eqs. (9.89) and (9.90) can be solved approximately by means of the small perturbations method.

Theorem 9.8 *Assume that $\delta\sigma^2 \ll \sigma^2$; then the initial upper and lower prices of a call option with strike price K admit approximations given by the formulas*

$$\hat{\mathbb{C}}^* = (S_0 - K)\Phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right) + \sigma\sqrt{T}\phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right) + \frac{\delta\sigma^2}{2\sigma}\sqrt{T}\phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right), \quad (9.91)$$

$$\hat{\mathbb{C}}_* = (S_0 - K)\Phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right) + \sigma\sqrt{T}\phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right) - \frac{\delta\sigma^2}{2\sigma}\sqrt{T}\phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right), \quad (9.92)$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $\Phi(x) = \int_{-\infty}^x \phi(y)dy$.

Proof Both formulas (9.91) and (9.92) are derived in the same way using the method of small perturbations from the PDE theory. Deriving only (9.91), we represent $v(t, x)$ as follows:

$$v(t, x) = v_0(t, x) + v_1(t, x)\delta\sigma^2 + v_2(t, x)(\delta\sigma^2)^2 + \dots \quad (9.93)$$

Let us make a change of variable $\theta = \sigma^2(T - t)$ and obtain $v(t, x) = V(\theta, x)$ and $\frac{\partial v}{\partial t} = -\sigma^2 \frac{\partial V}{\partial \theta}$.

Hence, the HJB-equation (9.89) is transformed to the following one:

$$\frac{\partial V}{\partial \theta} = \frac{1}{2} \frac{\partial^2 V}{\partial x^2} + \frac{\delta\sigma^2}{2\sigma^2} \left| \frac{\partial^2 V}{\partial x^2} \right|, \quad V(0, x) = (x - K)^+. \quad (9.94)$$

To solve (9.94), we rewrite (9.93) in new variables ignoring terms of the order greater one:

$$V(\theta, x) = V_0(\theta, x) + V_1(\theta, x)\delta\sigma^2 + o(\delta\sigma^2).. \quad (9.95)$$

Plugging (9.95) into Eq. (9.94), we get equations for V_0 and V_1 :

$$\begin{aligned}\frac{\partial V_0}{\partial \theta} &= \frac{1}{2} \frac{\partial^2 V_0}{\partial x^2}, \quad V_0(0, x) = (x - K)^+, \\ \frac{\partial V_1}{\partial \theta} &= \frac{1}{2} \frac{\partial^2 V_1}{\partial x^2} + \frac{1}{2\sigma^2} \left| \frac{\partial^2 V_0}{\partial x^2} \right|, \quad V_1(0, x) = 0.\end{aligned}\tag{9.96}$$

The first equation in (9.96) is exactly the Bachelier equation, and we know its solution (see, for example, (9.57)). Moreover, we have for derivatives of V_0 that

$$\begin{aligned}\frac{\partial V_0}{\partial x} &= \Phi\left(\frac{x - K}{\sqrt{\theta}}\right) + (x - K)\phi\left(\frac{x - K}{\sqrt{\theta}}\right) \frac{1}{\sqrt{\theta}} \\ &\quad + (x - K)\phi\left(\frac{x - K}{\sqrt{\theta}}\right) \frac{-x + K}{\sqrt{\theta}} = \Phi\left(\frac{x - K}{\sqrt{\theta}}\right), \\ \frac{\partial^2 V_0}{\partial x^2} &= \phi\left(\frac{x - K}{\sqrt{\theta}}\right) \frac{1}{\sqrt{\theta}} \geq 0.\end{aligned}$$

Therefore, the second equation in (9.96) is transformed to the next one

$$\frac{\partial V_1}{\partial \theta} = \frac{1}{2} \frac{\partial^2 V_1}{\partial x^2} + \frac{1}{2\sigma^2 \sqrt{\theta}} \phi\left(\frac{x - K}{\sqrt{\theta}}\right),$$

for which we find that $V_1(\theta, x) = \frac{\sqrt{\theta}}{2\sigma^2} \phi(\theta, x)$.

Finally, we arrive at conclusion that

$$\begin{aligned}v(0, S_0) &\approx V(\sigma^2 T, S_0) = (S_0 - K)\Phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right) + \sigma\sqrt{T}\phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right) \\ &\quad + \frac{\delta\sigma^2}{2\sigma}\sqrt{T}\phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right),\end{aligned}$$

which proves the formula (9.91). $\square$

Problem 9.4 Consider the Black-Scholes model with stochastic volatility and zero interest rate

$$\begin{aligned}dS_t &= S_t(\mu dt + \sigma_t dW_t), \quad S_0 > 0, \\ \sigma_t^2 &= s\sigma^2 + (-1)^{N_t} \delta\sigma^2, \quad \delta\sigma^2 < \sigma^2, \quad t \in [0, T].\end{aligned}\tag{9.97}$$

Using a similar reasoning as in Theorem 9.7, derive the following HJB-equations for the upper and lower call option prices

$$\frac{\partial v}{\partial t} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 v}{\partial x^2} + \frac{1}{2} \left| \frac{\partial^2 v}{\partial x^2} \right| x^2 \delta \sigma^2 = 0,$$

$$v(T, x) = (x - K)^+;$$

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 u}{\partial x^2} - \frac{1}{2} \left| \frac{\partial^2 u}{\partial x^2} \right| x^2 \delta \sigma^2 = 0,$$

$$u(T, x) = (x - K)^+.$$

Problem 9.5 Assuming in the model (9.97) that $\delta \sigma^2 \ll \sigma^2$, find the first-order approximations for initial lower and upper prices $\mathbb{C}_*$ and $\mathbb{C}^*$ of call option, respectively:

$$\begin{aligned} \hat{\mathbb{C}}^* &= S_0 \Phi \left(\frac{\ln S_0/K + \sigma^2 T/2}{\sigma \sqrt{T}} \right) - K \Phi \left(\frac{\ln S_0/K - \sigma^2 T/2}{\sigma \sqrt{T}} \right) \\ &\quad + \frac{K \delta \sigma^2}{2\sigma^2} \sigma \sqrt{T} \phi \left(\frac{\ln S_0/K - \sigma^2 T/2}{\sigma \sqrt{T}} \right), \\ \hat{\mathbb{C}}_* &= S_0 \Phi \left(\frac{\ln S_0/K + \sigma^2 T/2}{\sigma \sqrt{T}} \right) - K \Phi \left(\frac{\ln S_0/K - \sigma^2 T/2}{\sigma \sqrt{T}} \right) \\ &\quad - \frac{K \delta \sigma^2}{2\sigma^2} \sigma \sqrt{T} \phi \left(\frac{\ln S_0/K - \sigma^2 T/2}{\sigma \sqrt{T}} \right). \end{aligned}$$

Chapter 10

General Theory of Stochastic Processes Under “Usual Conditions”



Chapter 9 is devoted to a systematic exposition of a continuous time version of stochastic analysis under “usual conditions” with its standard notions like a stochastic basis, filtration, stopping times, random sets, predictable and optional sigma-algebras, etc. It is shown how the discrete time martingale theory and a pure continuous time theory of diffusion processes are generalized for so-called cadlag processes. Using the predictable notion of a compensator, the fundamental Doob-Meyer theorem is formulated for the class of sub- and supermartingales of class D. The full version of stochastic integration of predictable processes with respect to square-integrable martingale is developed. Moreover, different decompositions of such martingales are proved as well as the Kunita-Watanabe inequality. It is shown how the theory can be extended with the help of localization procedures (local martingales, processes with locally integrable variation, semimartingales). The Ito formula is proved for semimartingales. Stochastic differential equations (SDEs) with respect to semimartingales are studied including the existence and uniqueness of solutions of such equations with the Lipschitz coefficients (see [15, 16, 30, 32, 35, 43, 48, 51], and [52]).

10.1 Basic Elements of Martingale Theory

We will operate here with a complete probability space $(\Omega, \mathcal{F}, P)$ equipped with a nondecreasing family of σ -algebras $\mathcal{F}_t \subseteq \mathcal{F}$ satisfying the “usual conditions”:

- (1) $\mathcal{F}_s \subseteq \mathcal{F}_t$, $s \leq t$, $\mathcal{F}_t = \mathcal{F}_{t+} = \bigcap_{\epsilon > 0} \mathcal{F}_{t+\epsilon}$.
- (2) $\mathcal{F}_t$ contains all the P -null sets of $\mathcal{F}$.

Let us call such a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F})_{t \geq 0}, P)$ *standard*.

In stochastic analysis, a special place is occupied by the set of *stopping times*. That is why we study this notion in a more general setting as before.

Definition 10.1 A nonnegative random variable $\tau : \Omega \rightarrow \mathbb{R}_+ \cup \{\infty\}$ is a stopping time, if for any $t \geq 0$,

$$\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t.$$

We also define a σ -algebra

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty = \sigma(\cup_{t \geq 0} \mathcal{F}_t) : A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$$

as a set of all events occurred before τ .

Let us formulate some properties of stopping times (s.t.) as the following problem.

Problem 10.1

(1) If τ and σ are s.t.'s, then

$$\tau \vee \sigma = \max\{\tau, \sigma\} \text{ and}$$

$$\tau \wedge \sigma = \min\{\tau, \sigma\} \text{ are s.t.'s.}$$

(2) If $(\tau_n)_{n=1,2,\dots}$ is a monotone sequence of s.t.'s, then

$$\tau = \lim_{n \rightarrow \infty} \tau_n \text{ is an s.t.}$$

(3) If τ is an s.t., then $\mathcal{F}_\tau$ is a σ -algebra.

(4) For two s.t.'s $\tau \leq \sigma$ we have $\mathcal{F}_\tau \subseteq \mathcal{F}_\sigma$.

(5) Let τ be an s.t. and $A \in \mathcal{F}_\tau$, then

$$\tau_A = \begin{cases} \tau & \text{on } A, \\ \infty & \text{on } A^c \end{cases}$$

is an s.t.

(6) Let τ be an s.t., then there exists a monotonic sequence of s.t.'s $\tau_n > \tau$ such that $\lim_{n \rightarrow \infty} \tau_n = \tau$ (a.s.).

Definition 10.2 An s.t. τ is called *predictable*, if there exists a nondecreasing sequence $(\tau_n)_{n=1,2,\dots}$ of s.t.'s such that

$$\lim_{n \rightarrow \infty} \tau_n = \tau \text{ (a.s.)}, \quad \tau_n < \tau \text{ (a.s.) on } \{\tau > 0\}.$$

In this case, we say that $(\tau_n)_{n=1,2,\dots}$ *announces* τ . Moreover, we denote $\mathcal{F}_{\tau-} = \sigma(\cup_{n=1}^{\infty} \mathcal{F}_{\tau_n})$ as a collection of events *occurring strictly before* τ .

Problem 10.2

- (1) If $A \in \mathcal{F}_{\tau-}$, then τ_A is a predictable s.t.
 (2) If a.s. $\sigma < \tau$, then

$$\mathcal{F}_\sigma \subseteq \mathcal{F}_{\tau-} \subseteq \mathcal{F}_\tau.$$

Definition 10.3 For an s.t. τ , we define a subset of $\mathbb{R}_+ \times \Omega$

$$[\![\tau]\!] = \{(t, \omega) : t = \tau(\omega) < \infty\}$$

as the *graph* of τ .

Definition 10.4 If the graph $[\![\tau]\!]$ can be embedded to a countable union of graphs of predictable s.t.'s, then an s.t. τ is *accessible*. It means that $\Omega = \cup_n A_n$, and on each element A_n of such partition, τ is announced by a sequence of s.t.'s $(\tau_{n,m})_{m=1,2,\dots}$. So, an s.t. τ will be *predictable* if the sequence $(\tau_{n,m})$ can be chosen without dependence on n .

If an s.t. τ such that $P(\omega : \tau = \sigma < \infty) = 0$ for every predictable s.t. σ , then τ is *totally inaccessible*.

Let us note that every s.t. τ can be decomposed as follows. There exists a unique (up to zero probability) set $A \in \mathcal{F}_\tau$ such that τ_A is accessible and τ_{A^c} is totally inaccessible, $A \subseteq \{\omega : \tau(\omega) < \infty\}$.

We need to connect these notion with the notion of a stochastic process. We understand a stochastic process $X = X_t(\omega) = X(t, \omega)$ as a mapping

$$X : \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}^d.$$

For simplicity, we consider the case $d = 1$. Another stochastic process Y is a *modification* of X , if for arbitrary $t \geq 0$,

$$P\{\omega : X_t \neq Y_t\} = 0.$$

Two stochastic processes X and Y are *indistinguishable*, if

$$P\{\omega : X_t(\omega) = Y_t(\omega) \text{ for all } t\} = 1.$$

Example 10.1 To demonstrate a difference between the notion “modification” and “indistinguishability,” we give a standard example. Take $X_t := 0$ and

$$Y_t = \begin{cases} 0, & t \neq \tau, \\ 1, & t = \tau, \end{cases}$$

where an r.v. τ is exponentially distributed with parameter $\lambda > 0$. Then for fixed t_0 , we have $P(X_{t_0} = Y_{t_0}) = P(\tau \neq t_0) = 1$, but $P(X_t = Y_t \text{ for all } t) = 0$.

Definition 10.5 If X is a measurable mapping from $(\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \times \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then such a stochastic process is *measurable*.

If for $t \in \mathbb{R}_+$ an r.v. X_t is $\mathcal{F}_t$ -measurable, then the process X is *adapted* (to filtration $(\mathcal{F}_t)$).

For each $t \geq 0$, we can consider a restriction of X to the set $[0, t] \times \Omega$. If such restriction is $\mathcal{B}([0, t]) \times \mathcal{F}_t$ -measurable, then the process X is called *progressively measurable*. A family of sets $A \subseteq \mathbb{R}_+ \times \Omega$ such that the indicator process $X(t, \omega) = I_A(t, \omega)$ is progressively measurable is a σ -algebra of progressively measurable set Π .

In the case of a progressively measurable process X , we have the following important property: X_τ is $\mathcal{F}_\tau$ -measurable for any s.t. τ , where X_∞ must be $\mathcal{F}_\infty$ -measurable.

Let us note that for any progressively measurable set A , a random variable

$$D_A(\omega) = \inf\{t : (t, \omega) \in A\};$$

$D_A(\omega) = \infty$ if the set $\{\cdot\} = \emptyset$ is a stopping time which is called the *debut* of A .

One can note that every adapted and measurable process X admits a progressively measurable modification. It is convenient to operate with a stochastic process X , almost all trajectories of which are right-continuous and admit left limits at each time. Such processes are called *cadlag*. It turns out that such adapted cadlag process X is progressively measurable. This fact follows from the next standard considerations. For $t > 0, n = 1, 2, \dots, i = 0, 1, \dots, 2^n - 1, s \leq t$, we put $X_0^{(n)}(\omega) = X_0(\omega)$:

$$X_s^{(n)}(\omega) = X_{\frac{(i+1)t}{2^n}}(\omega) \text{ for } \frac{it}{2^n} < s \leq \frac{(i+1)t}{2^n}. \quad (10.1)$$

According to (10.1), we get $X^{(n)}$ which is $\mathcal{B}(0, t) \times \mathcal{F}_t$ -measurable, and by the right continuity $\lim_{n \rightarrow \infty} X_s^{(n)}(\omega) = X_s(\omega)$ for $(s, \omega) \in [0, t] \times \Omega$. Hence, X is progressively measurable.

Further, define the random variables

$$\begin{aligned} \tau_{1,0} = 0, \quad \tau_{1,1} = \inf(t > 0 : |\Delta X_t| \geq 1), \quad \tau_{1,2} = \inf(t > \tau_{1,1} : |\Delta X_t| \geq 1), \dots \\ \tau_{k,0} = 0, \dots, \tau_{k,n} = \inf\left(t > \tau_{k,n-1} : \frac{1}{k} \leq |\Delta X_t| < \frac{1}{k-1}\right), \dots, \end{aligned} \quad (10.2)$$

where as usual $\Delta X_t = X_t - X_{t-}$.

These random variables are stopping times because (X_t) and (X_{t-}) are progressively measurable. Moreover, the set

$$U = \{(t, \omega) : \Delta X_t(\omega) \neq 0\} \subseteq \cup_{k \geq 1, n \geq 1} \llbracket \tau_{k,n} \rrbracket$$

and even can be embedded to the union of graphs $\llbracket \sigma_n \rrbracket$ and $\llbracket \tau_n \rrbracket$, where (σ_n) are predictable s.t.'s, (τ_n) are totally inaccessible s.t.'s such that $P(\sigma_n = \sigma_m < \infty) = 0$ and $P(\tau_n = \tau_m < \infty) = 0$, $n \neq m$.

Besides Π , we need to introduce two additional σ -algebras on the space $[0, \infty) \times \Omega$. We shall do it with the help of the notion of *stochastic intervals*. These are examples of random set related to the stopping times σ and τ :

$$\llbracket \sigma, \tau \rrbracket = \{(t, \omega) : \sigma(\omega) \leq t < \tau(\omega)\},$$

$$\llbracket \sigma, \tau \rrbracket = \{(t, \omega) : \sigma(\omega) \leq t \leq \tau(\omega) < \infty\},$$

and so on, and also their graphs $\llbracket \sigma \rrbracket = \llbracket \sigma, \sigma \rrbracket$, $\llbracket \tau \rrbracket = \llbracket \tau, \tau \rrbracket$.

The filtration $(\mathcal{F}_t)_{t \geq 0}$ induces in $[0, \infty) \times \Omega$ besides Π the *predictable* and the *optional* σ -algebras $\mathcal{P}$ and $\mathcal{O}$:

$$\mathcal{P} = \sigma\{\llbracket 0_A \rrbracket, \llbracket 0, \tau \rrbracket : A \in \mathcal{F}_0, \tau \text{ is an s.t.}\}, \llbracket 0_A \rrbracket = \{0\} \times A,$$

$$\mathcal{O} = \sigma\{\llbracket 0, \tau \rrbracket : \tau \text{ is an s.t.}\}.$$

The processes which are measurable w.r. to $\mathcal{P}$ (respectively, $\mathcal{O}$) are called *predictable* (*optional*).

Problem 10.3 Let (τ_i) be a nondecreasing sequence of s.t.'s and bounded random functions ϕ_0^* and ϕ_i be $\mathcal{F}_0$ and $\mathcal{F}_{\tau_i}$ -measurable, respectively. Define a stochastic process

$$X_t(\omega) \phi_0^*(\omega) I_{\{0\}}(t) + \sum_{i=0}^{n-1} \phi_i(\omega) I_{\llbracket \tau_i, \tau_{i+1} \rrbracket}(t, \omega). \quad (10.3)$$

Prove that $\mathcal{P}$ is generated by all processes of type (10.3).

Remark 10.1 One can prove that $\mathcal{P}$ is generated by processes of type (10.3) with *deterministic* $\tau_i = t_i$. Moreover, $\mathcal{P}$ is generated by all left-continuous (continuous) adapted processes.

It is clear that for the left-continuous process X and a predictable s.t. τ $X_\tau I_{\{\tau < \infty\}} - \mathcal{F}_{\tau-}$ -measurable, and therefore such a statement will be true for all predictable processes and times.

Remark 10.2 If A is a predictable set ($A \in \mathcal{P}$), then its debut is an s.t. Moreover, D_A is predictable and $A \setminus \llbracket D_A, \infty \rrbracket \in \mathcal{P}$, if $\llbracket D_A \rrbracket \subseteq A$.

Remark 10.3 As far as the times $\tau_{k,n}$ of the n -th jump of size $|\Delta X_t| \in \left[\frac{1}{k}, \frac{1}{k-1}\right)$, for an adapted predictable process X (see (10.2)) they are predictable, $U = \cup_{n \geq 1, k \geq 1} \llbracket \tau_{k,n} \rrbracket$, and $X_{\tau_{n,k}} - \mathcal{F}_{\tau_{n,k}-}$ -measurable, according to Remark 10.1.

To go further, we will study two classes of stochastic processes: *processes with finite variation* and *martingales*.

Definition 10.6

- (1) The process $A = (A_t)_{t \geq 0}$ is *increasing*, if $A_0 = 0$ (a.s.), adapted and cadlag, $A_s \leq A_t$ (a.s.) for $s \leq t$.
- (2) The process $B = (B_t)_{t \geq 0}$ is the *process with finite variation*, if $B_0 = 0$ (a.s.), adapted and cadlag; for each $\omega \in \Omega$ the trajectory $B_\cdot(\omega)$ has finite variation on each compact interval.

Each process with finite variation can be decomposed into difference of two increasing processes and vice versa. For any bounded (nonnegative) Borel function $f(s)$, one can define the Lebesgue-Stieltjes integral $\int_0^t f(s) dB_s(\omega)$. We denote $\int_0^t f(s) |dB_s|$ as the integral of f under the variation of B , and $\int_0^t |dB_s|$ will be equal to the variation of B on $[0, t]$.

Definition 10.7 We call an increasing process A *integrable* if $\mathbf{E} \int_0^t dA_s < \infty$. Similarly, a process B has *integrable variation* if $\mathbf{E} \int_0^\infty |dB_s| < \infty$. The corresponding classes of such processes are denoted by $\mathcal{A}^+$ and $\mathcal{A}$.

The process B with finite variation can be decomposed as follows:

$$B_t = B_t^c + \sum_{s \leq t} \Delta B_s, \quad (10.4)$$

where B^c is a continuous process with finite variation and $\sum_{s \leq t} |\Delta B_s| \leq \int_0^t |dB_s| < \infty$.

Moreover, in decomposition (10.4), its discontinuous part $\sum_{s \leq t} \Delta B_s = \sum_{n \geq 1} \Delta B_{\tau_n} I_{\{\tau_n \leq t\}}$ with some sequence of s.t.'s $(\tau_n)_{n \geq 1}$. If the process B is predictable, then τ_n are predictable too, and $\Delta B_{\tau_n} - \mathcal{F}_{\tau_n-}$ -measurable. Hence, any *predictable process with finite variation* admits the form (10.4) as

$$B_t = B_t^c + \sum_{n \geq 1} \phi_n I_{\{\tau_n \leq t\}}, \quad (10.5)$$

where $\phi_n - \mathcal{F}_{\tau_n-}$ -measurable and $\sum_{n \geq 1} |\phi_n| I_{\{\tau_n \leq t\}}$ exist for all $t \in \mathbb{R}_+$.

In fact, if B satisfies (10.5), then it is a predictable process with finite variation.

Definition 10.8 A *martingale* is an adapted integrable process $M = (M_t)_{t \geq 0}$ such that

$$\mathbf{E}(M_t | \mathcal{F}_s) = M_s \quad (\text{a.s.}) \quad \text{for all } t \geq s \geq 0. \quad (10.6)$$

If in (10.6) the “=” is changed by “ $\geq$ ” (respectively, “ $\leq$ ”), then we have a *submartingale* (*supermartingale*).

Without loss of generality, we can always count these processes *cadlag*, because under “usual conditions,” every submartingale admits the right-continuous modification if and only if its expected value is right-continuous.

We can transform any martingale $(M_t)_{t \geq 0}$ to a submartingale $(\phi(M_t))_{t \geq 0}$ with the help of a convex function ϕ such that $\mathbf{E}|\phi(M_t)| < \infty$, $t \in \mathbb{R}_+$, using *Jensen's inequality*.

As in discrete time, we can define for the given process $X = (X_t)_{t \geq 0}$ the following notions. Let $a < b$ be real numbers and F be a finite subset of $\mathbb{R}_+$. For each (restricted to F) trajectory of process $(X_t(\omega))_{t \geq 0}$, one can define the number of upcrossings $U_F(a, b, X(\omega))$ of the interval $[a, b]$. To do it, we take

$$\tau_1(\omega) = \inf(t \in F : X_t(\omega) \leq a) \text{ and for } j = 1, 2, \dots$$

$$\sigma_j(\omega) = \inf(t \in F : t \geq \tau_j(\omega), X_t(\omega) > b),$$

$$\tau_{j+1}(\omega) = \inf(t \in F : t \geq \sigma_j(\omega), X_t(\omega) < a)$$

and define $U_F(a, b, X(\omega))$ the largest j such that $\sigma_j(\omega) < \infty$.

For an infinite set $G \subseteq \mathbb{R}_+$, one can define

$$U_G(a, b, X(\omega)) = \sup\{U_F(a, b, X(\omega)) : F \subseteq G, F \text{ is finite}\}.$$

Let us formulate the following *Doob inequalities* for continuous time submartingales.

Theorem 10.1 Assume $X = (X_t)_{t \geq 0}$ is a submartingale, $[t_1, t_2] \subseteq \mathbb{R}_+$, $\lambda > 0$. Then

- (1) $P\{\omega : \sup_{t_1 \leq t \leq t_2} X_t \geq \lambda\} \leq \frac{\mathbf{E}X_{t_2}^+}{\lambda}.$
- (2) $P\{\omega : \inf_{t_1 \leq t \leq t_2} X_t \leq -\lambda\} \leq \frac{\mathbf{E}X_{t_2}^+ \mathbf{E}X_{t_1}}{\lambda}.$
- (3) $\mathbf{E}U_{[t_1, t_2]}(a, b, X(\omega)) \leq \frac{\mathbf{E}X_{t_2}^+ + |a|}{b - a}.$
- (4) $\mathbf{E}\left(\sup_{t_1 \leq t \leq t_2} X_t\right)^p \leq \left(\frac{p}{p-1}\right) \mathbf{E}X_{t_2}^p, \quad p > 1,$

for a nonnegative (X_t) with $\mathbf{E}X_t^p < \infty$.

The derivation of inequalities of the above theorem is based on limiting arguments, right-continuity of X , and discrete time versions of these results.

Let us pay attention to other important results of martingale theory.

Theorem 10.2 If $X = (X_t)_{t \geq 0}$ is a submartingale with $\sup_{t \geq 0} \mathbf{E}X_t^+ < \infty$, then there exists an r.v. $X_\infty - \mathcal{F}_\infty$ -measurable and integrable such that $X_\infty(\omega) = \lim_{t \rightarrow \infty} X_t(\omega)$ (a.s.).

Proof For $n = 1, 2, \dots$ and $a < b \in \mathbb{R}$, we have from Theorem 9.1 that

$$\mathbf{E}U_{[0,n]}(a, b, X_{\cdot}(\omega)) \leq \frac{\mathbf{E}X_n^+ + |a|}{b - a}. \quad (10.7)$$

Taking the limit as $n \rightarrow \infty$ in (10.7), we get

$$\mathbf{E}U_{[0,\infty]}(a, b, X_{\cdot}(\omega)) \leq \frac{\sup_t \mathbf{E}X_t^+ + |a|}{b - a}. \quad (10.8)$$

Denote $A_{a,b} = \{\omega : U_{[0,\infty)}(a, b, X_{\cdot}(\omega)) = \infty\}$ and find from (10.8) that $P(A_{a,b}) = 0$. Hence, $P(A) = P(\cup_{a,b \in \mathbb{Q}} A_{a,b}) = 0$. The set A contains the set $\{\omega : \limsup_{t \rightarrow \infty} X_t(\omega) > \liminf_{t \rightarrow \infty} X_t(\omega)\}$.

Now, for $\omega \in \Omega \setminus A$, there exists

$$X_{\infty}(\omega) = \lim_{t \rightarrow \infty} X_t(\omega) \quad (a.s.)$$

By $\mathbf{E}|X_t| = 2\mathbf{E}X_t^+ - \mathbf{E}X_t \leq 2\sup_t \mathbf{E}X_t^+ - \mathbf{E}X_0$ and Fatou's lemma, we obtain that $\mathbf{E}|X_{\infty}| < \infty$. $\square$

Corollary 10.1 *Let $X = (X_t)_{t \geq 0}$ be a martingale. Then the following statements are equivalent:*

- (1) $(X_t)_{t \geq 0}$ is uniformly integrable.
- (2) X_t converges to X_{∞} in L^1 .
- (3) There exists $X_{\infty} \in L^1$ such that $X_t = \mathbf{E}(X_{\infty} | \mathcal{F}_t)$.

Corollary 10.2 *Let $X = (X_t)_{t \geq 0}$ be a nonnegative supermartingale. Then there exists an integrable r.v. $X_{\infty} = \lim_{t \rightarrow \infty} X_t$ (a.s.).*

Corollary 10.3 *Let $X = (X_t)_{t \in [0, \infty]}$ be a submartingale on extended real line $\mathbb{R}_+ \cup \{\infty\}$. Then for stopping times, $\sigma \leq \tau$*

$$\mathbf{E}(X_{\tau} | \mathcal{F}_{\sigma}) \geq X_{\sigma} \quad (a.s.) \quad (10.9)$$

Proof Consider the following sequence of discrete stopping times:

$$\sigma_n(\omega) = \begin{cases} \sigma(\omega), & \sigma = \infty, \\ \frac{kn}{2}, & \frac{k-1}{2^n} \leq \sigma(\omega) < \frac{k}{2^n}. \end{cases}$$

Similar sequence is constructed for τ , and by construction and conditions of the theorem, $\sigma_n \leq \tau_n$, $n = 1, 2, \dots$. Applying (10.9) with σ_n and τ_n , we derive (10.9) by taking the limit as $n \rightarrow \infty$. $\square$

Corollary 10.4 Let $X = (X_t)_{t \geq 0}$ be a submartingale and $\sigma \leq \tau$ be stopping times. Then

- (1) $(X_{\tau \wedge t})_{t \geq 0}$ is a submartingale w.r.to $(\mathcal{F}_t)$.
- (2) $\mathbb{E}(X_{\tau \wedge t} | \mathcal{F}_\sigma) \geq X_{\sigma \wedge t}$ (a.s.) for all $t \geq 0$.

Corollary 10.5 If $M = (M_t)_{t \geq 0}$ is a uniformly integrable martingale and τ is a predictable stopping time, then $\mathbb{E}(\Delta M_\tau | \mathcal{F}_{\tau-}) = 0$.

To prove it, we consider a sequence of stopping times $(\tau_n)_{n=1,2,\dots}$ announcing τ . Applying Theorem 10.2 to this sequence, we get $M_{\tau_n} = \mathbb{E}(M_\infty | \mathcal{F}_{\tau_n}) = \mathbb{E}(M_\tau | \mathcal{F}_{\tau_n})$. Taking the limit in this equality, we obtain (a.s.):

$$M_{\tau-} = \lim_{n \rightarrow \infty} M_{\tau_n} = \lim_{n \rightarrow \infty} \mathbb{E}(M_\tau | \mathcal{F}_{\tau_n}) = \mathbb{E}(M_\tau | \mathcal{F}_{\tau-}).$$

Denote by $\mathcal{M}$ the class of *uniformly integrable martingales*.

Assume $M \in \mathcal{M}$ and $A \in \mathcal{A}^+$; then the process $X = M - A$ is a supermartingale, satisfying the *condition (D)*, i.e., the class of random variables $\{X_\tau I_{\{\tau < \infty\}} : \tau \text{ is an s.t.}\}$ is uniformly integrable. Hence, such a construction leads to a supermartingale $X \in D$. The inverse statement is the *famous Meyer's theorem*:

Any supermartingale $X \in D$ admits the following *decomposition of Doob-Meyer*: for $t \geq 0$ (a.s.),

$$X_t = M_t - A_t, \tag{10.10}$$

where $M \in \mathcal{M}$ and A is a *predictable* process from $\mathcal{A}^+$. The decomposition (10.10) is unique in the class of predictable increasing processes.

Remark 10.4 If a supermartingale $X \in D$ and B is the predictable increasing process from the Doob-Meyer decomposition, then we can calculate the jumps of B as follows. First of all, the moments of jumps of B are predictable. For any such a moment τ , we can apply Corollary 10.4 with the uniformly integrable martingale $X = M + B$ and get

$$0 = \mathbb{E}(\Delta M_\tau | \mathcal{F}_{\tau-}) = \mathbb{E}(\Delta X_\tau + \Delta B_\tau | \mathcal{F}_{\tau-}) = \mathbb{E}(\Delta X_\tau | \mathcal{F}_{\tau-}) + \Delta B_\tau,$$

and hence $\Delta B_\tau = -\mathbb{E}(\Delta X_\tau | \mathcal{F}_{\tau-})$.

Let us note that for every $A \in \mathcal{A}^+$, the process $(-A)$ is a supermartingale of class (D).

Applying to $(-A)$ its Doob-Meyer decomposition, we find a *unique predictable* process $B \in \mathcal{A}^+$ such that $B - A \in \mathcal{M}$.

Definition 10.9 The process B is called the *compensator* of A .

These considerations can be extended to the class $\mathcal{V}$ of *processes with integrable variation*. Let $A \in \mathcal{V}$, then there exists a *unique predictable* process $B \in \mathcal{V}$ with the property $B - A \in \mathcal{M}$. We will call B the *compensator* of A .

Remark 10.5

- (a) If τ is a totally inaccessible stopping time and ϕ is an integrable $\mathcal{F}_\tau$ -measurable function, then the *compensator* of $A_t = \phi I_{\{\tau \leq t\}}$ is a *continuous* process from class $\mathcal{V}$.
- (b) If τ is predictable and ϕ is an integrable $\mathcal{F}_\tau$ -measurable function, then the *compensator* of $A_t = \phi I_{\{\tau \leq t\}}$ is the process $B_t = \mathbf{E}(\phi | \mathcal{F}_{\tau-}) I_{\{\tau \leq t\}}$.

The rest of this section is devoted to *square-integrable martingales*.

Definition 10.10 A martingale M is *square integrable* if $\sup_t \mathbf{E} M_t^2 < \infty$. The class of such martingales is denoted of $\mathcal{M}^2$.

We count some properties of such martingales in the following theorem.

Theorem 10.3 Let $M \in \mathcal{M}^2$; then the next properties are fulfilled:

- (1) $(M_t^2)_{t \geq 0}$ is a submartingale.
- (2) There exists $M_\infty = \lim_{t \rightarrow \infty} M_t$ (a.s.) and in L^2 .
- (3) $\mathbf{E} \sup_t M_t^2 \leq 4 \sup_t \mathbf{E} M_t^2 = 4 \mathbf{E} M_\infty^2$.
- (4) $\mathbf{E} \sum_t (\Delta M_t)^2 \leq \liminf_{(t_1 < \dots < t_n)} \mathbf{E} \sum (M_{t_i} - M_{t_{i-1}})^2 = \mathbf{E} M_\infty^2 < \infty$.

Proof Statements (1)–(3) come from Jensen’s inequality and Theorems 10.1 and 10.2. The last one 4) follows from the Fatou lemma, where $\liminf$ is provided via directed sets if subdivisions $(t_1 < \dots < t_n)$. $\square$

Let us note that by the Doob inequality, M^2 is a submartingale from class (D), and hence by the Doob-Meyer decomposition, there exists a unique predictable $A \in \mathcal{A}^+$ such that $M^2 - A \in \mathcal{M}$.

The process A is denoted by $\langle M, M \rangle$ and is called a *quadratic characteristic* (*compensator*) of $M \in \mathcal{M}^2$.

Remark 10.6 We can cover the jumps of $M \in \mathcal{M}^2$ by a sequence of predictable s.t.’s $(\tau_n)_{n=1,2,\dots}$ and a sequence of totally inaccessible s.t.’s $(\sigma_n)_{n=1,2,\dots}$.

For each $n = 1, 2, \dots$ we define processes $C_t^n = \Delta M_{\tau_n} I_{\{\tau_n \leq t\}}$ and $D_t^n = \Delta M_{\sigma_n} I_{\{\sigma_n \leq t\}}$.

By Remark 10.5 in the case of C_t^n , we have $\mathbf{E}(\Delta M_{\tau_n} | \mathcal{F}_{\tau_n-}) I_{\{\tau_n \leq t\}} = 0 \cdot I_{\{\tau_n \leq t\}} = 0$, and hence it is a square integrable martingale. Another process D_t^n has a compensator $\tilde{D}_t^n$, which is a continuous process with integrable variation. So, $\hat{D}_t^n = D_t^n - \tilde{D}_t^n \in \mathcal{M}$ and $\Delta \hat{D}_{\sigma_n}^n = \Delta M_{\sigma_n}$.

To investigate the structure of class $\mathcal{M}^2$, we need several auxiliary facts, which we formulate as Lemmas.

Lemma 10.1 Let A and $B \in \mathcal{A}$, $A - B$ be a martingale, and ϕ be a bounded (nonnegative) predictable function. Then

$$\mathbf{E} \int_0^\infty \phi_s d(A - B)_s = 0. \quad (10.11)$$

Proof It is clear that relation (10.11) is fulfilled for the step function (see (10.3) and Remark 10.1)

$$\phi_t = \phi(t) = \phi_0^* I_{\{0\}}^{(t)} + \sum_{i=0}^{n-1} \phi_i(\omega) I_{(t_i, t_{i+1}]}(t),$$

and extension to predictable bounded functions is straightforward. $\square$

Lemma 10.2 *If $a : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is the right-continuous increasing function, $a(0) = 0$, and $c(s) = \inf(t : a(t) > s)$, then $c(s)$ is the increasing right-continuous satisfying the equality*

$$\int_0^\infty f(s) da(s) = \int_0^\infty f(c(s)) I_{\{c(s) < \infty\}} ds \quad (10.12)$$

for any bounded (nonnegative) measurable function f .

Again, the *proof* is standard. We check (10.12) for $f(s) = I_{(0,t]}(s)$ and so on.

Lemma 10.3 *For any bounded martingale M and process $A \in \mathcal{A}$, the following equality is true:*

$$\mathbf{E} A_\infty M_\infty = \mathbf{E} \int_0^\infty M_s dA_s. \quad (10.13)$$

Proof Using (10.12) with A and c , we obtain (10.13) from the following chain of equalities:

$$\begin{aligned} \mathbf{E} \int_0^\infty M_s dA_s &= \mathbf{E} \int_0^\infty M_{c_s} I_{\{c_s < \infty\}} ds \\ &= \int_0^\infty \mathbf{E} M_{c_s} I_{\{c_s < \infty\}} ds \\ &= \int_0^\infty \mathbf{E} (M_\infty I_{\{c_s < \infty\}}) ds \\ &= \mathbf{E} \int_0^\infty M_\infty I_{\{c_s < \infty\}} ds \\ &= \mathbf{E} \int_0^\infty M_\infty dA_s \\ &= \mathbf{E} M_\infty dA_\infty. \end{aligned}$$

$\square$

The claim of the next Lemma was proved before for discrete time martingales.

Lemma 10.4 *Let $(L_t)_{t \in [0, \infty]}$ be an uniformly integrable process, $L_0 = 0$. If $\mathbf{E}L_\sigma = 0$ for any stopping time σ , then L is a martingale.*

Proof We define an s.t.

$$\sigma = \begin{cases} t & \text{on } A, \\ \infty & \text{on } A^c, \end{cases}$$

where A is a fixed set from $\mathcal{F}_t$. Then we obtain that $\int_A L_t dP + \int_{A^c} L_\infty dP = 0$, which can be transformed for $\sigma = \infty$ to the equality $\int_A L_\infty dP + \int_{A^c} L_\infty dP$, i.e., (L_t) is a martingale. $\square$

Lemma 10.5 *For $M \in \mathcal{M}^2$ and any predictable s.t. τ , the process $\Delta C_t = \Delta M_\tau I_{\{\tau \leq t\}} \in \mathcal{M}^2$, and for any $N \in \mathcal{M}^2$, the process*

$$L_t = C_t N_t - \Delta C_\tau \Delta N_\tau I_{\{\tau \leq t\}} \in \mathcal{M}.$$

Proof The first claim follows from Remark 10.6. Further, for the process (L_t) , we have

$$\sup_t |L_t| \leq \sup_t |C_t| \sup_t |N_t| + |\Delta C_\tau| |\Delta N_\tau| \in L^1,$$

and for any s.t. σ and $N_{t \wedge \sigma} = N_t^\sigma$ by Lemma 10.3

$$\begin{aligned} \mathbf{E}C_\infty N_\infty^\sigma &= \mathbf{E}C_\sigma N_\sigma = \mathbf{E} \int_0^\infty N_s^\sigma dC_s = \mathbf{E}N_\tau^\sigma \Delta C_\tau \\ &= \mathbf{E}\Delta N_\tau^\sigma \Delta C_\tau = \mathbf{E}\Delta N_\tau \Delta C_\tau I_{\{\sigma \leq \tau\}}, \end{aligned}$$

where the equality $\mathbf{E}N_{\sigma-}^\tau \Delta C_\sigma = \mathbf{E}[N_{\sigma-}^\tau \mathbf{E}(\Delta C_\sigma | \mathcal{F}_{\sigma-})] = 0$ was used. We arrive at the conclusion that $\mathbf{E}L_\tau = 0$ for any s.t. τ . Now the claim follows from Lemma 10.4. $\square$

Lemma 10.6 *Let $M \in \mathcal{M}^2$ and σ be a totally inaccessible s.t. Let $D_t = \Delta M_\sigma I_{\{\sigma \leq t\}}$ with compensator $\tilde{D}_t$ and $\hat{D}_t = D_t - \tilde{D}_t$. Then:*

- (1) *The process $\hat{D} \in \mathcal{M}^2$ and $\mathbf{E}\hat{D}_\infty^2 \leq 5\mathbf{E}(\Delta M_\sigma)^2$.*
- (2) *For any $N \in \mathcal{M}^2$, the process*

$$L_t = D_t N_t - \Delta D_\sigma \Delta N_\sigma I_{\{\sigma \leq t\}} \in \mathcal{M}.$$

Proof

- (1) It is sufficient to consider $D_t = \phi I_{\{\sigma \leq t\}}$ with nonnegative function ϕ , $\mathcal{F}_\sigma$ -measurable and from L^2 . Otherwise, one can consider separately ΔM_σ^+ and ΔM_σ^- . For function $|\phi| \leq a$, we have

$$\begin{aligned} \mathbf{E}\tilde{D}_\infty^2 &= 2\mathbf{E} \int_0^\infty \tilde{D}_s d\tilde{D}_s = 2\mathbf{E} \int_0^t \tilde{D}_s dD_s \\ &\leq 2\mathbf{E} \int_0^\infty \tilde{D}_\infty dD_s = 2\mathbf{E}\phi \tilde{D}_\infty \leq 2a\mathbf{E}\tilde{D}_\infty < \infty, \end{aligned}$$

where the formula $f^2(\infty) = 2 \int_0^\infty f(s)df(s)$ is used. Hence, $\tilde{D}_\infty \in L^2$ and $\mathbf{E}\tilde{D}_\infty^2 \leq 2\mathbf{E}\phi \tilde{D}_\infty \leq 2\|\phi\|_{L^2}\|\tilde{D}_\infty\|_{L^2}$ and $\mathbf{E}\tilde{D}_\infty^2 \leq 4\mathbf{E}\phi^2$ for a nonnegative bounded function ϕ .

In general case, we consider $\phi^n = \phi \wedge n$ and processes $\tilde{D}_t^n$ and $D_t^n = \phi^n I_{\{\sigma \leq t\}}$. Similarly, $\tilde{D}_t^{n+1} - \tilde{D}_t^n$ is the compensator of the increasing continuous process; we have

$$\mathbf{E}(\tilde{D}_\infty^n)^2 \leq 4\mathbf{E}(\phi^n)^2 \leq 4\mathbf{E}\phi^2 < \infty. \quad (10.14)$$

The nondecreasing sequence $\tilde{D}_\infty^n$ converges, and according to (10.14), the limit B_∞ will be in L^2 , as in case $B_t = \lim_n \tilde{D}_t^n$. Each sample function $\tilde{D}_t^n(\omega)$ converges to $B_t(\omega)$ uniformly, as $0 \leq \tilde{D}_t^{n+1} - \tilde{D}_t^n \leq \tilde{D}_t^{n+1} - \tilde{D}_t^\infty$. Hence, the process B_t is continuous, and by taking the limit in L^1 in the equality $\mathbf{E}(D_t^n - \tilde{D}_t^n | \mathcal{F}_s) = D_s^n - \tilde{D}_s^n$, $s \leq t$, we arrive at the conclusion that $D_t - B_t \in \mathcal{M}$. Therefore, $B_t = \tilde{D}_t$, and $\mathbf{E}\tilde{D}_\infty^2 = \mathbf{E}B_\infty^2 \leq 4\mathbf{E}\phi^2 < \infty$ and the martingale $\hat{D} = D - \tilde{D} \in \mathcal{M}^2$.

- (2) As in Lemma 10.5, the process L is uniformly integrable. Taking a stopping time τ and applying Lemma 10.3 to $\hat{D}$ and N^τ , we obtain

$$\begin{aligned} \mathbf{E}(D_\tau - \tilde{D}_\tau)N_\tau &= \mathbf{E} \left[\int_0^\infty N_t^\tau d(D - \tilde{D})_t \right] = \mathbf{E}N_\sigma^\tau \Delta D_\sigma - \mathbf{E} \int_0^\infty N_t^\tau d\tilde{D}_t \\ &= \mathbf{E}N_\sigma^\tau \Delta D_\sigma - \mathbf{E} \left[\int_0^\infty N_{t-}^\tau d\tilde{D}_t \right] = \mathbf{E} [\Delta N_\sigma \Delta D_\sigma I_{\{\sigma \leq t\}}]. \end{aligned}$$

Finally, applying Lemma 10.4, we show that L is a martingale. $\square$

Definition 10.11 A martingale $M \in \mathcal{M}^2$ is called *purely discontinuous*, if $M_0 = 0$, and if for any continuous martingale $N \in \mathcal{M}^2$, their product $M \cdot N$ is a martingale.

Let us denote $\mathcal{M}^{2,c}$ and $\mathcal{M}^{2,d}$ subclasses of $\mathcal{M}^2$ of continuous and purely discontinuous martingales. Now we are ready to formulate the key theorem for square integrable martingales.

Theorem 10.4 Any martingale $M \in \mathcal{M}^2$ admits the following unique decomposition:

$$M = M^c + M^d, \quad (10.15)$$

where $M^c \in \mathcal{M}^{2,c}$ and $M^d \in \mathcal{M}^{2,d}$.

Let us call M^c in (10.15) the *continuous part* of M and M^d the *compensated sum of the jumps* of M or the *purely discontinuous part* of M .

Proof Using Remark 10.6 with Lemmas 10.5 and 10.6, we can see that $C^n \hat{D}^n$, $C^n \hat{D}^m$, $C^n C^m$, $\hat{D}^n \hat{D}^m$ ($m \neq n$) are martingales. Denote for $N = 1, 2, \dots$, $Y_t^N = \sum_{i=1}^N C_t^i + \sum_{i=1}^N \hat{D}_t^i$ and find from Lemmas 10.5 and 10.6 that for $N \leq N'$

$$\begin{aligned} \mathbf{E}(Y_\infty^{N'} - Y_\infty^N)^2 &= \mathbf{E} \left[\sum_{N+1}^{N'} (C_\infty^n)^2 + \sum_{N+1}^{N'} (\hat{D}_\infty^n)^2 \right] \\ &\leq 5\mathbf{E} \left[\sum_{N+1}^{N'} \left((\Delta M_{\tau_n})^2 + (\Delta M_{\sigma_n})^2 \right) \right]. \end{aligned} \quad (10.16)$$

It follows from (10.16) and Theorem 9.4 that

$$\mathbf{E} \sup_t (Y_t^{N'} - Y_t^N)^2 \leq 5\mathbf{E} \left[\sum_{N+1}^{N'} \left((\Delta M_{\tau_n})^2 + (\Delta M_{\sigma_n})^2 \right) \right].$$

It implies $\sum_N \mathbf{E} \sup_t |Y_t^{N+1} - Y_t^N|^2 < \infty$ for some subsequence (Y^N) , which we denoted by Y^N again. By the Borel-Cantelli lemma, there exists $Y_t = \lim_N Y_t^N$ (a.s.) uniformly over t , and it is a limit in L^2 too. So, the process Y_t is cadlag with the same jumps as M , and $Y \in \mathcal{M}^2$. $\square$

Now for a continuous martingale $X \in \mathcal{M}^2$, the process XY^N is a martingale by Lemmas 10.5 and 10.6. The convergence Y_t^N to Y_t in L^2 implies a convergence $X_t Y_t^N$ to $X_t Y_t$ in L^1 . Hence, it is a martingale, and $M^d = Y$ is a purely discontinuous martingale. Finally, $M^c = M - M^d$ is a continuous martingale, and we get (10.15).

To prove the *uniqueness* of such decomposition, we assume that there are two decompositions $M = M^c + M^d = \bar{M}^c + \bar{M}^d$, and $M^c - \bar{M}^c = \bar{M}^d - M^d$ is continuous and purely discontinuous. Hence, $(M_t^d - \bar{M}_t^d)^2 = \mathbf{E}(M_0^d - \bar{M}_0^d)^2 = 0$ for all $t \geq 0$.

Definition 10.12 Let $M \in \mathcal{M}^2$, $M_0 = 0$, M^c and M^d are its continuous and purely discontinuous parts of M . Define $[M, M]_t = \langle M^c, M^c \rangle_t + \sum_{s \leq t} (\Delta M_s)^2$ and call it a quadratic brackets of M .

We can note that $(M_t^c)^2 - \langle M^c, M^c \rangle_t$, $(M_t^d)^2 - \sum_{s \leq t} (\Delta M_s)^2$ and $M^c M^d$ belong to $\mathcal{M}$. Hence, $M^2 - [M, M] \in \mathcal{M}$, and $\langle M, M \rangle$ is the compensator of $[M, M]$.

Remark 10.7 If $M_0 \neq 0$, then we can think of M_0 as the jump of M at $t = 0$ and define

$$[M, M] = \langle M^c - M_0, M^c - M_0 \rangle + M_0^2 + \sum_{s \leq t} (\Delta M_s)^2.$$

Remark 10.8 Let us list some properties such as quadratic characteristics of square integrable martingales.

Define for $M, N \in \mathcal{M}$, $M_0, N_0 = 0$, their *joint* quadratic characteristics (brackets, compensators):

$$\begin{aligned} \langle M, N \rangle_t &= \frac{1}{2} (\langle M + N, M + N \rangle_t - \langle M, M \rangle_t - \langle N, N \rangle_t), \\ [M, N]_t &= \frac{1}{2} ([M + N, M + N]_t - [M, M]_t - [N, N]_t). \end{aligned}$$

- (1) The process $\langle M, N \rangle$ is the unique predictable process such that $MN - \langle M, N \rangle \in \mathcal{M}$.
- (2) For any stopping time τ , we have

$$\langle M^\tau, N \rangle = \langle M, N^\tau \rangle = \langle M^\tau, N^\tau \rangle = \langle M, N \rangle^\tau,$$

and similar equalities are true for the quadratic brackets.

- (3) $[M, N] = \langle M^c, N^c \rangle + \sum_{s \leq t} \Delta M_s \Delta N_s$.
- (4) $MN - [M, N] \in \mathcal{M}$.

Analyzing class $\mathcal{M}$ of square integrable martingales, we observe that a Wiener process W does not belong to $\mathcal{M}$, because $\mathbf{E}W_t^2 = t \rightarrow \infty$ as $t \rightarrow \infty$. In the next section, it will be shown how to extend $\mathcal{M}$ to include this key process to a bigger class for which a very nice theory can be developed.

10.2 Extension of Martingale Theory by Localization of Stochastic Processes

We start with the following localization procedure. Let $\mathcal{H}$ be a class of stochastic processes $X = (X_t)_{t \geq 0}$. Define $\mathcal{H}_{loc}$ the set of stochastic processes $Y = (Y_t)_{t \geq 0}$ such that for each of them, we can find a sequence of stopping times $\tau_n \uparrow \infty$, $n \rightarrow \infty$, such that the process $Y^{\tau_n} = (Y_{t \wedge \tau_n})_{t \geq 0}$ is in $\mathcal{H}$ for any $n = 1, 2, \dots$. The sequence $(\tau_n)_{n=1,2,\dots}$ is called *localizing (fundamental)* for the process Y . For

example, in the case of classes $\mathcal{M}_{loc}$ and $\mathcal{M}_{loc}^2$, we will call the corresponding processes *local martingales* and *locally square-integrable martingales*.

Definition 10.13 We say that an s.t. τ *reduces* $M \in \mathcal{M}_{loc}$ if $M^\tau \in \mathcal{M}$.

Problem 10.4 Prove that for $M \in \mathcal{M}_{loc}$, the following statements are true:

- (1) An s.t. σ reduces $M \Leftrightarrow M^\sigma \in \mathcal{D}$.
- (2) If an s.t. τ reduces M and an s.t. $\sigma \leq \tau$, then σ reduces M .
- (3) If s.t.'s σ and τ reduce M , then $\max(\sigma, \tau) = \sigma \vee \tau$ reduces M .

Hint: to prove (3), one can use the equality

$$M^{\sigma \vee \tau} = M^\sigma + M^\tau - M^{\sigma \wedge \tau}.$$

The next theorem shows that the localization does not expand the class of local martingales.

Theorem 10.5 *If M is locally a local martingale, then M is a local martingale.*

Proof Denote a localizing sequence $(\tau_n)_{n=1,2,\dots}$ such that M^{τ_n} is a local martingale. Denote by $\mathcal{R}$ the set of all stopping times τ reducing M , and $\gamma = \text{esssup}_{\tau \in \mathcal{R}} \tau$. There exists a sequence $\sigma_n \in \mathcal{R}$, which converges to γ almost surely. Let us make the sequence $(\sigma_n)_{n=1,2,\dots}$ nondecreasing with the help of (3) of Problem 10.4. By definition, γ majorates $\tau_n \rightarrow \infty$ ($n \rightarrow \infty$), and, hence, $\gamma = \infty$ (a.s.), and $(\sigma)_{n \geq 1}$ is fundamental for M . $\square$

Problem 10.5

- (1) Let B be a process with locally integrable variation, i.e., $B \in \mathcal{A}_{loc}$. Prove that there exists a unique predictable process $A \in \mathcal{A}_{loc}$ such that $B - A \in \mathcal{M}_{loc}$. We call A a *compensator* of B .
- (2) Prove that the predictable process B with finite variation has a *locally bounded variation*.
- (3) A process with finite variation has a compensator $\Leftrightarrow$ its variation is locally integrable.

Remark 10.9 There is a simple way to check that the variation is locally integrable. Namely, if for the process A with finite variation there exists a localizing sequence (τ_n) such that $\mathbf{E} \sup_{s \leq \tau_n} |\Delta A_s| < \infty$, $n = 1, 2, \dots$, then the variation of A is locally integrable. To verify this test, we define a sequence of s.t.'s

$$\sigma_n = \tau_n \wedge \inf \left(t : \int_0^t |dA_s| \geq n \right), n = 1, 2, \dots$$

and find that

$$\int_{[0, \sigma_n]} |dA_s| \leq n + |\Delta A_{\sigma_n}| \leq n + \sup_{s \leq \tau_n} |\Delta A_s| \in L^1.$$

The next result plays a key role in many constructions for local martingales. That is why it is called a *fundamental lemma* in the literature.

Lemma 10.7 *Let $M \in \mathcal{M}_{loc}$, then*

- (1) $M_t^* = \sup_{s \leq t} |M_t| \in \mathcal{A}_{loc}^+$.
- (2) $M_t = U_t + V_t$, where $U \in \mathcal{M}_{loc}$, $|\Delta U_t| \leq 1$, $t \geq 0$, $V \in \mathcal{M}_{loc} \cap \mathcal{V}$.

Proof

- (1) Take a localizing sequence of s.t.'s $(\sigma_n)_{n=1,2,\dots}$ for M , and make elements of this sequence finite as follows $\sigma_n \rightarrow \sigma_n \wedge n$, $n = 1, 2, \dots$. Further, define s.t.'s

$$\tilde{\sigma}_n = \sigma_n \wedge \inf\{t : |M_t| \geq n\} \leq \sigma_n, \quad n = 1, 2, \dots,$$

and obtain that $M_{\tilde{\sigma}_n}$ is integrable by the uniform integrability of $(M_{\sigma_n \wedge t})_{t \geq 0}$. On stochastic interval $\llbracket 0, \tilde{\sigma}_n \rrbracket$, we have $|M_t| \leq n$, and $M_{\tilde{\sigma}_n}^* \leq n + |M_{\tilde{\sigma}_n}|$ belongs to L^1 . Hence, $(M_t^*)_{t \geq 0}$ is locally integrable.

- (2) Denote $(A_t) = \sum_{s \leq t} \Delta M_s I_{\{|\Delta M_s| > 1/2\}}$, which is finite for almost all ω . Using this process and $(\tilde{\sigma}_n)$, we define a new sequence of s.t.'s

$$\tau_n = \tilde{\sigma}_n \wedge \inf\{t : \int_0^t |dA_s| \geq n\} \leq \sigma_n, \quad n = 1, 2, \dots$$

It is a localizing sequence such that

$$\int_{\llbracket 0, \tau_n \rrbracket} |dA_t| \leq n + |\Delta A_{\tau_n}| \leq n + 2M_{\tau_n}^* \in L^1.$$

Therefore, there is a compensator B for A , and $V = A - B \in \mathcal{M}_{loc} \cap \mathcal{V}$. Because of predictability of B , its jumps happen at predictable times τ only. Stopping processes by sequence $\tilde{\sigma}_n$, we get now

$$\begin{aligned} |\Delta B_{\tau}^{\tilde{\sigma}_n}| &= \left| \mathbf{E} \left(\Delta A_{\tau}^{\tilde{\sigma}_n} | \mathcal{F}_{\tau-} \right) \right| \\ &= \left| \mathbf{E} \left(\Delta M_{\tau}^{\tilde{\sigma}_n} - \Delta M_{\tau}^{\tilde{\sigma}_n} I_{\{|\Delta M_{\tau}| < 1/2\}} | \mathcal{F}_{\tau-} \right) \right| \leq 0 + 1/2. \end{aligned}$$

Hence, the jumps of $U_t = M_t - V_t$ verify the following inequalities:

$$|\Delta U_t| \leq |\Delta(M - A)_t| + |\Delta B_t| \leq 1.$$

Moreover, the sequence of stopping times defined as $\tilde{\tau}_n = \inf\{t : |U_t| \geq n\}$ is localizing, and $|U_t^{\tilde{\tau}_n}| \leq n + 1$, $n = 1, 2, \dots$. $\square$

Below, we use this fundamental lemma to prove a decomposition of any local martingale into the sum of continuous and purely discontinuous if for any continuous local martingale $N \in \mathcal{M}_{loc}$, their product $MN \in \mathcal{M}_{loc}$.

Lemma 10.8 *Let $M \in \mathcal{M}_{loc} \cap \mathcal{V}$. Then:*

- (1) $V_t \sum_{s \leq t} \Delta M_s$ is a process with locally integrable variation and $M_t = V_t - \tilde{V}_t$, $t \geq 0$.
- (2) For any bounded continuous martingale N , the product $MN \in \mathcal{M}_{loc}$.

Proof By Lemma 10.7, the process $M_t^* \in \mathcal{A}_{loc}^+$. By Remark 10.9, the variation of the process V is locally integrable. Therefore, there is a compensator $\tilde{V}$ for V , and the local martingale $M - (V - \tilde{V})$ is continuous, and its variation is locally integrable according to Lemma 10.7 and Remark 10.9. It means $M - (V - \tilde{V})$ admits a compensator, but by the continuity of $M - (V - \tilde{V})$, it has to be zero as $M - (V - \tilde{V}) \in \mathcal{M}_{loc}$. As a result, $M = V - \tilde{V}$.

- (2) Let N be a bounded continuous martingale and let the variations of V and $\tilde{V}$ be already integrable (otherwise, we can choose a localizing sequence of s.t.'s and work with stopped processes). For an s.t. τ with the help of Lemmas 10.1 and 10.3, we obtain

$$\begin{aligned}
 \mathbf{E} M_\tau N_\tau &= \mathbf{E} M_\infty N_\infty^\tau = \mathbf{E} \int_0^\infty N_s^\tau dM_s \\
 &= \mathbf{E} \int_0^\infty N_s^\tau dV_s - \mathbf{E} \int_0^\infty N_s^\tau d\tilde{V}_s \\
 &= \mathbf{E} \int_0^\infty N_s^\tau dV_s - \mathbf{E} \int_0^\infty N_s^\tau dV_s \\
 &= 0.
 \end{aligned}$$

By Lemma 10.4, we arrive at the conclusion that MN is a martingale. □

Theorem 10.6 *Let $M \in \mathcal{M}_{loc}$; then*

$$M_t = M_t^c + M_t^d, \quad (10.17)$$

where $M^c \in \mathcal{M}_{loc}^c$ (continuous local martingales) and $M^d \in \mathcal{M}_{loc}^d$ (a class of purely discontinuous local martingales). The decomposition (10.17) is unique.

Proof According to the fundamental lemma, we represent $M = N + U$ with a locally bounded local martingale N and a martingale with finite variation U . Taking a localizing sequence of s.t.'s (τ_n) , we get N^{τ_n} as a bounded martingale for each $n = 1, 2, \dots$. Applying Theorem 10.5, we obtain $N^{\tau_n} = (N^{\tau_n})^c + (N^{\tau_n})^d$, and

using the uniqueness of such decomposition, we obtain that $(N^{\tau_n})^c = (N^{\tau_{n+1}})^c$ on $\llbracket 0, \tau_n \rrbracket$. Now putting

$$N^c = \sum_{n=1}^{\infty} (N^{\tau_n})^c I_{\llbracket \tau_{n-1}, \tau_n \rrbracket},$$

we get a continuous local martingale and a purely discontinuous local martingale $N^d = N - N^c$. As a result, we arrive at the following decomposition $M = N^c + (N^d + U)$, where $N^d + U$ is a purely discontinuous local martingale by Lemma 10.8.

As far as the uniqueness of (10.17) is concerned, we can assume that $M = X^c + X^d = M^c + M^d$. Then the equality $M^c - X^c = X^d - M^d$ leads to conclusion that locally $M^c - X^c = X^d - M^d = 0$ by Theorem 10.5. $\square$

Remark 10.10 For $M \in \mathcal{M} \in \mathcal{M}_{loc}^2$, we can apply the Doob-Meyer decomposition (for M^2) locally and construct its *compensator* $\langle M, M \rangle \in \mathcal{A}_{loc}^*$, which is predictable and $M^2 - \langle M, M \rangle \in \mathcal{M}_{loc}$. In particular, such compensator is well-defined for any continuous local martingale because $\mathcal{M}_{loc}^c \subseteq \mathcal{M}_{loc}^2$.

The exact extension of this approach to the whole class of local martingales is not possible. Nevertheless, one can define for $M \in \mathcal{M}_{loc}$ a similar *quadratic characterization* $[M, M]$ as follows:

$$[M, M]_t = \langle M^c, M^c \rangle_t + \sum_{s \leq t} (\Delta M_s)^2,$$

where M^c is a continuous part of M .

The formula (10.2) shows that the process $[M, M]$ is increasing, $M^2 - [M, M] \in \mathcal{M}_{loc}$, but it is not predictable anymore.

To provide the correctness of the definition of process $[M, M]$ by formula (10.2), we need to prove that $\sum_{s \leq t} (\Delta M_s)^2$ converges for all t .

Again, we use the fundamental lemma to represent $M = U + V$. Further, $\sum_{s \leq t} |\Delta V_s|$ converges, and therefore $\sum_{s \leq t} |\Delta V_s|^2$ converges too. Taking a localizing sequence (τ_n) for U , we obtain that $\mathbf{E} \sum_{s \leq \tau_n} |\Delta U_s|^2 < \infty$, and for almost all ω ,

$$\sum_{s \leq \tau_n} |\Delta M_s| \leq 2 \sum_{s \leq \tau_n} |\Delta U_s|^2 + 2 \sum_{s \leq \tau_n} |\Delta V_s|^2$$

converges.

Let us join together classes of local martingales and processes with finite variation to create a bigger family of stochastic processes.

Definition 10.14 A process X is called a *semimartingale*, if

$$X = X_0 + M + A, \tag{10.18}$$

where X_0 is a $\mathcal{F}_0$ -measurable r.v., $M \in \mathcal{M}_{loc}$, $A \in \mathcal{V}$.

The next auxiliary result explains when the semimartingale admits a unique representation in its definition.

Lemma 10.9 *Assume X is a semimartingale with bounded jumps $|\Delta X_t| \leq a < \infty$. Then X can be represented in a unique way as*

$$X = X_0 + M + A, \quad (10.19)$$

where X_0 is an $\mathcal{F}_0$ -measurable r.v., $M \in \mathcal{M}_{loc}$, A is a predictable process with finite variation.

Proof Let us write the semimartingale $X = X_0 + N + B$ with $N \in \mathcal{M}_{loc}$ and $B \in \mathcal{V}$ and with the uniquely defined r.v. X_0 - $\mathcal{F}_0$ -measurable. We can estimate the jumps of B as follows:

$$|\Delta B_t| \leq |\Delta N_t| + |\Delta X_t| \leq 2N_t^* + a, \quad t \geq 0.$$

□

So, the increasing process $Y_t = \sup_{s \leq t} |\Delta B_s| \in \mathcal{A}_{loc}^+$, and by Remark 10.9, the variation of B is also locally integrable, and, hence, there exists its compensator A . Let us put $M = N + B - A \in \mathcal{M}_{loc}$ and find that $X = X_0 + M + A$. If there are two representations (10.19) for $X = X_0 + M + A = X_0 + M' + A'$, then $A - A' \in \mathcal{M}_{loc}$, and therefore A' is a compensator of A and by predictability $A = A'$.

Now we are ready to prove that the class of semimartingales cannot be extended by localization procedures.

Theorem 10.7 *Any stochastic process X which is locally a semimartingale is a semimartingale.*

Proof Denote by $(\tau_n)_{n=1,2,\dots}$ a localizing sequence of stopping times, i.e., X^{τ_n} is a semimartingale for each $n = 1, 2, \dots$. Define the process $Y_t = \sum_{s \leq t} \Delta X_s I_{\{|\Delta X_s| \geq 1\}}$ and note that Y has finite variation. Therefore, the difference $X^{\tau_n} - Y^{\tau_n}$ is a semimartingale with jumps bounded by one. Let us use Lemma 10.9 and write the unique representation $X^{\tau_n} - Y^{\tau_n} = X_0 + M^{(n)} + A^{(n)}$ with $M^{(n)} \in \mathcal{M}_{loc}$, $A^{(n)} \in \mathcal{V}$ and predictable. Further, using the uniqueness of such representation, we obtain on $\llbracket 0, \tau_n \rrbracket$ that

$$A^{(n)} = A^{(n+1)} \quad \text{and} \quad M^{(n)} = M^{(n+1)}.$$

So, we get the processes $A = \sum_n A^{(n)} I_{\llbracket \tau_{n-1}, \tau_n \rrbracket}$ and $M = \sum_n M^{(n)} I_{\llbracket \tau_{n-1}, \tau_n \rrbracket}$. □

The process A is predictable process with finite variation, M is locally a local martingale, and according to Theorem 10.6, it belongs to class $\mathcal{M}_{loc}$. As a result, we conclude that X is a semimartingale.

Another result of this type is presented in the following theorem.

Theorem 10.8 *Let X be a cadlag adapted process, and there exists a sequence of s.t.'s (τ_n) and a sequence of semimartingales $Y^{(n)}$ such that:*

- (1) $\lim_{n \rightarrow \infty} \tau_n = \infty$ (a.s.).
- (2) $X = Y^{(n)}$ on $\llbracket 0, \tau_n \rrbracket$, $n = 1, 2, \dots$

Then X is a semimartingale.

Proof For each $n = 1, 2, \dots$ we have

$$X_{t \wedge \tau_n} = Y_{t \wedge \tau_n}^{(n)} - Y_{\tau_n}^{(n)} I_{\{\tau_n \leq t\}} + X_{\tau_n} I_{\{\tau_n \leq t\}},$$

and therefore X^{τ_n} is a semimartingale. Now, if we can transform (τ_n) to an increasing sequence, then we prove a semimartingale property of X according to Theorem 10.8. The trick to make it is standard. If τ and σ are s.t.'s, and X^σ and X^τ are semimartingales, then $X^{\sigma \wedge \tau}$ and $X^{\sigma \vee \tau} = X^\sigma + X^\tau - X^{\sigma \wedge \tau}$ are both semimartingales too. $\square$

Representation (10.18) for a semimartingale X may not be unique. Nevertheless, there is a continuous martingale component of X which does not depend on such representation. It gives a possibility to define a quadratic bracket of X that is a very helpful characteristic and tool for semimartingales. It is provided by the next theorem.

Theorem 10.9 *Let X be a semimartingale with representation (10.18).*

Then:

- (1) M^c in (10.18) does not depend on the particular representation (10.18).
- (2) $\sum_{s \leq t} (\Delta X_s)^2$ converges a.s. for all $t \geq 0$.

Proof (1) Indeed, if $X = X_0 + M + A = X_0 + N + B$ with $M, N \in \mathcal{M}_{loc}$, $A, B \in \mathcal{V}$, then the difference $M - N \in \mathcal{M}_{loc} \cap \mathcal{V}$, and according to Lemma 10.8, $M - N$ is purely discontinuous and $(M - N)^c = 0$. Therefore, $M^c = N^c$, and we denote this component X^c and call it the *continuous local martingale part* of X . The proof of (2) is provided in the same way as for local martingales. $\square$

Definition 10.15 Let X be a semimartingale. Then the increasing process

$$[X, X]_t = \langle X^c, X^c \rangle_t + \sum_{s \leq t} (\Delta X_s)^2, \quad t \geq 0 \quad (10.20)$$

is well-defined and is called the *quadratic bracket* of X . If we have another semimartingale $Y = (Y_t)_{t \geq 0}$, we can define a joint bracket

$$[X, Y]_t = \langle X^c, Y^c \rangle_t + \sum_{s \leq t} \Delta X_s \Delta Y_s \quad (10.21)$$

with similar properties.

Example 10.2 Let X be a supermartingale, then it is a semimartingale. To show it, we use Lemma 10.9. We consider the stepped supermartingale $X_t^n = X_{t \wedge n}$, $n = 1, 2, \dots$, which can be represented as follows:

$$X_t^n = \mathbf{E}(X_n | \mathcal{F}_t) + X_t^n - \mathbf{E}(X_n | \mathcal{F}_t) = \mathbf{E}(X_n | \mathcal{F}_t) + Z_t,$$

where Z is a nonnegative supermartingale. For arbitrary s.t. σ , we obtain $\mathbf{E}|Z_\sigma| < \infty$. Moreover, define s.t.’s

$$\tau_n = \inf(t : Z_t^* \geq n) \wedge n, \quad Z_t^* = \sup_{s \leq t} |Z_s|$$

and on $[[0, \tau_n]]$, we have $Z_t^* \leq n + |Z_{\tau_n}| \in L^1$, which certifies that $Z \in D$, and by the Doob-Meyer decomposition, (10.10) can be rewritten as $Z = M - A$ with a martingale M and a predictable process A with finite variation. Hence, Z is a semimartingale, and X is also a semimartingale.

Example 10.3 Assume X is a cadlag process with independent increments. Then $Z_t = \sum_{s \leq t} \Delta X_s I_{\{\Delta X_s \geq 1\}}$ is the process with independent increments and with finite variation. Define $Y = X - Z$ and observe that it has jumps bounded by one, and again, the process with independent increments. Hence, $Y_t - \mathbf{E}Y_t$ is a martingale, and we arrive at the conclusion: X is a semimartingale if and only if $\mathbf{E}Y_t$ has the finite variation.

Example 10.4 Let X be an integrable right-continuous adapted process defined on $[0, \infty]$. We call it a *quasimartingale* on $[0, \infty]$, if

$$\mathbf{Var}(X) = \sup_{0 \leq t_0 < \dots < t_n \leq \infty} \mathbf{E} \left[\sum_{i=0}^{n-1} |\mathbf{E}(X_{t_{i+1}} - X_{t_i} | \mathcal{F}_{t_i})| \right] < \infty.$$

The quasimartingale X admits the *Fisk decomposition* $X = M + Y - Z$, where M is a martingale, and Y and Z are nonnegative supermartingales on $[0, \infty]$. Hence, it is a semimartingale.

Now we will discuss the question about semimartingales and change of equivalent probability measures.

Theorem 10.10 If P and Q are equivalent probability measures on a measurable space $(\Omega, \mathcal{F})$, then each semimartingale with respect to P will be a semimartingale w.r. to Q .

Proof is based on the following lemmas.

Lemma 10.10 Let N be a positive local martingale w.r. to Q . Then N^{-1} is a semimartingale w.r. to Q .

Proof If $N_t \geq a > 0$, then $N_t^{-1} \leq a^{-1}$, and therefore $\mathbf{E}N_t^{-1} < \infty$, $t \geq 0$. Note that x^{-1} is convex. By the Jensen inequality, N_t^{-1} will be a submartingale, and,

hence, it is a semimartingale w.r. to Q by Example 10.2. In the general case, we consider a sequence of convex functions $f_n(x)$ such that $f_n(x) = \frac{1}{x}$ on $[1/n, \infty)$ and $[0, 1/n]$ the graph of f_n is the tangent to the curve $y = 1/x$ at the point $1/n$. For each $t \geq 0$, we have $\mathbf{E}f_n(N_t) < \infty$, which implies that $Y_n = f_n(N)$ is a submartingale w.r. to Q for each $n = 1, 2, \dots$. Taking s.t.'s $\tau_n = \inf(t : N_t \leq 1/n)$, we get a localizing sequence such that $Y_n = 1/N$ on $\llbracket 0, \tau_n \rrbracket$. So, by Theorem 10.9, $1/N$ is a semimartingale. $\square$

Lemma 10.11 *The product of semimartingales is a semimartingales.*

Proof Let us prove it (it is enough) for X^2 . We use $X = M + V$, where M is a locally bounded martingale, and V is a process with finite variation, as stated in the fundamental lemma.

Taking a localizing sequence $(\tau_n)_{n=1,2,\dots}$, we get M^{τ_n} as a bounded martingale, V^{τ_n} with bounded variation on $\llbracket 0, 1/n \rrbracket$. Let us show that $(M^{\tau_n} + V^{\tau_n})^2$ is a semimartingale. The term $(M^{\tau_n})^2$ is a submartingale. As far as $(V^{\tau_n})^2$ is concerned, we consider a subdivision $0 \leq t_0 < t_1 < \dots < t_n = t$ of $[0, t]$ and find that

$$\begin{aligned} \sum_i \left| (V_{t_{i+1}}^{\tau_n})^2 - (V_{t_i}^{\tau_n})^2 \right| &= \sum_i |V_{t_{i+1}}^{\tau_n} + V_{t_i}^{\tau_n}| |V_{t_{i+1}}^{\tau_n} - V_{t_i}^{\tau_n}| \\ &\leq 2 \int_0^t |dV_s^{\tau_n}| \left(\sum_i |V_{t_{i+1}}^{\tau_n} - V_{t_i}^{\tau_n}| \right) \\ &\leq 2 \left(\int_0^t |dV_s^{\tau_n}| \right)^2. \end{aligned}$$

To treat the term $M^{\tau_n} V^{\tau_n}$, we note that $M_t^{\tau_n} = \mathbf{E}(M_{t_n}^+ | \mathcal{F}_t) - \mathbf{E}(M_{t_n}^- | \mathcal{F}_t)$ and V^{τ_n} is the difference of two increasing processes. It means that we can consider only the case of nonnegative M^{τ_n} and V^{τ_n} . The process $M_t^{\tau_n} (V_t^{\tau_n} - \Delta V_{\tau_n} I_{\{\tau_n \leq t\}}) = M_t^{\tau_n} U_t^{\tau_n}$ is a submartingale, and $M^{\tau_n} V^{\tau_n} = M^{\tau_n} U^{\tau_n}$ on $\llbracket 0, \tau_n \rrbracket$. The application of Theorem 10.9 completes the proof of Lemma. $\square$

Let us come back to the proof of Theorem. Denote $Z_\infty = \frac{dQ}{dP}$ on $(\Omega, \mathcal{F})$ and find that $Z_t = \mathbf{E}^P(Z_\infty | \mathcal{F}_t)$ is a martingale w.r. to P and Z_∞ is positive P and Q -a.s. Obviously, X is a local martingale w.r. to $P \Leftrightarrow XZ^{-1}$ is a local martingale w.r. to Q . Writing X as the product $\frac{X}{Z} Z$, we get the statement from Lemmas 10.10 and 10.11.

Remark 10.11 We defined $[X, X]$ by (10.20). It turns out that one can prove that $[X, X]_t = P - \lim_n \sum_1^n (X_{t_i} - X_{t_{i-1}})^2$, and it is invariant regarding equivalent changes of measures. Similar fact is true also for $[X, Y]$ (see (10.21)).

10.3 On Stochastic Calculus for Semimartingales

We extend here Ito's calculus to bigger classes of processes defined on a standard stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$. A scheme of such extension will be presented sequentially for square-integrable martingales, local martingales, and semimartingales.

As we know (see, for example, Corollary 10.1 and Remark 10.7), any $M \in \mathcal{M}^2$ is described as $M_t = \mathbf{E}(M_\infty | \mathcal{F}_t)$, where $M_\infty \in L^2(\Omega, \mathcal{F}, P)$. It creates a possibility to think about $\mathcal{M}^2$ as a Hilbert space with the scalar product $(M, N) = \mathbf{E}M_\infty N_\infty$ and the norm $\|M\| = (\mathbf{E}M_\infty^2)^{1/2}$ for $M, N \in \mathcal{M}^2$.

Let us fix $M \in \mathcal{M}^2$ and define

$$S^2(M) = \left\{ \phi : \phi \text{ is predictable, } \mathbf{E} \int_0^\infty \phi_s^2 d[M, M]_s < \infty \right\}$$

with the norm $\|\phi\|_{S^2(M)} = (\mathbf{E} \int_0^\infty \phi_s^2 d[M, M]_s)^{1/2}$.

In this space, we consider a subspace $S_{step}^2(M)$ of predictable processes of the form

$$\phi = \phi_0^* I_{\{0\}} + \sum_{i=1}^{n-1} \phi_i I_{(t_i, t_{i+1}]},$$

where $\phi_0^* - \mathcal{F}_0$ -measurable, $\phi_i - \mathcal{F}_{t_i}$ -measurable, $i = 1, 2, \dots, n$.

For $\phi \in S_{step}^2(M)$, we define the *stochastic integral of ϕ w.r. to M* as the process of $\mathcal{M}^2$:

$$(\phi \circ M)_t = \int_0^\infty I_{(s, t]}(s) \phi_s dM_s = \sum_{i=1}^n \phi_i (M_{t_{i+1} \wedge t} - M_{t_i \wedge t}), \quad t \geq 0. \quad (10.22)$$

Using relation (10.22), we can easily obtain the following *properties* of $\phi \circ M$:

- (1) *Isometry*: $\|\phi \circ M\| = \|\phi\|_{S^2(M)}$.
- (2) *Continuity* of $(\phi \circ M)_t$ if M is continuous.
- (3) *Indistinguishability* of jumps $\Delta(\phi \circ M)_t$ and $\phi_t \Delta M_t$, $t > 0$.

The subspace $S_{step}^2(M)$ is dense in $S^2(M)$, and as in the case of stochastic integration w.t. to a Wiener process, one can extend the definition of $\phi \circ M$ to the whole space $S^2(M)$ with the properties (1)–(3) above.

Now we want to provide a more tight connection of $\phi \circ M$ with quadratic brackets of M and $\phi \circ M$. Creating such a connection, the following *Kunita-Watanabe inequality* is useful.

Lemma 10.12 *Let M and $N \in \mathcal{M}^2$, and ϕ and ψ be bounded measurable processes. Then*

$$\int_0^\infty |\phi_s| |\psi_s| |d[M, N]_s| \leq \left(\int_0^\infty \phi_s^2 d[M, M]_s \right)^{1/2} \left(\int_0^\infty \psi_s^2 d[N, N]_s \right)^{1/2}. \quad (10.23)$$

To prove (10.23), which is an analog of the Cauchy-Schwartz inequality, we consider for $\lambda \in \mathbb{R}$ and $s \leq t$ the difference

$$[M + \lambda N, M + \lambda N]_t - [M + \lambda N, M + \lambda N]_s.$$

It is nonnegative, and, hence,

$$|[M, N]_t - [M, N]_s|^2 \leq ([M, M]_t - [M, M]_s) ([N, N]_t - [N, N]_s).$$

Take $\phi = \sum_{i=1}^n \phi_i I_{(t_i, t_{i+1}]}$, $\psi = \sum_{j=1}^m \psi_j I_{(t_j, t_{j+1}]}$ with bounded r.v.'s ϕ_i and ψ_j , we obtain from the above inequality for brackets (10.23). After this, we extend (10.23) to the class of all bounded measurable functions ϕ and ψ using limit arguments, as usual.

Theorem 10.11 *For $M \in \mathcal{M}^2$ and $\phi \in S^2(M)$ the stochastic integral $\phi \circ M$ is a unique $L \in \mathcal{M}^2$ such that for any $N \in \mathcal{M}^2$ and any predictable bounded ψ :*

- (1) $[L, N] = \phi \circ [M, N]$
- (2) $\psi \circ [\phi \circ M, N] = \psi \phi \circ [M, N] = [\psi \phi \circ M, N]$ and $\psi \circ (\phi \circ M) = \psi \phi \circ M$
- (3) $(\phi \circ M)^c = \phi \circ M^c$, $(\phi \circ M)^d = \phi \circ M^d$.

In particular, for an s.t. τ and process $\psi = I_{\llbracket 0, \tau \rrbracket}$, we have equalities

$$\begin{aligned} (\phi \circ M)_\tau &= \int_0^\tau \phi_s dM_s = (\psi \circ (\phi \circ M))_\infty = ((\psi \phi) \circ M)_\infty \\ &= \int_0^\infty I_{\llbracket 0, \tau \rrbracket} \phi_s dM_s. \end{aligned}$$

Proof By the properties of $\phi \circ M$ and Lemma 10.12, the mapping $\phi \rightarrow \mathbf{E}((\phi \circ M)_\infty N_\infty - \int_0^\infty \phi_s d[M, N]_s)$ is continuous on $S^2(M)$ for any $N \in \mathcal{M}^2$.

Further, $I_t = (\phi \circ M)_t N_t - \int_0^t \phi_s d[M, N]_s$ is an uniformly integrable process, and for any s.t. τ , $\mathbf{E}I_\tau = 0$. Hence, by Lemma 10.4, the process (I_t) is a martingale, and we arrive at the statement (1). As far as the uniqueness, assume L and $L' \in \mathcal{M}^2$ and $[L, N] = \phi \circ [M, N]$ for any $N \in \mathcal{M}$. Then we have $[L - L', N] = 0$ and $[L - L', L - L'] = 0$. Hence, $(L - L')^2$ is a martingale, and we obtain equality $L = L'$. Other statements (2) and (3) are almost obvious. $\square$

Now we show the way of construction of stochastic integral for $M \in \mathcal{M}_{loc}$ and a *locally bounded predictable process* ϕ , i.e., $\phi_{\tau_n} I_{(\tau_n > 0)}$ is bounded for a localizing sequence of s.t.'s $(\tau_n)_{n=1,2,\dots}$.

To define such a stochastic integral $\phi \circ M$, we use the fundamental lemma, when $M = U + V$ with $U \in \mathcal{M}_{loc}^2$; V is a local martingale with bounded variation. Then for a localizing sequence $(\tau_n)_{n=1,2,\dots}$ for ϕ , U , V , we define

$$(\phi \circ M)_t = \int_0^t \phi_s^{\tau_n} dU_s^{\tau_n} + \int_0^t \phi_s^{\tau_n} dV_s^{\tau_n}$$

on $\llbracket 0, \tau_n \rrbracket$. To continue the construction, we need the following lemma.

Lemma 10.13

- (1) If V is a martingale with integrable variation and ϕ is predictable with $\mathbf{E} \int_0^\infty |\phi_s| dV_s| < \infty$, then the Lebesgue-Stieltjes integral $\int_0^t \phi_s dV_s \in \mathcal{M}$.
- (2) If $V \in \mathcal{M}^2$ and ϕ is bounded, then the Stochastic integral $(\phi \circ V)_t$ and the Lebesgue-Stieltjes integral $\int_0^t \phi_s dV_s$ are indistinguishable.

Proof We prove statement (1) only because second statement can be proved in a similar manner. Define $S'(V)$ the set of predictable processes ϕ such that $\mathbf{E} \int_0^\infty |\phi_s| dV_s| < \infty$. Then $\|\phi\|_{S'(V)} = \mathbf{E} \int_0^\infty |\phi_s| dV_s|$ defines the norm of this space. If $\|\phi^n - \phi\|_{S'(V)} \rightarrow 0$, $n \rightarrow \infty$, then

$$\begin{aligned} \mathbf{E} \left\{ \sup_t \left| \int_0^t \phi_s^n dV_s - \int_0^t \phi_s dV_s \right| \right\} &\leq \mathbf{E} \int_0^\infty |\phi_s^n - \phi_s| dV_s| \\ &\leq \|\phi^n - \phi\|_{S'(V)} \rightarrow 0, n \rightarrow \infty. \end{aligned}$$

Using these findings, we make a transition from step-predictable functions to all functions of space $S'(V)$. □

Coming back to the definition of $\phi \circ M$ by relation (10.3), one can conclude with the help of Lemma 10.13 that $\phi \circ M \in \mathcal{M}_{loc}$ and its definition does not depend on the decomposition $M = U + V$. Moreover, we arrive at statements which are similar to (1)–(3) of Theorem 10.11 if we replace class $\mathcal{M}^2$ by $\mathcal{M}_{loc}$.

The final step of construction of stochastic integral is the case of semimartingale. Assume ϕ is a predictable locally bounded process and X is a semimartingale of the form

$$X = X_0 + M + A, \tag{10.24}$$

where X_0 is a finite $\mathcal{F}_0$ -random variable (r.v.), $M \in \mathcal{M}_{loc}$, $A \in \mathcal{V}$.

Let us define

$$(\phi \circ X)_t = (\phi \circ M)_t + \int_0^t \phi_s dA_s, \quad (10.25)$$

where $\int_0^t \phi_s dA_s$ is the Lebesgue-Stieltjes integral.

According to Lemma 10.13, the definition (10.25) of $(\phi \circ X)$ does not depend on the decomposition (10.24). As a result, let us formulate the list of natural properties of $\phi \circ X$:

- (1) $\phi \circ X$ is a semimartingale.
- (2) $(\phi \circ X)^c = \phi \circ X^c$.
- (3) jumps $\Delta(\phi \circ X)_t$ and $\phi_t \Delta X_t$ are indistinguishable.
- (4) $(\phi \circ X)_\tau = \int_0^\infty I_{[0, \tau]} \phi_s dX_s = \int_0^\tau \phi_s dX_s$ for any finite s.t. τ .

Remark 10.12 It is interesting to note that there is no hope to go beyond semimartingales in extension of stochastic integration.

Let $X^1, \dots, X^d$ be d semimartingales taking the values in $\mathbb{R}$. It is convenient to collect them to a d -dimensional process $X_t = (X_t^1, \dots, X_t^d)$, called a d -valued semimartingale.

Let $F : \mathbb{E} = \mathbb{R}^d \rightarrow \mathbb{R}$, $d \geq 1$, be a twice continuously differentiable function. Then one can transform X with the help of this smooth function. It turns out that the transformed process $F(X_t)$ will be a semimartingale, satisfying the next *generalized Ito's formula*:

$$\begin{aligned} F(X_t) &= F(X_0) + \sum_{i=1}^d \int_0^t D^i F(X_{s-}) dX_s^{i,c} \\ &\quad + \frac{1}{2} \sum_{1 \leq i, j \leq d} \int_0^t D^i D^j F(X_{s-}) d[X^{i,c}, X^{j,c}]_s \\ &\quad + \sum_{s \leq t} \left(F(X_s) - F(X_{s-}) - \sum_{i=1}^d D^i F(X_{s-}) \Delta X_s^i \right), \end{aligned} \quad (10.26)$$

where $D^i F(x) = \frac{\partial F(x)}{\partial x_i} = \frac{\partial F(x^1, \dots, x^d)}{\partial x_i}$, $D^i D^j F(x) = \frac{\partial^2 F(x)}{\partial x^i \partial x^j}$, $i, j = 1, \dots, d$.

Remark 10.13 Let us note that processes $D^i F(X_{s-})$ and $D^i D^j F(X_{s-})$ are predictable and locally bounded, and therefore all integrals in (10.26) are well-defined.

Defining the localizing sequence $\tau_n = \inf\{t : |X_t| \geq n\}$, we obtain

that $|X_t|$ and $|X_{t-}| \leq n$ on $\llbracket 0, \tau_n \rrbracket$, $n = 1, 2, \dots$. Further, letting $K_n = \sup_{|x| \leq n} \sum_{i,j} |D^i D^j F(x)|$, we get with the help of the Taylor decomposition that

$$\left| \sum_{s < \tau_n} (F(X_s) - F(X_{s-}) - \sum_{i=1}^d D^i F(X_{s-}) \Delta X_s^i) \right| \leq \frac{K_n}{2} \sum_i \sum_{s < \tau_n} |\Delta X_s^i|^2.$$

Hence, for almost all ω , both series

$$\sum_{s < \tau_n} |F(X_s) - F(X_{s-}) - \sum_{i=1}^d D^i F(X_{s-}) \Delta X_s^i|$$

and

$$\sum_{s \leq \tau_n} |F(X_s) - F(X_{s-}) - \sum_{i=1}^d D^i F(X_{s-}) \Delta X_s^i|$$

converge, and we can conclude that

$$\sum_{s \leq t} |F(X_s) - F(X_{s-}) - \sum_{i=1}^d D^i F(X_{s-}) \Delta X_s^i|$$

converges for all $t > 0$.

This observation generates the idea that the formula (10.26) can be proved without big technical difficulties in between moments of jumps, i.e., in a *pure continuous case*.

This is the reason that we prove (10.26) for a one-dimensional continuous semimartingale $X_t = X_0 + M_t + A_t$, $M \in \mathcal{M}_{loc}^c$, $A \in \mathcal{V}$. Using a localization procedure, we can reduce the proof to the case when continuous processes M , A , $[M, M]$ are bounded as well as X_0 , F , F' and F'' , and the upper bound for $|F|$, $|F'|$, $|F''|$ is denoted by C .

Let us use the Taylor decomposition

$$F(y) - F(x) = (y - x)F'(x) + \frac{1}{2}(y - x)^2 F''(x) + o(x, y),$$

where $o(x, y) \leq \epsilon(|y - x|)|y - x|^2$ and ϵ is a nondecreasing function with $\lim_{t \rightarrow 0} \epsilon(t) = 0$.

Take $a > 0$ and define a sequence of s.t.'s $(\tau_n)_{n=0,1,\dots}$ as follows:

$$\begin{aligned} \tau_0 = 0, \dots, \tau_{i+1} = t \wedge (\tau_i + a) \wedge \inf\{s > \tau_i : M_s - M_{\tau_i} > a, \\ [M, M]_s - [M, M]_{\tau_i} > a \text{ or } |A_s - A_{\tau_i}| > a\}, \dots \end{aligned}$$

Obviously, for each $\omega : \tau_i(\omega) = t$ except for a finite number of i , and we can write the following equality using the Taylor formula above:

$$\begin{aligned} F(X_t) - F(X_0) &= \sum_i (F(X_{\tau_{i+1}}) - F(X_{\tau_i})) \\ &= \sum_i \left[F'(X_{\tau_i})(X_{\tau_{i+1}} - X_{\tau_i}) + \frac{1}{2} F''(X_{\tau_{i+1}})(X_{\tau_{i+1}} - X_{\tau_i})^2 \right. \\ &\quad \left. + o(X_{\tau_i}, X_{\tau_{i+1}}) \right]. \end{aligned} \quad (10.27)$$

For the martingale part of (10.27), we have

$$\begin{aligned} &\mathbf{E} \left[\int_0^t F'(X_s) dM_s - \sum_i F'(X_{\tau_i})(M_{\tau_{i+1}} - M_{\tau_i}) \right]^2 \\ &= \mathbf{E} \left[\sum_i \int_{\tau_i}^{\tau_{i+1}} (F'(X_s) - F'(X_{\tau_i}))^2 d[M, M]_s \right] \\ &\leq \mathbf{E} \left(\sup_i |F'(X_s) - F'(X_{\tau_i})|^2 \cdot [M, M]_t \right) \rightarrow 0 \end{aligned} \quad (10.28)$$

as $a \rightarrow 0$.

The term of (10.27) with process A can be treated similarly, and $\sum_i F'(X_{\tau_i})(A_{\tau_{i+1}} - A_{\tau_i})$ converges to $\int_0^t F'(X_s) dA_s$ in L^1 -sense.

The second term in (10.27), we represent as a sum of three components.

For the first component, we have

$$\begin{aligned} \left| \sum_i F''(X_{\tau_i})(A_{\tau_{i+1}} - A_{\tau_i})^2 \right| &\leq C \sup_i |A_{\tau_{i+1}} - A_{\tau_i}| \int_0^t |dA_s| \\ &\leq Ca \int_0^t |dA_s| \rightarrow 0, \quad a \rightarrow 0. \end{aligned} \quad (10.29)$$

For the second component, we obtain

$$\begin{aligned} \left| \sum_i F''(X_{\tau_i})(A_{\tau_{i+1}} - A_{\tau_i})(M_{\tau_{i+1}} - M_{\tau_i}) \right| &\leq C \sup_i |M_{\tau_{i+1}} - M_{\tau_i}| \int_0^t |dA_s| \\ &\leq Ca \int_0^t |dA_s| \rightarrow 0, \quad a \rightarrow 0. \end{aligned} \quad (10.30)$$

Taking into account that $M^2 - [M, M]$ is a martingale, we get for the third component that

$$\begin{aligned}
 & \mathbf{E} \left[\left(\sum_i F''(X_{\tau_i})(M_{\tau_{i+1}} - M_{\tau_i})^2 - \sum_i F''(X_{\tau_i})([M, M]_{\tau_{i+1}} - [M, M]_{\tau_i}) \right)^2 \right] \\
 &= \sum_i \mathbf{E} \left[(F''(X_{\tau_i}))^2 \left((M_{\tau_{i+1}} - M_{\tau_i})^2 - ([M, M]_{\tau_{i+1}} - [M, M]_{\tau_i}) \right)^2 \right] \\
 &\leq 2C^2 \mathbf{E} \left\{ \sup_i |M_{\tau_{i+1}} - M_{\tau_i}|^2 M_t^2 \right\} \\
 &\quad + 2C^2 \mathbf{E} \left\{ \sup_i |[M, M]_{\tau_{i+1}} - [M, M]_{\tau_i}| \cdot [M, M]_t \right\} \\
 &\leq 2C^2 a^2 \mathbf{E} M_t^2 + 2C^2 a \mathbf{E} [M, M]_t \rightarrow 0, \quad a \rightarrow 0.
 \end{aligned} \tag{10.31}$$

The remaining term in (10.27) can be treated as follows:

$$\begin{aligned}
 \mathbf{E} \sum_i o(X_{\tau_i}, X_{\tau_{i+1}}) &\leq \mathbf{E} \left[\sum_i (X_{\tau_{i+1}} - X_{\tau_i})^2 \epsilon(|X_{\tau_{i+1}} - X_{\tau_i}|) \right] \\
 &\leq \mathbf{E} \left[2\epsilon(2a) \sum_i (A_{\tau_{i+1}} - A_{\tau_i})^2 + 2\epsilon(2a) \sum_i (M_{\tau_{i+1}} - M_{\tau_i})^2 \right] \\
 &\leq 2\epsilon(2a) \mathbf{E} \left\{ a \int_0^t |dA_s| + M_t^2 \right\} \rightarrow 0, \quad a \rightarrow 0.
 \end{aligned} \tag{10.32}$$

Putting together relations (10.28)–(10.32), we complete the proof.

Now we are ready to consider stochastic differential equations with respect to semimartingale. The main result is contained in the next theorem.

Theorem 10.12 *Let $N = M + A$ be a semimartingale, $M \in \mathcal{M}_{loc}$, $A \in \mathcal{V}$. Assume that the function $F : \mathbb{R}_+ \times \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ satisfies the conditions:*

- (1) $f(s, \omega, \cdot)$ is a Lipschitz function with a constant K .
- (2) $f(s, \cdot, x)$ is $\mathcal{F}_s$ -measurable.
- (3) $f(\cdot, \omega, \cdot)$ is a continuous function.

Then the stochastic differential equation (SDE) w.r. to N

$$X_t(\omega) = X_0(\omega) + \int_0^t f(s, \omega, X_{s-}(\omega)) dN_s(\omega) \tag{10.33}$$

has a unique solution as a cadlag adapted process with the initial value $X_0(\omega) - \mathcal{F}_0$ -measurable finite random variable.

Proof Obviously, (1)–(3) guarantee that the integral on the right-hand side of (10.33) is well-defined. The proof includes a couple of steps. We start with the assumption that processes

$$[M, M]_t \text{ and } \int_0^t |dA_s| \text{ are bounded by } b > 0 \quad (10.34)$$

and

$$|f(t, \omega, 0)| \leq c \text{ for all } t, \omega. \quad (10.35)$$

Let H be a class of cadlag processes X such that $X^* = \sup_t |X_t| \in L^2$, $X_0 = 0$. Define the norm in this space $\|X\| = \|X^*\|_{L^2}$.

Denote for $x \in H$

$$U(X)_t = \int_0^t f(s, \omega, X_{s-}) dN_s.$$

□

Lemma 10.14 *The process $U(x) \in H$ if $X \in H$. Moreover, for processes X and $Y \in H$, we have*

$$\|U(X) - U(Y)\| \leq h\|X - Y\|, \quad h = K(2\sqrt{b} + b). \quad (10.36)$$

Proof For zero process, we have

$$U(0)_t = \int_0^t f(s, \omega, 0) dM_s + \int_0^t f(s, \omega, 0) dA_s = L_t + V_t.$$

Process $L \in \mathcal{M}_{loc}$, and by (10.34) and (10.35)

$$[L, L]_\infty = \int_0^\infty f^2(s, \omega, 0) d[M, M]_s \leq c^2 b, \quad (10.37)$$

and $\mathbf{E}(L^*)^2 \leq 4c^2 b$ by Doob's inequality. □

For process V_t , we get from (10.34)–(10.35) that

$$V^* \leq \int_0^\infty |dV_s| \leq \int_0^\infty |f(s, \omega, 0)| |dA_s| \leq cb, \quad (10.38)$$

and hence $\mathbf{E}(V^*)^2 \leq c^2 b^2$. Relations (10.37) and (10.38) show that $U(X) \in H$. Further, denote $Z = X - Y$ and find that

$$\begin{aligned} U(X)_t - U(Y)_t &= \int_0^t (f(X_{s-}) - f(Y_{s-})) dM_s + \int_0^t (f(X_{s-}) - f(Y_{s-})) dA_s \\ &= L'_t + V'_t, \quad L' \in \mathcal{M}_{loc}, \quad V' \in \mathcal{V}. \end{aligned}$$

Using similar reasonings as in (10.37) and (10.38), we obtain for processes L' and V' that

$$[L', L']_\infty = \int_0^\infty (f(X_{s-}) - f(Y_{s-}))^2 d[M, M]_s \leq K^2 b (Z^*)^2, \quad (10.39)$$

$$\mathbf{E}((L')^*)^2 \leq 4K^2 b \mathbf{E}(Z^*)^2 \text{ and}$$

$$V^* \leq \int_0^\infty K |Z_s| |dA_s| \leq K b Z^*. \quad (10.40)$$

It follows from (10.39) and (10.40) that

$$\|U(X) - U(Y)\| \leq K \|X - Y\| (b + 2\sqrt{b}),$$

and we get (10.36).

Lemma 10.15 *Assume conditions (10.34) and (10.35) are satisfied, and $h = K(b + 2\sqrt{b}) < 1$. Then there exists a unique cadlag adapted process X solving the equation*

$$X_t = \int_0^t f(s, \omega, X_{s-}) dN_s.$$

Proof By $h < 1$, there exists a unique solution X from space H . Let us note from conditions (1) and (2) of the theorem that

$$|\Delta N_s| \leq |\Delta M_s| + |\Delta A_s| = (\Delta[M, M]_s)^{1/2} + |\Delta A_s| \leq \sqrt{2b} + b.$$

So, if Z is a cadlag adapted process, $Z_t = \int_0^t f(s, \omega, Z_{s-}) dN_s$, we can define stopping times $\tau_n^Z = \inf(t : |Z_t| \geq n)$, $n = 1, 2, \dots$, and get

$$\begin{aligned} \Delta Z_{\tau_n^Z} &= f(\tau_n^Z, Z_{\tau_n^Z-}) \Delta N_{\tau_n^Z} \text{ and} \\ |\Delta Z_{\tau_n^Z}| &\leq (c + nK)(2b + \sqrt{2b}). \end{aligned}$$

Hence, the process Z is locally bounded and locally in space H . By uniqueness in H , we obtain $Z = X$. $\square$

Lemma 10.16 Assume that condition (10.34) is fulfilled and $h = K(b + 2\sqrt{b}) < 1$, then there exists a unique cadlag adapted process X solving equation

$$X_t = \int_0^t f(s, \omega, X_{s-}) dN_s.$$

Proof Let $\tau_n = \inf(t : |f(t, \omega, 0)| \geq n)$, $n = 1, 2, \dots$ and $f_n(t, \omega, x) = f(t, \omega, x)I_{\{0 < t \leq \tau_n\}}$. Obviously, functions f_n satisfy (10.35) with $c = n$, and each equation $Y_t^n = \int_0^t f_n(s, \omega, Y_{s-}^n) dN_s$ has a unique solution. By $f_{n+1} = f_n$ on $\llbracket 0, \tau_n \rrbracket$, we have $Y^n = Y^{n+1}$ on $\llbracket 0, \tau_n \rrbracket$.

Further, an adapted process X solves the equation $X_t = \int_0^t f(s, \omega, X_{s-}) dN_s \Leftrightarrow$ for each τ_n , we have $X_t^{\tau_n} = \int_0^t f(s, \omega, X_{s-}^{\tau_n} N_s^{\tau_n})$. Let us define now the process X as follows: $X = Y^n$ on $\llbracket 0, \tau_n \rrbracket$, $X_0 = 0$, and find the unique solution. $\square$

Lemma 10.17 Assume a semimartingale N has jumps $|\Delta N_t| \leq b/4$ and both $b, h = K(b + 2\sqrt{b}) < 1$. Then the equation $X_t = \int_0^t f(s, \omega, X_{s-}) dN_s$ has one and only one solution.

Proof Under the conditions of Lemma, we can write $N = M + A$, $M \in \mathcal{M}_{loc}$ and A is a predictable process with locally bounded variation. Without loss of generality, we can assume that $M \in \mathcal{M}$ and predictable process A has integrable variation. Then

$$|\Delta A_\tau| = |\mathbf{E}(\Delta N_\tau | \mathcal{F}_{\tau-}) - \mathbf{E}(\Delta M_\tau | \mathcal{F}_{\tau-})| = |\mathbf{E} \Delta N_\tau | \mathcal{F}_{\tau-})| \leq b/4,$$

and we have a boundedness of $\Delta[M, M]_t = (\Delta M_t)^2$ and $|\Delta A_t|$ by $b/4$. Denote $D_t = \int_0^t |dA_s|$ and put $\tau_0 = 0, \dots, \tau_n = \inf(t > \tau_{n-1} : D_t - D_{\tau_{n-1}} \geq b/2)$. We have $\int_{\llbracket \tau_n, \tau_{n+1} \rrbracket} d[M, M]_s \leq b$ and $\int_{\llbracket \tau_n, \tau_{n+1} \rrbracket} |dA_s| \leq b$ for all $n = 1, 2, \dots$.

By Lemma 10.16, there is one and only one solution on $\llbracket 0, \tau_1 \rrbracket$; denote it by X^1 . At the interval $\llbracket \tau_1, \tau_2 \rrbracket$, we have equation

$$X_t = X_{\tau_1}^1 + \int_{\llbracket \tau_1, t \rrbracket} f(s, \omega, X_{s-}) dN_s.$$

Obviously, X solves (10.3) $\Leftrightarrow Y = X - X_{\tau_1}^1$ is a solution of the equation

$$Y_t = \int_0^t f(s, \omega, Y_{s-} + X_{\tau_1}^1) dN_s.$$

So, we get a unique solution X^2 at this interval and so on. $\square$

Let us come back to the proof of the theorem. We assume that b as in Lemma 10.17, and we have a nondecreasing sequence of s.t.'s $\tau_1 \leq \tau_2 \leq \dots \leq \tau_n \leq \dots$ at which $|\Delta N_t| > b/4$. Construct the following semimartingale:

$$N_t^1 = M_t I_{(t < \tau_1)} + N_{\tau_1} I_{(t \geq \tau_1)}.$$

A cadlag process X is a solution of (10.33) on $\llbracket 0, \tau_1 \rrbracket \Leftrightarrow$ the process $X_t^1 = X_t I_{(t < \tau_1)} + X_{\tau_1} I_{(t \geq \tau_1)}$ is a solution on $\llbracket 0, \tau_1 \rrbracket$ of the equation $X_t^1 = X_0 \int_0^t f(s, \omega, X_{s-}^1) dN_s^1$. On the interval $\llbracket 0, \tau_1 \rrbracket$, the process N^1 satisfies the conditions of Lemma 10.17, and hence the last SDE has a unique cadlag adapted solution X^1 on $\llbracket 0, \tau_1 \rrbracket$. We can continue this construction on $\llbracket \tau_1, \tau_2 \rrbracket$ and so on.

Example 10.5 Let us consider the linear SDE:

$$Z_t = 1 + \int_0^t Z_s dN_s. \quad (10.41)$$

We know from Theorem 10.12 that a unique solution of the Eq.(10.41) exists. It turns out that it admits a special form, which is similar to the Girsanov exponent:

$$Z_t = \exp \left\{ N_t - \frac{1}{2} [N^c, N^c]_t \right\} \prod_{s \leq t} (1 + \Delta N_s) e^{-\Delta N_s}. \quad (10.42)$$

The form of solution in (10.42) is called a *stochastic exponent* or the Doleans exponent, denoted by $\mathcal{E}_t(N)$. To prove this, we denote $Y_t = N_t - \frac{1}{2} [N^c, N^c]_t$ and $X_t = \prod_{s \leq t} (1 + \Delta N_s) e^{-\Delta N_s}$, and apply Ito's formula to the function $F(Y_t, X_t) = \exp(Y_t) X_t$.

Problem 10.6 Prove the properties of stochastic exponents w.r. to semimartingales:

- (1) Let $\mathcal{E}_t(N)$ be a stochastic exponent w.r. to a semimartingale N with $\Delta N_t \neq -1$, $t \geq 0$. Then $\frac{1}{\mathcal{E}_t(N)} = \mathcal{E}_t(-N^*)$, where $N_t^* = N_t - [N^c, N^c]_t - \sum_{s \leq t} \frac{(\Delta N_s)^2}{1 + \Delta N_s}$. stochastic exponent w.r. to a semimartingale N with $\Delta N_t \neq -1$, $t \geq 0$. Then $\frac{1}{\mathcal{E}_t(N)} = \mathcal{E}_t(-N^*)$, where $N_t^* = N_t - [N^c, N^c]_t - \sum_{s \leq t} \frac{(\Delta N_s)^2}{1 + \Delta N_s}$.
- (2) If N in $\mathcal{M}_{loc}$, then $\mathcal{E}_t(N) \in \mathcal{M}_{loc}$.
- (3) $\mathcal{E}_t(N) = 0$ for $(t, \omega) \in \llbracket \tau, \infty \rrbracket$, where $\tau = \inf(t : \Delta N_t = -1)$.
- (4) $\mathcal{E}_t(U) \mathcal{E}_t(V) = \mathcal{E}_t(U + V + [U, V])$, where U and V are semimartingales. This property is called the *rule of multiplication of stochastic exponents*.

Remark 10.14 Equation (10.33) admits further generalizations. One can prove an analog of Theorem 10.12 for equation

$$dX_t(\omega) = b(t, \omega, X_{t-}) dA_t + \sigma(t, \omega, X_{t-}) dM_t, \quad (10.43)$$

where $M \in \mathcal{M}_{loc}$, $A \in \mathcal{V}$, both coefficients b and σ are Lipschitz's functions, and $b(t, \omega, x)$ is optional and $\sigma(t, \omega, x)$ is predictable as functions of (t, ω) .

Let us note that the Eq.(10.43) can also be considered for a multidimensional semimartingale $X = A + M$ and for matrix-valued coefficients b and σ .

Remark 10.15 Equation (10.33) admits further important results. One can prove an analog of comparison Lemma 9.1 for stochastic differential equations (10.43) with a predictable increasing process $A = (A_t)_{t \geq 0}$.

Moreover, the Ito's formula (10.26) admits a further generalization to a class of convex functions F . The corresponding formula is similar to (10.26) where the second term on the right-hand side is replaced by a continuous process V from $\mathcal{V}$. The process V admits a representation with the help of a *local time* of X .

10.4 The Doob-Meyer Decomposition: Proof and Related Remarks

The leading idea of this short book as it was stated in preface dictates the way of exposition avoiding too long proofs and too many technical details. That is why the proof of the Doob-Meyer decomposition was omitted in Sect. 10.1. However, the proof of this fundamental fact of stochastic analysis must be presented in an appropriate manner.

It is necessary to note that the initial Doob-Meyer decomposition in continuous time was given by Meyer in the form

$$X_t = X_0 + M_t + A_t, \quad t \geq 0, \quad (10.44)$$

where $X = (X_t)_{t \geq 0}$ is a submartingale of class (D) , $M = (M_t)_{t \geq 0}$, $M_0 = 0$, is a martingale, and $A = (A_t)_{t \geq 0}$, $A_0 = 0$, is a “natural” increasing process.

The notion “natural” means that for every bounded cadlag martingale $m = (m_t)_{m \geq 0}$ on a standard stochastic basis $(\Omega, \mathcal{F}, (F_t)_{t \geq 0}, P)$, the following equality is satisfied

$$\mathbf{E} \int_0^t m_s dA_s = \mathbf{E} \int_0^t m_{s-} dA_s. \quad (10.45)$$

This characterization of A looks very artificial in comparison with discrete time, where A is predictable. Nevertheless, since that time, there were several proofs of the Doob-Meyer decomposition in these terms, including the proof of Rao who used the limiting arguments for transition from discrete time to continuous time decomposition. Doleans was the first to note that the notions “natural” and “predictable” are equivalent in (10.44).

This section is devoted to the proof of (10.44) in such a new fashion, and as in many other books, we work with this form only. Nevertheless, before providing the

proof which is based on the paper [8], we want to compare notions “natural” and “predictable”.

First of all, one can show that $\mathbf{E}m_t A_t = \mathbf{E} \int_0^t m_s dA_s$, and therefore the condition (10.45) is equivalent to the next one:

$$\mathbf{E}m_t A_t = \mathbf{E} \int_0^t m_s dA_s. \quad (10.46)$$

Problem 10.7 Prove (for discrete time) that an increasing process A is natural $\iff A$ is predictable, i.e., $A_n - F_{n-1}$ -measurable, $n \geq 1$.

Hint: We rewrite (10.46) for the discrete time case as follows:

$$\mathbf{E}m_n A_n = \mathbf{E} \sum_1^n m_{k-1} \Delta A_k.$$

It follows from this condition that $\mathbf{E}m_n(A_n - \mathbf{E}(A_n|F_{n-1})) = 0$ and even $\mathbf{E}|A_n - \mathbf{E}(A_n|F_{n-1})| = 0$, which leads to predictability of A .

The opposite implication follows from the equality $\mathbf{E} \sum_1^n A_k \Delta m_k = 0$ and the representation

$$\sum_1^n A_k \Delta m_k = A_n m_n - m_{n-1} \Delta A_n - m_{n-2} \Delta A_{n-1} - \dots$$

which leads to $\mathbf{E}A_n m_n = \mathbf{E} \sum_1^n m_{k-1} \Delta A_k$.

Let us show that the decomposition (10.44) is unique in the class of natural processes A . To prove it, we assume that there exist two such decompositions for X with martingales M', M'' and with the natural increasing processes A', A'' . Hence, $M'_t + A'_t = M''_t + A''_t$ for all $t \geq 0$. Define process $N_t = A'_t - A''_t = M'_t - M''_t$, which is a martingale with finite variation.

Further, for every bounded martingale $m = (m_t)_{t \geq 0}$, we have by (10.46) that

$$\mathbf{E}m_t.(A'_t - A''_t) = \mathbf{E} \int_0^t m_s dN_s = \lim_{n \rightarrow \infty} \mathbf{E} \sum_j^{k_n} m_{t_{j-1}^{(n)}} (N_{t_j^{(n)}} - N_{t_{j-1}^{(n)}}),$$

where $(t_k^{(n)})_{k=0, \dots, k_n}$ is a sequence of subdivision of $(0, T]$ with diameter $\Delta t_k^{(n)} \rightarrow 0, n \rightarrow \infty$. Obviously, $\mathbf{E}m_{t_{j-1}^{(n)}} \Delta N_{t_j^{(n)}} = 0$, and, hence, $\mathbf{E}m_t.(A'_t - A''_t) = 0, t \geq 0$.

To finish the proof, we take an arbitrary integrable random variable ζ and define a martingale $\zeta_t = \mathbf{E}(\zeta|F_t)$. Then $\mathbf{E}\zeta.(A'_t - A''_t) = 0$ for every such random variable, which leads to the conclusion that $A'_t = A''_t$ (a.s) for all t .

Let us prove the Doob-Meyer decomposition (10.44) assuming for simplicity that $t \in [0, 1]$.

Denote by $D = \{D_1, D_2, \dots, D_n, \dots\}$, where $n = 1, 2, \dots$, $D_n = \{0, \frac{1}{2^n}, \frac{2}{2^n}, \dots, \frac{j}{2^n}, \dots\}$, the dyadic numbers of the interval $[0, 1]$.

Consider the process $X = (X_t)_{t \geq 0}$ on this discrete time, and write the Doob-Meyer decomposition for $t \in D_n$:

$$X_t = M_t^n + A_t^n, \quad (10.47)$$

$$A_t^n - A_{t-\frac{1}{2^n}}^n = \mathbf{E}(X_t - X_{t-\frac{1}{2^n}} | F_{t-\frac{1}{2^n}}),$$

$$M_t^n = X_t - A_t^n.$$

Here, $M^n = (M_t^n)_{t \in D_n}$ is a martingale, and $A^n = (A_t^n)_{t \in D_n}$ is increasing and predictable with respect to $(F_t)_{t \in D_n}$.

The idea is to take a limit in (10.47) and, as a result, to obtain (10.44).

To make such limiting transition carefully, we need the following lemma.

Lemma 10.18 *Assume $(f_n)_{n \geq 1}$ is a uniformly integrable sequence of random variables on the probability space (Ω, F, P) . Then there exists $g_n \in \text{conv}\{f_n, f_{n+1}, \dots\}$, which converges in space $L^1(\Omega, F, P)$.*

We will use a version of the Komlos Lemma here:

If $(f_n)_{n \geq 1}$ is a bounded sequence on a Hilbert space, then for each n , one can pick $g_n \in \text{conv}\{f_n, f_{n+1}, \dots\}$ such that $(g_n)_{n \geq 1}$ will converge in the norm $\|\cdot\|_2$ of this space.

To Prove this lemma, we define the finite number

$$a = \sup_n \inf\{\|g\|_2 : g \in \text{conv}\{f_n, f_{n+1}, \dots\}\}.$$

For each n , we can take some $g_n \in \text{conv}\{f_n, f_{n+1}, \dots\}$ with the norm $\|g_n\|_2 \leq a + \frac{1}{n}$. For a fixed $\varepsilon > 0$, and a big enough n , we obtain for all $m, k \geq n$ that

$$\|(g_k + g_m)/2\|_2 > a - \varepsilon.$$

We get that (g_n) is fundamental by

$$\|(g_k - g_m)\|_2^2 = 2\|g_k\|_2^2 + 2\|g_m\|_2^2 - \|(g_k + g_m)\|_2^2 \leq 4(a + \frac{1}{n})^2 - 4(a - \varepsilon)^2,$$

and therefore this sequence converges.

Proof of Lemma 10.18: We truncate f_n and denote truncated functions by $f_n^{(i)} = f_n \cdot I_{\{|f_n| \leq i\}}$, $n = 1, 2, \dots$. Then for every i , we get a bounded sequence $(f_n^{(i)})_{n \geq 1}$ for which the Komlos lemma can be applied in the following manner.

Problem 10.8 For each $n = 1, 2, \dots$ there exist convex weights $(\lambda_n^n, \dots, \lambda_{N_n}^n)$ such that $\sum_{j=n}^{N_n} \lambda_j^n f_j^{(i)}$ converge on space $L^2(\Omega)$ for each $i = 1, 2, \dots$

Hint: Use the Komlos lemma to find convex weights $\lambda_n^n, \dots, \lambda_{N_n}^n$ with the convergent sum $\sum_{j=n}^{N_n} \lambda_j^n f_j^{(i)}$ as $n \rightarrow \infty$ for $i = 1, 2, \dots, m$ and apply a diagonalization procedure to get a desirable result.

Now using uniform integrability of (f_n) , we obtain that $f_n^{(i)}$ converges uniformly in n to f_n in $L^1(\Omega)$ as $i \rightarrow \infty$ and therefore uniformly in n :

$$\sum_{j=n}^{N_n} \lambda_j^n f_j^{(i)} \rightarrow \sum_{j=n}^{N_n} \lambda_j^n f_j \quad \text{in } L^1(\Omega) \quad \text{as } i \rightarrow \infty.$$

Lemma 10.19 *The sequence $(M_1^n)_{n \geq 1}$ is uniformly integrable.*

Proof Without loss of generality, we can put $X_1 = 0$ and $X_t \leq 0$ for $t \leq 1$, because there is an obvious transition from X_t to $X_t - \mathbf{E}(X_1 | F_t)$. In such a case, $M_1^n = -A_1^n$ and

$$X_\tau^n = -\mathbf{E}(A_1^n | F_\tau) + A_\tau^n \quad (10.48)$$

for every s.t. τ w.r. to $(F_t)_{t \in D_n}$.

Let us derive from assumption $X \in (D)$ that $(A_1^n)_{n \geq 1}$ is uniformly integrable. Define the following s.t.:

$$\tau_n(c) = \inf\{(j-1)/2^n : A_{\frac{j}{2^n}}^n > c\} \wedge 1, \quad c > 0,$$

and note

$$\{\tau_n(c) < 1\} \subseteq \{\tau_n\left(\frac{c}{2}\right) < 1\}, \quad A_{\tau_n(c)}^n \leq c.$$

It follows from (10.48) that

$$X_{\tau_n(c)} \leq -\mathbf{E}(A_1^n | F_{\tau_n(c)}) + c.$$

Thus,

$$\begin{aligned} \mathbf{E}A_1^n \cdot I_{\{A_1^n > c\}} &= \mathbf{E}\mathbf{E}(A_1^n | F_{\tau_n(c)}) \cdot I_{\{\tau_n(c) < 1\}} \\ &\leq c \cdot P(\tau_n(c) < 1) - \mathbf{E}X_{\tau_n(c)} \cdot I_{\{\tau_n(c) < 1\}} \end{aligned}$$

and by (10.48),

$$\begin{aligned} \mathbf{E}(-X_{\tau_n(\frac{c}{2})}) \cdot I_{\{\tau_n(\frac{c}{2}) < 1\}} &= \mathbf{E}(A_1^n - A_{\tau_n(\frac{c}{2})}^n) \cdot I_{\{\tau_n(\frac{c}{2}) < 1\}} \\ &\geq \mathbf{E}(A_1^n - A_{\tau_n(c)}^n) \cdot I_{\{\tau_n(c) < 1\}} \\ &\geq \frac{c}{2} P\{\tau_n(c) < 1\}. \end{aligned}$$

These inequalities imply that

$$\mathbf{E}A_1^n \cdot I_{\{A_1^n > c\}} \leq -2\mathbf{E}X_{\tau_n(\frac{c}{2})}^n \cdot I_{\{\tau_n(c) < 1\}} - \mathbf{E}X_{\tau_n(c)}^n \cdot I_{\{\tau_n(c) < 1\}}. \quad (10.49)$$

We also note

$$P\{\tau_n(c) < 1\} = P\{A_1^n > c\} \leq \frac{1}{c} \mathbf{E}A_1^n = -\frac{1}{c} \mathbf{E}(M_1^n) = -\frac{1}{c} \mathbf{E}X_0$$

and conclude that

$$\lim_{c \rightarrow \infty} P(\tau_n(c) < 1) = 0 \quad \text{uniformly in } n.$$

Assumption $X \in (D)$ together with (10.49) implies that $(A_1^n)_{n \geq 1}$ is uniformly integrable as well as $(M_1^n)_{n \geq 1} = (X_1 - A_1^n)_{n \geq 1}$.

Now we are ready for a limiting transition as $n \rightarrow \infty$ in (10.47).

First, we extend M^n to the whole interval $[0, 1]$ by setting $M_t^n = \mathbf{E}(M_1^n | F_t)$. According to Lemmas 10.18 and 10.19, there exist $M \in L^1(\Omega)$ and convex weights $(\lambda_j^n)_{j=n, \dots, N_n}$ such that $M_1^n(\lambda) = \sum_{j=n}^{N_n} \lambda_j^n M_1^j$ converges to M on space $L^1(\Omega)$. Moreover, by Jensen's inequality for each $t \in [0, 1]$,

$$M_t^n(\lambda) \rightarrow M_t = \mathbf{E}(M_1 | F_t) \quad \text{as } n \rightarrow \infty.$$

Second, we extend A^n to $[0, 1]$ by setting $A^n = \sum_{t \in D_n} A_t^n \cdot I_{(t - \frac{1}{2n}, t]}$, $t \in [0, 1]$, $n = 1, 2, \dots$, and denote $A^n(\lambda) = \sum_{j=n}^{N_n} \lambda_j^n A^j$.

Then the cadlag process

$$A_t = X_t - M_t, \quad \text{where } t \in [0, 1],$$

satisfies the following relations

$$A_t^n(\lambda) = (X_t - M_t^n(\lambda)) \rightarrow (X_t - M_t) = A_t \quad \text{in } L^1(\Omega) \quad \text{for } t \in D, n \rightarrow \infty.$$

So, we find a subsequence $A_t^{n_k}(\lambda) \rightarrow A_t$ (a.s.), which we identify with $A_t^n(\lambda) \rightarrow A_t$ (a.s.). These standard considerations lead us to conclude that A is (a.s.) increasing on D and therefore on the whole interval $[0, 1]$ by its right-continuity. $\square$

Lemma 10.20 *Process A is predictable.*

Proof: In fact, we can restrict ourself by the proof that for all $t \in [0, 1]$ (a.s.),

$$\limsup_n A_t^n(\lambda) = A_t, \quad (10.50)$$

because the process A^n and $A^n(\lambda)$ are left-continuous and adapted, and hence they are predictable.

Problem 10.9 Let f_n and f be increasing functions from $[0, 1]$ to R^1 such that f is right-continuous and $f(t) = \lim_{n \rightarrow \infty} f_n(t)$, $t \in D$. Prove that

$$\limsup_n f_n(t) \leq f(t) \quad \text{for all } t \in [0, 1], \quad (10.51)$$

$$\lim_n f_n(t) = f(t), \quad \text{if } f \text{ is continuous at } t \in [0, 1]. \quad (10.52)$$

Hint: By the right-continuity ($t_k \in D$, $t_k > t$),

$$f(t) = \lim_{k \rightarrow \infty} f(t_k) = \lim_{k \rightarrow \infty} (\lim_{n \rightarrow \infty} f_n(t_k)) \geq \lim_n \sup f_n(t),$$

since $\lim_n \sup f_n(t_k) \geq \lim_n \sup f_n(t)$.

By (10.51) and (10.52), the relation (10.50) can be broken at discontinuity points of A only. The process A is cadlag, and, as was shown in Sect. 10.1, the points of jumps of A can be exhausted by a countable sequence of stopping times (see (10.2) and related remarks). So, it is sufficient for (10.50) that $\lim_n \sup A_\tau^n(\lambda) = A_\tau$ for all s.t. τ .

By (10.51), $\lim_n A_\tau^n(\lambda) \leq A_\tau$ and $A_\tau^n(\lambda) \leq A_1^n(\lambda)$ which converges to A , on space $L^1(\Omega)$. Hence, by Fatou's lemma

$$\liminf_n \mathbf{E} A_\tau^n \leq \limsup_n \mathbf{E} A_\tau^n(\lambda) \leq \mathbf{E} \limsup_n A_\tau^n(\lambda) \leq \mathbf{E} A_\tau.$$

To prove that $\lim_n \mathbf{E} A_\tau^n = \mathbf{E} A_\tau$ we set $\sigma_n = \inf\{t \in D_n : t \geq \tau\}$. Then $\sigma_n \downarrow \tau$ as $n \rightarrow \infty$ and $A_\tau^n = A_{\sigma_n}^n$.

The process X belongs to class (D) and therefore

$$\mathbf{E} A_\tau^n = \mathbf{E} A_{\sigma_n}^n = \mathbf{E} X_{\sigma_n} - \mathbf{E} M_0 \rightarrow \mathbf{E} X_\tau - \mathbf{E} M_0 = \mathbf{E} A_\tau.$$

Lemma 10.20 is proved, and hence the decomposition (10.44) is true with a martingale M and an increasing predictable process A .

Chapter 11

General Theory of Stochastic Processes in Applications



The main goal of this chapter is to show how the general theory developed before can be applied to mathematical finance and statistics of random processes. In the area of mathematical finance, a semimartingale financial market model is introduced. Applying to this general model the technique of stochastic exponents, the fundamental questions of arbitrage and completeness of such a market are studied. These results have a number of corollaries for modeling and option pricing (Black-Scholes model and formula, Cox-Ross-Rubinstein model and formula, etc.). In the area of statistics of random processes, the technique developed above gives a possibility to introduce semimartingale models. It is shown that classical discrete time and continuous time models of stochastic approximation are embedded in a semimartingale scheme. Moreover, it is proved that semimartingale stochastic approximation procedures are strong consistent and asymptotically normal under very wide conditions. In case of semimartingale regression, the structural least-squares estimates are strong consistent and their sequential versions satisfy the important fixed accuracy property (see [9, 10, 18, 20, 32, 39, 47, 49], and [60]).

11.1 Stochastic Mathematical Finance

Suppose that besides the original measure on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$, we are given a measure $\tilde{P}$ locally equivalent to P and with local density $(Z_t)_{t \geq 0}$. The equivalence implies strict positivity of Z_t (P -a.s.), $t \geq 0$. Therefore, one can define a local martingale $N = (N_t)_{t \geq 0}$ with respect to P as a stochastic integral to P as a stochastic integral $N_t = \int_0^t Z_s^{-1} dZ_s$. It leads to a stochastic exponent form of $Z_t = \mathcal{E}_t(N)$. We already know that a local martingale is transformed into a semimartingale (not a local martingale) under an equivalent change of measure.

The next lemma contains a result of this type. To formulate it, we denote $\mathcal{M}_{loc}(P)$ and $\mathcal{M}_{loc}(\tilde{P})$ classes of local martingales w.r.t. P and $\tilde{P}$ correspondently.

Lemma 11.1 *Let $X = (X_t)_{t \geq 0}$ be a semimartingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$, and then*

$$XZ \in \mathcal{M}_{loc}(P) \Rightarrow X \in \mathcal{M}_{loc}(\tilde{P}). \quad (11.1)$$

The proof of Lemma 11.1 can be given using a standard scheme. First, it is proved for martingales. Second, using localization, one can extend it to local martingales. The first step is given with the help of change of probability in conditional expectations, which was stated before for a discrete time. A continuous time version of such change of measure is proved in the same way.

In fact, we prove (11.1) below in the framework of a semimartingale model of financial markets. We define a (B, S) -market as a collection of two positive semimartingales B and S on given stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$. The values of B_t and S_t are interpreted as the prices of a *non-risky asset* B and a *risky asset* S . A pair $\pi_t = (\beta_t, \gamma_t)_{t \geq 0}$ of stochastic processes with a predictable second component is defined as a *portfolio* of investment strategy. The portfolio has a capital $X_t^\pi = X_t^\pi(x) = \beta_t B_t + \gamma_t S_t$, where $X_0^\pi = \beta_0 B_0 + \gamma_0 S_0 = x \in \mathbb{R}$, $t \geq 0$.

The ratio $X_t = \frac{S_t}{B_t}$ is called the discounted price of S , and the ratio $\frac{X_t^\pi}{B_t}$ is called the discounted capital of the strategy π at the time $t \geq 0$.

In the set of all strategies, we distinguish those portfolios π such that

$$\frac{X_t^\pi}{B_t} = \frac{X_0^\pi}{B_0} + \int_0^t \gamma_u d\left(\frac{S}{B}\right)_u. \quad (11.2)$$

We call them *self-financing* and denote their collection SF . We say that (B, S) -market admits an *arbitrage* at time $T > 0$ if there exists $\tilde{\pi} \in SF$ such that $X_0^{\tilde{\pi}} = 0$, $X_t^{\tilde{\pi}} \geq 0$, $t \leq T$, P -a.s., and $P(\omega : X_T^{\tilde{\pi}}(\omega) > 0) > 0$. Such a strategy $\tilde{\pi}$ is called the *arbitrage strategy*. Any probability measure $\tilde{P}$ locally equivalent to P is called a *local martingale measure*, if the discounted price $\left(\frac{S_t}{B_t}\right) \in \mathcal{M}_{loc}(\tilde{P})$.

Denote the set of such measures $\mathcal{M}(X, P)$. It is well-known that for a semimartingale (B, S) -market the *absence of arbitrage* is characterized as $\mathcal{M}(X, P) \neq \emptyset$.

By a *contingent claim* with exercise time T , we understand any nonnegative $\mathcal{F}_T$ -measurable random variable f . We remark that if such a claim f represents a payoff of an option, then the option is called a *European option*. A strategy $\pi \in SF$ is said to be a *hedge* (hedging strategy) for f (for the option with the payoff f) if $X_T^\pi(x) \geq f$ (a.s.) for some initial capital x . If in some class of hedging strategies π , there is a strategy π^* such that $X_t^{\pi^*} \leq X_t^\pi$ (a.s.) for all $t \in [0, T]$, then π^* is called a *minimal hedge* (in this class). Usually, the minimal hedge coincides with a *replicating strategy* π^* for which $X_T^{\pi^*} = f$ (a.s.). In this case, f is called *attainable*. The (B, S) -market is said to be *complete* if any contingent claim is attainable.

For many financial markets, this notion is equivalent to the *uniqueness* of a local martingale measure P^* , i.e., $\mathcal{M}(X, P) = \{P^*\}$. Otherwise, the market is called *incomplete*. Correspondingly, pricing of an option with payoff f in complete and incomplete markets is achieved with the help of martingale measures as follows:

$$\mathbb{C}(f) = \mathbf{E}^* \frac{f}{B_T} \text{ and } \mathbb{C}^*(f) = \sup_{\tilde{P} \in \mathcal{M}(X, P)} \tilde{\mathbf{E}} \frac{f}{B}, \quad (11.3)$$

where $\mathbb{C}(f)$ is the *fair price* and $\mathbb{C}^*(f)$ is the so-called *upper price* of such an option.

So, we can see that an investigation of conditions under which $\tilde{P} \in \mathcal{M}(X, P)$ is extremely important for mathematical finance. Let us show how it can be done for a semimartingale (B, S) -market.

Suppose that a (B, S) -market is determined by the two equations

$$\begin{aligned} B_t &= B_0 + \int_0^t B_u - dh_u, \quad \Delta h_u > -1, \\ S_t &= S_0 + \int_0^t S_u - dH_u, \quad \Delta H_u > -1, \end{aligned} \quad (11.4)$$

where h and H are given semimartingales.

Using stochastic exponents, we can rewrite (11.4) as follows:

$$B_t = B_0 \mathcal{E}_t(h) \text{ and } S_t = S_0 \mathcal{E}_t(H).$$

The problem in the model (11.4) is to determine conditions under which a measure P^* equivalent to P takes the process $X = \frac{S}{B}$ into a local martingale, i.e.,

$$P^* \in \mathcal{M}(X, P) \Leftrightarrow X = \frac{S}{B} \in \mathcal{M}_{loc}(P^*). \quad (11.5)$$

Let us check first when measure P is a martingale one for the market (11.4). Using the properties of stochastic exponents, we have

$$\begin{aligned} X_t &= X_0 \mathcal{E}_t(H) \mathcal{E}_t^{-1}(h) = X_0 \mathcal{E}_t(H) \mathcal{E}_t(-h^*) \\ &= X_0 \mathcal{E} \left(H - h + \langle h^c, h^c \rangle + \sum \frac{(\Delta h)^2}{1 + \Delta h} - \langle H^c, h^c \rangle - \sum \frac{\Delta H \Delta h}{1 + \Delta h} \right) \\ &= X_0 \mathcal{E}_t \left(H - h + \langle h^c, h^c - H^c \rangle + \sum \frac{\Delta h (\Delta h - \Delta H)}{1 + \Delta h} \right). \end{aligned} \quad (11.6)$$

Denote

$$\Psi_t(h, H) = H_T - h_t + \langle h^c, h^c - H^c \rangle_t + \sum_{s \leq t} \frac{\Delta h(\Delta h - \Delta H)}{1 + \Delta h}$$

and rewrite (11.6) in the form of the following stochastic differential equation:

$$X_t = X_0 + \int_0^t X_{s-} d\Psi_s(h, H). \quad (11.7)$$

Hence, $X \in \mathcal{M}_{loc}(P)$ if $\Psi(h, H) \in \mathcal{M}_{loc}(P)$. These considerations generate the idea how to find conditions to provide that $X \in \mathcal{M}_{loc}(P^*)$ for a local martingale measure P^* with the local density

$$Z_t = \frac{dP_t^*}{dP_t} = \mathcal{E}_t(N), \quad N_t = \int_0^t Z_{s-}^{-1} dZ_s \in \mathcal{M}_{loc}(P).$$

We just need to recognize when $XZ = X\mathcal{E}(N) \in \mathcal{M}_{loc}(P)$, and apply (11.1) of Lemma 11.1.

In view (11.6)–(11.7), we have

$$X_t \mathcal{E}_t(N) = X_0 \mathcal{E}_t(\Psi(h, H)) \mathcal{E}_t(N),$$

and by the properties of stochastic exponents, obtain that

$$X_t \mathcal{E}_t(N) = X_0 \mathcal{E}_t(\Psi(h, H, N)), \quad (11.8)$$

where

$$\begin{aligned} \Psi_t(h, H, N) &= H_t - h_t + N_t + \langle (h - N)^c, (h - H)^c \rangle_t \\ &\quad + \sum_{s \leq t} \frac{(\Delta h_s - \Delta N_s)(\Delta h_s - \Delta H_s)}{1 + \Delta h_s}. \end{aligned}$$

Using relation (11.8), we arrive to the following theorem.

Theorem 11.1 *Let $X = \frac{S}{B}$ in the model (11.4). Then the following claims are true:*

- (1) *If $\Psi(h, H) \in \mathcal{M}_{loc}(P)$, then $X \in \mathcal{M}_{loc}(P)$.*
- (2) *If $\Psi(h, H, N) \in \mathcal{M}_{loc}(P)$, then $X \in \mathcal{M}_{loc}(P^*)$.*

The above theorem presents a convenient methodology of finding of martingale measures for the model (11.4). Let us demonstrate this methodology for several partial cases of (11.4).

Example 11.1 The Black-Scholes model:

$$\begin{aligned} dB_t &= rB_t dt, \quad B_0 = 1, \\ dS_t &= S_t(\mu dt + \sigma dW_t), \quad S_0 > 0, \end{aligned} \quad (11.9)$$

where $r, \mu \in \mathbb{R}_+, \sigma > 0$, $W = (W_t)_{t \geq 0}$ is a Wiener process.

In the model (11.9), $h_t = rt$, $H_t = \mu t + \sigma W_t$. Due to W is the sole source of randomness, we take N in the form $N_t = \phi W_t$. To use Theorem 11.1, we find that

$$\begin{aligned} \Psi(h, H, N) &= \mu t + \sigma W_t - rt + \phi W_t + \phi \sigma t \\ &= (\mu - r + \phi \sigma)t + (\sigma + \phi)W_t. \end{aligned}$$

Therefore, the condition (2) of Theorem 11.1 is satisfied if $\mu - r + \phi \sigma = 0$ or $\phi = -\frac{\mu-r}{\sigma}$, and we can construct the local density

$$Z_t = \frac{dP_t^*}{dP_t} = \mathcal{E}_t(N) = \exp \left\{ -\frac{\mu-r}{\sigma} W_t - \frac{1}{2} \left(\frac{\mu-r}{\sigma} \right)^2 t \right\}. \quad (11.10)$$

Using the Girsanov exponent (11.10), we find the martingale measure P^* under which $W_t^* = W_t + \frac{\mu-r}{\sigma} t$ will be a Wiener process by the Girsanov theorem. So, if $f = (S_T - K)^+$ is a contingent claim, then its initial price $\mathbb{C}(f)$ can be calculated as follows:

$$\mathbb{C}(f) = \mathbf{E}^* \frac{(S_T - K)^+}{e^{rT}} = S_0 \Phi(d_+) - K e^{-rT} \Phi(d_-), \quad (11.11)$$

$d_{\pm} = \frac{\ln(S_0/K) + (r \pm \sigma^2/2)T}{\sigma\sqrt{T}}$, and we again arrive to the Black-Scholes formula.

Example 11.2 The Merton model:

$$\begin{aligned} dB_t &= rB_t dt, \quad B_0 = 1, \\ dS_t &= S_t(\mu dt - v d\Pi_t), \end{aligned} \quad (11.12)$$

where $r, \mu \in \mathbb{R}_+, v < 1$, $\Pi = (\Pi_t)_{t \geq 0}$ is a Poisson process with parameter $\lambda > 0$.

In the model (11.12) $h_t = rt$, $H_t = \mu t - v\Pi_t$, and the martingale N_t will be chosen as $N_t = \psi(\Pi_t - \lambda t)$. Further,

$$\begin{aligned} \Psi(h, H, N) &= \mu t - v\Pi_t - rt + \psi(\Pi_t - \lambda t) - \psi v\Pi_t \\ &= (\mu - r - v\lambda - \psi v\lambda)t - v(\Pi_t - \lambda t) + \psi(\Pi_t - \lambda t) - \psi v(\Pi_t - \lambda t). \end{aligned}$$

Hence, the condition (2) of Theorem 11.1 is fulfilled if $\mu - r - v\lambda - \psi v\lambda = 0$ or $\psi = \frac{\mu-r}{v\lambda} - 1$ is the unique solution. The uniqueness means that a martingale

measure P^* is also unique, and its local density has the following exponential form

$$Z_t = \frac{dP_t^*}{dP_t} = \mathcal{E}_t(N) = \exp[(\lambda - \lambda^*)t + (\ln \lambda^* - \ln \lambda)\Pi_t], \quad (11.13)$$

where $\lambda^* = \frac{\mu-r}{v}$ is a parameter of Π_t under measure P^* .

Using (11.13), we calculate a call option price $\mathbb{C}$ in the model (11.12) as follows.

$$\mathbb{C} = \mathbf{E}^* \frac{(S_T - K)^+}{e^{rT}}.$$

It is clear that

$$\begin{aligned} \frac{S_T}{B_T} &= \frac{S_0}{B_0} \exp[-v\Pi_T + v\lambda^*T] \prod_{t \leq T} (1 - v\Delta\Pi_t) e^{v\Delta\Pi_t} \\ &= \frac{S_0}{B_0} \exp[\Pi_T \ln(1-v) + v\lambda^*T]. \end{aligned} \quad (11.14)$$

Using (11.14), we find that

$$\begin{aligned} \mathbb{C} &= \mathbf{E}^* \left(\frac{S_T}{B_T} - \frac{K}{B_T} \right)^+ \\ &= \sum_{n=0}^{\infty} e^{-\lambda^*T} \frac{(\lambda^*T)^n}{n!} \left(S_0 e^{n \ln(1-v) + v\lambda^*T} - B_T^{-1}K \right)^+. \end{aligned} \quad (11.15)$$

Denote

$$\begin{aligned} n_0 &= \inf \left[n : S_0 e^{n \ln(1-v) + v\lambda^*T} \geq \frac{K}{B_T} \right] \\ &= \left\lceil \frac{\ln(K/S_0) - \mu T}{\ln(1-v)} \right\rceil, \\ \Psi_p(x, y) &= \sum_{n=x}^{\infty} e^{-y} \frac{y^n}{n!}, \end{aligned}$$

and find from (11.15) that

$$\begin{aligned} \mathbb{C} &= S_0 \sum_{n=n_0}^{\infty} e^{-\lambda^*T + \lambda^*vT} e^{n \ln(1-v)} \frac{(\lambda^*T)^n}{n!} - K e^{-rT} \Psi_p(n_0, \lambda^*T) \\ &= S_0 \sum_{n=n_0}^{\infty} e^{-\lambda^*(v-1)T} \frac{(\lambda^*(1-v)T)^n}{n!} - K e^{-rT} \Psi_p(n_0, \lambda^*T) \\ &= S_0 \Psi_p(n_0, \lambda^*(1-v)T) - K e^{-rT} \Psi_p(n_0, \lambda^*T). \end{aligned} \quad (11.16)$$

Formula (11.16) is called the *Merton formula* for the price of a call option in the model (11.12).

Putting together models (11.9) and (11.12), we get a *jump-diffusion market model*:

$$\begin{aligned} B_t &= r B_t dt, \quad B_0 = 1, \\ dS_t &= S_{t-} (\mu dt + \sigma dW_t - \nu d\Pi_t), \quad S_0 > 0. \end{aligned} \quad (11.17)$$

To calculate process $\Psi(h, H, N)$ in this case, we choose $N_t = \phi W_t + \psi \Pi_t$ and find that

$$\Psi_t(h, H, N) = (\mu - r - \nu\lambda + \phi\sigma - \psi\lambda)t + \text{martingale}.$$

Hence, $\mu - r + \phi\sigma - \lambda\nu(1 + \psi) = 0$ to make $\Psi(h, H, N)$ a martingale. But this equation has infinitely many solutions (ϕ, ψ) , and therefore, the model (11.15) admits infinitely many martingale measures, i.e., the market (11.17) is *incomplete*.

Let us consider a discrete time model, called a *binomial market* or the *Cox-Ross-Rubinstein model*. We show that such a model is embedded in the model (11.4).

Example 11.3 The model we are talking represents a kind of Binomial random walk:

$$\begin{aligned} \Delta B_n &= B_n - B_{n-1} = r B_{n-1}, \quad B_0 > 0, \\ \Delta S_n &= S_n - S_{n-1} = \rho_n S_{n-1}, \quad S_0 > 0, \end{aligned} \quad (11.18)$$

where $(\rho_n)_{n=1,2,\dots}$ is a sequence of independent random variables taking two values $b > a$ with probabilities p and $q = 1 - p$, $p \in (0, 1)$.

Assume also that $-1 < a < r < b$. Putting $B_t = B_n$ on $[n, n + 1)$ (we do the same with $S_t, \mathcal{F}_t, \dots$), we transform (11.18) to the model (11.4). This standard procedure allows us to apply the theory developed for the semimartingale model.

In the case under consideration,

$$\Delta h_n = r, \quad \Delta H_n = \rho_n, \quad \Delta N_n = \psi_n(\rho_n - \mu), \quad \mu = \mathbf{E}\rho_n,$$

and

$$(1 + r)\Delta\Psi_n(h, H, N) = \Delta H_n - \Delta h_n + \Delta N_n + \Delta N_n \Delta H_n.$$

Hence, the martingality property Ψ_n means that

$$\mathbf{E}(\rho_n - r + \psi_n(\rho_n - \mu) + \psi_n(\rho_n - \mu)\rho_n | \mathcal{F}_{n-1}) = 0,$$

which leads to

$$\psi_n = -\frac{\mu - r}{\sigma^2},$$

where $\sigma^2 = \mathbf{Var}(\rho_n)$.

As a result, we can construct a local density of a martingale measure P^* here in the form of stochastic exponent

$$Z_n = \mathcal{E}_n \left(-\frac{\mu - r}{\sigma^2} \sum (\rho_n - \mu) \right). \quad (11.19)$$

As in previous examples, we want to derive the price of a call option $(S_N - K)^+$, where $N \geq 1$. According to the general theory, the price $\mathbb{C}$ can be calculated as

$$\begin{aligned} \mathbb{C} &= \mathbf{E}^*(1+r)^{-N} (S_N - K)^+ \\ &= (1+r)^{-N} \mathbf{E}^* (S_N - K)^+ \\ &= (1+r)^{-N} \mathbf{E} \mathcal{E}_N \left(-\frac{\mu - r}{\sigma^2} \sum (\rho_k - \mu) \right) (S_N - K) I_{\{S_N > K\}}, \end{aligned} \quad (11.20)$$

where we used (11.19).

Define $k_0 = \min \{k \leq N : S_0(1+b)^k(1+a)^{N-k} > K\}$ and find that $k_0 \ln(1+b) + (N - k_0) \ln(1+a) > \ln \frac{K}{S_0}$ or $k_0 = \left\lceil \ln \frac{K}{S_0(1+a)^N} / \ln \frac{1+b}{1+a} \right\rceil + 1$, where $[x]$ is an integer part of $x \in \mathbb{R}$.

Denote $p^* = \frac{r-a}{b-a}$ and derive for the term with K in (11.20) using elementary equalities $\mu = p(b-a) + a$, $\sigma^2 = (b-a)^2 p(1-p)$ that

$$\begin{aligned} & (1+r)^{-N} \mathbf{E} \mathcal{E}_N \left(-\frac{\mu - r}{\sigma^2} \sum (\rho_k - \mu) \right) K I_{\{S_N > K\}} \\ &= K (1+r)^{-N} \sum_{k=k_0}^N \binom{N}{k} \left(1 - \frac{\mu - r}{\sigma^2} (b - \mu) \right)^k \\ & \quad \times p^k \left(1 - \frac{\mu - r}{\sigma^2} (a - \mu) \right)^{N-k} (1-p)^{N-k} \\ &= K (1+r)^{-N} \sum_{k=k_0}^N \binom{N}{k} \left(\frac{p^*}{p} \right)^k p^k \left(\frac{1-p^*}{1-p} \right)^{N-k} (1-p)^{N-k} \\ &= K (1+r)^{-N} \sum_{k=k_0}^N \binom{N}{k} (p^*)^k (1-p^*)^{N-k}. \end{aligned} \quad (11.21)$$

To calculate the term with $S_N = S_0 \mathcal{E}_N(\sum \rho_k)$ in (11.20), we use a multiplication rule of stochastic exponents:

$$\begin{aligned}
 & (1+r)^{-N} \mathbf{E} \mathcal{E}_N \left(-\frac{\mu-r}{\sigma^2} \sum (\rho_k - \mu) \right) S_N I_{\{S_N > K\}} \\
 &= S_0 (1+r)^{-N} \mathbf{E} \mathcal{E}_N \left(-\frac{\mu-r}{\sigma^2} \sum (\rho_k - \mu) + \sum \rho_k \right. \\
 & \quad \left. - \frac{\mu-r}{\sigma^2} \sum (\rho_k - \mu) \rho_k \right) I_{\{S_N > K\}} \\
 &= S_0 (1+r)^{-N} \sum_{k=k_0}^N \binom{N}{k} \left(1 - \frac{\mu-r}{\sigma^2} (b-\mu) + b - \frac{\mu-r}{\sigma^2} (b-\mu)b \right)^k p^k \\
 & \quad \times \left(1 - \frac{\mu-r}{\sigma^2} (a-\mu) + a - \frac{\mu-r}{\sigma^2} (a-\mu)a \right)^{N-k} (1-p)^{N-k} \\
 &= S_0 (1+r)^{-N} \sum_{k=k_0}^N \binom{N}{k} \left(\frac{p^*}{p} (1+b) \right)^k \\
 & \quad \times p^k \left(\frac{1-p^*}{1-p} (1+a) \right)^{N-k} (1-p)^{N-k} \\
 &= S_0 (1+r)^{-N} \sum_{k=k_0}^N \binom{N}{k} \left(p^* \frac{1+b}{1+r} \right)^k \left((1-p^*) \frac{1+a}{1+r} \right)^{N-k}.
 \end{aligned} \tag{11.22}$$

Introducing the notations $\tilde{p} = \frac{1+b}{1+r} p^*$ and $\mathbb{B}(j, N, p) = \sum_{k=j}^N \binom{N}{k} p^k (1-p)^{N-k}$ and putting together (11.21)–(11.22), we get from (11.20) the *Cos-Ross-Rubinstein formula* for the initial price of a call option

$$\mathbb{C} = S_0 \mathbb{B}(k_0, N, \tilde{p}) - K (1+r)^{-N} \mathbb{B}(k_0, N, p^*).$$

11.2 Stochastic Regression Analysis

Proposed and developed below an extension of classical regression models and techniques is based on those theoretical findings that were delivered in the previous chapter, and can be called a *stochastic regression analysis*.

We start with the classical problem of *stochastic approximation*. It consists of a construction of a stochastic sequence θ_n or a stochastic process θ_t that converges in

some probabilistic sense to unique root $\theta \in \mathbb{R}$ of the regression equation

$$R(\theta) = 0, \quad (11.23)$$

where R is a regression function.

In classical theory, a solution to (11.23) is given by the *Robbins-Monro procedure*

$$\theta_n = \theta_{n-1} - \gamma_n y_n, \quad n = 1, 2, \dots, \quad (11.24)$$

where the sequence of observations y_n is such that

$$y_n = R(\theta_{n-1}) + \xi_n, \quad (11.25)$$

$(\xi_n)_{n=1,2,\dots}$ is a sequence of independent random variables or martingale differences, and $(\gamma_n)_{n=1,2,\dots}$ is a numerical positive sequence converging to zero.

The convergence (a.s.) of the procedure (11.24)–(11.25) for a continuous linearly bounded regression function $R(x)$ such that

$$R(x)(x - \theta) > 0 \text{ for all } x \neq \theta$$

is guaranteed by finiteness of the variance (conditional variance) of observation errors ξ_n and the following conditions:

$$\sum_{n=1}^{\infty} \gamma_n = \infty, \quad (11.26)$$

$$\sum_{n=1}^{\infty} \gamma_n^2 < \infty. \quad (11.27)$$

Diffusion analogues of (11.24)–(11.25) and (11.26)–(11.27) are

$$d\theta_t = -\gamma_t R(\theta_t) dt - \gamma_t dW_t, \quad (11.28)$$

$$\int_0^{\infty} \gamma_s ds = \infty, \quad (11.29)$$

$$\int_0^{\infty} \gamma_s^2 ds < \infty, \quad (11.30)$$

where W_t is a Wiener process and γ_t is a positive deterministic function tending to zero as $t \rightarrow \infty$.

Under conditions (11.29)–(11.30), the procedure θ_t converges to θ as $t \rightarrow \infty$.

The leading idea to generalize (11.24)–(11.25) and (11.28) consists in the possibility of describing such stochastic algorithms as strong solutions of some special classes of stochastic differential equations with respect to semimartingales.

It leads to a generalized Robbins-Monro procedure as a process θ_t satisfying the stochastic differential equation

$$\theta_t = \theta_0 - \int_0^t \gamma_s R(\theta_{s-}) da_s - \int_0^t \gamma_s dm_s, \quad (11.31)$$

where θ_0 is a finite $\mathcal{F}_0$ -measurable random variable, a predictable process $a \in \mathcal{A}_{loc}^+$, $m \in \mathcal{M}_{loc}^2$, and γ is a predictable process decreasing to zero (a.s.) as $t \rightarrow \infty$.

To simplify the demonstration how martingale methods work here, we consider a linear case only: $R(x) = \beta(x - \theta)$, $\beta > 0$.

In this case, (11.31) is reduced to

$$\theta_t - \theta = \theta_0 - \theta - \int_0^t \gamma_s \beta(\theta_{s-} - \theta) da_s - \int_0^t \gamma_s dm_s. \quad (11.32)$$

Let us assume that (a.s.)

$$\int_0^\infty \gamma_s da_s = \infty, \quad (11.33)$$

$$\int_0^\infty \gamma_s^2 d\langle m, m \rangle_s < \infty. \quad (11.34)$$

Define the following stochastic exponent

$$\mathcal{E}_t(-\beta\gamma \cdot a) = \mathcal{E}_t\left(-\beta \int \gamma_s da_s\right),$$

and assume that $\beta\gamma_t \Delta a_t < 1$ to provide that $\mathcal{E}_t(-\beta\gamma \cdot a) > 0$ (a.s.).

Applying the Ito formula to $\mathcal{E}_t^{-1}(-\beta\gamma \cdot a)(\theta_t - \theta)$, we arrive from (11.32) to

$$\theta_t - \theta = \mathcal{E}_t(-\beta\gamma \cdot a)(\theta_0 - \theta) - \mathcal{E}_t(-\beta\gamma \cdot a) \int_0^t \gamma_s \mathcal{E}_s^{-1}(-\beta\gamma \cdot a) dm_s. \quad (11.35)$$

The first term in the right-hand side of (11.35) converges to zero (a.s.) as $t \rightarrow \infty$ because of $\mathcal{E}_t(-\beta\gamma \cdot a) \rightarrow 0$ (a.s.), $t \rightarrow \infty$. Second term in (11.35) is treated by the arguments of the large numbers law for square-integrable martingales. Using (11.34), we have that (a.s.)

$$\int_0^\infty \mathcal{E}_s^2(-\beta\gamma \cdot a) \gamma_s^2 \mathcal{E}_s^2(-\beta\gamma \cdot a) d\langle m, m \rangle_s = \int_0^\infty \gamma_s^2 d\langle m, m \rangle_s < \infty.$$

Hence, the second term of (11.35) converges to zero (a.s.) as $t \rightarrow \infty$. As a result, we get the convergence (a.s.) $\theta_t \rightarrow \theta$ as $t \rightarrow \infty$.

To investigate an asymptotic normality of the procedure (11.32), we simplify the situation assuming that a_t is a deterministic function and m_t is a Gaussian

martingale satisfying conditions:

$$\langle m, m \rangle_t = \sigma^2 a_t, \quad \sigma^2 > 0, \quad \gamma_t = \frac{\alpha}{1 + a_t}, \quad \alpha > 0, \quad \beta\alpha < 1,$$

$$a_t \uparrow, \quad t \rightarrow \infty, \quad \sum_{0 \leq s < \infty} \left(\frac{\Delta a_s}{1 + a_s} \right)^2 < \infty.$$

Under these conditions, we have

$$\mathcal{E}_t(-\beta\gamma \cdot a) = \mathcal{E}_t(-\beta\alpha) \sim (1 + a_t)^{-\beta\alpha} \text{ as } t \rightarrow \infty.$$

Multiplying (11.35) by $(1 + a_t)^{1/2}$, we obtain

$$(1 + a_t)^{1/2}(\theta_t - \theta) = (1 + a_t)^{1/2}\mathcal{E}_t(-\beta\alpha)(\theta_0 - \theta) \\ - (1 + a_t)^{1/2}\mathcal{E}_t(-\beta\alpha) \int_0^t \frac{\alpha}{1 + a_s} \mathcal{E}_s^{-1}(-\beta\alpha) dm_s. \quad (11.36)$$

Let us note that

$$(1 + a_t)^{1/2}\mathcal{E}_t(-\beta\alpha) \sim (1 + a_t)^{1/2-\beta\alpha} \text{ as } t \rightarrow \infty,$$

and therefore, under $2\beta\alpha > 1$, we obtain that the first term of (11.36) converges to zero (a.s.) as $t \rightarrow \infty$.

The second term of (11.36) has a Gaussian distribution, and hence, we need to calculate the asymptotic value of the variance of this term:

$$(1 + a_t)\mathcal{E}_t^2(-\beta\alpha) \int_0^t \mathcal{E}_s^2(-\beta\alpha) \frac{\alpha^2 \sigma^2}{1 + a_s} da_s \rightarrow \frac{\alpha^2 \sigma^2}{2\beta\alpha - 1} \text{ as } t \rightarrow \infty. \quad (11.37)$$

Finally, we can arrive to conclusion that

$$(1 + a_t)^{1/2}(\theta_t - \theta) \xrightarrow[t \rightarrow \infty]{d} N\left(0, \frac{\alpha^2 \sigma^2}{2\beta\alpha - 1}\right). \quad (11.38)$$

As a consequence of (11.38), we get the well-known classical results from the theory of stochastic approximation for *discrete* and *continuous* time (11.24)–(11.25) and (11.28):

$$n^{1/2}(\theta_n - \theta) \xrightarrow[n \rightarrow \infty]{d} N\left(0, \frac{\alpha^2 \sigma^2}{2\beta\alpha - 1}\right),$$

$$t^{1/2}(\theta_t - \theta) \xrightarrow[t \rightarrow \infty]{d} N\left(0, \frac{\alpha^2 \sigma^2}{2\beta\alpha - 1}\right).$$

Another important problem of regression analysis is the problem of estimation of the unknown parameter of a linear regression function.

Suppose we observe the process X_t having the following structure

$$X_t = \int_0^t f_s da_s \theta + M_t, \quad (11.39)$$

where $a \in \mathcal{A}_{loc}^+$ and predictable, $M \in \mathcal{M}_{loc}^2$, a predictable function f_t such that $\int_0^t f_s^2 da_s = F_t < \infty$ (a.s.), $t \geq 0$, $\theta \in \mathbb{R}$ is an unknown parameter.

Define the following structural least-squares estimate as follows:

$$\theta_t = F_t^{-1} \int_0^t f_s dX_s, \quad (11.40)$$

where we assume that $F_t > 0$ (a.s.) and $F_t \in \mathcal{A}_{loc}^+$ and predictable.

Assuming that $F_t \rightarrow \infty$ (a.s.) as $t \rightarrow \infty$, we rewrite (11.40):

$$\begin{aligned} \theta_t &= F_t^{-1} \int_0^t f_s^2 da_s \theta + F_t^{-1} \int_0^t f_s dM_s \\ &= \theta + F_t^{-1} \int_0^t f_s dM_s. \end{aligned} \quad (11.41)$$

Obviously, we can study the asymptotic behavior of θ_t with the help of the large numbers law. To apply such LNL, we find from (11.41) that (a.s.)

$$\int_0^\infty F_s^{-2} d \left\langle \int_0^t f_s dM_s, \int_0^t f_s dM_s \right\rangle = \int_0^\infty F_s^{-2} f_s^2 d \langle M, M \rangle_s < \infty. \quad (11.42)$$

So, if $F_t \rightarrow \infty$ (a.s.), $t \rightarrow \infty$, and (11.42) is fulfilled, then $\theta_t \rightarrow \theta$ (a.s.) $t \rightarrow \infty$.

Let us continue our study of estimates (11.40), by considering their *sequential analog*. To do this in the model (11.39), we assume that

$$\frac{d \langle M, M \rangle_t}{da_t} \leq \xi \gamma_t, \quad (11.43)$$

$$\int_0^t \gamma_s^{-1} f_s^2 da_s \in \mathcal{A}_{loc}^+ \text{ and predictable,}$$

where ξ is a positive r.v. and γ_t is a positive predictable process.

Next, for $H > 0$, we put

$$\tau_H = \inf \left\{ t : \int_0^t \gamma_s^{-1} f_s^2 da_s \geq H \right\}$$

with $\tau_H = \infty$ if the set in $\{\cdot\}$ is empty.

On the set $\{\tau_H < \infty\}$, we define a random variable β_H by the relation

$$\int_{(0, \tau_H)} \gamma_s^{-1} f_s^2 da_s + \beta_H \gamma_{\tau_H}^{-1} f_{\tau_H}^2 \Delta a_{\tau_H} = H, \quad (11.44)$$

and we put $\beta_H = 0$ on the set $\{\tau_H = \infty\}$. Then $\beta_H \in [0, 1)$, and it is a $\mathcal{F}_{\tau_H-}$ -measurable random variable.

Let us define the following *sequential* least-squares estimate

$$\hat{\theta}_H = H^{-1} \left[\int_{(0, \tau_H)} \gamma_s^{-1} f_s dX_s + \beta_H \gamma_{\tau_H}^{-1} f_{\tau_H} \Delta X_{\tau_H} \right]. \quad (11.45)$$

The next theorem shows a nice property of $\hat{\theta}_H$ called a *fixed accuracy*.

Theorem 11.2 Suppose (11.43)–(11.45) are fulfilled, $\mathbf{E}\xi < \infty$, and $\int_0^\infty \gamma_s^{-1} f_s^2 da_s = \infty$ (a.s.). Then $P\{\omega : \tau_H < \infty\} = 1$, $\mathbf{E}\hat{\theta}_H = \theta$, and

$$\mathbf{Var}\hat{\theta}_H \leq H^{-1} \mathbf{E}\xi. \quad (11.46)$$

Proof The finiteness (a.s.) of τ_H follows from the relation $\{\tau_H \leq T\} = \left\{ \omega : \int_0^T \gamma_s^{-1} f_s^2 da_s \geq H \right\}$. Using (11.45), we find that

$$\hat{\theta}_H = \theta + H^{-1} N_{\tau_H}, \quad (11.47)$$

where $N_t = \int_0^t I_{(s < \tau_H)} \gamma_s^{-1} f_s dM_s + I_{\{t = \tau_H\}} \beta_H \gamma_{\tau_H}^{-1} f_{\tau_H} \Delta M_{\tau_H}$.

Since the process N_t is a stochastic integral with respect to $M \in \mathcal{M}_{loc}^2$, by the properties of stochastic integrals, we obtain that

$$\langle N, N \rangle_t = \int_0^t I_{(s < \tau_H)} \gamma_s^{-1} f_s^2 d\langle M, M \rangle_s + I_{\{t = \tau_H\}} \beta_H^2 \gamma_{\tau_H}^{-2} f_{\tau_H}^2 \Delta \langle M, M \rangle_{\tau_H}. \quad (11.48)$$

Hence, by (11.44), we get

$$\langle N, N \rangle_{\tau_H} = \xi \left[\int_{(0, \tau_H)} \gamma_s^{-1} f_s^2 da_s + \beta_H \gamma_{\tau_H}^{-1} f_{\tau_H}^2 \Delta a_{\tau_H} \right] = \xi H,$$

and consequently, $N_{t \wedge \tau_H} \in \mathcal{M} \cap \mathcal{M}^2$ and

$$\mathbf{E}N_{\tau_H} = 0, \quad \mathbf{E}N_{\tau_H}^2 = H\mathbf{E}\xi,$$

which leads to (11.46). □

Example 11.4 Let us consider the first order autoregression model:

$$x_t = \theta x_{t-1} + e_t, \quad x_0 = 0, \quad t = 1, 2, \dots,$$

where $(e_t)_{t=1,2,\dots}$ is a martingale difference w.r.t. $(\mathcal{F}_t)$ and $\mathbf{E}(e_t^2 | \mathcal{F}_{t-1}) \leq \xi$, $\mathbf{E}\xi < \infty$.

Then (11.43) is fulfilled with $\gamma_t = 1$, and the estimate $\hat{\theta}_H$ takes the form

$$\begin{aligned} \hat{\theta}_H &= H^{-1} \left[\sum_{k=1}^{\tau_H-1} x_k x_{k-1} + \beta_H x_{\tau_H} x_{\tau_H-1} \right], \\ \tau_H &= \inf \left\{ n : \sum_{k=1}^n x_k^2 \geq H \right\}, \\ \beta_H &= x_{\tau_H}^{-2} \left[H - \sum_{k=1}^{\tau_H-1} x_k^2 \right]. \end{aligned}$$

Embedding the present model in the general model (11.39) in a standard way, we get

$$\mathbf{E}\hat{\theta}_H = \theta, \quad \mathbf{Var}\hat{\theta}_H \leq H^{-1}\mathbf{E}\xi.$$

Chapter 12

Basic Elements of Optional Stochastic Analysis



This chapter provides an overview of optionality and predictability, general stopping times, processes of finite variations, optional martingales and semimartingales, along with their integration with respect to random measures. The chapter is ended by the canonical decomposition of optional semimartingales and the change of variables formulas (see [1, 14, 24, 26, 41, 42], and [38]).

12.1 Building Blocks of Optional Stochastic Analysis

Let a triplet $(\Omega, \mathcal{F}, \mathbf{P})$ be a complete probability space, i.e., if $A \subseteq B \in \mathcal{F}$ and $P(B) = 0$, then $A \in \mathcal{F}$. This probability space equipped with a nondecreasing family of σ -algebras (filtration/information flow) $\mathbf{F} = (\mathcal{F}_t)_{t \in [0, \infty)}$, where $\mathcal{F}_t \subseteq \mathcal{F}$, $\mathcal{F}_s \subseteq \mathcal{F}_t$, $s \leq t$, forms a stochastic basis—a fundamental notion in stochastic analysis. Usually, the following conditions on the filtration $\mathbf{F}$:

- $\mathbf{F}$ is right-continuous, i.e., $\mathcal{F}_t = \mathcal{F}_{t+}$ for all t , where $\mathcal{F}_{t+} = \bigcap_{t < s} \mathcal{F}_s$
- $\mathbf{F}$ is complete, i.e., each $\mathcal{F}_t$ contains P -null sets of $\mathcal{F}$

are collectively known as “usual conditions” on the filtration $\mathbf{F}$.

The theory of stochastic processes is well-developed under “usual conditions.” This theory is widely applied in different areas such as mathematical finance, mathematical statistics, risk theory, stochastic differential equations, and others. Although “usual conditions” are considered as a golden standard in stochastic analysis, one can easily give an example illustrating the existence of a stochastic basis where such assumptions do not hold.

Example 12.1 Suppose

$$X_t = \mathbf{1}_{t > t_0} \mathbf{1}_A,$$

where A is $\mathcal{F}$ -measurable with $0 < P(A) < 1$. Let $\mathcal{F}_t$ be a natural filtration of X , $\mathcal{F}_t = \sigma(X_s, s \leq t)$.

Then $\mathbf{F}$ is not right-continuous at t_0 , i.e., $A \notin \mathcal{F}_{t_0}$, but $A \in \mathcal{F}_{t_0+}$, and it is not possible to make it right-continuous in a useful way.

A reasonable question arises: Why are “usual conditions” so predominant in the studying of random processes? The majority of important results in the stochastic analysis are proved under “usual conditions,” and for many of them, these conditions are vital. Here we provide a brief introduction to stochastic analysis without “usual conditions.”

Let us introduce $\mathcal{O}(\mathbf{F})$ and $\mathcal{P}(\mathbf{F})$ as the optional and predictable σ -algebras on $(\Omega, [0, \infty))$, respectively. $\mathcal{O}(\mathbf{F})$ is generated by all $\mathbf{F}$ -adapted processes whose trajectories are right-continuous and have left limits (or cadlag, RCLL). $\mathcal{P}(\mathbf{F})$ is generated by all $\mathbf{F}$ -adapted processes whose trajectories are left-continuous and have right limits.

A random process $X = (X_t)$, $t \in [0, \infty)$, is said to be *optional* if it is $\mathcal{O}(\mathbf{F})$ -measurable. In general, an optional process has right and left limits but is not necessarily right- or left-continuous in $\mathbf{F}$.

A random process (X_t) , $t \in [0, \infty)$, is predictable if X is $\mathcal{P}(\mathbf{F})$ -measurable. As well, in general, a predictable process has right and left limits but may not necessarily be right- or left-continuous in $\mathbf{F}$.

Denote $\mathcal{P}$ and $\mathcal{O}$ sets of predictable and optional processes, respectively. For either optional or predictable processes, the following processes can be defined: $X_- = (X_{t-})_{t \geq 0}$ and $X_+ = (X_{t+})_{t \geq 0}$, $\Delta X = (\Delta X_t)_{t \geq 0}$ such that $\Delta X_t = X_t - X_{t-}$ and $\Delta^+ X = (\Delta^+ X_t)_{t \geq 0}$ such that $\Delta^+ X_t = X_{t+} - X_t$.

Since the random processes that we are going to work with possibly have left and right limits only, it is necessary to introduce the following definition.

Definition 12.1 A random process (X_t) , $t \in [0, \infty)$, is called *strongly predictable* if X is $\mathcal{P}(\mathbf{F})$ -measurable and X_+ is $\mathcal{O}(\mathbf{F})$ -measurable. Denote a space of strongly predictable processes $\mathcal{P}_s$.

Since in general $\mathbf{F} \neq \mathbf{F}_+ = (\mathcal{F}_{t+})_{t \geq 0}$, where $\mathcal{F}_{t+} = \cap_{t < s} \mathcal{F}_s$, there are two distinct notions of stopping times (s.t.’s) with respect to $\mathbf{F}$ and $\mathbf{F}_+$ on unusual stochastic basis, i.e., on a stochastic basis without “usual conditions.”

Definition 12.2 The random variable $T : \Omega \rightarrow [0, \infty]$ is a stopping time (s.t.) if the set $\{T \leq t\} \in \mathcal{F}_t$ for all $t \in [0, \infty)$.

Definition 12.3 The random variable $T : \Omega \rightarrow [0, \infty]$ is a wide (broad) sense stopping time if the set $\{T \leq t\} \in \mathcal{F}_{t+}$ for all $t \in [0, \infty)$.

As on a usual stochastic basis, there are notions of predictable and totally inaccessible stopping times on an unusual stochastic basis.

Definition 12.4 A stopping time T is called predictable if there exists a sequence of wide sense stopping times (S_n) , $n \in \mathbb{N}$, such that $\lim S_n = T$ a.s. and $S_n < T$ a.s. on the set $\{T > 0\}$ for all $n \in \mathbb{N}$. Correspondently, a wide sense stopping time T is called totally inaccessible wide sense s.t. if $P(S = T < \infty) = 0$ for every stopping time S .

Analogously, it is possible to define predictable wide sense stopping times, but it turns out that this is unnecessary because the set of predictable s.t.'s and the set of predictable wide sense s.t.'s are equal. Denote by $\mathcal{T}$ and $\mathcal{T}_+$ a set of s.t.'s and a set of wide sense s.t.'s, respectively; $\mathcal{T}^p$, $\mathcal{T}^i$, $\mathcal{T}_+^i$ —the set of predictable s.t.'s, the set of totally inaccessible s.t.'s, and totally inaccessible wide sense s.t.'s, respectively. It is immediately seen that $\mathcal{T}^p \subseteq \mathcal{T} \subseteq \mathcal{T}_+$.

On unusual stochastic basis, defined above three canonical types of stopping times have the following properties. Predictable stopping times, $S \in \mathcal{T}^p$, are such that $(S \leq t)$ is $\mathcal{F}_{t-}$ -measurable for all t ; totally inaccessible stopping times, $T \in \mathcal{T}^i$, are such that $(T \leq t)$ is $\mathcal{F}_t$ -measurable for all t . However, we note that $(T < t)$ is not necessarily $\mathcal{F}_t$ -measurable since $\mathcal{F}_t$ is not right-continuous; finally, totally inaccessible stopping times in the broad sense, $U \in \mathcal{T}_+^i$, are such that $(U \leq t)$ is $\mathcal{F}_{t+}$ -measurable for all t , but since $\mathcal{F}_{t+}$ is right-continuous, and $(U < t)$ is also $\mathcal{F}_{t+}$ -measurable.

Definition 12.5 Suppose S and T are maps $\Omega \rightarrow [0, \infty]$ and $S \leq T$ a.s. The stochastic interval denoted by $[S, T[$ is the set

$$\{(t, \omega) \in [0, \infty[\times \Omega : S(\omega) \leq t < T(\omega)\}.$$

The stochastic intervals $[S, T]$, $]S, T]$ and $]S, T[$ are defined similarly. The stochastic interval

$$[T, T] = \{(t, \omega) \in [0, \infty[\times \Omega : T(\omega) = t\}$$

is denoted by $[T]$ and is called the graph of T . We use the notation introduced above for wide sense stopping times. Whenever we use this notation for the deterministic times, we mean a usual interval of real numbers.

Definition 12.6 The sequence of $(T_n)_{n \in \mathbb{N}} \subset \mathcal{T}_+$ exhausts the jumps of the process $(X_t)_{t \in [0, \infty)}$ if for any $T \in \mathcal{T}_+$ for which the set $[T] \cap (\cup_n [T_n]) = \emptyset$ we have $\Delta X_T = \Delta^+ X_T = 0$ a.s. on $\{t < \infty\}$.

Below, a fundamental result that gives foundations for many results in the optional stochastic analysis is presented. In particular, it is used in the canonical decomposition of optional semimartingales.

Theorem 12.1 Suppose $X = (X_t)_{t \in [0, \infty)}$ is an optional process whose paths have left and right limits a.s. (rlll or *ladlag*). Then there exist sequences (S_n) , (T_n) , and (U_n) , $n \in \mathbb{N}$, of predictable s.t.'s, totally inaccessible s.t.'s, and totally inaccessible wide sense s.t.'s, respectively, exhausting all jumps of the process X and having the

following properties: The graphs of these s.t.'s are mutually nonintersecting within each sequence.

Let us consider the processes which do not vary a lot, i.e., processes having paths of finite variation for almost all ω . We begin with the definitions of several spaces of such processes.

Definition 12.7 A process $A = (A_t)$, $t \in \mathbb{R}_+$, is of finite variation if it has finite variation on every segment $[0, t]$, $t \in \mathbb{R}_+$ a.s., that is, $\mathbf{Var}(A)_t < \infty$, for all $t \in \mathbb{R}_+$ a.s. where

$$\mathbf{Var}(A)_t = \sum_{0 \leq s < t} |\Delta^+ A_s| + \int_{]0, t]} |dA_s^r|.$$

Denote by $\mathcal{V}$ the set of all $\mathbf{F}$ -adapted processes of finite variation.

A process $A = (A_t)_{t \geq 0}$ is increasing if it is nonnegative, $\mathbf{F}$ -adapted, and its trajectories do not decrease. Let $\mathcal{V}^+$ be the collection of increasing processes, and hence, $\mathcal{V}^+ \subset \mathcal{V}$. Further, an increasing process A is integrable if $\mathbf{E}A_\infty < \infty$. The collection of such processes is denoted by $\mathcal{A}^+$. Correspondently, a process $A = (A_t)_{t \geq 0}$ of finite variation belongs to the space $\mathcal{A}$ of integrable finite variation processes if $\mathbf{E}[\mathbf{Var}(A)_\infty] < \infty$.

A process $X = (X_t)_{t \geq 0}$ belongs to the space $\mathcal{J}_{loc}$ if there is a localizing sequence of wide sense s.t.'s, (R_n) , $n \in \mathbb{N}$, $R_n \in \mathcal{T}_+$, $R_n \uparrow \infty$ a.s. such that $X\mathbf{1}_{[0, R_n]} \in \mathcal{J}$ for all n , where $\mathcal{J}$ is a space of processes and $\mathcal{J}_{loc}$ is an extension of $\mathcal{J}$ by localization.

In general, the spaces $\mathcal{V}$, $\mathcal{A}$, $\mathcal{A}^+$ can be extended to $\mathcal{V}_{loc}$, $\mathcal{A}_{loc}$, $\mathcal{A}_{loc}^+$, respectively, by localization. It is well-known that $\mathcal{V} = \mathcal{V}_{loc}$, and the relationships $\mathcal{A} \subseteq \mathcal{A}_{loc} \subseteq \mathcal{V}$ and $\mathcal{A}^+ \subseteq \mathcal{A}_{loc}^+ \subseteq \mathcal{V}^+$ hold.

A finite variation process A can be decomposed to $A = A^r + A^g = A^c + A^d + A^g$ where A^c is continuous, A^r is right-continuous, A^d is discrete right-continuous, and A^g is discrete left-continuous such that

$$A_t^d = \sum_{0 < s \leq t} \Delta A_s \text{ and } A_t^g = \sum_{0 \leq s < t} \Delta^+ A_s,$$

where the series converge absolutely.

Definition 12.8 A process $M = (M_t)_{t \geq 0}$ is an (square integrable) optional martingale (supermartingale, submartingale) if:

- M is $\mathcal{O}(\mathbf{F})$ -measurable.
- There exists an (square) integrable $\mathcal{F}_\infty$ -measurable random variable $\hat{M}$ such that

$$M_T = \mathbf{E}[\hat{M} | \mathcal{F}_T]$$

(respectively, $M_T \geq \mathbf{E}[\hat{M} | \mathcal{F}_T]$, $M_T \leq \mathbf{E}[\hat{M} | \mathcal{F}_T]$) a.s. on the set $(T < \infty)$ for any $T \in \mathcal{T}$.

Let $\mathcal{M}$ (resp. $\mathcal{M}^2$) denote the set of optional martingales (resp. square-integrable optional martingales). The space $\mathcal{M}$ is extended to a space of local optional martingale, $\mathcal{M}_{loc}$, and the space $\mathcal{M}^2$ is extended to a space locally square-integrable optional martingales, $\mathcal{M}_{loc}^2$, respectively.

Definition 12.9 A process $M = (M_t)_{t \geq 0}$ is called an optional local (locally square integrable) martingale if there exists a sequence $(R_n, M^{(n)})$, $n \in \mathbb{N}$, where $R_n \in \mathcal{T}_+$, $R_n \uparrow \infty$ a.s. and $M^{(n)}$ is a (square integrable) optional martingale, such that $M = M^{(n)}$ on the stochastic interval $[0, R_n]$ and the random variable M_{R_n+} is integrable for any $n \in \mathbb{N}$.

If $M \in \mathcal{M}_{loc}$, then it can be decomposed into

$$M = M^r + M^g \text{ where } M^r = M^c + M^d,$$

M^c is continuous, M^d is right-continuous, and M^g is left-continuous optional local martingales. M^d and M^g are orthogonal to each other and to any continuous (local) martingale. Moreover, M^d and M^g can be written as

$$M_t^d = \sum_{0 < s \leq t} \Delta M_s \text{ and } M_t^g = \sum_{0 \leq s < t} \Delta^+ M_s.$$

Now, we discuss the extension of the quadratic variation process for the square-integrable optional martingales.

Lemma 12.1

- (1) Suppose $X \in \mathcal{M}^2$. There exists a unique increasing strongly predictable process $\langle X \rangle \in \mathcal{A}$ such that $\mathbf{E}X_T^2 = \mathbf{E}\langle X \rangle_T$ for every s.t. T , or, equivalently, $X^2 - \langle X \rangle \in \mathcal{M}$, where $\langle X \rangle = \langle X^g \rangle + \langle X^c \rangle + \langle X^d \rangle$.
- (2) If $X, Y \in \mathcal{M}^2$, then there exists a unique strongly predictable process $\langle X, Y \rangle \in \mathcal{A}$ such that $XY - \langle X, Y \rangle \in \mathcal{M}$, where

$$\langle X, Y \rangle = \frac{1}{2}[\langle X + Y \rangle - \langle X \rangle - \langle Y \rangle].$$

Suppose $X \in \mathcal{M}^2$ and X^c is its continuous part. We define $[X, X]$ to be the process

$$[X, X]_t = \langle X^c \rangle_t + \sum_{s \leq t} (\Delta X_s)^2 + \sum_{s < t} (\Delta^+ X_s)^2, \quad t \in \mathbb{R}_+.$$

It is not difficult to show that $[X, X]$ is increasing, $\mathbf{F}$ -adapted and integrable, and that $X^2 - [X, X] \in \mathcal{M}$. Moreover, we define

$$[X, Y] = \frac{1}{2}([X + Y, X + Y] - [X, X] - [Y, Y])$$

for $X, Y \in \mathcal{M}^2$, with $XY - [X, Y] \in \mathcal{M}$. Further, we get that

$$[X, Y]_t = \langle X^c, Y^c \rangle_t + \sum_{s \leq t} \Delta X_s \Delta Y_s + \sum_{s < t} \Delta^+ X_s \Delta^+ Y_s.$$

Lemma 12.2 *A nonnegative optional local martingale X is a supermartingale.*

Proof Let $X \in \mathcal{M}_{loc}$, $X \geq 0$. Then by definition of optional local martingale, there exist $X^n \in \mathcal{M}$, $X^n \geq 0$, and $R_n \in \mathcal{T}_+$, $R_n \uparrow \infty$ a.s. such that for any $n \geq 1$: $X = X^n \mathbf{1}_{[0, R_n]}$. Next, for any $t \geq s$ and $A \in \mathcal{F}_s$, we have

$$\begin{aligned} \mathbf{E}X_t \mathbf{1}_A &= \lim_{n \rightarrow \infty} \mathbf{E}X_t \mathbf{1}_A \mathbf{1}_{(t \leq R_n)} \\ &= \lim_{n \rightarrow \infty} \mathbf{E}X_t^n \mathbf{1}_A \mathbf{1}_{(t \leq R_n)} \\ &\leq \lim_{n \rightarrow \infty} \mathbf{E}X_t^n \mathbf{1}_A \mathbf{1}_{(s \leq R_n)} \\ &= \lim_{n \rightarrow \infty} \mathbf{E}X_s^n \mathbf{1}_A \mathbf{1}_{(s \leq R_n)} \\ &= \mathbf{E}X_s \mathbf{1}_A. \end{aligned}$$

Hence, the process X_t is a nonnegative supermartingale (in usual sense). $\square$

On an unusual stochastic basis, the most general processes that have good properties to work with are optional semimartingales. Optional semimartingales are linear combinations of finite variation processes and local optional martingales. Since local optional martingales and processes of finite variation are in general neither right-continuous nor left-continuous, optional semimartingales possess the same continuity characteristics. Therefore, their structure is complicated.

Definition 12.10 The stochastic process X is called an optional semimartingale if

$$X = X_0 + M + A,$$

where $M \in \mathcal{M}_{loc}$, $A \in \mathcal{V}$, $A_0 = M_0 = 0$ and X_0 is an $\mathcal{F}_0$ -measurable finite random variable. Moreover, an optional semimartingale X is called special optional semimartingale if the above decomposition exists with a strongly predictable process $A \in \mathcal{A}_{loc}$.

Let $\mathcal{S}$ denote the set of optional semimartingales and $\mathcal{S}_p$ the set of special optional semimartingales. If $X \in \mathcal{S}_p$, then the semimartingale decomposition is unique. By the optional martingale decomposition and the decomposition of

predictable processes, we can decompose a semimartingale as follows:

$$\begin{aligned}
 X &= X_0 + X^r + X^g, \text{ with} \\
 X^r &= A^r + M^r, \\
 M^r &= M^c + M^d, \\
 A^r &= A^c + A^d, \\
 X^g &= A^g + M^g,
 \end{aligned} \tag{12.1}$$

where A^c , A^d , and A^g are finite variation processes that are continuous, discrete right-continuous, and discrete left-continuous, respectively; $M^c \in \mathcal{M}_{loc}^c$, $M^d \in \mathcal{M}_{loc}^d$, $M^g \in \mathcal{M}_{loc}^g$ is a continuous, a discrete right-continuous and a left-continuous local martingale, respectively. This decomposition is useful for defining integration with respect to optional semimartingales.

12.2 Optional Stochastic Calculus

A stochastic integral with respect to optional semimartingale is defined as

$$\begin{aligned}
 \varphi \circ X_t &= \int_{[0,t]} \varphi_s dX_s = \int_{]0,t]} \varphi_{s-} dX_s^r + \int_{[0,t[} \varphi_s dX_{s+}^g, \text{ where} \\
 \int_{]0,t]} \varphi_{s-} dX_s^r &= \int_{]0,t]} \varphi_{s-} dA_s^r + \int_{]0,t]} \varphi_{s-} dM_s^r \text{ and} \\
 \int_{[0,t[} \varphi_s dX_{s+}^g &= \int_{[0,t[} \varphi_s dA_{s+}^g + \int_{[0,t[} \varphi_s dM_{s+}^g.
 \end{aligned}$$

The stochastic integral with respect to the finite variation processes or strongly predictable processes A^r over $]0, t]$ and A^g over $[0, t[$ is interpreted as usual, in the Lebesgue sense. The integral $\int_{]0,t]} \varphi_{s-} dM_s^r$ over $]0, t]$ is the usual stochastic integral with respect to cadlag local martingale, whereas $\int_{[0,t[} \varphi_s dM_{s+}^g$ over $[0, t[$ is the Gal'chuk stochastic integral with respect to left-continuous local martingale. In general, the stochastic integral with respect to optional semimartingale X can be defined as a bilinear form $(\varphi^r, \varphi^g) \circ X_t$, $\varphi^r \in \mathcal{P}$ and $\varphi^g \in \mathcal{O}$ such that

$$\begin{aligned}
 Y_t &= (\varphi^r, \varphi^g) \circ X_t = \varphi^r \cdot X_t^r + \varphi^r \odot X_t^g, \\
 \varphi^r \cdot X^r &= \int_{]0,t]} \varphi_s^r dX_s^r, \quad \varphi^g \odot X^g = \int_{[0,t[} \varphi_s^g dX_{s+}^g,
 \end{aligned}$$

where Y is again an optional semimartingale. Note that the stochastic integral over optional semimartingales is defined on a much larger space of integrands, the product space of predictable and optional processes, $\mathcal{P} \times \mathcal{O}$. From now on, we are going to use the operator “ $\circ$ ” to denote the stochastic optional integral, the operator “ $\cdot$ ” to denote the regular stochastic integral with respect to RCLL semimartingales, and the operator “ $\odot$ ” for the Galchuk stochastic integral $g \odot X^g$ with respect to left-continuous semimartingales.

The properties of optional stochastic integral are the followings.

First, isometry is satisfied with

$$\left(f^2 \cdot [X^r, X^r]\right)^{1/2} \in \mathcal{A}_{loc} \text{ and } \left(g^2 \odot [X^g, X^g]\right)^{1/2} \in \mathcal{A}_{loc}.$$

The quadratic variations are defined as

$$\begin{aligned} [X, X] &= [X^r, X^r] + [X^g, X^g] \text{ where} \\ [X^r, X^r]_t &= \langle X^c, X^c \rangle_t + \sum_{0 \leq s \leq t} (\Delta X_s)^2 \text{ and} \\ [X^g, X^g]_t &= \sum_{0 \leq s < t} (\Delta^+ X_s)^2; \end{aligned}$$

linearity is also satisfied with $(f^1 + f^2, g^1 + g^2) \circ X_t = (f^1, g^1) \circ X_t + (f^2, g^2) \circ X_t$ for any (f^1, g^1) and (f^2, g^2) in the space $\mathcal{P}(\mathbf{F}) \times \mathcal{O}(\mathbf{F})$; $\Delta^+ X^g$ is $\mathcal{O}(\mathbf{F}_+)$ -measurable with its martingale part satisfying $\mathbf{E}[\Delta^+ M_T^g \mathbf{1}_{(T < \infty)} | \mathcal{F}_T] = 0$ a.s. for any stopping time T in the broad sense and ΔX^r is $\mathcal{O}(\mathbf{F})$ -measurable with its martingale part satisfying $\mathbf{E}[\Delta M_T^r \mathbf{1}_{(T < \infty)} | \mathcal{F}_T] = 0$ a.s. for any stopping time T .

Moreover, orthogonality is as such that $X^r \perp X^g$ are orthogonal, in the sense that their product is an optional local martingale. Also, differentials are independent: $\Delta Y = f \Delta X^r$ and $\Delta^+ Y = g \Delta^+ X^g$; lastly, for any semimartingale Z , the quadratic projection is $[Y, Z] = f \cdot [X^r, Z^r] + g \odot [X^g, Z^g]$.

Consider the space $(E, \mathcal{E})$, where $E = \mathbb{R} \setminus \{0\}$ and $\mathcal{E} = \mathcal{B}(E)$ is the Borel σ -algebra in E . Define the spaces

$$\begin{aligned} \tilde{\Omega} &= \Omega \times \mathbb{R}_+ \times E, \quad \tilde{E} = \mathbb{R}_+ \times E, \quad \tilde{\mathcal{E}} = \mathcal{B}(\mathbb{R}_+) \times \mathcal{E} \\ \tilde{\mathcal{O}}(\mathbf{F}) &= \mathcal{O}(\mathbf{F}) \times \mathcal{E}, \quad \tilde{\mathcal{O}}(\mathbf{F}_+) = \mathcal{O}(\mathbf{F}_+) \times \mathcal{E}, \quad \tilde{\mathcal{P}} = \mathcal{P} \times \mathcal{E}. \end{aligned}$$

Definition 12.11 A nonnegative random set function $\mu(\omega, \Gamma)$, $\omega \in \Omega$, $\Gamma \in \tilde{\mathcal{E}}$, is called a random measure on $\tilde{\mathcal{E}}$ if $\mu(\cdot, \Gamma) \in \mathcal{F}$ for any $\Gamma \in \tilde{E}$ and $\mu(\omega, \cdot)$ is a σ -finite measure on $(\tilde{E}, \tilde{\mathcal{E}})$ for each $\omega \in \Omega$.

A random measure is called integer-valued if $\mu(\omega, \Gamma) \in \{0, 1, \dots, +\infty\}$ and $0 \leq \mu(\omega, \{t\} \times E) \leq 1$ for all (ω, Γ) .

For a nonnegative function $f \in \mathcal{F} \times \tilde{\mathcal{E}}$ and measure μ , let us form the process $f * \mu$, where

$$f * \mu_t = \int_{[0,t] \times E} f(\omega, s, x) \mu(\omega, ds, dx), \quad t < \infty.$$

The random measure μ is $O(\mathbf{F}_+)$ -optional if the process $f * \mu$ is $O(\mathbf{F}_+)$ -measurable for any nonnegative $\tilde{O}(\mathbf{F})$ -measurable f . Similarly, the random measure μ is $O(\mathbf{F})$ -optional (for brevity, optional) if $f * \mu \in \mathcal{O}$ for any nonnegative $\tilde{O}(\mathbf{F})$ -measurable f . The random measure μ is predictable if $f * \mu \in \mathcal{P}$ for any nonnegative $\tilde{\mathcal{P}}$ -measurable f .

Let μ be an optional measure. On the space $(\tilde{\Omega}, \tilde{O}(\mathbf{F}))$, we consider

$$\mathbf{E} f * \mu_\infty, \quad f \in \tilde{O}(\mathbf{F}), \quad f \geq 0.$$

Lemma 12.3 *Let the optional measure μ be such that the measure $\mathbf{E} f * \mu_\infty$ is $\tilde{\mathcal{P}}$ - σ -finite (i.e., the restriction of $\mathbf{E} f * \mu_\infty$ to $(\tilde{\Omega}, \tilde{\mathcal{P}})$ is σ -finite). Then there exists a unique (up to indistinguishability) predictable measure $\nu = \nu(\omega, dt, dx)$ such that for any function $f \in \tilde{\mathcal{P}}$, $f \geq 0$, one has $\mathbf{E} f * \mu_\infty = \mathbf{E} f * \nu_\infty$.*

The measure ν can be written in the form $\nu(\omega, dt, dx) = dA_t(\omega)K(\omega, t, dx)$, where A is an increasing predictable right-continuous process, and $K(\omega, t, dx)$ is the kernel of the space $(\Omega \times \mathbb{R}_+, \mathcal{P})$ into $(E, \mathcal{E})$.

If the measure μ does not load any predictable s.t.'s whatever, then the same is true for ν , and the process $f * \nu$ is continuous for any $f \in \tilde{\mathcal{P}}$, $f \geq 0$. Moreover, for any $S \in \mathcal{T}^p$ and any $f \in \tilde{\mathcal{P}}$, $f \geq 0$, on $(S < \infty)$,

$$\mathbf{E} \left[\int_E f(S, x) \mu(\{S\}, dx) \middle| \mathcal{F}_{S-} \right] = \int_E f(S, x) \nu(\{S\}, dx).$$

The process $f * \nu$ is the dual projection for the process $f * \mu$, $f \in \tilde{\mathcal{P}}$, $f \geq 0$. If $f * \mu \in \mathcal{A}_{loc}$, then the process

$$f * \mu - f * \nu \in \mathcal{M}_{loc}.$$

In the case of integer-valued μ , outside a set of P -null measure, $0 \leq \nu(\omega, \{t\} \times E) \leq 1$ for all $t \in \mathbb{R}_+$.

The measure ν is called the compensator of the measure μ .

Let η be an $O(\mathbf{F}_+)$ -optional measure. On the space $(\tilde{\Omega}, \tilde{O}(\mathbf{F}_+))$, we consider

$$\mathbf{E} f * \mu_\infty, \quad f \in \tilde{O}(\mathbf{F}_+), \quad f \geq 0.$$

Lemma 12.4 *Let the $O(\mathbf{F}_+)$ -optional measure η be such that the measure $\mathbf{E}f * \eta_\infty$ is $\tilde{O}$ - σ -finite. Then:*

- (a) *There exists a unique (up to indistinguishability) optional measure $\lambda = \lambda(\omega, dt, dx)$ such that for any function $f \in \tilde{O}(\mathbf{F})$, $f \geq 0$, one has $\mathbf{E}f * \eta_\infty = \mathbf{E}f * \lambda_\infty$.*

The measure λ can be written in the form $\lambda(\omega, dt, dx) = dA_t(\omega)K(\omega, t, dx)$, where A is an increasing optional right-continuous process, and $K(\omega, t, dx)$ is the kernel of the space $(\Omega \times \mathbb{R}_+, O)$ into $(E, \mathcal{E})$.

- (b) *If the measure η does not load any s.t.'s whatever, then the same is true for λ , and the process $f * \lambda$ is continuous for any $f \in \tilde{O}(\mathbf{F})$, $f \geq 0$.*
- (c) *The process*

$$f * \lambda_+ = \left(\int_{[0, t] \times E} f \lambda(ds, dx) \right)$$

is the dual optional projection for the process

$$f * \eta_+ = \left(\int_{[0, t] \times E} f \eta(ds, dx) \right)$$

for $f \in \tilde{O}(\mathbf{F})$, $f \geq 0$. In particular, for any $T \in \mathcal{T}$ and any $f \in \tilde{O}(\mathbf{F})$, $f \geq 0$ on $(T < \infty)$

$$\mathbf{E} \left[\int_E f(T, x) \eta(\{T\}, dx) \middle| \mathcal{F}_T \right] = \mathbf{E} \left[\int_E f(T, x) \lambda(\{T\}, dx) \middle| \mathcal{F}_T \right] \text{ a.s..}$$

If $f * \eta_+ \in \mathcal{A}_{loc}$, then the process

$$f * \eta_t - f * \lambda_t = \int_{[0, t] \times E} f \eta(ds, dx) - \int_{[0, t] \times E} f \lambda(ds, dx) \in \mathcal{M}_{loc}.$$

- (d) *If the measure η is integer-valued, then there exists a modification of the measure λ such that outside a set of P -null measure, $0 \leq \lambda(\omega, \{t\} \times E) \leq 1$ for any $t \in \mathbb{R}_+$.*

The measure λ is called the *compensator* of the $O(\mathbf{F}_+)$ -optional measure η .

Let Y be a one-dimensional optional semimartingale and $(S_n)_{n \geq 1}$, $(T_n)_{n \geq 1}$, $(U_n)_{n \geq 1}$ be sequences of predictable, totally inaccessible stopping times and totally inaccessible wide sense stopping times, respectively, exhausting all jumps of the process Y , i.e., the set $\{\Delta Y \neq 0\} \cup \{\Delta^+ Y \neq 0\}$, such that the graphs of these stopping times do not intersect within each sequence. Define integer random measures on $(\mathbb{R}_+ \times E, \tilde{\mathcal{E}})$

$$\mu^d(\Gamma) = \sum_{n \geq 1} \mathbf{1}_\Gamma(T_n, \beta_{T_n}^d), \quad \mu^g(\Gamma) = \sum_{n \geq 1} \mathbf{1}_\Gamma(U_n, \beta_{U_n}^g),$$

$$\begin{aligned}
p^d(\Gamma) &= \sum_{n \geq 1} \mathbf{1}_\Gamma(S_n, \beta_{S_n}^d), & p^g(\Gamma) &= \sum_{n \geq 1} \mathbf{1}_\Gamma(S_n, \beta_{S_n}^g), \\
\eta(\Gamma) &= \sum_{n \geq 1} \mathbf{1}_\Gamma(T_n, \beta_{T_n}^g),
\end{aligned}$$

where $\mathbf{1}_\Gamma(\cdot)$ is an indicator function of a set $\Gamma \in \tilde{\mathcal{E}}$, $\beta_t^d = \Delta Y_t$ if $\Delta Y_t \neq 0$ and $\beta_t^d = \delta$ if $\Delta Y_t = 0$, $\beta_t^g = \Delta^+ Y_t$ if $\Delta^+ Y_t \neq 0$, $\beta_t^g = \delta$ if $\Delta^+ Y_t = 0$, $t > 0$.

Under the *unusual* conditions on probability space admits the following *canonical decomposition*:

$$\begin{aligned}
Y_t &= Y_0 + a_t + m_t + \int_{]0,t] \times E} u \mathbf{1}_{(|u| \leq 1)} d(\mu^d - \nu^d) + \int_{[0,t[\times E} u \mathbf{1}_{(|u| \leq 1)} d(\mu^g - \nu^g) \\
&\quad + \int_{]0,t] \times E} u \mathbf{1}_{(|u| > 1)} d\mu^d + \int_{[0,t[\times E} u \mathbf{1}_{(|u| > 1)} d\mu^g \\
&\quad + \int_{]0,t] \times E} u dp^d + \int_{[0,t[\times E} u dp^g + \int_{[0,t[\times E} u d\eta
\end{aligned}$$

or in short notation

$$Y = Y_0 + a + m + \sum_{j=r,g} \left[u \mathbf{1}_{(|u| \leq 1)} * (\mu^j - \nu^j) + u \mathbf{1}_{(|u| > 1)} * \mu^j + u * p^j \right] + u * \eta, \quad (12.2)$$

where Y_0 is $\mathcal{F}_0$ -measurable random variable, $a \in \mathcal{A}_{loc}$, $a_0 = 0$, and $m \in \mathcal{M}_{loc}^2$, $m_0 = 0$ (a and m are both continuous); ν^j are the compensators of μ^j .

The Ito's formula is a basic instrument in stochastic calculus. In optional settings, we have a generalization of the Ito's formula which will be called also a change of variables formula.

Theorem 12.2 *Suppose X is an n -dimensional optional semimartingale, i.e.,*

$$X = (X^1, \dots, X^n),$$

where X^i is an optional semimartingale, $i = 1, \dots, n$, and $f(x) = f(x^1, \dots, x^n)$ is a twice continuously differentiable function on $\mathbb{R}^n$. Then the process $f(x)$ is an optional semimartingale, and for all $t \in \mathbb{R}_+$

$$f(X_t) = f(X_0) + \sum_{i=1}^n \int_{]0,t]} \frac{\partial f}{\partial x_i}(X_{s-}) d(A^{ir} + M^{ir})_s$$

$$\begin{aligned}
& + \frac{1}{2} \sum_{i,j=1}^n \int_{[0,t]} \frac{\partial^2 f}{\partial x_i \partial x_j} f(X_{s-}) d\langle M^{ic}, M^{jc} \rangle_s \\
& + \sum_{0 \leq s \leq t} \left[f(X_s) - f(X_{s-}) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X_{s-}) \Delta X_s^i \right] \\
& + \sum_{i=1}^n \int_{[0,t]} \frac{\partial f}{\partial x_i}(X_s) d(A^{ig} + M^{ig})_{s+} \\
& + \sum_{0 \leq s < t} [f(X_{s+}) - f(X_s) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X_s) \Delta^+ X_s^i],
\end{aligned} \tag{12.3}$$

where A^{ir} , M^{ir} , M^{ic} , A^{ig} , and M^{ig} are from the decomposition of X^i in (12.1).

After comparing Ito's formula with the above changing of variables formula, we can notice that there are two additional last terms in (12.3) that appeared due to jumps from the right.

Let us now consider the following optional semimartingale:

$$X = X_0 + f \cdot a + g \cdot m + \sum_{j=r,g} \left[Uh_j * (\mu^j - \nu^j) + (k_j + l_j) * p^j \right] + (r + w) * \eta, \tag{12.4}$$

where X_0 is $\mathcal{F}_0$ -measurable random variable; $U = U(u) = \mathbf{1}_{(|u| \leq 1)}(u)$; $a \in \mathcal{A}_{loc}$ is continuous increasing, $m, \mu^j, \nu^j, p^j, \eta$ as in (12.2). Further, for convenience, we use the notation $f = f(\omega, t, x)$, $g = g(\omega, t, x)$, $h_d = h(\omega, t, u, x)$, $h_g = h(\omega, t, u, x)$ and similarly for k_j, l_j, w and r whenever this does not lead to confusion.

Assume the following:

- For $j = r, g$,

$$\begin{aligned}
& |f(X)| \cdot a \in \mathcal{A}_{loc}, \\
& g(X) \cdot \langle m \rangle \in \mathcal{A}_{loc}, \\
& |h(X)|^2 * \nu^j \in \mathcal{A}_{loc}, \\
& |l_j(X)| * p^j \in \mathcal{A}_{loc}, \\
& \left[|k_j(X)|^2 * p^j \right]^{1/2} \in \mathcal{A}_{loc}, \\
& |r(X)| * \eta \in \mathcal{A}_{loc}, \\
& \left[|w(X)|^2 * \eta \right]^{1/2} \in \mathcal{A}_{loc},
\end{aligned}$$

and $E[k_r(S, \beta_S^d, X_0)|\mathcal{F}_{S-}] = 0$ a.s. for any predictable stopping time S on $\{S < \infty\}$ and $E[k_g(T, \beta_T^g, X_0)|\mathcal{F}_T] = 0$, $E[w(T, \beta_T^g, X_0)|\mathcal{F}_T] = 0$ a.s., for any stopping time T on $\{T < \infty\}$.

- $f(\omega, s, x)$ and $g(\omega, s, x)$ are defined on $(\Omega \times \mathbb{R}_+ \times \mathbb{R})$ and $\mathcal{P} \times \mathcal{B}(\mathbb{R})$ -measurable.
- $Uh_d(\omega, s, u, x)$ is defined on $(\Omega \times \mathbb{R}_+ \times E \cap (|u| \leq 1) \times \mathbb{R})$ and $\mathcal{P} \times \mathcal{B}(E \cap (|u| \leq 1)) \times \mathcal{B}(\mathbb{R})$ -measurable.
- $Uh_g(\omega, s, u, x)$ is defined on $(\Omega \times \mathbb{R}_+ \times E \cap (|u| \leq 1) \times \mathbb{R})$ and $\mathcal{O} \times \mathcal{B}(E \cap (|u| \leq 1)) \times \mathcal{B}(\mathbb{R})$ -measurable.
- $k_d(\omega, s, u, x)$, $l_d(\omega, s, u, x)$ are defined on $(\Omega \times \mathbb{R}_+ \times E \times \mathbb{R})$ and $\tilde{\mathcal{P}} \times \mathcal{B}(\mathbb{R})$ -measurable.
- $k_g(\omega, s, u, x)$, $l_g(\omega, s, u, x)$, $r(\omega, s, u, x)$, $w(\omega, s, u, x)$ are defined on $(\Omega \times \mathbb{R}_+ \times E \times \mathbb{R})$ and $\tilde{\mathcal{O}} \times \mathcal{B}(\mathbb{R})$ -measurable.

These integrability and measurability assumptions guarantee well-posedness of the integrals in (12.4).

Theorem 12.3 *Let X be an optional semimartingale given in (12.4). Let $f(x) = f(x^1, \dots, x^n)$ be a twice continuously differentiable function on $\mathbb{R}^n$. Then the process $f(x)$ is an optional semimartingale, and for all $t \in \mathbb{R}_+$, it has the following representation*

$$\begin{aligned}
 f(X_t) = & f(X_0) + \sum_{i=1}^n \int_{[0,t]} \frac{\partial f}{\partial x_i}(X_{s-}) d(a^i + m^i)_s \\
 & + \frac{1}{2} \sum_{i,j=1}^n \int_{[0,t]} \frac{\partial^2 f}{\partial x_i \partial x_j} f(X_{s-}) d\langle m, m \rangle_s \\
 & + [f(X_- + h_d) - f(X_-)] U * (\mu^d - \nu^d)_t \\
 & + [f(X + h_g) - f(X)] U * (\mu^g - \nu^g)_t \\
 & + \left[f(X_- + h_d) - f(X_-) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X_{s-}) h_d^i \right] U * \nu_t^d \\
 & + \left[f(X + h_g) - f(X) - \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X_s) h_g^i \right] U * \nu_t^g \\
 & + [f(X_- + (k_d + l_d)) - f(X_-)] * p_t^d \\
 & + [f(X + (k_g + l_g)) - f(X)] * p_t^g \\
 & + [f(X + (r + w)) - f(X)] * \eta_t.
 \end{aligned}$$

Chapter 13

Optional Stochastic Differential Equations and Their Applications



This chapter is devoted to study existence and uniqueness of solutions of SDEs w.r.t. optional semimartingales as well as their comparison property. A special attention is paid to linear SDEs and their solutions as stochastic exponents and logarithms. It is shown how this technique is exploited for financial modeling and pricing (see [1, 4, 21, 25, 29], and [34]).

13.1 Existence and Uniqueness of Solutions of Optional SDEs

Let us begin with a short exposition of results on existence and uniqueness of solutions of optional SDEs.

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space equipped with a filtration $\mathbf{F} = (\mathcal{F}_t)_{t \geq 0}$, which is not complete, right- or left-continuous.

Let Y be a one-dimensional optional semimartingale which has the representation (12.2). Consider the following SDE with respect to Y :

$$X = X_0 + f(X) \cdot a + g(X) \cdot m + (r + w)(X) * \eta + \sum_{j=r,g} \left[U h_j(X) * (\mu^j - \nu^j) + (k_j + l_j)(X) * p^j \right], \quad (13.1)$$

where X_0 is $\mathcal{F}_0$ -measurable random variable; $U = U(u) = \mathbf{1}_{(|u| \leq 1)}(u)$; $a \in \mathcal{A}_{loc}$ is continuous increasing, $m, \mu^j, \nu^j, p^j, \eta$ as in (12.2). Further, for convenience, we use the notation $f(X) = f(\omega, t, X_{t-})$, $g(X) = g(\omega, t, X_{t-})$, $h_d(X) = h(\omega, t, u, X_{t-})$, $h_g(X) = h(\omega, t, u, X_t)$ and similarly for k_j, l_j, w and r whenever this does not lead to confusion.

To guarantee well-posedness of the integrals in (13.1), we make the following.

Assumption A

- For $j = r, g$,

$$\begin{aligned}
 &|f(X)| \cdot a \in \mathcal{A}_{loc}, \\
 &(g(X))^2 \cdot \langle m \rangle \in \mathcal{A}_{loc}, \\
 &|h(X)|^2 * v^j \in \mathcal{A}_{loc}, \\
 &|l_j(X)| * p^j \in \mathcal{A}_{loc}, \\
 &\left[|k_j(X)|^2 * p^j\right]^{1/2} \in \mathcal{A}_{loc}, \\
 &|r(X)| * \eta \in \mathcal{A}_{loc}, \\
 &\left[|w(X)|^2 * \eta\right]^{1/2} \in \mathcal{A}_{loc},
 \end{aligned}$$

and $E[k_r(S, \beta_S^d, X_0)|\mathcal{F}_{S-}] = 0$ a.s. for any predictable stopping time S on $\{S < \infty\}$ and $E[k_g(T, \beta_T^g, X_0)|\mathcal{F}_T] = 0$, $E[w(T, \beta_T^g, X_0)|\mathcal{F}_T] = 0$ a.s., for any stopping time T on $\{T < \infty\}$.

- $f(\omega, s, x)$ and $g(\omega, s, x)$ are defined on $(\Omega \times \mathbb{R}_+ \times \mathbb{R})$ and are $\mathcal{P} \times \mathcal{B}(\mathbb{R})$ -measurable.
- $h_d(\omega, s, u, x)$, $k_d(\omega, s, u, x)$, $l_d(\omega, s, u, x)$ are defined on $(\Omega \times \mathbb{R}_+ \times E \times \mathbb{R})$ and $\tilde{\mathcal{P}} \times \mathcal{B}(\mathbb{R})$ -measurable.
- $h_g(\omega, s, u, x)$, $k_g(\omega, s, u, x)$, $l_g(\omega, s, u, x)$, $r(\omega, s, u, x)$, $w(\omega, s, u, x)$ are defined on $(\Omega \times \mathbb{R}_+ \times E \times \mathbb{R})$ and $\tilde{\mathcal{O}} \times \mathcal{B}(\mathbb{R})$ -measurable.

For convenience, we state here sufficient conditions for the existence and the uniqueness of the strong solution of (13.1). Suppose Assumption A holds. We say that the functions $f, g, h_j, k_j, l_j, r, w$ in (13.1) satisfy the $\mathbf{L}(\mathbf{Y}, \mathbf{X}_0)$ conditions if:

(L1) There exist nonnegative functions $F, G, H^j, L^j, K^j, R, W$, $j = r, g$, such that:

- $F(\omega, s)$, $G(\omega, s)$ are $\mathcal{P}$ -measurable; $H^d(\omega, s, u)$ is $\mathcal{P} \times \mathcal{B}(E \cap (|u| \leq 1))$ -measurable;
 $H^g(\omega, s, u)$ is $\mathcal{O} \times \mathcal{B}(E \cap (|u| \leq 1))$ -measurable;
 $L^d(\omega, s, u)$, $K^d(\omega, s, u)$ are $\tilde{\mathcal{P}}$ -measurable;
 $L^g(\omega, s, u)$, $K^g(\omega, s, u)$, $R(\omega, s, u)$, $W(\omega, s, u)$ are $\tilde{\mathcal{O}}$ -measurable.
- $F \cdot a_t + G \cdot \langle m \rangle_t + \sum_{j=r,g} [H^j U * v^j + (K^j + L^j) * \lambda^j] + (R + W) * \zeta < \infty$ a.s. for any $t > 0$, where λ^j and ζ are compensators of p^j and η , respectively.
- For any $x, y \in \mathbb{R}$, $s \in \mathbb{R}_+$ and $j = r, g$,

$$\begin{aligned}
 &|f(x) - f(y)| \cdot a_t \leq (F|x - y|) \cdot a_t \\
 &(g(x) - g(y))^2 \cdot \langle m \rangle_t \leq (G|x - y|^2) \cdot \langle m \rangle_t
 \end{aligned}$$

$$\begin{aligned}
|h_j(x) - h_j(y)| * v_t^j &\leq \left(H^j |x - y| \right) * v_t^j, \\
|l_j(x) - l_j(y)| * \lambda_t^j &\leq \left(L^j |x - y| \right) * \lambda_t^j, \\
|k_j(x) - k_j(y)|^2 * \lambda_t^j &\leq \left(K^j |x - y|^2 \right) * \lambda_t^j, \\
|r(x) - r(y)| * \zeta_t &\leq (R|x - y|) * \zeta_t, \\
|w(x) - w(y)|^2 * \zeta_t &\leq \left(W|x - y|^2 \right) * \zeta_t.
\end{aligned}$$

(L2)

$$\begin{aligned}
&(g(X_0))^2 \cdot \langle m \rangle_t + f(X_0) \cdot a_t + \left[r(X_0) + (w(X_0))^2 \right] * \zeta \\
&+ \sum_{j=r,g} \left[(h_j(X_0))^2 U * v^j + \left[(k_j(X_0))^2 + l_j(X_0) \right] * \lambda^j \right] < \infty
\end{aligned}$$

a.s. for any $t > 0$.

Theorem 13.1 *Let Y be an optional semimartingale and suppose that $f, g, h_j, k_j, l_j, r, w$ satisfy the $\mathbf{L}(\mathbf{Y}, \mathbf{X}_0)$ conditions. Then the strong solution of (13.1) exists and is unique.*

Remark 13.1 Results in Theorem 13.1 can be easily generalized to Eq. (13.1) with an additional term $\mathbf{1}_{\{|u|>1\}} h'_j * \mu^j$ due to its simple structure.

Let us study the comparison theorem for solutions of optimal SDEs. For the sake of brevity, here we want to compare solutions of the following SDEs in the form of

$$X_t^i = X_0^i + f^i(X^i) \cdot a_t + g_i(X^i) \cdot m_t + \sum_{j=r,g} \left[U h_j^i(X^i) * (\mu^j - v^j)_t \right], \quad i = 1, 2, \quad (13.2)$$

with initial condition $X_0^i = x_0^i$.

Denote

$$\begin{aligned}
F_i(z) &:= \int_{x_0^i}^z \frac{dx}{g_i(x)}, \\
\tilde{f}^i(z) &:= \frac{f^i(z)}{g_i(z)} - \frac{1}{2} g_i'(z) \alpha - \sum_{j=r,g} \int_E \left[\frac{h_j^i(z, u)}{g_i(z)} \right] U \tilde{v}^j(du), \\
\tilde{h}_j^i(z) &:= \int_z^{z+h_j^i(z)} \frac{dx}{g_i(x)},
\end{aligned}$$

where α and $\tilde{v}^j$ are given below in **Assumption B**:

(B1) (Structural conditions): There exist densities

$$\begin{aligned}\alpha &= \frac{d\langle m \rangle_t}{da_t}, \\ v^d(\omega, (0, t], \Gamma) &= \int_{0+}^t \tilde{v}^d(\omega, s, \Gamma) da_s, \\ v^g(\omega, [0, t), \Gamma) &= \int_0^{t-} \tilde{v}^g(\omega, s, \Gamma) da_{s+}.\end{aligned}$$

(B2) g_i $i = 1, 2$, is positive and continuously differentiable in z such that for any $s \in \mathbb{R}_+$, $z \in \mathbb{R}$

$$F_1(s, z) \geq F_2(s, z). \quad (13.3)$$

(B3) For any $s \in \mathbb{R}_+$, $u \in E$, $z, y \in \mathbb{R}$, $y \geq z$

$$\begin{aligned}\tilde{f}^1(s, z) &\leq \tilde{f}^2(s, y), \\ \tilde{h}_j^1(s, u, z) &\leq \tilde{h}_j^2(s, u, y), \\ z + \tilde{h}_j^2(s, u, z) &\leq y + \tilde{h}_j^2(s, u, y).\end{aligned}$$

(B4) $\tilde{f}^2(z)$ and $\tilde{h}_j^2(z)$ are Lipschitz continuous.

Theorem 13.2 Suppose that f^i, g_i, h_j^i in (13.2) satisfy $\mathbf{L}(\mathbf{Y}, \mathbf{X}_0^i)$, $i = 1, 2$, and Assumption B holds. Then there exist unique strong solutions X^1 and X^2 , and $X^1 \leq X^2$.

Proof We transform the processes X^i with the help of the change of variables formula and structural conditions, $i = 1, 2$:

$$\begin{aligned}\tilde{X}_t^i &:= \int_{x_0^i}^{X_t^i} \frac{dx}{g_i(x)} \\ &= \frac{f^i(X_{t-}^i)}{g_i(X_{t-}^i)} \cdot a_t + m_t - \frac{1}{2} g_i'(X_{t-}^i) \cdot \langle m \rangle_t \\ &\quad + \int_{X_{t-}^i}^{X_{t-}^i + h_d^i(X_{t-}^i)} \frac{dx}{g_i(x)} U * \mu_t^d + \int_{X_t^i}^{X_t^i + h_g^i(X_t^i)} \frac{dx}{g_i(x)} U * \mu_t^g \\ &\quad - \frac{h_d^i(X_{t-}^i)}{g_i(X_{t-}^i)} U * v_t^d - \frac{h_g^i(X_t^i)}{g_i(X_t^i)} U * v_t^g \\ &= \tilde{f}^i(F_i^{-1}(\tilde{X}_{t-}^i)) \cdot a_t + m_t + \tilde{h}_d^i(F_i^{-1}(\tilde{X}_{t-}^i)) U * \mu_t^d + \tilde{h}_g^i(F_i^{-1}(\tilde{X}_t^i)) U * \mu_t^g.\end{aligned}$$

From the condition (B2), it follows that $F_1^{-1}(z) \leq F_2^{-1}(z)$, and, consequently, for any Z

$$\tilde{f}^1(F_1^{-1}(z)) \leq \tilde{f}^2(F_2^{-1}(z)), \quad (13.4)$$

$$\tilde{h}_j^1(F_1^{-1}(z)) \leq \tilde{h}_j^2(F_2^{-1}(z)) \quad (13.5)$$

$$z + \tilde{h}_j^2(s, u, z) \leq y + \tilde{h}_j^2(s, u, y), \quad (13.6)$$

by applying (B3).

Functions $\tilde{f}^2(F_2^{-1}(x))$ and $\tilde{h}_j^2(F_2^{-1}(x))$ are obviously Lipschitz continuous since F_2^{-1} is continuously differentiable transformation and (B4).

Because

$$\tilde{h}_j^1(F_1^{-1}(z)) \neq \tilde{h}_j^2(F_2^{-1}(z)),$$

in general, we notice that

$$\begin{aligned} Y^+ &:= (\tilde{X}_t^1 - \tilde{X}_t^2)^+ \\ &= \mathbf{1}_{Y_- > 0} \left[\tilde{f}^1(F_1^{-1}(\tilde{X}_{t-}^1)) - \tilde{f}^2(F_2^{-1}(\tilde{X}_{t-}^1)) \right. \\ &\quad \left. + \tilde{f}^2(F_2^{-1}(\tilde{X}_{t-}^1)) - \tilde{f}^2(F_2^{-1}(\tilde{X}_{t-}^2)) \right] \cdot a_t \\ &\quad + \left[\left(Y + \tilde{h}_d^1(F_1^{-1}(\tilde{X}_{t-}^1)) - \tilde{h}_d^2(F_2^{-1}(\tilde{X}_{t-}^2)) \right)^+ - Y^+ \right] U * \mu_t^d \\ &\quad + \left[\left(Y + \tilde{h}_g^1(F_1^{-1}(\tilde{X}_t^1)) - \tilde{h}_g^2(F_2^{-1}(\tilde{X}_t^2)) \right)^+ - Y^+ \right] U * \mu_t^g \\ &= I_1 + I_2 + I_3. \end{aligned}$$

By (13.4) and Lipschitz continuity, $I_1 \leq \text{const.} Y^+ \cdot a_t$. Next, using (13.5), (13.6), and Lipschitz continuity, we get $I_2 \leq \text{const.} Y^+ U * \mu_t^d$ and $I_3 \leq \text{const.} Y^+ U * \mu_t^g$. Consequently, by Gronwall Lemma, we prove that $\tilde{X}^1 \leq \tilde{X}^2$.

This, together with (B2), implies that

$$\int_{x_0^2}^{X_t^1} \frac{dx}{g_2(x)} \leq \int_{x_0^1}^{X_t^1} \frac{dx}{g_1(x)} \leq \int_{x_0^2}^{X_t^2} \frac{dx}{g_2(x)}.$$

Since $g_2(x) > 0$, we conclude that $X^1 \leq X^2$. □

Remark 13.2 Proceeding with the same technique for jump measures as in the above proof, Theorem 13.2 can be directly extended to solutions of (13.1).

Let us give an example.

Example 13.1 Let X^1 and X^2 follow Eq. (13.2) with $m_t = W_t$, $a_t = t$. For $A \in \mathcal{B}(\mathbb{R}_+)$ and $\Gamma \in \mathcal{E}$, the Poisson random measures μ^d and μ^g are defined as

$$\begin{aligned}\mu^d(A \times \Gamma) &:= \# \left\{ (t, \Delta L_t^d) \in A \times \Gamma \mid t > 0 \text{ such that } \Delta L_t^d \neq 0 \right\} \\ \mu^g(A \times \Gamma) &:= \# \left\{ (t, \Delta^+ L_t^g) \in A \times \Gamma \mid t > 0 \text{ such that } \Delta^+ L_t^g \neq 0 \right\},\end{aligned}$$

where L_t^1 and L_t^2 are a Poisson process and a left-continuous modification of a Poisson process with constant intensities $\gamma^d = 1$ and $\gamma^g = 2$, respectively, and compensators $\nu^d = \gamma^d t$ and $\nu^g = \gamma^g t$. Furthermore, we assume that L^1 and L^2 are independent. We have

$$\begin{aligned}f^1(z) &= 0, & g_1(z) &= 1, & x_0^1 &= 0; \\ f^2(z) &= \frac{0.15 \cos(z)}{(1 - 0.3 \sin(z))^3}, & g_2(z) &= (1 - 0.3 \sin(z))^{-1}, & x_0^2 &= 0.\end{aligned}$$

Firstly, we consider the case with no jumps, i.e., $h_j^i = 0$, $i = 1, 2$. It is easy to check that all assumptions of Theorem 13.2 hold, and, thus, $X^1 \leq X^2$.

Next, let us assume that $h_j^i = 1$, $i = 1, 2$. By adding the compensated jumps with the same magnitude, we anticipate the same result as in the continuous case. However, condition (B3) of Theorem 13.2 is clearly not satisfied. Therefore, $X^1 \leq X^2$ is not necessarily true.

Finally, assume that $h_j^1 = 0.7$, $h_j^2 = 1$ and $f^1(z) = -1.8$, and other functions stay the same. Since

$$\begin{aligned}\tilde{f}^1(z) &= -3.9, \\ \tilde{f}^2(y) &= -3(1 - 0.3 \sin(y)) \\ \tilde{h}_j^1(z) &= \int_z^{z+0.7} \frac{dx}{1} = 0.7, \\ \tilde{h}_j^2(y) &= \int_y^{y+1} \frac{dx}{(1 - 0.3 \sin(x))^{-1}} = 1 + 0.3 \cos(y+1) - 0.3 \cos(y),\end{aligned}$$

then (B3) holds. In addition, $\tilde{f}^2$ and $\tilde{h}_j^2$ are obviously Lipschitz continuous. Thus, by Theorem 13.2, we obtain that $X^1 \leq X^2$.

13.2 Optional Stochastic Exponents and Logarithms

For optional semimartingales $X \in \mathcal{S}(\mathbb{F}, \mathbb{P})$, we define a linear optional SDE with respect to X :

$$Z_t = Z_0 + Z \circ X_t = Z_0 + \int_{0+}^t Z_{s-} dX_s^r + \int_0^{t-} Z_s dX_{s+}^g. \quad (13.7)$$

The next is the theorem that shows the existence and uniqueness of solution of Eq. (13.7) and gives a formula of such solution.

Theorem 13.3 *There exists a unique (up to indistinguishability) semimartingale $Z \in \mathcal{S}$ satisfying to (13.7).*

The process Z is given by the formula

$$Z_t = Z_0 \cdot \varepsilon_t(X), \quad (13.8)$$

where

$$\varepsilon_t(X) = \exp\left(X_t - \frac{1}{2}\langle X^c, X^c \rangle\right) \prod_{0 < s \leq t} (1 + \Delta X_s) e^{-\Delta X_s} \prod_{0 \leq s < t} (1 + \Delta^+ X_s) e^{-\Delta^+ X_s},$$

and $E(X) = (E_t(X))_{t \geq 0}$ is called a stochastic optional exponential of X .

Proof Set $U_0 = Z_0, V_0 = 1$,

$$U_t = Z_0 \prod_{0 < s \leq t} (1 + \Delta X_s) e^{-\Delta X_s}, \quad V_t = \prod_{0 \leq s < t} (1 + \Delta^+ X_s) e^{-\Delta^+ X_s} \quad t > 0,$$

and let us show that these products converge absolutely a.s. for any $t \in \mathbb{R}_+$.

Since there are almost surely a finite number of jumps of the form $|\Delta X| > 1/2$, $|\Delta^+ X| > 1/2$ on any time segment, the processes

$$\prod_{0 < s \leq t} (1 + \Delta X_s \mathbf{1}(A_s^c)) e^{-\Delta X_s \mathbf{1}(A_s^c)}, \quad \prod_{0 \leq s < t} (1 + \Delta^+ X_s \mathbf{1}(B_s^c)) e^{-\Delta^+ X_s \mathbf{1}(B_s^c)}$$

have a.s. finite variation on any segment, where A_s^c, B_s^c are the complements of the sets

$$A_s = \{|\Delta X_s| > \frac{1}{2}\}, \quad B_s = \{|\Delta^+ X_s| \leq \frac{1}{2}\}.$$

Set

$$U'_t = \prod_{0 < s \leq t} (1 + \Delta X_s \mathbf{1}(A_s)) e^{-\Delta X_s \mathbf{1}(A_s)}, \quad V'_t = \prod_{0 \leq s < t} (1 + \Delta^+ X_s \mathbf{1}(B_s)) e^{-\Delta^+ X_s \mathbf{1}(B_s)}.$$

Since the products contain strictly positive quantities, one can take their logarithm:

$$\log U'_t = \sum_{0 < s \leq t} [\log(1 + \Delta X_s \mathbf{1}(A_s)) - \Delta X_s \mathbf{1}(A_s)],$$

$$\log V_t^i = \sum_{0 \leq s < t} [\log(1 + \Delta^+ X_s \mathbf{1}(B_s)) - \Delta^+ X_s \mathbf{1}(B_s)].$$

These series a.s. converge absolutely since they are majorized by the series $c \sum_{0 < s \leq t} (\Delta X_s)^2$, $c \sum_{0 \leq s < t} (\Delta^+ X_s)^2$. Let us set

$$K_t = X_t - \frac{1}{2} \langle X^c, X^c \rangle_t, \quad F(x, y, z) = e^x y z,$$

and apply the change of variables formula to F . We have

$$\begin{aligned} Z_t = F(K_t, U_t, V_t) &= Z_0 + \int_{]0, t]} \left(Z_{s-} d(X_s^r - \frac{1}{2} \langle X^c, X^c \rangle_s) + e^{K_{s-}} V_{s-} dU_s \right) \\ &+ \frac{1}{2} \int_{]0, t]} Z_{s-} d \langle X^c, X^c \rangle_s + \int_{]0, t[} Z_s dX_{s+}^g + \int_{]0, t[} e^{K_s} U_s dV_{s+} \\ &+ \sum_{0 < s \leq t} \left[Z_s - Z_{s-} - Z_{s-} \Delta X_s - e^{K_{s-}} V_{s-} \Delta U_s \right] \\ &+ \sum_{0 \leq s < t} \left[Z_{s+} - Z_s - Z_s \Delta^+ X_s - e^{K_s} U_s \Delta^+ V_s \right]. \end{aligned} \quad (13.9)$$

Reducing similar terms and noting that

$$\begin{aligned} \int_{]0, t]} e^{K_{s-}} V_{s-} dU_s &= \sum_{0 < s \leq t} e^{K_{s-}} V_{s-} \Delta U_s, \\ \int_{]0, t[} e^{K_s} U_s dV_{s+} &= \sum_{0 \leq s < t} e^{K_s} U_s \Delta^+ V_s, \end{aligned}$$

and also that $\Delta Z_t = Z_t - \Delta X_t$, $\Delta^+ Z_t = Z_t \Delta^+ X_t$, we see that (13.9) coincides with (13.7).

Uniqueness. Let $\tilde{Z} \in \mathcal{S}$ be yet another solution of (13.7), $\tilde{Z} = Z_0$. Put $Y_t = e^{-K_t} \tilde{Z}_t$. From the change of variables formula, reducing similar terms, we get

$$Y_t = Z_0 + \sum_{0 < s \leq t} [e^{-K_s} \tilde{Z}_s - e^{-K_{s-}} - \tilde{Z}_{s-}] + \sum_{0 < s \leq t} [e^{-K_{s+}} \tilde{Z}_{s+} - e^{-K_s} \tilde{Z}_s],$$

or

$$Y_t = Z_0 + \sum_{0 < s \leq t} \Delta Y_s + \sum_{0 \leq s < t} \Delta^+ Y_s,$$

where the series converge absolutely a.s. It follows that Y has a.s. finite variation on any segment $[0, t]$.

Since

$$\Delta Y_s = Y_{s-} [e^{-\Delta^+ X_s} (1 + \Delta X_s) - 1], \quad \Delta^+ Y_s = Y_s [e^{-\Delta^+ X_s} (1 + \Delta^+ X_s) - 1],$$

the process Y is a solution of the equation

$$Y_t = Z_0 + \int_{0,t} Y_{s-} dA_s^r + \int_{0,t} Y_s dA_{s+}^g,$$

where

$$A_t^r = \sum_{0 < s \leq t} [e^{-\Delta X_s} (1 + \Delta X_s) - 1], \quad A_t^g = \sum_{0 \leq s < t} [e^{-\Delta^+ X_s} (1 + \Delta^+ X_s) - 1]. \quad (13.10)$$

are processes a.s. of bounded variation on any segment. The equations

$$Y = Y_0^r + \int_{0,t} Y_{s-} dA_s^r, \quad Y = Y_0^g + \int_{0,t} Y_s dA_{s+}^g,$$

have at most one solution. Consequently, Eq. (13.10) has at most one solution. It follows that Eq. (13.7) also has at most one solution. $\square$

Two useful properties of stochastic exponentials are the inverse and product formulas.

Problem 13.1 For $X, Y \in \mathcal{S}$, the product of stochastic exponent is calculated as

$$\mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(X + Y + [X, Y]).$$

Hint. Using the change of variables formula and the definition of stochastic exponent, we obtain

$$\begin{aligned} \mathcal{E}(X)\mathcal{E}(Y) &= 1 + \mathcal{E}(X) \circ \mathcal{E}(Y) + \mathcal{E}(Y) \circ \mathcal{E}(X) + [\mathcal{E}(X), \mathcal{E}(Y)] \\ &= 1 + \mathcal{E}(X)\mathcal{E}(Y) \circ Y + \mathcal{E}(X)\mathcal{E}(Y) \circ X + \mathcal{E}(X)\mathcal{E}(Y) \circ [X, Y] \\ &= 1 + \mathcal{E}(X)\mathcal{E}(Y) \circ (X + Y + [X, Y]) \\ &= \mathcal{E}(X + Y + [X, Y]). \end{aligned}$$

Lemma 13.1 The inverse of stochastic exponent for the semimartingale X is

$$\mathcal{E}^{-1}(X) = \mathcal{E}(-X^*),$$

where

$$X_t^* = X_t - \langle X^c, X^c \rangle_t - \sum_{0 < s \leq t} \frac{(\Delta X_s)^2}{1 + \Delta X_s} - \sum_{0 \leq s < t} \frac{(\Delta^+ X_s)^2}{1 + \Delta^+ X_s}.$$

Proof Let X^* be such that $\mathcal{E}(-X^*)\mathcal{E}(X) = \mathcal{E}(X - X^* - [X, X^*]) = 1$. Hence, $X - X^* - [X, X^*] = 0$ and

$$\begin{aligned} X^* &= X - \langle X^c \rangle_t + \sum_{s \leq t} (\Delta X_s)^2 + \sum_{s < t} (\Delta^+ X_s)^2 \\ &= X - \langle X^c \rangle_t + \sum_{0 < s \leq t} \frac{(\Delta X_s)^2 + (\Delta X_s)^3}{1 + \Delta X_s} + \sum_{0 \leq s < t} \frac{(\Delta^+ X_s)^2 + (\Delta^+ X_s)^3}{1 + \Delta^+ X_s} \\ &= X - \langle X^c \rangle_t + \sum_{0 < s \leq t} \frac{(\Delta X_s)^2}{1 + \Delta X_s} + \sum_{0 \leq s < t} \frac{(\Delta^+ X_s)^2}{1 + \Delta^+ X_s}. \end{aligned}$$

The sum of $(\Delta X)^3/(1 + \Delta X)$ is zero. To see this, consider

$$\sum \frac{(\Delta X)^3}{1 + \Delta X} = \left[\sum \frac{(\Delta X)^2}{1 + \Delta X}, \sum \Delta X \right] = 0$$

and as well, for $(\Delta^+ X)^3/(1 + \Delta^+ X)$. Hence,

$$X^* = X - \langle X^c \rangle_t + \sum_{0 < s \leq t} \frac{(\Delta X_s)^2}{1 + \Delta X_s} + \sum_{0 \leq s < t} \frac{(\Delta^+ X_s)^2}{1 + \Delta^+ X_s}.$$

The *optional stochastic logarithm* is defined as follows.

Let Y be a real-valued optional semimartingale such that the processes Y_- and Y do not vanish, and then the process

$$X_t = \frac{1}{Y} \circ Y_t = \int_{0+}^t \frac{1}{Y_{s-}} dY_s^r + \int_0^{t-} \frac{1}{Y_s} dY_s^g, \quad X_0 = 0, \quad (13.11)$$

also denoted by $X = \mathcal{L}og(Y)$, called the stochastic logarithm of Y . It is the unique semimartingale X such that $Y = Y_0 \mathcal{E}(X)$. Moreover, if $\Delta Y \neq -1$ and $\Delta^+ X \neq -1$, we also have

$$\begin{aligned} \mathcal{L}og Y_t &= \log \left| \frac{Y_t}{Y_0} \right| + \frac{1}{2Y^2} \circ \langle Y^c, Y^c \rangle_t - \sum_{0 < s \leq t} \left(\log \left| 1 + \frac{\Delta Y_s}{Y_{s-}} \right| - \frac{\Delta Y_s}{Y_{s-}} \right) \\ &\quad - \sum_{0 < s < t} \left(\log \left(1 + \frac{\Delta^+ Y_s}{Y_s} \right) - \frac{\Delta^+ Y_s}{Y_s} \right). \end{aligned} \quad (13.12)$$

It is important to note that the process Y does not need to be positive for $\mathcal{L}og(Y)$ to exist, in accordance with the fact that the stochastic exponential $\mathcal{E}(X)$ may take negative values.

The assumption that Y_- and Y do not vanish implies that $S_n = \inf\{t : |Y_t| \leq \frac{1}{n}\} \uparrow \infty$; hence, $1/Y_-$ is locally bounded; likewise, $T_n = \inf\{t : |Y_t| \leq \frac{1}{n}\} \uparrow \infty$; hence, $1/Y$ is also locally bounded and the stochastic integral in (13.11) makes sense. Let $\tilde{Y} = Y/Y_0$, then $\tilde{Y}_0 = 1$, and by (13.11), $X = (1/\tilde{Y}) \circ \tilde{Y}$. Thus,

$$1 + \tilde{Y} \circ X = 1 + \tilde{Y} \circ \left(\frac{1}{\tilde{Y}} \circ \tilde{Y} \right) = 1 + \left(\frac{1}{\tilde{Y}} \right) \circ \tilde{Y} = \tilde{Y},$$

or $\tilde{Y} = \mathcal{E}(X)$. Furthermore, $\Delta X = \Delta Y/Y_- \neq -1$ and $\Delta X = \Delta^+ Y/Y \neq -1$. Let $\tilde{X}$ be any other semimartingale satisfying $Y = \mathcal{E}(\tilde{X})$. With $\tilde{Y} = Y/Y_0$ yields $\tilde{Y} = \mathcal{E}(\tilde{X})$. So, $\tilde{Y} = 1 + \tilde{Y} \circ \tilde{X}$ or $Y = Y_0 + \tilde{Y} \circ \tilde{X}$. Since $X_0 = 0$, we have

$$\tilde{X} = \frac{\tilde{Y}}{\tilde{Y}} \circ \tilde{X} = \frac{1}{\tilde{Y}} \circ (Y - Y_0) = \frac{1}{\tilde{Y}} \circ Y = X,$$

and we get uniqueness.

To prove representation (13.12), we apply the change of variable formula to optional semimartingale $\log |Y|$. Since the log explodes at 0, we consider: For each n , the C^2 function f_n on $\mathbb{R}$, with $f_n(x) = \log |x|$ when $|x| \geq 1/n$. Then for all n and $t < T_n$ and $t < S_n$, we have

$$\begin{aligned} \log |Y_t| &= \log |Y_0| + \frac{1}{Y} \circ Y_t - \frac{1}{2Y^2} \circ \langle Y^c, Y^c \rangle_t + \sum_{0 < s \leq t} \left(\Delta \log |Y_s| - \frac{\Delta Y_s}{Y_{s-}} \right) \\ &\quad + \sum_{0 \leq s < t} \left(\Delta^+ \log |Y_s| - \frac{\Delta^+ Y_s}{Y_s} \right). \end{aligned}$$

This, together with (13.11), yields (13.12) for $t < T_n$ and $t < S_n$. Since $T_n \uparrow \infty$ and $S_n \rightarrow \infty$, then we obtain (13.12) everywhere. $\square$

Now, we present some of the properties of stochastic logarithms.

Lemma 13.2

- (a) If X is a semimartingale satisfying $\Delta X \neq -1$ and $\Delta^+ X \neq -1$, then $\mathcal{L}og(\mathcal{E}(X)) = X - X_0$.
- (b) If Y is a semimartingale such that Y and Y_- do not vanish, then $\mathcal{E}(\mathcal{L}og(Y)) = Y/Y_0$.
- (c) For any two optional semimartingales X and Z , we get the following identities:

$$\mathcal{L}og(XZ) = \mathcal{L}og X + \mathcal{L}og Z + [\mathcal{L}og X, \mathcal{L}og Z],$$

and

$$\mathcal{L}og\left(\frac{1}{X}\right) = 1 - \mathcal{L}og(X) - \left[X, \frac{1}{X}\right].$$

Proof

(a)

$$\mathcal{L}og(\mathcal{E}(X)) = \mathcal{E}(X)^{-1} \circ \mathcal{E}(X) = \frac{\mathcal{E}(X)}{\mathcal{E}(X)} \circ X = X - X_0.$$

(b) Let $Z = \mathcal{E}(\mathcal{L}og(Y))$; then by (a) we find that

$$\mathcal{L}og(Z) = \mathcal{L}og(\mathcal{E}(\mathcal{L}og(Y))) = \mathcal{L}og(Y) - \mathcal{L}og(Y_0) = \mathcal{L}og(Y/Y_0).$$

Therefore, $Z = Y/Y_0$.

(c) By the integral representation of the stochastic logarithm, we find

$$\mathcal{L}og(XZ) = \frac{1}{XZ} \circ (X \circ Z + Z \circ X + [X, Z]) = \mathcal{L}og X + \mathcal{L}og Z + [\mathcal{L}og X, \mathcal{L}og Z].$$

Using integration by parts and the integral definition of the stochastic logarithm,

$$\mathcal{L}og\left(\frac{1}{X}\right) = X \circ \frac{1}{X} = 1 - \frac{1}{X} \circ X - \left[X, \frac{1}{X}\right] = 1 - \mathcal{L}og(X) - \left[X, \frac{1}{X}\right].$$

□

A generalization of the stochastic exponential integral equation (13.7) is the nonhomogeneous linear stochastic equation, $X = G + X \circ H$. This equation has a natural application in finance: G is a stochastic cash flow, H is the interest rate of the money market account, and X is the time value of the cash flow accumulated in the money market account. Here we will give the solution of the nonhomogeneous linear stochastic integral equation to the case where G and H are optional semimartingales.

Theorem 13.4 *Consider the nonhomogeneous linear stochastic equation:*

$$X_t = G_t + \int_0^t X_s dH_s = G_t + \int_{0+}^t X_{s-} dH_s^r + \int_0^{t-} X_s dH_{s+}^g, \quad (13.13)$$

where $G = (G_t)_{t \geq 0} \in \mathcal{S}(F, P)$ is an optional semimartingale. The solution of (13.13) admits the form

$$X_t = \mathcal{E}_t(H) \left[G_0 + \int_0^t \mathcal{E}_s(H)^{-1} d\tilde{G}_s \right], \quad (13.14)$$

where

$$d\tilde{G}_t = dG_t - [G, \tilde{H}]_t, \quad H_t = H_t^c + \sum_{0 < s \leq t} \frac{\Delta H_s}{1 + \Delta H_s} + \sum_{0 \leq s < t} \frac{\Delta^+ H_s}{1 + \Delta^+ H_s}.$$

Proof Let us consider a solution of the following form $X_t = \mathcal{E}_t(H)Z_t$, where Z_t is related to G_t . The differential of X_t is

$$dX_t = d(\mathcal{E}_t(H)Z_t) = X_t dH_t + \mathcal{E}_t(H)\{dZ_t + d[H, Z]_t\}. \quad (13.15)$$

Comparing Eqs. (13.13) to (13.15), we find that

$$dG_t = \mathcal{E}_t(H)\{dZ_t + d[H, Z]_t\}.$$

To solve this equation for Z_t , we choose

$$\tilde{H}_t = H_t^c + \sum_{0 < s \leq t} \frac{\Delta H_s}{1 + \Delta H_s} + \sum_{0 \leq s < t} \frac{\Delta^+ H_s}{1 + \Delta^+ H_s}$$

and compute the quadratic variation of G with $\tilde{H}$. The quadratic variation for optional processes belongs to the space $\mathcal{A}_{loc}$ and is defined as

$$[G, \tilde{H}]_t = \langle G^c, \tilde{H}^c \rangle_t + \sum_{0 < s \leq t} \Delta G_s \Delta \tilde{H}_s + \sum_{0 \leq s < t} \Delta^+ G_s \Delta^+ \tilde{H}_s.$$

Hence,

$$\begin{aligned} d[G, \tilde{H}]_t &= \mathcal{E}_t(H) \left\{ d[Z, \tilde{H}]_t + d[H, Z]_t, \tilde{H}_t \right\} \\ &= \mathcal{E}_t(H) \left\{ d[Z^c, H^c]_t + \frac{\Delta H_t \Delta Z_t}{1 + \Delta H_t} + \frac{\Delta^+ H_t \Delta^+ Z_t}{1 + \Delta^+ H_t} \right. \\ &\quad \left. + \frac{(\Delta H_t)^2}{1 + \Delta H_t} \Delta Z_t + \frac{(\Delta^+ H_t)^2}{1 + \Delta^+ H_t} \Delta^+ Z_t \right\} \\ &= \mathcal{E}_t(H) \left\{ d[Z^c, H^c]_t + \Delta H_t \Delta Z_t \left(\frac{1 + \Delta H_t}{1 + \Delta H_t} \right) \right. \\ &\quad \left. + \Delta^+ H_t \Delta^+ Z_t \left(\frac{1 + \Delta^+ H_t}{1 + \Delta^+ H_t} \right) \right\} \\ &= \mathcal{E}_t(H) \left\{ d[Z^c, H^c]_t + \Delta H \Delta Z + \Delta^+ H_t \Delta^+ Z_t \right\} \\ &= \mathcal{E}_t(H) d[Z, H]_t. \end{aligned}$$

Then we calculate Z in the following way:

$$\begin{aligned} dG_t &= \mathcal{E}_t(H)\{dZ_t + d[H, Z]_t\}, \\ \mathcal{E}_t(H)^{-1}dG_t &= dZ_t + d[H, Z]_t, \\ &= dZ_t + \mathcal{E}_t(H)^{-1}d[G, \tilde{H}]_t, \\ dZ_t &= \mathcal{E}_t(H)^{-1}[dG_t - d[G, \tilde{H}]_t]. \end{aligned}$$

Note that

$$\begin{aligned} [H, Z], \tilde{H} &= \left[\langle H^c, Z^c \rangle + \sum_{s \leq \cdot} \Delta H \Delta Z + \sum_{s < \cdot} \Delta^+ H \Delta^+ Z, \tilde{H} \right] \\ &= \left[\langle H^c, Z^c \rangle, \tilde{H} \right] + \left[\left(\sum_{s \leq \cdot} \Delta H \Delta Z \right), \tilde{H} \right] + \left[\left(\sum_{s < \cdot} \Delta^+ H \Delta^+ Z \right), \tilde{H} \right] \end{aligned}$$

$[\langle H^c, Z^c \rangle, \tilde{H}] = 0$ because $\langle H^c, Z^c \rangle$ is continuous locally bounded variation processes and $\tilde{H}$ is a semimartingale. Then,

$$\begin{aligned} [H, Z], \tilde{H} &= \left[\left(\sum_{s \leq \cdot} \Delta H \Delta Z \right), \tilde{H} \right] + \left[\sum_{s < \cdot} \left(\Delta^+ H \Delta^+ Z \right), \tilde{H} \right] \\ &= \sum_{s \leq t} (\Delta H_s)^2 \Delta Z_s + \sum_{s < t} (\Delta^+ H_s)^2 \Delta^+ Z_s. \end{aligned}$$

□

Another important application of stochastic exponent is the *Gronwall lemma*.

In stochastic analysis, the Gronwall lemma is essential, and many extensions have been proposed. Here we will extend the Gronwall lemma to optional semimartingales in *unusual* probability spaces.

Lemma 13.3 *Let X be an optional process and H be an optional increasing process and C_t a constant such that*

$$X_t \leq C_t + \int_0^t X_s dH_s = C_t + \int_{0+}^t X_{s-} dH_s + \int_0^{t-} X_s dH_{s+}$$

for all $t \in [0, \infty)$.

Then,

$$X_t \leq C_t \mathcal{E}_t(H).$$

Proof Let

$$N_t = C_t + \int_0^t X_s dH_s - X_t,$$

and then $N_t \geq 0$ for all t . Therefore,

$$X_t = C_t - N_t + \int_0^t X_s dH_s$$

is a nonhomogeneous stochastic integral equation whose solution is given by

$$X_t = \mathcal{E}_t(H) \left[G_0 + \int_0^t \mathcal{E}_s(H)^{-1} d\tilde{G}_s \right], \quad (13.16)$$

where

$$\begin{aligned} G_t &= C_t - N_t, \quad \tilde{G}_t = G_t - [G, \tilde{H}]_t, \\ \tilde{H}_t &= H_t^c + \sum_{0 \leq s \leq t} \frac{\Delta H_s}{1 + \Delta H_s} + \sum_{0 \leq s < t} \frac{\Delta^+ H_s}{1 + \Delta^+ H_s}. \end{aligned}$$

Since H is increasing, then $\Delta H_s \geq 0$ and $\Delta^+ H_s \geq 0$, and therefore, $\tilde{H}$ is increasing. Hence, $[G, \tilde{H}] = 0$, $G_t = G_t - N_t \leq C_t$ for all t , since $N_t \geq 0$.

Knowing all this, we can write (13.16) as

$$\begin{aligned} X_t &= \mathcal{E}_t(H) \left[G_0 + \int_0^t \mathcal{E}_s(H)^{-1} d\tilde{G}_s \right] \\ &= \mathcal{E}_t(H) \left[G_0 + \int_0^t \mathcal{E}_s(H)^{-1} dG_s \right] \\ &\leq C_t \mathcal{E}_t(H), \end{aligned}$$

where $\int_0^t \mathcal{E}_s(H)^{-1} dG_s \leq 0$ and $G_0 \leq C_0$. □

13.3 Financial Modeling Using Optional Semimartingales

Consider a market on *unusual* probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, P)$. Let us also introduce the filtration $\mathbb{F}_+ = (\mathcal{F}_{t+})_{t \geq 0}$ and its completion under P , $\mathbb{F}^* = (\mathcal{F}_t^P)_{t \geq 0}$. Our market consists of two types of securities x and X and a portfolio $\pi = (\eta, \xi)$ which is composed of the optional processes η and ξ . η is the volume of the reference

asset x , while ξ is the volume of the security X . Suppose $x_t > 0$ and $X_t \geq 0$ for all $t \geq 0$ and write the ratio process $R_t = X_t/x_t$. Then, the value of the portfolio is

$$Y_t = \eta_t + \xi_t R_t. \quad (13.17)$$

$Y \in \mathcal{O}(\mathbb{F})$ is a real-valued optional semimartingale that is r.l.l.l. Furthermore, we restrict the portfolio, π , to be self-financing; that is, we must have

$$Y_t = Y_0 + \xi \circ R_t. \quad (13.18)$$

Reconciling Eqs. (13.17) and (13.18), we get

$$C_t = \eta_t + R_t \circ \xi_t + [\xi, R]_t = C_0,$$

where C_t is the consumption process with its initial value C_0 . Since the ratio process R is an optional semimartingale, then ξ evolves in the space $\mathcal{P}(\mathbb{F}) \times \mathcal{O}(\mathbb{F})$ with the predictable part determining the volume of R^r and the optional part determining the volume of R^s . R^r is the right-continuous part of X/x and R^s is its left-continuous part. Also, η belongs to the space $\mathcal{O}(\mathbb{F})$. Furthermore, for the integral in Eq. (13.18) to be well-defined, ξ must be R -integrable,

$$\int_0^\infty \xi_s^2 d[R, R]_s \in \mathcal{A}_{loc}.$$

In addition to the process Y , let $Y^+ = (Y_{t+})_{t \geq 0}$, which is $\mathcal{O}(\mathbb{F}^*)$. Y^+ is r.c.l.l. semimartingale on $(\Omega, \mathcal{F}, \mathbb{F}^*, P)$. The role of Y^+ will be discussed when we discuss no free lunch in optional markets.

Remark 13.3 It is also possible to write the market portfolio as

$$Z_t = \eta_t x_t + \xi_t X_t,$$

where $Y = Z/x$. However, in this case, the volume of the reference asset, η , is a process that also evolves in $\mathcal{P}(\mathbb{F}) \times \mathcal{O}(\mathbb{F})$. Therefore, a variety of portfolios are possible on unusual financial markets. $\square$

Here we present methods for finding local martingale deflators for these markets. A *local martingale deflator* is a strictly positive supermartingale multiplier used to transform the value process of a portfolio to a supermartingale (i.e., a local martingale). It is important for finding local martingale measures and hence fair prices and hedging strategies for contingent claims.

Suppose that the dynamics of securities in our market follow the stochastic exponentials

$$X_t = X_0 \mathcal{E}_t(H), \quad x_t = x_0 \mathcal{E}_t(h),$$

where x_0 and X_0 are $\mathcal{F}_0$ -measurable r.v.s, and $h = (h_t)_{t \geq 0}$ and $H = (H_t)_{t \geq 0}$ are optional semimartingales admitting the representations

$$h_t = h_0 + a_t + m_t, \quad H_t = H_0 + A_t + M_t, \quad (13.19)$$

with respect to P . Processes $a = (a_t)_{t \geq 0}$ and $A = (A_t)_{t \geq 0}$ are locally bounded variation and predictable, and $m = (m_t)_{t \geq 0}$ and $M = (M_t)_{t \geq 0}$ are optional local martingales.

First, we shall study, when the ratio process, R , is a local optional martingale with respect to the initial measure P , i.e., when $R \in \mathcal{M}_{loc}(\mathbb{F}, P)$. For $t \geq 0$. The ratio R in exponential form can be written as

$$R_t = \frac{X_t}{x_t} = R_0 \mathcal{E}(H)_t \mathcal{E}^{-1}(h)_t,$$

and using the properties of stochastic exponentials, we find

$$R_t = R_0 \mathcal{E}(H_t) \mathcal{E}_t^{-1}(h_t) = R_0 \mathcal{E}(H - h_t^* - [H, h^*]_t),$$

where

$$\begin{aligned} H_t - h_t^* - [H, h^*]_t &= H_t - h_t + \langle h^c, h^c \rangle_t + \sum_{0 < s \leq t} \frac{(\Delta h_s)^2}{1 + \Delta h_s} + \sum_{0 \leq s < t} \frac{(\Delta^+ h_s)^2}{1 + \Delta^+ h_s} \\ &\quad - \langle H^c, h^c \rangle_t - \sum_{0 < s \leq t} \Delta H_s \Delta h_s^* - \sum_{0 < s \leq t} \Delta^+ H_s \Delta^+ h_s^* \\ &= H_t - h_t + \langle h^c, h^c \rangle_t + \sum_{0 < s \leq t} \frac{(\Delta h_s)^2}{1 + \Delta h_s} + \sum_{0 \leq s < t} \frac{(\Delta^+ h_s)^2}{1 + \Delta^+ h_s} \\ &\quad - \langle H^c, h^c \rangle_t - \sum_{0 < s \leq t} \Delta H_s \frac{\Delta h_s}{1 + \Delta h_s} - \sum_{0 \leq s < t} \Delta^+ H_s \frac{\Delta^+ h_s}{1 + \Delta^+ h_s} \\ &= H_t - h_t + \langle h^c, h^c - H^c \rangle_t + \sum_{0 < s \leq t} \frac{\Delta h_s (\Delta h_s - \Delta H_s)}{1 + \Delta h_s} \\ &\quad + \sum_{0 \leq s < t} \frac{\Delta^+ h_s (\Delta^+ h_s - \Delta^+ H_s)}{1 + \Delta^+ h_s}. \end{aligned}$$

Let

$$J^g = \sum_{0 \leq s < t} \frac{\Delta^+ h_s (\Delta^+ h_s - \Delta^+ H_s)}{1 + \Delta^+ h_s}, \quad J^d = \sum_{0 < s \leq t} \frac{\Delta h_s (\Delta h_s - \Delta H_s)}{1 + \Delta h_s},$$

and write

$$\Psi(h_t, H_t) = H_t - h_t + \langle h^c, h^c - H^c \rangle_t + J_t^d + J_t^g.$$

Then the ratio R satisfies

$$R_t = R_0 + \int_{0+}^t R_{s-} d\Psi_s + \int_0^t R_s d\Psi_{s+}.$$

Considering the properties of stochastic integrals, we get

$$\Psi(h, H) \in \mathcal{M}_{loc}(P, \mathbb{F}) \Rightarrow R \in \mathcal{M}_{loc}(P, \mathbb{F}),$$

and if $\Delta^+ \Psi \neq -1$ and $\Delta \Psi \neq -1$, then

$$\Psi(h, H) \in \mathcal{M}_{loc}(P, \mathbb{F}) \Leftrightarrow R \in \mathcal{M}_{loc}(P, \mathbb{F}).$$

Given the decomposition (13.19) of h and H , one can write

$$\begin{aligned} \Psi(h, H) &= (A - a) + \langle m^c, m^c - M^c \rangle + (M - m) + J^d + J^g \\ &= (A - a) + \langle m^c, m^c - M^c \rangle + \tilde{J}^d + \tilde{J}^g + (M - m) \\ &\quad + (J^g - \tilde{J}^g) + (J^d - \tilde{J}^d), \end{aligned} \tag{13.20}$$

where $(M - m) + (J^g - \tilde{J}^g) + (J^d - \tilde{J}^d) \in \mathcal{M}_{loc}(P, \mathbb{F})$. Thus, $\Psi(h, H) \in \mathcal{M}_{loc}(P, \mathbb{F})$ if and only if

$$(A - a) + \langle m^c, m^c - M^c \rangle + \tilde{J}^d + \tilde{J}^g = 0.$$

If Ψ is a local optional martingale, then R is a local optional martingale, and we are done. Otherwise, we have to find a strictly positive transformation $Z \in \mathcal{M}_{loc}(P, \mathbb{F})$ that will render $ZR \in \mathcal{M}_{loc}(P, \mathbb{F})$. Z is known as the local martingale transform (deflator). Since our market ratio process R is positive by definition, it would not make financial sense to search for a local martingale transform Z that leads to $Z_t R_t \leq 0$ for some t when $R_t > 0$ for all t . Therefore, we will restrict ourselves to a set of possible local martingale transforms that are strictly positive, i.e., $Z > 0$ a.s. P . For a strictly positive Z , we can define $N \in \mathcal{M}_{loc}(P, \mathbb{F})$ with

$N = \mathcal{L}og(Z) = Z^{-1} \circ Z$ or $Z = \mathcal{E}(N)$). To find N , we have the following theorem to so:

Theorem 13.5 *Given $R = R_0 \mathcal{E}_t(\Psi(h, H))$, where $\Psi(h, H)$ is as in Eq. (13.20), and $Z = \mathcal{E}(N)$, where $Z, N \in \mathcal{M}_{loc}(P, \mathbb{F})$ and $Z > 0$, we have $ZR \in \mathcal{M}_{loc}(P, \mathbb{F})$ is a local optional martingale if and only if*

$$(A - a) + \langle m^c - N^c, m^c - M^c \rangle + \tilde{K}^d + \tilde{K}^g = 0,$$

where $\tilde{K}^d, \tilde{K}^g$ are the compensators of the processes

$$K^d = \sum_{0 < s \leq t} \frac{(\Delta h_s - \Delta N_s)(\Delta h_s - \Delta H_s)}{1 + \Delta h_s},$$

$$K^g = \sum_{0 \leq s < t} \frac{(\Delta^+ h_s - \Delta^+ N_s)(\Delta^+ h_s - \Delta^+ H_s)}{1 + \Delta^+ h_s}.$$

Proof Suppose $Z_t = \mathcal{E}(N)_t \in \mathcal{M}_{loc}(P, \mathbb{F})$, $Z_t > 0$ for all t such that $ZR \in \mathcal{M}_{loc}(P, \mathbb{F})$; then

$$\begin{aligned} ZR &= R_0 \mathcal{E}(N) \mathcal{E}(\Psi(h, H)) = \mathcal{E}(N + \Psi(h, H) + [N, \Psi(h, H)]) \\ &= R_0 \mathcal{E}_t(\Psi(h, H, N)), \end{aligned}$$

where

$$\begin{aligned} \Psi(h, H, N) &= N_t + H_t - h_t + \langle h^c, h^c - H^c \rangle_t + J_t^d + J_t^g \\ &\quad + [N, H] - [N, h] + [N, J^d] + [N, J^g] \\ &= N_t + H_t - h_t + \langle h^c, h^c - H^c \rangle_t + J_t^d + J_t^g \\ &\quad + \langle N^c, H^c \rangle_t + \sum_{0 < s \leq t} \Delta N_s \Delta H_s + \sum_{0 \leq s < t} \Delta^+ N_s \Delta^+ H_s \\ &\quad - \langle N^c, h^c \rangle_t - \sum_{0 < s \leq t} \Delta N_s \Delta h_s - \sum_{0 \leq s < t} \Delta^+ N_s \Delta^+ h_s \\ &\quad + \sum_{0 < s \leq t} \Delta N_s \frac{\Delta h_s (\Delta h_s - \Delta H_s)}{1 + \Delta h_s} \\ &\quad + \sum_{0 \leq s < t} \Delta^+ N_s \frac{\Delta^+ h_s (\Delta^+ h_s - \Delta^+ H_s)}{1 + \Delta^+ h_s}. \end{aligned}$$

Hence,

$$\begin{aligned}
\Psi(h, H, N) &= N_t + H_t - h_t + \langle h^c - N^c, h^c - H^c \rangle_t \\
&\quad + \sum_{0 < s \leq t} \frac{\Delta h_s (\Delta h_s - \Delta H_s)}{1 + \Delta h_s} + \Delta N_s (\Delta H_s - \Delta h_s) \\
&\quad + \Delta N_s \frac{\Delta h_s (\Delta h_s - \Delta H_s)}{1 + \Delta h_s} \\
&\quad + \sum_{0 < s \leq t} \frac{\Delta^+ h_s (\Delta^+ h_s - \Delta^+ H_s)}{1 + \Delta^+ h_s} + \Delta^+ N_s (\Delta^+ H_s - \Delta^+ h_s) \\
&\quad + \Delta^+ N_s \frac{\Delta^+ h_s (\Delta^+ h_s - \Delta^+ H_s)}{1 + \Delta^+ h_s}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Psi(h, H, N) &= N_t + H_t - h_t + \langle h^c - N^c, h^c - H^c \rangle_t \\
&\quad + \sum_{0 < s \leq t} \frac{(\Delta h_s - \Delta N_s)(\Delta h_s - \Delta H_s)}{1 + \Delta h_s} \\
&\quad + \sum_{0 < s \leq t} \frac{(\Delta^+ h_s - \Delta^+ N_s)(\Delta^+ h_s - \Delta^+ H_s)}{1 + \Delta^+ h_s}.
\end{aligned}$$

Let

$$\begin{aligned}
K^d &= \sum_{0 < s \leq t} \frac{(\Delta h_s - \Delta N_s)(\Delta h_s - \Delta H_s)}{1 + \Delta h_s}, \\
K^g &= \sum_{0 \leq s \leq t} \frac{(\Delta^+ h_s - \Delta^+ N_s)(\Delta^+ h_s - \Delta^+ H_s)}{1 + \Delta^+ h_s},
\end{aligned}$$

and write

$$\Psi(h, H, N) = N_t + H_t - h_t + \langle h^c - N^c, h^c - H^c \rangle_t + K^d + K^g.$$

So, if $\Psi(h, H, N) \in \mathcal{M}_{loc}(P, \mathbb{F})$, then $ZR \in \mathcal{M}_{loc}(P, \mathbb{F})$. And if $\Delta^+ \Psi(h, H, N) \neq -1$ and $\Delta \Psi(h, H, N) \neq -1$, then $\Psi(h, H, N) \in \mathcal{M}_{loc}(P, \mathbb{F}) \Leftrightarrow ZR \in \mathcal{M}_{loc}(P, \mathbb{F})$. Now, let us take into consideration the decomposition of H and h and write

$$\Psi(h, H, N) = (A - a) + (M - m + N) + \langle (m - N)^c, (m - M)^c \rangle + K^d + K^g.$$

So, $\Psi(h, H, N)$ is a local optional martingale under P if

$$(A - a) + \langle m^c - N^c, m^c - M^c \rangle + \tilde{K}^d + \tilde{K}^g = 0, \quad (13.21)$$

where $\tilde{K}^d$ and $\tilde{K}^g$ are the compensators of K^d and K^g , respectively. $\square$

By finding all $N \in \mathcal{M}_{loc}(P, \mathbb{F})$ such that the above Eq. (13.21) is valid and $\mathcal{E}(N) > 0$, we find the set of all appropriate local optional martingale deflators Z such that ZR is a local optional martingale. Note that if Z is a local martingale deflator such that ZR is a local martingale, then it is true for all self-financing strategies π .

Theorem 13.6 *If Z is a local martingale transform of R , that is, ZR is a local optional martingale, and $\pi = (\eta, \xi)$ is a self-financing portfolio which is R -integrable, then ZY_t^π is a local optional martingale.* $\square$

Proof Let Z be a local martingale deflator of R , $Z > 0$, and let $\pi = (\eta, \xi)$ be self-financing and R -integrable, and then $Y_t^\pi = Y_0 + \xi \circ R_t$, and $Z_t Y_t^\pi$ can be written as

$$\begin{aligned} d(Z_t Y_t^\pi) &= Z_t dY_t^\pi + dZ_t Y_t^\pi + d[Z_t, Y_t^\pi] \\ &= Z_t \xi_t dR_t + dZ_t (\eta_t + \xi_t R_t) + d[Z_t, Y_t^\pi] \\ &= Z_t \xi_t dR_t + dZ_t \xi_t R_t + dZ_t \eta_t + \xi_t d[Z_t, R_t] \\ &= \xi_t (Z_t dR_t + dZ_t R_t + d[Z_t, R_t]) + dZ_t \eta_t \\ &= \xi_t d(Z_t R_t) + dZ_t \eta_t. \end{aligned}$$

$\square$

This leads us to the following result

$$Z_t Y_t^\pi = \xi \circ Z_t R_t + \eta \circ Z_t.$$

$\eta \circ Z_t$ and $\xi \circ Z_t R_t$ are local optional martingales. Therefore, their sum $Z_t Y_t^\pi$ is a well-defined local optional martingale. Note that we have implicitly used the fact that η is bounded, i.e., comes from the fact that π is a self-financing and also that

$$\int_0^\infty \xi_s^2 d[ZR]_s \in A_{loc}.$$

Remark 13.4 On the other hand, if we know that there exist a Z such that ZY^π is a local optional martingale, then what can we say about the portfolio π and the product ZR ? It is reasonable to suppose that $Z = \mathcal{E}(N) > 0$, π self-financing, ξ is R -integrable and η is bounded. In this case, $\xi \circ Z_t R_t = Z_t Y_t^\pi - \eta \circ Z_t$ is a sum of two local optional martingales and therefore a local optional martingale itself, for any optional process ξ , in particular, for $\xi = 1$; therefore, ZR is a local optional martingale. $\square$

Remark 13.5 If $Z \in \mathcal{M}_{loc}(P, \mathbb{F})$ and $Z > 0$, then one can define $\tilde{P} = \int_{\Omega} Z_t dP_t$ is a new measure equivalent to P , i.e., $\tilde{P} \stackrel{loc}{\sim} P$, and $Z_t = \frac{d\tilde{P}_t}{dP_t}$. $\square$

Now, is there a way to construct N knowing Eq. (13.21). We make an educated guess, choosing N as

$$N = [\alpha_c \ \alpha_d \ \alpha_g] \circ [m^c \ m^d \ m^g]^T + [\beta_c \ \beta_d \ \beta_g] \circ [M^c \ M^d \ M^g]^T, \quad (13.22)$$

where $\alpha_c, \beta_c, \alpha_d, \beta_d \in \mathcal{P}(\mathbb{F})$ and $\alpha_g, \beta_g \in \mathcal{O}(\mathbb{F})$. Note, with α 's and β 's, the local optional martingales (m, M) span a subspace of the local optional martingales space $\mathcal{M}_{loc}(P, \mathbb{F})$. So if N takes the representation above (13.22), then what is the solution of Eq. (13.22)? Substituting for N in (13.21) by Eq. (13.22), we get

$$\begin{aligned} \langle m^c - N^c, m^c - N^c \rangle &= (1 - 2\alpha_c + \alpha_c^2) \langle m^c, m^c \rangle + \beta_c^2 \langle M^c, M^c \rangle, \\ K^d &= \sum_{0 < s \leq t} \frac{((1 - \alpha_r) \Delta m_s^r - \beta_r \Delta M_s^r)(\Delta m_s^r - \Delta M_s^r)}{1 + \Delta m_s^r}, \\ K^g &= \sum_{0 < s \leq t} \frac{(1 - \alpha_g) \Delta m_s^g - \beta_g \Delta M_s^g (\Delta^+ m_s^g - \Delta^+ M_s^g)}{1 + \Delta^+ m_s^g}. \end{aligned}$$

The compensator $(\tilde{K}^d, \tilde{K}^g)$ of K^d, K^g is not easy to compute in general and will have to be evaluated on a problem-by-problem basis.

Here we demonstrate how the coparison property of solutions of optional SDE's works to find boundaries of option prices in case of the constant elasticity of variance (CEV) market model. We extend it to a jump-diffusion CEV model. CEV mode is often used in mathematical finance to capture leverage effects and stochasticity of volatility. It is also widely used by practitioners in the financial industry for modeling equities and commodities. Consider a version of the jump-diffusion CEV model, where the stock price is said to satisfy the following stochastic integral equation:

$$\begin{aligned} S_t &= \rho \int_0^t S_{s-} ds + \sigma S_t^\alpha \cdot W_t + S_{t-} U * \left(\mu^1 - \nu^1 \right)_t + S_t U * \left(\mu^2 - \nu^2 \right)_t, \\ S_0 &= s, \end{aligned} \quad (13.23)$$

where ρ and σ are constants. W_t is a Wiener process, $\mu^1 - \nu^1$ is a compensated measure of left jumps, and $\mu^2 - \nu^2$ is a compensated measure of right jumps. For $B \in \mathcal{B}(\mathbb{R}_+)$ and $\Gamma \in \mathcal{E}$, the jump measures are defined as follows:

$$\begin{aligned} \mu^1(B \times \Gamma) &:= \# \left\{ (t, \Delta L_t^1) \in B \times \Gamma \mid t > 0 \text{ such that } \Delta L_t^1 \neq 0 \right\} \\ \mu^2(B \times \Gamma) &:= \# \left\{ (t, \Delta^+ L_t^2) \in B \times \Gamma \mid t > 0 \text{ such that } \Delta^+ L_t^2 \neq 0 \right\}, \end{aligned}$$

where L_t^1 and L_t^2 are a Poisson process and a left-continuous modification of a Poisson process with constant intensities γ^1 and γ^2 , respectively, and L^1 and L^2 are independent. Hence, $v^1 = \gamma^1 t$ and $v^2 = \gamma^2 t$.

Consider the function

$$F(x) = \frac{1}{\sigma} \int_s^x u^{-\alpha} du = \frac{x^{1-\alpha} - s^{1-\alpha}}{\sigma(1-\alpha)}, \text{ and find}$$

$$F'(x) = \frac{x^{-\alpha}}{\sigma}, \quad F''(x) = \frac{-\alpha x^{-\alpha-1}}{\sigma},$$

where $0 < \alpha < 1$.

Denote $X_t = F(S_t)$ and apply the change of variables formula. We have

$$\begin{aligned} X_t &= F'(S) \left[\rho \int_0^t S_{s-} ds + \sigma S^\alpha \cdot W_t \right] + \int_0^t \frac{\sigma^2}{2} F''(S_{s-}) S_{s-}^{2\alpha} ds \\ &\quad + [F(2S_{t-}) - F(S_{t-})] U * (\mu^1 - v^1)_t \\ &\quad + [F(2S_t) - F(S_t)] U * (\mu^2 - v^2)_t \\ &\quad + [F(2S_{t-}) - F(S_{t-}) - S_{t-} F'(S_{t-})] U * v_t^1 \\ &\quad + [F(2S_t) - F(S_t) - S_t F'(S_t)] U * v_t^2 \\ &= \int_0^t \left[k S_{s-}^{1-\alpha} - \frac{\sigma \alpha}{2} S_{s-}^{\alpha-1} \right] ds + W_t \\ &\quad + c S_{t-}^{1-\alpha} U * (\mu^1 - v^1)_t + c S_t^{1-\alpha} U * (\mu^2 - v^2)_t \\ &= \int_0^t \left[k \left(X_{s-} \sigma(1-\alpha) + s^{1-\alpha} \right) - \frac{\sigma \alpha}{2} \left(X_{s-} \sigma(1-\alpha) + s^{1-\alpha} \right)^{-1} \right] ds + W_t \\ &\quad + c \left(X_{t-} \sigma(1-\alpha) + s^{1-\alpha} \right) U * (\mu^1 - v^1)_t \\ &\quad + c \left(X_t \sigma(1-\alpha) + s^{1-\alpha} \right) U * (\mu^2 - v^2)_t, \end{aligned}$$

where $k := \frac{\rho}{\sigma} + (2^{1-\alpha} - 1 - \frac{1}{\sigma})\lambda_1 U + (2^{1-\alpha} - 1 - \frac{1}{\sigma})\lambda_2 U$ and $c := \frac{2^{1-\alpha}-1}{\sigma(1-\alpha)}$.

With the comparison theorem, we can give an estimate of the process X_t from above by a new process Y_t , satisfying the equation,

$$\begin{aligned} Y_t &= \int_0^t \left[k \left(Y_{s-} \sigma(1-\alpha) + s^{1-\alpha} \right) \right] ds + W_t \\ &\quad + c \left(Y_{t-} \sigma(1-\alpha) + s^{1-\alpha} \right) U * (\mu^1 - v^1)_t \\ &\quad + c \left(Y_t \sigma(1-\alpha) + s^{1-\alpha} \right) U * (\mu^2 - v^2)_t, \\ Y_0 &= 0. \end{aligned}$$

□

The process Y is an *Ornstein-Uhlenbeck process with left and right jumps*. The explicit solution for the above nonhomogeneous linear stochastic equation is given by the following formula:

$$Y_t = \mathcal{E}_t(H) \left[\mathcal{E}(H)^{-1} \tilde{G}_t \right],$$

where $\mathcal{E}$ is an optional stochastic exponent and

$$\begin{aligned} H_t &= k\sigma(1-\alpha)t + (2^{1-\alpha} - 1)U * (\mu^1 - \nu^1)_t + (2^{1-\alpha} - 1)U * (\mu^2 - \nu^2)_t, \\ \tilde{G}_t &= s^{1-\alpha} \left(k - c(1 - 2^{\alpha-1})\gamma^1 U - c(1 - 2^{\alpha-1})\gamma^2 U \right)_t \\ &\quad + c(s2)^{1-\alpha} U * \left[(\mu^1 - \nu^1)_t + (\mu^2 - \nu^2)_t \right] + W_t. \end{aligned}$$

Applying the comparison theorem to X_t and Y_t yields that $Y_t \geq X_t = F(S_t)$ a.s. Since $F(x)$ is monotonically increasing function, we have

$$S_t \leq F^{-1}(Y_t) \quad a.s. \quad (13.24)$$

Now let us consider a function f with an option payoff $f(S_T)$, where f is increasing. Assuming zero interest rates, the price of such option is given by $\tilde{\mathbf{E}}f(S_T)$ for an appropriate martingale measure $\tilde{\mathbf{P}}$. Using inequality (13.24), we have that $\tilde{\mathbf{E}}f(S_T) \leq \tilde{\mathbf{E}}f(F^{-1}(Y_T))$, and thus we obtain an estimate for the option price for which $\tilde{\mathbf{E}}f(F^{-1}(Y_T))$ is easier to compute.

In the remaining part of this section, we briefly discuss the *uniform Doob-Meyer decomposition* (UDMD), commonly known as the *optional decomposition*. It is a fundamental result in stochastic analysis, which is exploited in the area of mathematical finance for the construction of superhedging strategies in incomplete markets. It asserts that if, for a non-empty set, $\mathbb{Q}(X)$, of equivalent local martingale measures for the semimartingale with respect to all measures in $\mathbb{Q}(X)$, then there exists a unique predictable stochastic integrated φ such that the difference $Y - \varphi \cdot X$ is a decreasing optional process for all measures in $\mathbb{Q}(X)$. In contrast to the classical Doob-Meyer decomposition, the process $Y - \varphi \cdot X$ is optional and is not uniquely determined. Moreover, UDMD is *universal*, meaning that it holds simultaneously for all probability measures in $\mathbb{Q}(X)$. Formally, it can be presented as the following theorem.

Theorem 13.7 *Let X be an $\mathbb{R}^d$ -valued r.l.l. optional semimartingale on the unusual stochastic basis $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t), P)$. Let $\mathbb{Q}(X)$ be the set of all probability measures $Q \sim P$ such that X is a Q -local optional martingale. Suppose that $\mathbb{Q}(X) \neq \emptyset$ and Y is a nonnegative process. Then Y is a Q -local optional supermartingale for each $Q \in \mathbb{Q}(X)$ if and only if there exists a unique optional process φ integrable with respect to X and an adapted optional increasing process C with $C_0 = 0$ such that $Y = Y_0 + \varphi \cdot X - C$ for each $Q \in \mathbb{Q}(X)$. $\square$*

Chapter 14

Optional Semimartingales for Stochastic Regression Analysis and Risk Theory



This chapter shows how optional processes work in the area of regression analysis and risk theory. We introduce an optional regression model and study the LS-estimates and sequential least-squares (LS)-estimates in this framework. We derive a very general estimate of the ruin probability for an insurer company when its capital is formally on the base of an optional semimartingale. We show how these general results work for several classical models of regression analysis and risk theory (see [5, 6, 53], and [58]).

14.1 Optional Regression Analysis

Suppose that on the fixed stochastic basis $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbf{P})$ without “usual conditions”, we observe a one-dimensional process X .

Suppose the process X has the following form:

$$X_t = f \circ a_t \theta + M_t, \quad (14.1)$$

where $f \circ a_t$ is an optional stochastic integral such that

$$f \circ a_t = \int_{[0,t]} f_s^r da_s^r + \int_{[0,t]} f_s^g da_{s+}^g.$$

$a = a^r + a^g \in \mathcal{A}_{loc}^+ \cap \mathcal{P}_s$, $M \in \mathcal{M}_{loc}$, f_t is a bilinear pair $f_t = (f_t^r, f_t^g)$, $f_t^r \in \mathcal{P}$, $f_t^g \in \mathcal{O}$, and $\theta \in \mathbb{R}$ is the unknown parameter which we need to estimate.

As the estimator of θ in the model (14.1), we consider the statistic

$$\theta_t = F_t^{-1}(f \circ X_t), \quad (14.2)$$

where $F_t := f^2 \circ a_t \in \mathcal{A}_{loc}^+ \cap \mathcal{P}_s$ is assumed to be nonzero (a.s.). This assumption is not restrictive because further we suppose that $F_t \rightarrow \infty$ (a.s.) to provide strong consistency of θ_t .

The structure of the estimator (14.2) is similar to estimator obtained by the method of least squares (LS) in classical regression analysis. Therefore, θ_t will be called the *structural LS-estimator* of θ . It is well-known how to study its asymptotic behavior with the help of the strong law of large numbers (SLLN). Here we develop a very general form of SLLN for optimal local martingales using an optimal stochastic Kronecker's Lemma.

To prove Kronecker's Lemma in optional setting, we need the following result on sets of convergence of optional martingales. In what follows, we denote $\tilde{D}$ the compensator of some increasing process D .

Lemma 14.1 *If $Y \in \mathcal{M}_{loc}$, then*

$$(\tilde{D}_\infty < \infty) \subseteq (Y \rightarrow) \text{ a.s.},$$

where

$$D_t = \langle Y^c \rangle_t + \sum_{0 < s \leq t} \frac{(\Delta Y_s)^2}{1 + |\Delta Y_s|} + \sum_{0 \leq s < t} \frac{(\Delta^+ Y_s)^2}{1 + |\Delta^+ Y_s|}, \quad (14.3)$$

and $(Y \rightarrow)$ is the set, on which there exists a finite random variable $Y_\infty(\omega) = \lim_{t \rightarrow \infty} Y_t(\omega) < \infty$. $\square$

Now we present the following generalization of Kronecker's Lemma.

Lemma 14.2 *For processes $N \in \mathcal{M}_{loc}$ and $A \in \mathcal{V}^+ \cap \mathcal{P}_s$, the following relation holds*

$$(A_\infty = \infty) \cap (Y_t \rightarrow) \subseteq (A_t^{-1} N_t \rightarrow 0) \text{ (a.s.) } (t \rightarrow \infty),$$

where

$$Y_t = \int_{]0,t]} (1 + A_s)^{-1} dN_s^r + \int_{[0,t[} (1 + A_{s+})^{-1} dN_{s+}^g. \quad (14.4)$$

Proof From (14.4), it is easy to see that

$$\int_{]0,t]} (1 + A_s) dY_s^r + \int_{[0,t[} (1 + A_{s+}) dY_{s+}^g = N_t - N_0. \quad (14.5)$$

Using integration by parts formula, we obtain

$$(1 + A_t) Y_t = \int_{]0,t]} (1 + A_s) dY_s^r + \int_{[0,t[} (1 + A_{s+}) dY_{s+}^g + \int_{]0,t]} Y_s - dA_s^r + \int_{[0,t[} Y_s dA_{s+}^g.$$

Then from (14.5), we conclude that

$$\frac{N_t}{1 + A_t} = \frac{N_0 + Y_t}{1 + A_t} + \frac{1}{1 + A_t} (A_t Y_t - \int_{[0,t]} Y_{s-} dA_s^r - \int_{[0,t]} Y_s dA_{s+}^g).$$

Since $\sup_{t \geq 0} |Y_t| < \infty$ on the set $(A_\infty = \infty) \cap (Y \rightarrow)$, we have that $(1 + A_t)^{-1}(N_0 + Y_t) \rightarrow 0$ a.s. as $t \rightarrow \infty$. On the other hand, we have for $u < t, v < t$

$$\begin{aligned} & \frac{1}{1 + A_t} \left| A_t Y_t - \int_{[0,t]} Y_{s-} dA_s^r - \int_{[0,t]} Y_s dA_{s+}^g \right| \\ &= \frac{1}{1 + A_t} \left| \int_{[0,t]} (Y_t - Y_{s-}) dA_s^r + \int_{[0,t]} (Y_t - Y_s) dA_{s+}^g \right| \\ &\leq \frac{1}{1 + A_t} \left(\int_{[0,u]} |Y_t - Y_{s-}| dA_s^r + \int_{[u,t]} (|Y_\infty - Y_t| + |Y_\infty - Y_{s-}|) dA_s^r \right. \\ &\quad \left. + \int_{[0,v]} |Y_t - Y_s| dA_{s+}^g + \int_{[v,t]} (|Y_\infty - Y_t| + |Y_\infty - Y_s|) dA_{s+}^g \right) \\ &\leq 2 \sup_{s \geq 0} |Y_s| (1 + A_t)^{-1} (A_u^r + A_v^g) + |Y_\infty - Y_t| \\ &\quad + (1 + A_t)^{-1} \int_{[u,t]} |Y_\infty - Y_{s-}| dA_s^r + (1 + A_t)^{-1} \int_{[v,t]} |Y_\infty - Y_s| dA_{s+}^g. \end{aligned} \tag{14.6}$$

Using the fact that $\sup_{s \geq 0} |Y_s| < \infty$ on the set $(Y \rightarrow)$, we can choose for sufficiently large t appropriate values u and v on the set $(A_\infty = \infty) \cap (Y \rightarrow)$ to make the right side of (14.6) tend to zero. $\square$

Remark 14.1 We proved Lemma 14.2 for the process $N \in \mathcal{M}_{loc}$, but the proof also works for optional semimartingales under related conditions. $\square$

We will show now that the estimator θ_t in (14.2) is strongly consistent. The proof of the strong consistency is based on the SLLN in optional case.

Let $N \in \mathcal{M}_{loc}$ and

$$N_t = N_t^c + \int_{[0,\infty]} \int_{\mathbb{R}_0} x d(\mu^r - \nu^r)_s + \int_{[0,\infty]} \int_{\mathbb{R}_0} x d(\mu^g - \nu^g)_{s+} \tag{14.7}$$

be the canonical decomposition of N , where $\mathbb{R}_0 = \mathbb{R} \setminus \{0\}$, N^c be a continuous part of N , μ^r and μ^g be random measures of right and left jumps of N , and ν^r and ν^g be their respective compensators.

Theorem 14.1 Let $A \in \mathcal{V}^+ \cap \mathcal{P}_s$ and $A_\infty = \infty$ a.s. If $N \in \mathcal{M}_{loc}$ and for some $q \in [1, 2]$

$$\begin{aligned} \int_{[0, \infty]} \frac{d\langle N^c \rangle_s}{(1 + A_s)^2} + \int_{[0, \infty]} \int_{\mathbb{R}_0} |1 + A_s|^{-q} |x|^q d\nu_s^r \\ + \int_{[0, \infty[} \int_{\mathbb{R}_0} |1 + A_{s+}|^{-q} |x|^q d\nu_{s+}^g < \infty, \end{aligned} \quad (14.8)$$

then

$$A_t^{-1} N_t \rightarrow 0 \text{ a.s. as } t \rightarrow \infty.$$

Proof Using the fact that

$$\frac{x^2}{1 + |x|} \leq |x|^q, \quad q \in [1, 2],$$

we get for $q \in [1, 2]$

$$\begin{aligned} \tilde{D}_\infty &= \langle Y^c \rangle_\infty + \sum_{s \leq \infty} \frac{\widetilde{(\Delta Y_s)^2}}{1 + |\Delta Y_s|} + \sum_{s < \infty} \frac{\widetilde{(\Delta^+ Y_s)^2}}{1 + |\Delta^+ Y_s|} \\ &\leq \langle Y^c \rangle_\infty + \sum_{s \leq \infty} \widetilde{(\Delta Y_s)^q} + \sum_{s < \infty} \widetilde{(\Delta^+ Y_s)^q}, \end{aligned}$$

where Y as defined in (14.4). Thus, from (14.8), it follows that

$$\begin{aligned} \tilde{D}_\infty &\leq \int_{[0, \infty]} \frac{d\langle N^c \rangle_s}{(1 + A_s)^2} + \int_{[0, \infty]} \int_{\mathbb{R}_0} |1 + A_s|^{-q} |x|^q d\nu_s^r \\ &\quad + \int_{[0, \infty[} \int_{\mathbb{R}_0} |1 + A_{s+}|^{-q} |x|^q d\nu_{s+}^g < \infty. \end{aligned}$$

By Lemma 14.2, $A_t^{-1} N_t \rightarrow 0$ a.s. as $t \rightarrow \infty$. □

Theorem 14.2 Suppose for the model (14.1) that $F_\infty = \infty$ and for some $q \in [1, 2]$

$$\begin{aligned} \int_{[0, \infty]} \frac{(f_s^r)^2 d\langle M^c \rangle_s}{(1 + F_s)^2} + \int_{[0, \infty]} \int_{\mathbb{R}_0} |1 + F_s^r|^{-q} |f_s^r|^q |x|^q d\nu_s^r \\ + \int_{[0, \infty[} \int_{\mathbb{R}_0} |1 + F_s^g|^{-q} |f_s^g|^q |x|^q d\nu_{s+}^g < \infty. \end{aligned} \quad (14.9)$$

Then $\theta_t \rightarrow \theta$ (a.s.) as $t \rightarrow \infty$. □

Proof It is sufficient to note that

$$\theta_t - \theta = A_t^{-1} N_t,$$

where $A_t := F_t$ and $N_t := f \circ M_t$. By Theorem 14.1, we obtain the statement of the theorem. $\square$

Let us consider the model (14.1) with $M \in \mathcal{M}_{loc}^2(\mathbb{R})$. We assume that there exists a nonnegative random variable ξ such that

$$\frac{d\langle M \rangle_t}{da_t} \leq \xi, \quad f^2 \circ a_t \in \mathcal{A}_{loc}^+ \cap \mathcal{P}_s. \quad (14.10)$$

Next, for fixed H , we define

$$\tau_H = \inf\{t : f^2 \circ a_t \geq H\}, \quad (14.11)$$

with $\tau_H = \infty$ if the corresponding set is empty. We assumed that processes $f^r \in \mathcal{P}$, $f^g \in \mathcal{O}$, and $a \in \mathcal{P}_s$, consequently, and hence τ_H is a wide sense stopping time.

On the set $\{\tau_H < \infty\}$, we define a random variable β_H by the relation

$$f^2 \circ a_{\tau_H-} + \beta_H \left((f_{\tau_H}^r)^2 \Delta a_{\tau_H} + (f_{\tau_H}^g)^2 \Delta^+ a_{\tau_H} \right) = H, \quad (14.12)$$

and on $\tau_H = \infty$, we put $\beta_H = 0$. Then $\beta_H \in [0, 1]$ and is a $\mathcal{F}_{\tau_H}$ -measurable random variable.

We consider the following statistic as an estimator of θ

$$\hat{\theta}_H = H^{-1} \left[f^2 \circ X_{\tau_H-} + \beta_H \left((f_{\tau_H}^r)^2 \Delta X_{\tau_H} + (f_{\tau_H}^g)^2 \Delta^+ X_{\tau_H} \right) \right]. \quad (14.13)$$

The next theorem shows that the statistic defined by means of (14.10)–(14.13) is an unbiased estimator of θ and has the property of guaranteed accuracy, i.e., bounded variance.

Theorem 14.3 Suppose that assumptions (14.10) hold, $\mathbf{E}\xi < \infty$, and

$$\mathbf{P}\{f^2 \circ a_\infty = \infty\} = 1. \quad (14.14)$$

Then for all $H > 0$,

$$\mathbf{P}\{\tau_H < \infty\} = 1, \quad \mathbf{E}\hat{\theta}_H = \theta, \quad \mathbf{Var}\hat{\theta}_H \leq H^{-1}\mathbf{E}\xi.$$

Proof First,

$$\mathbf{P}\{\tau_H = \infty\} = \mathbf{P}\{f^2 \circ a_\infty < H\}$$

$$\begin{aligned}
&= 1 - \mathbf{P}\{f^2 \circ a_\infty \geq H\} \\
&\leq 1 - \mathbf{P}\{f^2 \circ a_\infty = \infty\} \\
&= 1 - 1 = 0.
\end{aligned}$$

Thus,

$$\mathbf{P}\{\tau_H < \infty\} = 1 - \mathbf{P}\{\tau_H = \infty\} = 1.$$

Next, using (14.1) and (14.13), we obtain

$$\begin{aligned}
\hat{\theta}_H &= H^{-1} \left[f^2 \circ X_{\tau_H-} + \beta_H \left((f_{\tau_H}^r)^2 \Delta X_{\tau_H} + (f_{\tau_H}^g)^2 \Delta^+ X_{\tau_H} \right) \right] \\
&= H^{-1} \left[(f^2 \circ a_{\tau_H-}) \theta + \beta_H \left((f_{\tau_H}^r)^2 \Delta a_{\tau_H} + (f_{\tau_H}^g)^2 \Delta^+ a_{\tau_H} \right) \theta \right. \\
&\quad \left. + f^2 \circ M_{\tau_H-} + \beta_H \left((f_{\tau_H}^r)^2 \Delta M_{\tau_H} + (f_{\tau_H}^g)^2 \Delta^+ M_{\tau_H} \right) \right] \\
&= \theta + H^{-1} N_{\tau_H},
\end{aligned}$$

where

$$N_t = \mathbf{1}_{\{t < \tau_H\}} f^2 \circ M_t + \mathbf{1}_{\{t = \tau_H\}} \beta_H \left((f_{\tau_H}^r)^2 \Delta M_{\tau_H} + (f_{\tau_H}^g)^2 \Delta^+ M_{\tau_H} \right)$$

Since the process N_t is a stochastic integral with respect to the optional square integrable local martingale M , by the properties of optional stochastic integrals, we have

$$\langle N \rangle_t = \mathbf{1}_{\{t < \tau_H\}} f^2 \circ \langle M \rangle_t + \mathbf{1}_{\{t = \tau_H\}} \beta_H^2 \left((f_{\tau_H}^r)^2 \Delta \langle M \rangle_{\tau_H} + (f_{\tau_H}^g)^2 \Delta^+ \langle M \rangle_{\tau_H} \right).$$

Hence, by (14.10) and (14.12), we get

$$\langle N \rangle_{\tau_H} \leq \xi \left[f^2 \circ a_{\tau_H-} + \beta_H \left((f_{\tau_H}^r)^2 \Delta a_{\tau_H} + (f_{\tau_H}^g)^2 \Delta^+ a_{\tau_H} \right) \right] = \xi H.$$

Consequently, $N_{t \wedge \tau_H}$ is an optional square integrable martingale, and therefore

$$\mathbf{E} N_{\tau_H} = 0, \quad \mathbf{E} N_{\tau_H}^2 \leq H \mathbf{E} \xi,$$

which proves the theorem. $\square$

Let us illustrate our findings with examples.

Example 14.1 (Risk Process) Consider the following risk process:

$$X_t = c \cdot t + \sigma W_t - a N_t^r + b N_t^g \quad (14.15)$$

where c, σ, a, b are positive constants, W is a Wiener process, and N^r and N^g are a Poisson process and left-continuous modification of a Poisson process, respectively. The constant c usually represents premium payments in risk theory, whereas a and b describe average value of claims and positive gains, respectively. The process σW_t is a random perturbation.

We can rewrite the process X_t as follows:

$$X_t = \theta t + M_t,$$

where $\theta := c - a\lambda^r + b\lambda^g$, $M_t := \sigma W_t - a(N_t^r - \lambda^r t) + b(N_t^g - \lambda^g t)$, λ^r and λ^g are jump intensities of N_t^r and N_t^g , respectively.

The structural LS-estimator of θ in the model (14.15) is

$$\theta_t = \frac{X_t}{t}. \quad (14.16)$$

The condition (14.9) of Theorem 3.2, i.e.,

$$\sigma^2 \int_{[0, \infty]} \frac{ds}{(1+s)^2} + a^q \lambda^r \int_{[0, \infty]} (1+s)^{-q} ds < \infty,$$

holds for any $q \in (1, 2]$. Thus, the estimator (14.16) is strongly consistent.

The sequential LS-estimators have the following form:

$$\hat{\theta}_H = \frac{X_{\tau_H}}{H}.$$

All assumptions of Theorem 14.3 are obviously satisfied, and $\xi = \sigma^2 + a^2\lambda^r + b^2\lambda^g$. Consequently,

$$\mathbf{P}\{\tau_H < \infty\} = 1, \quad \mathbf{E}\hat{\theta}_H = \theta, \quad \mathbf{Var}\hat{\theta}_H \leq \frac{\sigma^2 + a^2\lambda^r + b^2\lambda^g}{H}.$$

We usually want the process X_t to be positive, so the estimator $\theta_t = c - a\lambda^r + b\lambda^g > 0$. This assertion is called a net profit condition in risk theory.

Example 14.2 (Ornstein-Uhlenbeck Process) Let us consider Ornstein-Uhlenbeck processes

$$X_t = \int_0^t (\mu - X_{s-}) ds + M_t,$$

where μ is a positive constant, and $M \in \mathcal{M}_{loc}^2$.

We assume that

$$F_t := \int_0^t (\mu - X_{s-})^2 ds \in \mathcal{A}_{loc}^+ \cap \mathcal{P}_s \quad \text{and} \quad \frac{d\langle M \rangle_t}{dt} \leq \xi.$$

Then, the structural LS-estimator of θ is

$$\theta_t = \frac{(\mu - X_-) \circ X_t}{\int_0^t (\mu - X_{s-})^2 ds}, \quad (14.17)$$

and its sequential version has the following form:

$$\hat{\theta}_H = H^{-1} \left[(\mu - X_{s-})^2 \circ X_{\tau_H-} \right].$$

If the following condition

$$\int_{[0, \infty]} \frac{(\mu - X_{s-})^2 d\langle M^c \rangle_s}{(1 + F_s)^2} + \int_{[0, \infty]} \int_{\mathbb{R}_0} |1 + F_s^r|^{-q} |\mu - X_{s-}|^q |x|^q dv_s^r < \infty$$

holds for some $q \in [1, 2]$ and $F_\infty = \infty$, then by Theorem 3.2, the estimator (14.17) is strongly consistent.

Furthermore, by Theorem 14.3, we have

$$\mathbf{P}\{\tau_H < \infty\} = 1, \quad \mathbf{E}\hat{\theta}_H = \theta, \quad \mathbf{Var}\hat{\theta}_H \leq H^{-1} \mathbf{E}\xi.$$

Example 14.3 Consider a regression model with the centered Gaussian martingale $M \in \mathcal{M}_{loc}^2$ and a deterministic function f_t :

$$X_t = \int_0^t f_s ds \theta + M_t. \quad (14.18)$$

Then, the LS-estimator of θ is

$$\theta_t = \frac{f \circ X_t}{\int_0^t f_s^2 ds}. \quad (14.19)$$

It can be shown that strong consistency of (14.19) in (14.18) follows only from the assumption of $\int_0^\infty f_s^2 ds = \infty$.

We assume that

$$\frac{d\langle M \rangle_t}{dt} \leq \xi,$$

where ξ is constant. Note that in the case of centered Gaussian martingales, $\langle M \rangle_t = \mathbf{E}M_t^2 < \infty$ is a deterministic function. Then sequential LS-estimators have the

following form:

$$\hat{\theta}_H = H^{-1} \left[f^2 \circ X_{\tau_H-} \right].$$

If, in addition,

$$\mathbf{P} \left\{ \int_0^\infty f_s^2 ds = \infty \right\} = 1,$$

then, by Theorem 14.3, we have

$$\mathbf{P}\{\tau_H < \infty\} = 1, \quad \mathbf{E}\hat{\theta}_H = \theta, \quad \mathbf{Var}\hat{\theta}_H \leq H^{-1}\xi.$$

14.2 Optional Risk Model and Estimation of Ruin Probability

Let $(\Omega, \mathcal{F}, \mathbf{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbf{P})$ be an *unusual* probability space on which we consider a risk process. Elements of this process are the following components: premium payments, returns on investments, and payments of liabilities, costs, and claims. These elements are inherently random. So, let us consider a risk process whose flow can be summarized by the following equation:

$$R_t = u + B_t + N_t + D_t + L_t, \quad (14.20)$$

where $u > 0$ is the initial capital and $B_0 = W_0 = D_0 = L_0 = 0$.

The process B is a predictable process of finite variation characterizing a stable flow of income payments including premiums and other sources, N is a continuous local martingale representing a random perturbation, D and L are right-continuous and left-continuous jump processes, respectively. The process L may model, for example, the dividend payments or capital injections. The process D includes a sum of negative jumps representing accumulated claims. In addition, D may also consist of jumps formed by any non-anticipated shocks such as, for example, shocks related to the currency risk. Note also that the jump processes D and L are not restricted by their sign and can have positive or negative jumps in general. All these processes are optional and adapted to the filtration $\mathcal{F}_t$.

Let us consider a particular case of (14.20) is a capital process of some company at time t , and

$$R_t = u + ct + \sigma W_t + \sum_{k=1}^{N_t^r} Y_k - \sum_{i=1}^{N_t^s} Z_i,$$

where c and σ are some constant parameters, W is a Wiener process, N^r and N^g are a Poisson process and left-continuous modification of a Poisson process with intensities λ^r and λ^g , respectively, and Z_i and Y_k denote the left and right jump sizes, respectively, with some specified distribution. In this case, the process ct is included in the income process B , $N_t = \sigma W_t$, while $L_t = \sum_{i=1}^{N_t^r} Z_i$ and $\sum_{k=1}^{N_t^g} Y_k$ is a part of the process D .

Let $\mu^r(\omega, dt, dx)$ and $\mu^g(\omega, dt, dx)$ be random measures that describe jumps of the process D_t and L_t , respectively. We assume that

$$\int_{[0,t]} \int_{\mathbb{R}_0} x d\mu^r \in \mathcal{A}_{loc} \text{ and } \int_{[0,t]} \int_{\mathbb{R}_0} x d\mu^g \in \mathcal{A}_{loc}.$$

This assumption implies that the processes

$$\int_{[0,t]} \int_{\mathbb{R}_0} x d(\mu^r - \nu^r) \quad \text{and} \quad \int_{[0,t]} \int_{\mathbb{R}_0} x d(\mu^g - \nu^g)$$

are optional local martingales, where ν^r, ν^g are compensators of μ^r, μ^g , respectively.

To find an upper bound of probability of ruin, we first obtain the martingale characterization of optional semimartingales by means of stochastic exponents. Consider an optional semimartingale X (e.g., risk process) with the local characteristics $(a, \langle X^c \rangle, \nu^r, \nu^g)$ and the following representation:

$$\begin{aligned} X_t = u + a_t + X_t^c + \int_{[0,t]} \int_{|x| \leq 1} x d(\mu^r - \nu^r) + \int_{[0,t]} \int_{|x| \leq 1} x d(\mu^g - \nu^g) \\ + \int_{[0,t]} \int_{|x| > 1} x d\mu^r + \int_{[0,t]} \int_{|x| > 1} x d\mu^g, \end{aligned} \quad (14.21)$$

where $a_t \in \mathcal{P}_s$, $X_t^c \in \mathcal{M}_{loc}^c$, $\int_{[0,t]} \int_{|x| \leq 1} x d(\mu^r - \nu^r) \in \mathcal{M}_{loc}^r$ and $\int_{[0,t]} \int_{|x| \leq 1} x d(\mu^g - \nu^g) \in \mathcal{M}_{loc}^g$.

In particular, for the risk process R ,

$$a_t = B_t + \int_{[0,t]} \int_{|x| \leq 1} x d\nu^r + \int_{[0,t]} \int_{|x| \leq 1} x d\nu^g, \quad \langle X^c \rangle_t = \langle N \rangle_t.$$

We introduce an optional cumulant function for X :

$$\begin{aligned} G_t(z) = -za_t + \frac{z^2}{2} \langle X^c \rangle_t + \int_{[0,t]} \int_{\mathbb{R}_0} (e^{-zx} - 1 + zx \mathbf{1}_{(|x| \leq 1)}) d\nu_s^r \\ + \int_{[0,t]} \int_{\mathbb{R}_0} (e^{-zx} - 1 + zx \mathbf{1}_{(|x| \leq 1)}) d\nu_{s+}^g \end{aligned} \quad (14.22)$$

with the corresponding optional stochastic exponent $(\mathcal{E}(G(z)) = \mathcal{E}_t(G(z)))_{t \geq 0}$.

Let us discuss when the optional stochastic cumulant function $G_t(z)$ in (14.22) is well-defined. Let

$$I_+ = \int_{]0,t]} \int_0^\infty (e^{-zx} - 1 + zx \mathbf{1}_{(|x| \leq 1)}) dv_s^r,$$

$$I_- = \int_{]0,t]} \int_{-\infty}^0 (e^{-zx} - 1 + zx \mathbf{1}_{(|x| \leq 1)}) dv_s^r.$$

Using Taylor's formula, we have

$$\begin{aligned} I_- &\leq \int_{]0,t]} \int_{-1}^0 |e^{-zx} - 1 + zx| dv_s^r + \int_{]0,t]} \int_{-\infty}^{-1} |e^{-zx} - 1| dv_s^r \\ &\leq \frac{z^2}{2} \int_{]0,t]} \int_{-1}^0 x^2 dv_s^r + \int_{]0,t]} \int_{-\infty}^{-1} e^{-zx} dv_s^r; \\ I_+ &\leq \int_{]0,t]} \int_0^1 |e^{-zx} - 1 + zx| dv_s^r + \int_{]0,t]} \int_1^\infty |e^{-zx} - 1| dv_s^r \\ &\leq \frac{z^2}{2} \int_{]0,t]} \int_0^1 x^2 dv_s^r + v_t^r([1, +\infty)). \end{aligned}$$

The same inequalities hold for the integral $\int_{]0,t]} \int_{\mathbb{R}_0} (e^{-zx} - 1 + zx \mathbf{1}_{(|x| \leq 1)}) dv_{s+}^g$ in (14.22). We can see that $G_t(z)$ is well-defined, if, in addition to the assumption $v_t^j([1, +\infty)) < \infty$, $j = r, g$, there exists $z_0 > 0$ such that

$$\int_{]0,t]} \int_{-\infty}^{-1} e^{-zx} dv^r < \infty, \quad \int_{]0,t]} \int_{-\infty}^{-1} e^{-zx} dv^g < \infty \quad (14.23)$$

almost surely for all $t > 0$ and $0 < z \leq z_0$.

Denote

$$T(z) = \inf(t : |\mathcal{E}_t(G(z))| = 0)$$

and

$$Z_t = e^{-z(X_t - X_0)} \mathcal{E}_t^{-1}(G(z)) \mathbf{1}_{(|\mathcal{E}_t(G(z))| > 0)}.$$

The next theorem contains a key technical result for ruin probability estimation.

Theorem 14.4 *For every $z \in [0, z_0]$, the process $Z(z) = (Z_{t \wedge T(z)})_{t \geq 0}$ is an optional local martingale.* $\square$

The following lemmas present some technicalities from optional stochastic analysis.

Lemma 14.3 *Let X be an $\mathcal{O}(\mathbf{F}_+)$ -measurable process. There exists a unique (up to indistinguishability) process $U \in \mathcal{M}_{loc}^g$ with the property $\Delta^+ U = X$ if and only if the following conditions are satisfied:*

(a) *The $\mathcal{O}(\mathbf{F})$ -optional projection of X is zero.*

(b) *the process $(\sum_{s \leq t} X_s^2)^{1/2} \in \mathcal{A}_{loc}$.* □

Lemma 14.4 *Let $M \in \mathcal{M}_{loc}^g$ and $Y \in \mathcal{V} \cap \mathcal{O}$. Then*

$$\sum_{s < t} \Delta^+ Y_s \Delta^+ M_s = \int_{[0, t[} \Delta^+ Y_s dM_{s+}.$$

Proof We use a simple fact:

$$\Delta^+ U := \Delta^+ \left(\sum_{s < t} \Delta^+ Y_s \Delta^+ M_s - \int_{[0, t[} \Delta^+ Y_s dM_s \right) = 0.$$

Notice that integral in the above expression is well-defined. Let $X := \Delta^+ U$ as in the Lemma 14.3. It is easily seen that all sufficient conditions of Lemma 14.3 are satisfied and, thus, $U \in \mathcal{M}_{loc}^g$. On the other hand, since $\Delta^+ U = 0$, it follows that U is continuous. Therefore, $U = 0$ or

$$\sum_{s < t} \Delta^+ Y_s \Delta^+ M_s = \int_{[0, t[} \Delta^+ Y_s dM_{s+}.$$

□

Lemma 14.5 *If X is a semimartingale and $Y^r \in \mathcal{V} \cap \mathcal{P}$ and $Y^g \in \mathcal{V} \cap \mathcal{O}$, then $[X, Y] = \int_{[0, t]} \Delta Y_s dX_s^r + \int_{[0, t[} \Delta^+ Y_s dX_{s+}^g$ and*

$$X_t Y_t - X_0 Y_0 = \int_{[0, t]} Y_s dX_s^r + \int_{[0, t]} X_{s-} dY_s^r + \int_{[0, t[} Y_{s+} dX_{s+}^g + \int_{[0, t[} X_s dY_{s+}^g$$

Proof First, we obtain by Lemma 14.4

$$\begin{aligned} [X, Y]_t &= \sum_{s \leq t} \Delta Y_s \Delta X_s + \sum_{s < t} \Delta^+ Y_s \Delta^+ X_s \\ &= \sum_{s \leq t} \Delta Y_s \Delta (M^r + A^r)_s + \sum_{s < t} \Delta^+ Y_s \Delta^+ (M^g + A^g)_s \\ &= \int_{[0, t]} \Delta Y_s dM_s^r + \int_{[0, t]} \Delta Y_s dA_s^r \\ &\quad + \int_{[0, t[} \Delta^+ Y_s dM_{s+}^g + \int_{[0, t[} \Delta^+ Y_s dA_{s+}^g \end{aligned}$$

$$= \int_{]0,t]} \Delta Y_s dX_s^r + \int_{]0,t[} \Delta^+ Y_s dX_{s+}^g.$$

Next, using integration by parts, we get

$$\begin{aligned} X_t Y_t - X_0 Y_0 &= \int_{]0,t]} Y_{s-} dX_s^r + \int_{]0,t]} X_{s-} dY_s^r \\ &\quad + \int_{]0,t[} Y_s dX_{s+}^g + \int_{]0,t[} X_s dY_{s+}^g + [X, Y]_t \\ &= \int_{]0,t]} Y_s dX_s^r + \int_{]0,t]} X_{s-} dY_s^r \\ &\quad + \int_{]0,t[} Y_{s+} dX_{s+}^g + \int_{]0,t[} X_s dY_{s+}^g. \end{aligned}$$

□

Proof of Lemma 14.4 In all considerations below, we will fix the parameter z and write $T, G, \dots$ instead of $T(z), G(z), \dots$. In addition, we define $t := t \wedge T$ for convenience.

Next,

$$Z_t = e^{-z(X_t - X_0)} \mathcal{E}_t^{-1}(G),$$

where

$$\begin{aligned} \mathcal{E}_t^{-1}(G) &= \exp \left(- \int_{]0,t]} \frac{dG_s^r}{1 + \Delta G_s} - \int_{]0,t[} \frac{dG_{s+}^g}{1 + \Delta^+ G_s} \right) \\ &\quad \times \prod_{0 < s \leq t} \left(1 - \frac{\Delta G_s}{1 + \Delta G_s} \right) e^{\frac{\Delta G_s}{1 + \Delta G_s}} \prod_{0 \leq s < t} \left(1 - \frac{\Delta^+ G_s}{1 + \Delta^+ G_s} \right) e^{\frac{\Delta^+ G_s}{1 + \Delta^+ G_s}}. \end{aligned}$$

By this representation, it follows that $\mathcal{E}_t^{-1}(G)$ is the solution of Doleans equation:

$$\mathcal{E}_t^{-1}(G) = 1 - \int_{]0,t]} \mathcal{E}_{s-}^{-1}(G) \frac{dG_s^r}{1 + \Delta G_s} - \int_{]0,t[} \mathcal{E}_s^{-1}(G) \frac{dG_{s+}^g}{1 + \Delta^+ G_s} \quad (14.24)$$

Using (14.21), (14.22) and change of variables formula for optional semimartingales, we get

$$\begin{aligned} L_t &= e^{-rz(X_t - X_0)} \\ &= 1 + \int_{]0,t]} L_{s-} dG_s^r + \int_{]0,t[} L_s dG_{s+}^g - z \int_{]0,t]} L_{s-} dX_s^c \end{aligned}$$

$$\begin{aligned}
& + \int_{]0,t]} \int_{\mathbb{R}_0} L_{s-} (e^{-zx} - 1) d(\mu^r - \nu^r)_s \\
& + \int_{[0,t[} \int_{\mathbb{R}_0} L_s (e^{-zx} - 1) d(\mu^g - \nu^g)_{s+}.
\end{aligned} \tag{14.25}$$

By (14.24) and (14.25) and Lemma 14.5, we find

$$\begin{aligned}
Z_t &= L_t \mathcal{E}_t^{-1}(G) \\
&= 1 + \int_{]0,t]} \mathcal{E}_s^{-1}(G) dL_s^r + \int_{]0,t]} L_{s-} d(\mathcal{E}_s^{-1}(G))^r \\
&\quad + \int_{[0,t[} \mathcal{E}_{s+}^{-1}(G) dL_{s+}^g + \int_{[0,t[} L_s d(\mathcal{E}_{s+}^{-1}(G))^g \\
&= 1 + \int_{]0,t]} \frac{\mathcal{E}_{s-}^{-1}(G)}{1 + \Delta G_s} \left[L_{s-} dG_s^r - z L_{s-} dX_s^c + \int_{\mathbb{R}_0} L_{s-} (e^{-zx} - 1) d(\mu^r - \nu^r)_s \right] \\
&\quad + \int_{[0,t[} \frac{\mathcal{E}_s^{-1}(G)}{1 + \Delta^+ G_s} \left[L_s dG_{s+}^g + \int_{\mathbb{R}_0} L_s (e^{-zx} - 1) d(\mu^g - \nu^g)_{s+} \right] \\
&\quad - \int_{]0,t]} \mathcal{E}_{s-}^{-1}(G) L_{s-} \frac{dG_s^r}{1 + \Delta G_s} - \int_{[0,t[} \mathcal{E}_s^{-1}(G) L_s \frac{dG_{s+}^g}{1 + \Delta^+ G_s} \\
&= 1 - z \int_{]0,t]} \frac{Z_{s-}}{1 + \Delta G_s} dX_s^c + \int_{]0,t]} \int_{\mathbb{R}_0} \frac{Z_{s-} (e^{-zx} - 1)}{1 + \Delta G_s} d(\mu^r - \nu^r)_s \\
&\quad + \int_{[0,t[} \int_{\mathbb{R}_0} \frac{Z_s (e^{-zx} - 1)}{1 + \Delta^+ G_s} d(\mu^g - \nu^g)_{s+}.
\end{aligned}$$

This implies $(Z_{t \wedge T})_{t \geq 0}$ is an optional local martingale (for each $z \in [0, z_0]$). $\square$

Along with the stochastic exponent of the cumulant process $\mathcal{E}(G)$, there is a usual exponent of the cumulant process e^G . We know that if $\Delta G > -1$ and $\Delta^+ G > -1$, then $\mathcal{E}(G)$ can be represented as

$$\begin{aligned}
\mathcal{E}_t(G) &= \exp \left\{ G_t + \sum_{0 < s \leq t} (\log(1 + \Delta G_s) - \Delta G_s) \right. \\
&\quad \left. + \sum_{0 \leq s < t} (\log(1 + \Delta^+ G_s) - \Delta^+ G_{s+}) \right\}
\end{aligned}$$

From the above, we observe that if $\Delta G = 0$ and $\Delta^+ G = 0$, or equivalently, G is a continuous process, then

$$\mathcal{E}(G) = e^G.$$

Let us present sufficient conditions for the cumulant process G to be continuous.

Lemma 14.6

- (1) If $\Delta X_T = 0$ a.s. on the set $\{T < \infty\}$ for every predictable time T , i.e., $T \in \mathcal{T}^p$, then $\Delta G = 0$.
- (2) If $\Delta^+ X_T = 0$ a.s. on the set $\{T < \infty\}$ for every totally inaccessible time T , i.e., $T \in \mathcal{T}$, then $\Delta^+ G = 0$.

Now we are prepared to study probability of ruin of the optional semimartingale R introduced by formula (14.20). Given R , our main goal is to evaluate the ruin probability $\mathbf{P}(\tau < \infty)$, where $\tau = \inf\{t > 0 : R_t < 0\}$.

We assume that condition (14.23) holds and note that such risk process R is called as a *process with light-tailed negative jumps*.

Let us define the optional cumulant process for the risk process R : for all $t > 0$ and $z \in (0, z_0]$,

$$G_t(z) = -zB_t + \frac{z^2}{2} \langle N \rangle_t + \int_{[0,t]} \int_{\mathbb{R}_0} (e^{-zx} - 1) dv^r + \int_{[0,t]} \int_{\mathbb{R}_0} (e^{-zx} - 1) dv^g.$$

Define a process

$$M_t(z) = \exp[-z(R_t - u)] \mathcal{E}_t^{-1}(G(z)). \quad (14.26)$$

It follows from Theorem 14.4 that the process $M_t(z)$ in (14.26) is an optional local martingale for every z in $[0, z_0]$, if $\Delta G > -1$ and $\Delta^+ G > -1$. With the help of the Fatou Lemma, we can easily derive that a nonnegative optional local martingale is an optional supermartingale. Thus, since $M_0(z) = 1$, it follows that

$$\begin{aligned} 1 &\geq \mathbf{E}(M_{\tau \wedge t}(z)) \\ &= \mathbf{E}(M_\tau(z) \mathbf{1}_{\tau \leq t} + M_t(z) \mathbf{1}_{\tau > t}) \\ &\geq \mathbf{E}(M_\tau(z) \mathbf{1}_{\tau \leq t}) \\ &= \mathbf{E}(M_\tau(z) | \tau \leq t) \mathbf{P}(\tau \leq t) \end{aligned}$$

for every z in $[0, z_0]$ and $t > 0$.

Hence, for all $z \in [0, z_0]$,

$$\begin{aligned} \mathbf{P}(\tau \leq t) &\leq \frac{1}{\mathbf{E}(M_\tau | \tau \leq t)} \\ &= \frac{1}{\mathbf{E}(e^{-z(R_\tau - u)} \mathcal{E}_\tau^{-1}(G(z)) | \tau \leq t)}. \end{aligned}$$

Since $R_\tau \leq 0$ on $\{\tau \leq t\}$, we get

$$\mathbf{P}(\tau \leq t) \leq \frac{e^{-zu}}{\mathbf{E}(\mathcal{E}_\tau^{-1}[G(z)]|\tau \leq t)}, \quad (14.27)$$

for all $z \in [0, z_0]$.

If we want to find a better estimate of the upper bound of the ruin probability in (14.27), we are required to impose additional assumptions.

First, suppose that

$$-G'_\tau(0) = B_\tau + \int_{]0, \tau]} \int_{\mathbb{R}_0} x dv_s^r + \int_{]0, \tau]} \int_{\mathbb{R}_0} x dv_{s+}^g > 0. \quad (14.28)$$

Note that differentiation with respect to z under the integral sign with respect to v^j , $j = r, g$, in (14.28) is possible because zero is an interior point in the range of z -values for which the integral exists.

If we assume that $\Delta G = 0$ and $\Delta^+ G = 0$, then

$$\mathcal{E}_t(G(z)) = \exp(G_t(z)). \quad (14.29)$$

Now, using (14.29) and Jensen's inequality, we get from (14.27) that

$$\begin{aligned} \mathbf{P}(\tau \leq t) &\leq e^{-zu} \left[\mathbf{E} \left(\mathcal{E}_\tau^{-1}[G(z)] | \tau \leq t \right) \right]^{-1} \\ &\leq e^{-zu} \mathbf{E}(\mathcal{E}_\tau[G(z)] | \tau \leq t) \\ &= \mathbf{E}(\exp[-zu + G_\tau(z)] | \tau \leq t), \end{aligned} \quad (14.30)$$

for all $z \in [0, z_0]$ and $t > 0$.

The function $G_\tau(z)$ is a strictly convex function of z with $G_\tau(0) = 0$. Further, by the assumption (14.28), the function $\mathbf{E}(\exp[-zu + G_\tau(z)] | \tau \leq t)$ is a strictly convex function of z decreasing from 1 at $z = 0$. Furthermore, under assumption (14.29) with $z_0 = \infty$, the function $\mathbf{E}(\exp[-zu + G_\tau(z)] | \tau \leq t)$ increases to $+\infty$ as $z \rightarrow \infty$. Hence, there exists a unique $z^* \in [0, z_0]$ for which this function attains its minimum, and by (14.30), we obtain

$$\mathbf{P}(\tau \leq t) \leq \frac{e^{-z^*u}}{\mathbf{E}(\exp[-G_\tau(z^*)] | \tau \leq t)}. \quad (14.31)$$

It seems the estimate (14.31) is the best upper bound for ruin probability obtained from (14.27).

We can find a more explicit estimate of the ruin probability if we can choose z -value $z_t > 0$, for which the denominator of (14.27) equals one. Let us note that this z -value is not necessarily unique, and it often depends on t . In this case, we get

$$\mathbf{P}(\tau \leq t) \leq e^{-uz_t}.$$

Further, if z_t exists for all $t > 0$ and if $\hat{z} = \lim_{t \rightarrow \infty} z_t$ exists, then

$$\mathbf{P}(\tau < \infty) \leq e^{-u\hat{z}}.$$

Finally, we summarize our results in the following theorem.

Theorem 14.5 *Given the optional risk model (14.20), suppose assumptions (14.2) hold.*

1. *If $\Delta G > -1$ and $\Delta^+ G > -1$ then*

$$\mathbf{P}(\tau \leq t) \leq \frac{e^{-zu}}{\mathbf{E}(\mathcal{E}_\tau^{-1}[G(z)]|\tau \leq t)}$$

for all $z \in [0, z_0]$ and $t > 0$.

2. *If $G'_\tau(0) < 0$, $\Delta G = 0$ and $\Delta^+ G = 0$, then there exists a unique $z^* \in [0, z_0]$ for which the right-hand side of*

$$\mathbf{P}(\tau \leq t) \leq \frac{e^{-z^*u}}{\mathbf{E}(\exp[-G_\tau(z^*)]|\tau \leq t)}$$

attains its minimum.

3. *If $z_t > 0$ exists for which $\mathbf{E}(\mathcal{E}_\tau^{-1}[G(z_t)]|\tau \leq t) = 1$, then*

$$\mathbf{P}(\tau \leq t) \leq e^{-uz_t}.$$

Moreover, if z_t exists for all $t > 0$ and if $\hat{z} = \lim_{t \rightarrow \infty} z_t$ exists, then

$$\mathbf{P}(\tau < \infty) \leq e^{-u\hat{z}}. \quad (14.32)$$

Remark 14.2 Notice that the inequality (14.32) is an analog of the classical Cramer-Lundberg bound obtained in a very general optional setting.

Moreover, if the process D represents the accumulated claims (all jumps are downwards) while the process $L = 0$, then the condition (14.28) becomes

$$-G'_t(0) = B_t + \int_{[0,t]} \int_{-\infty}^0 x dv^r > 0 \text{ for all } t > 0$$

and is known as net-profit condition in risk theory. This means that the insurance company adopts the wise premium policy such that premiums follow the claims' intensity. $\square$

Chapter 15

Supplementary Problems



The list below contains problems which are related to all chapters of the book. Some of them are numerical and some others are pure theoretical, but in any case, they are help both students and instructors. Students can improve their understanding and scope. Instructors can transform most of the problems for teaching and examination purposes. The following references might be useful to create detailed solutions (see [7, 11, 13, 17, 18, 20, 27, 28, 30–32, 36, 37, 39, 40, 45–48, 50, 52, 60, 61], and [62]).

Problem 15.1 Let X be a standard normal random variable and Y be a Bernoulli random variable such that $P(Y = 1) = P(Y = -1) = \frac{1}{2}$, and X, Y are independent. Prove that

- (a) $Z = YX \sim N(0, 1)$.
- (b) Z and X are uncorrelated, but they are not independent.

Solution The claim (a) is obvious. To prove (b), we calculate $\text{cov}(X, Z) = \mathbb{E}(XZ) = \mathbb{E}X^2\mathbb{E}Y = 1 \cdot 0 = 0$. Further, $P(X \geq 1, Z \geq 1) = P(X \geq 1, Y = 1) = P(X \geq 1) \cdot P(Y = 1) = \frac{1}{2}P(X \geq 1)$. But $P(X \geq 1) \cdot P(Z \geq 1) = (P(X \geq 1))^2$ and $P(X \geq 1) \simeq 0.1587$, and, hence, $\frac{1}{2}P(X \geq 1) \neq (P(X \geq 1))^2$. ■

Problem 15.2 Assume that stochastic process $X_t = e^{W_t}$, where $W = (W_t)_{t \geq 0}$ is a standard Wiener process. Calculate the drift and diffusion coefficients of process $X = (X_t)_{t \geq 0}$:

$$b(x) = \lim_{h \rightarrow 0} h^{-1} [E(X_{t+h} - X_t | X_t = x)],$$

$$\sigma^2(x) = \lim_{h \rightarrow 0} h^{-1} \left[E \left((X_{t+h} - X_t)^2 | X_t = x \right) \right]$$

for $t, x \in \mathbb{R}_+^1$

□

Hint: Use the fact that

$$W_{t+h} - W_t \sim N(0, \sqrt{h}), \quad h > 0$$

■

Problem 15.3 Let $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$, and $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$. Then $\text{cov}(X_1, X_2) = \rho\sigma_1\sigma_2$ and $\text{corr}(X_1, X_2) = \rho$. □

Hint: Use the formula for bivariate normal distribution. ■

Problem 15.4 (Theorem of Slutsky) Let r.v.'s $X_n \xrightarrow[n \rightarrow \infty]{d} X$ and the sequence of real numbers $a_n \rightarrow a \in \mathbb{R}^1$. Then $X_n + a_n \xrightarrow[n \rightarrow \infty]{d} X + a$ and $a_n X_n \xrightarrow[n \rightarrow \infty]{d} aX$. □

Solution Let us note for $\forall \varepsilon > 0$, $x \in \mathbb{R}^1$, that

$$\begin{aligned} P(X_n + a_n \leq x) &= P(X_n + a_n \leq x, |a_n - a| < \varepsilon) + P(X_n + a_n \leq x, |a_n - a| \geq \varepsilon) \\ &\leq P(X_n \leq x - a + \varepsilon) + P(|a_n - a| \geq \varepsilon). \end{aligned}$$

It follows from here for all points $x - a + \varepsilon$ of continuity of distribution function F_x that

$$\limsup_n F_{X_n + a_n}(x) \leq F_x(x - a + \varepsilon).$$

So, by the arbitrary choice of $\varepsilon > 0$, we derive

$$\lim_{n \rightarrow \infty} F_{X_n + a_n}(x) = F_X(x - a) = F_{X+a}(x)$$

for all x at which F_{X+a} is continuous.

The second claim is proved in the same way. ■

Problem 15.5 Let $X_n \sim N(\mu_n, \sigma_n^2)$ and $\mu_n \rightarrow \mu$, $\sigma_n^2 \rightarrow \sigma^2$, $n \rightarrow \infty$. Then $X_n \xrightarrow[n \rightarrow \infty]{d} X \sim N(\mu, \sigma^2)$. □

Solution Denote $Z_n = \frac{X_n - \mu_n}{\sigma_n} \sim N(0, 1)$ and, hence, $Z_n \xrightarrow[n \rightarrow \infty]{d} Z \sim N(0, 1)$. Using Problem 15.4, we obtain that

$$X_n = \sigma_n Z_n + \mu_n \xrightarrow[n \rightarrow \infty]{d} X = \sigma Z + \mu.$$

■

Problem 15.6 (Borel-Cantelli Lemma) Let $(A_n)_{n=1,2,\dots}$ be a sequence of events, and $C = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$. Then

1. $P(C) = 0$, if $\sum_{n=1}^{\infty} P(A_n) < \infty$.
2. $P(C) = 1$, if $(A_n)_{n=1,2,\dots}$ are independent and $\sum_{n=1}^{\infty} P(A_n) = \infty$.

□

Solution

(a) We have $P(C) \leq P\left(\bigcup_{n=m}^{\infty} A_n\right) \leq \sum_{n=m}^{\infty} P(A_n)$ for each $n = 1, 2, \dots$. Take $\varepsilon > 0$ and find N_ε big enough that $\sum_{n=N_\varepsilon}^{\infty} P(A_n) < \varepsilon$. Hence, for all $N > N_\varepsilon$, we obtain $P(C) < \varepsilon$.

(b) Let us note that $P\left(\left(\bigcup_{n=m}^{\infty} A_n\right)^c\right) \leq P\left(\bigcup_{n=m}^{\infty} A_n^c\right) \leq P\left(\bigcup_{n=m}^{m+M} A_n^c\right)$ for any $M > 0$.

But (A_n) are independent, and, hence, $P\left(\left(\bigcup_{n=m}^{\infty} A_n\right)^c\right) \leq \prod_{n=m}^{m+M} (1 - P(A_n)) \leq \exp\left\{-\sum_{n=m}^{m+M} P(A_n)\right\}$, where we used the inequality $1 - x \leq e^{-x}$ for $x \in [0, 1]$. The claim follows from the inequality above for probabilities. ■

Problem 15.7 Let $\mu = \sum_{n=1}^{\infty} \alpha_n P_n$, where (P_n) and (α_n) are sequences of probability measures and positive numbers, respectively. Define $\nu = \sum_{n=1}^{\infty} \beta_n Q_n$, where (Q_n) and (β_n) are sequences of probability measures $Q_n \ll P_n$ and nonnegative numbers. Prove that $\nu \ll \mu$. □

Solution For any A with $\mu(A) = 0$, we have $\sum_{n=1}^{\infty} \alpha_n P_n(A) = \mu(A) = 0$ and, hence, $P_n(A) = 0$ for all n . So, $Q_n(A) = 0$ and therefore $\nu(A) = \sum_{n=1}^{\infty} \beta_n Q_n(A) = 0$, and we get $\nu \ll \mu$. ■

Problem 15.8 Let $([0, 1], \mathcal{B}(0, 1), l)$ be a Borel space with the Lebesgue probability measure l , and X and Y be random variables: $X(\omega) = 2\omega^2$ and

$$Y(\omega) = \begin{cases} 0, & \omega \in [0, \frac{1}{3}] \\ 2, & \omega \in [\frac{1}{3}, \frac{2}{3}] \\ 1, & \omega \in [\frac{2}{3}, 1] \end{cases}.$$

Find $\mathbb{E}(X|Y)$. □

Solution For $\omega \in [0, \frac{1}{3}]$, we have $\mathbb{E}(X|Y)(\omega) = \frac{\int_0^{\frac{1}{3}} x dP}{P([0, \frac{1}{3}])} = \frac{1}{\frac{1}{3}} \int_0^{\frac{1}{3}} 2\omega^2 d\omega = \frac{2}{27}$.
 Values of $\mathbb{E}(X|Y)$ on other sets $[\frac{1}{3}, \frac{2}{3}]$ and $[\frac{2}{3}, 1]$ can be determined in the same way: $\frac{14}{27}$ and $\frac{38}{27}$, correspondingly. ■

Problem 15.9 Let $(\varepsilon_n)_{n=1,2,\dots,N}$ be a sequence of independent random variables with values $+1$ and -1 taking with probability $\frac{1}{2}$. Define $X_n = (-1)^n \cos(\pi \sum_{k=1}^n \varepsilon_k)$, $n = 1, 2, \dots, N$ and prove that $(X_n)_{n=1 \dots N}$ is a martingale with respect to a natural filtration $\mathcal{F}_n = \mathcal{F}_n^\varepsilon = \mathcal{F}_n^X$. □

Solution We represent the sequence (X_n) as follows:

$$X_n = (-1)^n \frac{1}{2} \left[e^{i\pi \sum_{k=1}^n \varepsilon_k} + e^{-i\pi \sum_{k=1}^n \varepsilon_k} \right], \quad 1 = \sqrt{-1}.$$

Using independence of $(\varepsilon_n)_{1 \dots N}$, we have

$$\mathbb{E}(X_n | \mathcal{F}_{n-1}) = \frac{1}{2} (-1)^n \left[\mathbb{E} e^{i\pi \varepsilon_n} \cdot e^{i\pi \sum_{k=1}^{n-1} \varepsilon_k} + \mathbb{E} e^{-i\pi \varepsilon_n} \cdot e^{-i\pi \sum_{k=1}^{n-1} \varepsilon_k} \right].$$

Further, applying an obvious relation $\mathbb{E} e^{i\pi \varepsilon_n} = \frac{1}{2}(e^{i\pi} + e^{-i\pi}) = -1$, we obtain

$$\begin{aligned} \mathbb{E}(X_n | \mathcal{F}_{n-1}) &= \frac{(-1)^{n-1}}{2} \left[e^{i\pi \sum_{k=1}^{n-1} \varepsilon_k} + e^{-i\pi \sum_{k=1}^{n-1} \varepsilon_k} \right] \\ &= (-1)^{n-1} \cos\left(\pi \sum_{k=1}^{n-1} \varepsilon_k\right) = X_{n-1}. \end{aligned}$$
■

Problem 15.10 Values and joint distribution of random variables X and Y are given in the table

X \ Y	Y		
	-0.1	0	0.1
-0.2	0.1	0	0.4
0.1	0.3	0.1	0.1

Find marginal distributions of X and Y , average of Y , and $\mathbb{E}(Y|X)$. □

Solution We have from the table above that

$$P(X = -0.2) = 0.1 + 0.4 = 0.5.$$

$$P(X = 0.1) = 0.3 + 0.1 + 0.1 = 0.5$$

$$P(Y = -0.1) = 0.1 + 0.3 = 0.4.$$

$$P(Y = 0) = 0.1, \quad P(Y = 0.1) = 0.4 + 0.1 = 0.5,$$

which give us marginal distributions. We also derive from the above equalities that $\mathbb{E}Y = -0.1 \cdot 0.4 + 0.1 \cdot 0.5 = 0.01$.

To calculate the conditional expectation $\mathbb{E}(Y|X)$, we write

$$\mathbb{E}(Y|X) = \mathbb{E}(Y|X = -0.2) \cdot I_{\{X=-0.2\}} + \mathbb{E}(Y|X = 0.1) \cdot I_{\{X=0.1\}}.$$

Calculating

$$\begin{aligned} \mathbb{E}(Y|X = -0.2) &= -0.1 \cdot P(Y = -0.1|X = -0.2) + 0.1 \cdot P(Y = 0.1|X = -0.2) \\ &= \frac{-0.1 \cdot 0.1 + 0.1 \cdot 0.4}{0.5} = 0.06, \end{aligned}$$

and similarly

$$\mathbb{E}(Y|X = 0.1) = \frac{-0.1 \cdot 0.3 + 0.1 \cdot 0.1}{0.5} = -0.04,$$

we obtain

$$E(Y|X) = 0.06 \cdot I_{\{X=-0.2\}} - 0.04 \cdot I_{\{X=0.1\}}.$$

■

Problem 15.11 Let $(X_n)_{n=1,2,\dots}$ be a sequence of independent random variables such that $P(X_n = 1) = p$, $P(X_n = -1) = 1 - p$, $1 < p < \frac{1}{2}$. Show that the following stochastic sequences are martingales with respect to a natural filtration $(\mathcal{F}_n)_{n=1,2,\dots}$ generated by (X_n) :

(a) $M_n = \sum_{k=1}^n X_k - n \cdot (2p - 1).$

(b) $Y_n = M_n^2 - 4np(1 - p).$

(c) $Z_n = \left(\frac{1-p}{p}\right) \sum_{k=1}^n X_k.$

□

Hint: Check the martingale property. ■

Problem 15.12 Let X_0 be a random variable such that $P(X_0 = 2) = P(X_0 = 0) = \frac{1}{2}$. Define $X_n = n \cdot X_{n-1}$, $n = 1, 2, \dots$ and $M_n = X_n - \mathbb{E}X_n$. Prove that (M_n) is not a martingale with respect to the natural filtration $(\mathcal{F}_n)$. $\square$

Solution We observe that $\mathbb{E}(M_n | \mathcal{F}_{n-1}) = \mathbb{E}(X_n | \mathcal{F}_{n-1}) - n! = n \cdot X_{n-1} - n! = n(X_{n-1} - (n-1)!) = n \cdot M_{n-1} \neq M_{n-1}$. $\blacksquare$

Problem 15.13 Find a stochastic differential for the process $X_t = (\sqrt{123} + \frac{1}{2}W_t)^2$, where (W_t) is a Wiener process. $\square$

Solution Here $X_t = f(W_t)$ with the function $f(x) = (\sqrt{123} + \frac{1}{2}x)^2$. The first and second derivatives of this function are $f'(x) = 2(\sqrt{123} + \frac{1}{2}x)$, $f''(x) = 1$. Therefore, using the Ito formula, the stochastic differential of (X_t) is

$$\begin{aligned} dX_t &= f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt \\ &= 2(\sqrt{123} + \frac{1}{2}W_t)dW_t + \frac{1}{2}dt. \end{aligned}$$

 $\blacksquare$

Problem 15.14 Let $(N_t)_{t \geq 0}$ be a Poisson process with intensity $\lambda = 1$. Prove that $(N_t - t)^2 - t$ is a martingale with respect to natural filtration generated by (N_t) .

Calculate $\mathbb{E} \int_1^3 N_t dt \cdot \int_2^4 N_t dt$ $\square$

Hint: In the first case, check the martingale property. In the second case, the answer is $34\frac{1}{3}$. $\blacksquare$

Problem 15.15 Check whether the processes are martingales:

- (a) $X_t = W_t^3 - 3tW_t$.
- (b) $X_t = W_t + 287t$.
- (c) $X_t = e^{\frac{1}{2}} \sin(W_t)$,

where $(W_t)_{t \geq 0}$ is a Wiener process. $\square$

Solution (a) We have here $X_t = f(t, W_t)$ with the function $f(t, x) = x^3 - 3tx$. This function has the partial derivatives

$$\begin{aligned} \frac{\partial}{\partial x} f(t, x) &= 3x^2 - 3t, \quad \frac{\partial^2 f(t, x)}{\partial x^2} = 6x, \\ \frac{\partial}{\partial t} f(t, x) &= -3x. \end{aligned}$$

Therefore, using the Ito formula, we derive

$$dX_t = (3W_t^2 - 3t)dW_t + \left(-3W_t + \frac{1}{2}(6W_t)\right)dt = (3W_t^2 - 3t)dW_t.$$

Therefore, (X_t) is a martingale as a stochastic integral has a martingale property.

In the case of (b), we have $\mathbb{E}X_0 = 0$ and, for instance, $\mathbb{E}X_1 = 287$. This implies that (X_t) cannot be a martingale, since martingales have constant expectations.

For (c), we have $X_t = f(t, W_t)$ with function $f(t, x) = e^{\frac{t}{2}} \cdot \sin(x)$ and its partial derivatives:

$$\begin{aligned}\frac{\partial}{\partial x} f(t, x) &= e^{\frac{t}{2}} \cdot \cos(x), \quad \frac{\partial^2 f(t, x)}{\partial x^2} = -e^{\frac{t}{2}} \cdot \sin(x), \\ \frac{\partial}{\partial t} f(t, x) &= \frac{1}{2} e^{\frac{t}{2}} \cdot \sin(x).\end{aligned}$$

Using the Ito formula, we get

$$\begin{aligned}dX_t &= e^{\frac{t}{2}} \cdot \cos(W_t) dW_t + \left[\frac{1}{2} e^{\frac{t}{2}} \cdot \sin(W_t) + \frac{1}{2} (-e^{\frac{t}{2}} \cdot \sin(W_t)) \right] dt = \\ &= e^{\frac{t}{2}} \cdot \cos(W_t) dW_t,\end{aligned}$$

which certifies that (X_t) is a martingale. ■

Problem 15.16 Provide a condition on the mapping $\varphi : \mathbb{R}^1 \longrightarrow \mathbb{R}^1$ under which $\varphi(\tau)$ remains a stopping time, where τ is a stopping time. □

Solution Suppose the mapping φ meets the following conditions:

- (a) φ is injective.
- (b) $\varphi([0, \infty)) = [0, \infty)$.
- (c) φ order-preserving and $t \leq \varphi(t)$ for all $t \in [0, \infty)$.

Condition (a) ensures that the inverse mapping $\varphi^{-1} : \varphi(\mathbb{R}^1) \longrightarrow \mathbb{R}^1$ is defined. Condition (b) ensures that $\varphi^{-1}(t)$ is well-defined and positive for all $t \geq 0$. Condition (c) together with the previous two conditions (a) and (b) gives that whenever τ is a stopping time

$$\{\varphi(\tau) \leq t\} = \{\tau \leq \varphi^{-1}(t)\} \in \mathcal{F}_{\varphi^{-1}(t)} \subseteq \mathcal{F}_t.$$

For instance, if $\varphi : [0, \infty) \longrightarrow [0, \infty)$ is a strictly increasing function satisfying $t \leq \varphi(t)$ for all $t \geq 0$ and φ is differentiable with $\varphi'(t) \geq 1$ for all t , $\varphi(0) = 0$, then by the mean value theorem, we get $\varphi(t) \geq \varphi'(c) \cdot t \geq t$, and the above holds. ■

Problem 15.17 Consider the function $p(t, x, y) = \frac{1}{\sqrt{2\pi t}} \exp \left\{ -\frac{(x-y)^2}{2t} \right\}$, $x, y \in \mathbb{R}^1$, $t \in \mathbb{R}_t^1$, representing the transition density of a Wiener process (W_t) . Prove the $p(t, x, y)$ satisfies the PDE:

$$\frac{\partial}{\partial t} p(t, x, y) = \frac{1}{2} \frac{\partial^2}{\partial y^2} p(t, x, y).$$

□

Solution On one side, we have

$$\frac{\partial}{\partial t} p(t, x, y) = p(t, x, y) \left[-\frac{1}{2t} + \frac{(x-y)^2}{2t} \right].$$

On the other hand, differentiating with respect to y , we get

$$\begin{aligned} \frac{\partial^2}{\partial y^2} p(t, x, y) &= p \cdot \frac{(x-y)^2}{t} + p \cdot \left(-\frac{1}{t}\right) \\ &= p(t, x, y) \left[\frac{(x-y)^2}{t} - \frac{1}{t} \right]. \end{aligned}$$

Then it is clear that $p(t, x, y)$ satisfies the above differential equation. ■

Problem 15.18 Prove that every nonnegative local martingale is a supermartingale. □

Hint: Apply the Fatou lemma. ■

Problem 15.19 Let $(X_n)_{n=0,1,\dots,N}$ be a submartingale with $E(X_N^+)^p < \infty$, $p > 1$. Prove the Doob inequality

$$\mathbf{E}(\max_n X_n^+)^p \leq \left(\frac{p}{p-1}\right)^p \mathbf{E}(X_N^+)^p, \quad X_N^+ = \max(0, X_N).$$

□

Solution Denote $X_N^* = \max_{n \leq N} X_n^+$ and note that

$$\lambda \cdot \mathbf{P}(X_N^* > \lambda) \leq \mathbf{E}X_N^+ \cdot I_{\{X_N^* > \lambda\}}, \quad \lambda > 0.$$

Multiplying the above inequality by λ^{p-2} and integrating over $(0, \infty)$, we obtain

$$\int_0^\infty \lambda^{p-1} \mathbf{P}(X_N^* > \lambda) d\lambda \leq \mathbf{E}X_N^+ \cdot \int_0^{X_N^*} \lambda^{p-2} d\lambda \leq \frac{1}{p-1} \mathbf{E}X_N^+ \cdot (X_N^*)^{p-1}.$$

By $\int_0^\infty \lambda^{p-1} \mathbf{P}(X_N^* > \lambda) d\lambda = \frac{1}{p} \int_0^\infty \lambda^p d\mathbf{P}(X_N^* \leq \lambda)$, we get $\mathbf{E}(X_N^*)^p \leq \frac{p}{1-p} \mathbf{E}X_N^+ \cdot (X_N^*)^{p-1}$, and with the help of the Hölder-inequality, we derive

$$\mathbf{E}X_N^+ \cdot (X_N^*)^{p-1} \leq (\mathbf{E}(X_N^+)^p)^{\frac{1}{p}} (\mathbf{E}(X_N^*)^p)^{\frac{p-1}{p}}$$

and hence

$$\mathbf{E}(X_N^*)^p \leq \frac{p}{p-1} (\mathbf{E}(X_N^+)^p)^{\frac{1}{p}} (\mathbf{E}(X_N^*)^p)^{\frac{p-1}{p}}.$$

It leads to the desirable inequality. ■

Problem 15.20 Give an example of a non-right-continuous filtration and a martingale which is not right-continuous. □

Solution For a positive real number $a > 0$, define $\Omega = \{a, -a\}$, $\mathbf{P}(a) = \mathbf{P}(-a) = \frac{1}{2}$, and the Bernoulli random variable

$$X = \begin{cases} a & \text{with probability } \frac{1}{2}, \\ -a & \text{with probability } \frac{1}{2}. \end{cases}$$

Define $X_t = \begin{cases} 0 & \text{if } t \leq t_0, \\ X & \text{if } t > t_0 \end{cases}$, $t \geq 0$, and $F_t = \begin{cases} \{\emptyset, \Omega\} & \text{if } t \leq t_0, \\ F_X = \sigma(X) & \text{if } t > t_0 \end{cases}$. Process (X_t) is a martingale w.r. to (F_t) . Both (X_t) and (F_t) are not right-continuous. ■

Problem 15.21 Let $(A_n)_{n \geq 0}$ be a predictable $(d \times d)$ -matrix-valued sequence of random variables such that $\Delta A_n = A_n - A_{n-1}$ is positive-defined, $n \geq 1$, and $\lambda \min(A)$ and $\lambda \max(A)$ are minimal and maximal eigenvalues of A , respectively. Let $(M_n)_{n \geq 0}$ be a d -dimensional square-martingale with the quadratic characteristic $\langle\langle M, M \rangle\rangle_n = (\langle M^i, M^j \rangle_n)_{i,j=1,\dots,d}$. Prove an analog of lemma 7.1 for d -dimensional case: (a.s.)

$$\begin{aligned} & \{\omega : \lambda_{\min}(A_\infty) = \infty\} \cap \{\omega : N \rightarrow \infty\} \cap \{\omega : \limsup_n \frac{\lambda_{\max}}{\lambda_{\min}}(A_n) < \infty\} \\ & \subseteq \{\omega : A_n^{-1} M_n \rightarrow 0\}. \end{aligned}$$

Hint: Adapt the proof of lemma 7.1 to this multidimensional case. ■

Problem 15.22 Let (A_n) and (M_n) be as in the previous problem, $\lambda \min(A_n) \rightarrow \infty$ (a.s.), $\lim_n \sup \frac{\lambda_{\max}}{\lambda_{\min}}(A_n) < \infty$ (a.s.), and (a.s.).

$$\sum_{n=1}^{\infty} \text{tr}((A_n^{-1} \Delta \langle\langle M, M \rangle\rangle_n A_n^{-1})) < \infty.$$

Then (a.s.) $A_n^{-1} M_n \rightarrow 0$, $n \rightarrow \infty$.

Hint: Adapt the proof of theorem 7.1 to this multidimensional case. ■

Problem 15.23 Let us consider a polynomial transformation $P_n(t) \cdot Q_m(w_t)$ of a Wiener process $(w_t)_{t \geq 0}$, where P_n and Q_m are polynomials of degrees n and m relatively. Determine when this transformation leads to a martingale.

Hint: Use the Ito formula. ■

Problem 15.24 Let $(X_n)_{n=1,2,\dots}$ be a sequence of independent random variables with the density

$$f_a(x) = \begin{cases} e^{-ax} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases}$$

Find the parameter a to provide that $Z_n = \prod_{i=1}^n X_i$ is a martingale w.r. to a natural filtration $(F_n^X)_{n=1,2,\dots}$.

Hint: Use the conditions $\int_0^\infty f_a(x)dx = 1$ and $\mathbf{E}X_n = 1$. ■

Problem 15.25 Assume $X_n \sim N(\mu_n, \sigma_n^2)$, $n = 1, 2, \dots$, and $X_n \xrightarrow{L_2} X$, $n \rightarrow \infty$. Then $X \sim N(\mu, \sigma^2)$, where $\mu = \lim_{n \rightarrow \infty} \mu_n$, $\sigma^2 = \lim_{n \rightarrow \infty} \sigma_n^2$.

Solution It follows from the L_2 -convergence that $\mu_n \rightarrow \mu = \mathbf{E}X$ and $\text{Var} X_n \rightarrow \text{Var} X = \sigma^2$, $n \rightarrow \infty$. Hence, for an arbitrary $\lambda \in \mathbb{R}^1$, we obtain

$$\mathbf{E}e^{i\lambda X} = \lim_{n \rightarrow \infty} \mathbf{E}e^{i\lambda X_n} = \lim_{n \rightarrow \infty} e^{i\mu_n \lambda - \frac{\sigma_n^2}{2} \lambda^2} = e^{i\mu \lambda - \frac{\sigma^2}{2} \lambda^2}.$$

■

Problem 15.26 Let $b = b(x)$ and $\sigma = \sigma(x)$ be bounded functions from $\mathbb{R}^1$ to $\mathbb{R}^1$ such that $b \in \mathbb{C}^1(\mathbb{R}^1)$, $\sigma \in \mathbb{C}^2(\mathbb{R}^1)$. Assume $(W_t)_{t \geq 0}$ is a Wiener process. Then the Ito process

$$dX_t = b(X_t)dt + \sigma(X_t).dW_t$$

can be rewritten in the form of the Stratonovich stochastic integral

$$dX_t = (b(X_t) - \frac{1}{2}\sigma(X_t)\sigma'(X_t))dt + \sigma(X_t).dW_t.$$

Also, that for $X_t = W_t$ and $F \in \mathbb{C}^2(\mathbb{R}^1)$, the Ito formula $F(X_t) - F(X_0) = \int_0^t F'(X_s)dX_s + \frac{1}{2} \int_0^t F''(X_s)ds$ admits the following form in terms of Stratonovich integral $F(X_t) - F(X_0) = \int_0^t F'(X_s).dX_s$.

Hint: Use definitions of these integrals. ■

Problem 15.27 Let us consider the exponential transformation of a probability measure $\mathbf{P}_T$, $T > 0$ to a measure function $\mathbf{P}_T^*$ such that

$$\frac{d\mathbf{P}_T^*}{d\mathbf{P}_T} = e^{aW_T^- + bT + c} = Z_T,$$

where $(W_t)_{t \leq T}$ is a Wiener process. Determine parameters a , b , and c under which $\mathbf{P}_T^*$ will be a probability measure.

Hint: Use the condition $\mathbf{E}Z_T = 1$. ■

Problem 15.28 Prove that the first and second variations of a Wiener process $(W_t)_{t \geq 0}$ converge (a.s.) to infinity and to the length of time interval correspondingly.

Solution For a fixed $T > 0$, we define a subdivision $t_i^n = iT \cdot 2^{-n}$, $i \leq 2^n$, $n \geq 1$ of the interval $[0, T]$. Define

$$FV^n = \sum_{i=0}^{2^n-1} |W_{t_{i+1}^n} - W_{t_i^n}| = \sum_{i=0}^{2^n-1} |\Delta W_{t_i^n}|$$

and

$$SV^n = \sum_{i=0}^{2^n-1} (\Delta W_{t_i^n})^2, \quad n \geq 1.$$

For SV^n using properties of (W_t) , we have

$$\begin{aligned} \mathbf{E}(SV^n - T)^2 &= \mathbf{E} \sum_{i=0}^{2^n-1} \left((\Delta W_{t_i^n})^2 - T \cdot 2^{-n} \right)^2 \\ &= \sum_{i=0}^{2^n-1} \mathbf{E} \left((\Delta W_{t_i^n})^2 - T \cdot 2^{-n} \right)^2 \\ &= 2^n \cdot 2 \cdot (T \cdot 2^{-n})^2 \\ &= 2 \cdot T^2 \cdot 2^{-n} \end{aligned}$$

and

$$\begin{aligned} \left\| \sum_{n=1}^{\infty} |SV^n - T| \right\|_{L_2} &\leq \sum_{n=1}^{\infty} \|SV^n - T\|_{L_2} \\ &\leq \sqrt{2}T \cdot \sum_{n=1}^{\infty} 2^{-\frac{n}{2}} < \infty. \end{aligned}$$

Hence, $SV^n \rightarrow T$ (a.s.) as $n \rightarrow \infty$ by $\sum_{n=1}^{\infty} \|SV^n - T\|_{L_2} < \infty$ (a.s.).

Regarding FV^n for each $\omega \in \Omega$, we have $SV^n(\omega) \leq (\max_i |\Delta W_{t_i^n}(\omega)|) \cdot \sum_{i=0}^{2^n-1} |\Delta W_{t_i^n}(\omega)| \leq (\max_i |\Delta W_{t_i^n}(\omega)|) \cdot FV^n(\omega)$. It follows from this inequality for paths of $(W_t)_{t \geq 0}$ that $\lim_{n \rightarrow \infty} \inf FV^n(\omega)$ cannot be finite; otherwise $SV^n \xrightarrow[n \rightarrow \infty]{} 0$. ■

Problem 15.29 Let $(M_t)_{t \geq 0}$ be a continuous square integrable martingale with a finite first variation $FV_t(M)$. Show that $M_t = M_0$ (a.s.) for all $t \geq 0$.

Hint: Let (t_j^n) be a finite partition of $[0, t]$ with $\max_j \Delta t_j^n \rightarrow 0$ as $n \rightarrow \infty$. Then (a.s.)

$$\sum_j |\Delta M_{t_j^n}|^2 \leq FV_t(M(\omega)). \max_j |\Delta M_{t_j^n}(\omega)| \rightarrow 0, \quad n \rightarrow \infty,$$

and

$$\langle M, M \rangle_t = 0 \quad (a.s.).$$

Hence, $M_t = M_0$ (a.s.). ■

Problem 15.30 Let $M = (M_t)_{t \geq 0}$ be a continuous local martingale and

$$Z_t(M) = \exp\{M_t - \frac{1}{2}\langle M, M \rangle_t\}, \quad t \in [0, \infty].$$

Assume that (Krylov's condition)

$$\liminf_{\varepsilon \rightarrow 0} \varepsilon \cdot \log \mathbf{E} \exp\left\{\frac{1-\varepsilon}{2}\langle M, M \rangle_\infty\right\} < \infty.$$

Prove that $\mathbf{E}Z_\infty(M) = 1$, i.e., the exponential local martingale $(Z_t(M))_{t \geq 0}$ is uniformly integrable. Show that the condition above is wider than the Novikov condition $\mathbf{E}e^{\frac{1}{2}\langle M, M \rangle_\infty} < \infty$. □

Hint: First of all, we note that $\mathbf{E}Z_\infty(M) \leq 1$ as a consequence of a supermartingale property of $(Z_t(M))$ and the Fatou lemma. Using the Hölder inequality, we can derive for a constant $c > 0$ that

$$\begin{aligned} 1 &= \mathbf{E}Z_\infty((1-\varepsilon)M) = \mathbf{E}e^{(1-\varepsilon)(M_\infty - \frac{1}{2}\langle M, M \rangle_\infty)} \cdot e^{\frac{(1-\varepsilon)\varepsilon}{2}\langle M, M \rangle_\infty} \cdot I_{\{\langle M, M \rangle_\infty \leq c\}} \\ &\quad + \mathbf{E}e^{(1-\varepsilon)(M_\infty - \frac{1}{2}\langle M, M \rangle_\infty)} \cdot e^{\frac{(1-\varepsilon)\varepsilon}{2}\langle M, M \rangle_\infty} \cdot I_{\{\langle M, M \rangle_\infty > c\}} \\ &\leq (\mathbf{E}Z_\infty(M))^{1-\varepsilon} \cdot \left(\mathbf{E}e^{\frac{(1-\varepsilon)\varepsilon}{2}\langle M, M \rangle_\infty} \cdot I_{\{\langle M, M \rangle_\infty \leq c\}} \right)^\varepsilon \\ &\quad + (\mathbf{E}Z_\infty(M) \cdot I_{\{\langle M, M \rangle_\infty > c\}})^{1-\varepsilon} \cdot \left(\mathbf{E}e^{\frac{(1-\varepsilon)\varepsilon}{2}\langle M, M \rangle_\infty} \right)^\varepsilon. \end{aligned}$$

By taking the limit when $\varepsilon \rightarrow 0$ in the above inequality, we arrive at

$$1 \leq \mathbf{E}Z_\infty(M) + \text{const} \cdot \mathbf{E}Z_\infty(M) \cdot I_{\{\langle M, M \rangle_\infty > c\}}.$$

Taking the limit as $c \rightarrow \infty$, we get $\mathbf{E}Z_\infty(M) \geq 1$. ■

Problem 15.31 Show that every process (A_t) with finite variation can be represented as the difference of two increasing processes.

Hint: Use the representation

$$A_t = \frac{1}{2}(|A|_t + A_t - \frac{1}{2}(|A|_t - A_t)),$$

where $|A|_t$ is the variation of A on $[0, t]$. ■

Problem 15.32 Let $(X_t)_{t \geq 0}$ be a semimartingale and let (A_t) be a process of finite variation. Prove that

$$[X, A]_t = \sum_{s \leq t} (\Delta X_s)(\Delta A_s).$$

In particular, $[X, A] = 0$, if (A_t) or (X_t) is continuous.

Hint: Use the limiting arguments dividing $[0, t]$ by a subdivision $t_j^n = jt2^{-n}$, $j = 0, \dots, 2^n$, $n \geq 1$. ■

Problem 15.33 Prove the Levy characterization of a Wiener process (Remark 8.1).

Solution Let us prove that the following three statements are equivalent for a continuous martingale $(X_t)_{t \geq 0}$, $X_0 = 0$:

- (1) Process (X_t) is a standard Wiener process on the underlying stochastic basis.
- (2) Process $(X_t^2 - t)_{t \geq 0}$ is a martingale.
- (3) Process $(X_t)_{t \geq 0}$ has a quadratic variation $[X, X]_t = t$.

To prove (1) $\Rightarrow$ (2), we just observe that $\mathbf{E}|X_t^2 - t| < \infty$ for each $t \geq 0$ and for $s \leq t$, $\mathbf{E}(X_t^2 - t | F_s) = \mathbf{E}((X_t^2 - X_s^2) + X_s^2 - (t - s) - s | F_s) = X_s^2 - s$ by the properties of a Wiener process (X_t) .

For the proof of the second implication (2) $\Rightarrow$ (3), we note that $(X_t^2)_{t \geq 0}$ is a submartingale with the Doob-Meyer decomposition $X_t^2 = (X_t^2 - t) + t$. Hence, $[X, X]_t = \langle X, X \rangle_t = t$.

The third implication (3) $\Rightarrow$ (1) can be stated as follows. First, we prove that increments of $(X_t)_{t \geq 0}$ are Gaussian. Applying the Ito formula to $e^{i\lambda x + \frac{\lambda^2}{2}t} = f(t, x)$, $t \geq 0$, and $\lambda > 0$, we obtain

$$\begin{aligned} df(t, X_t) &= \frac{\partial}{\partial t} f(t, X_t)dt + \frac{\partial}{\partial x} f(t, X_t)dX_t + \frac{1}{2} \frac{\partial^2}{\partial x^2} f(t, X_t)d[X, X]_t \\ &= \frac{1}{2} \lambda^2 f(t, X_t)dt + i\lambda f(t, X_t)dX_t + \left(-\frac{\lambda^2}{2}\right) f(t, X_t)dt \\ &= i\lambda f(t, X_t)dX_t, \end{aligned}$$

and, therefore, $(f(t, X_t))_{t \geq 0}$ is a martingale. Further,

$$\mathbf{E}(e^{i\lambda X_t + \frac{1}{2}\lambda^2 t} | F_s) = e^{i\lambda X_s + \frac{1}{2}\lambda^2 s},$$

and

$$\mathbf{E}e^{i\lambda(X_t - X_s)} = \mathbf{E}(\mathbf{E}(e^{i\lambda(X_t - X_s)} | F_s)) = e^{-\frac{1}{2}\lambda^2(t-s)},$$

which means that $X_t - X_s \sim N(0, t - s)$, $t \geq s$.

To finish the proof, we show that the increments $X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent for any subdivision of time $0 \leq t_1 < t_2 < t_3 < \dots < t_{n-1} < t_n$. We have

$$\mathbf{E}(e^{i\lambda_1 X_{t_1} + i\lambda_2(X_{t_2} - X_{t_1}) + \dots + i\lambda_n(X_{t_n} - X_{t_{n-1}})}) = e^{-\frac{\lambda_1^2}{2}t_1} \times \dots \times e^{-\frac{1}{2}\lambda_n^2(t_n - t_{n-1})},$$

which states independence of increments of (X_t) . So, $(X_t)_{t \geq 0}$ is a Wiener process. ■

Problem 15.34 Prove that any submartingale $X = (X_t)_{t \geq 0}$ on a standard stochastic basis $(\Omega, F, (F_t)_{t \geq 0}, P)$ admits the right-continuous modification, if EX_t is right-continuous.

Hint: Using limiting arguments together with a submartingale property, we have $X_t \leq E(X_{t+} | F_t)$ (a.s.), $t \leq 0$. But $F_t = F_{t+}$; hence, $X_t \leq X_{t+}$ (a.s.), and from $EX_t = EX_{t+}$, we get the result. ■

Problem 15.35 Let $M = (M_t)_{t \geq 0}$ be a martingale on a stochastic basis $(\Omega, F, (F_t)_{t \geq 0}, P)$. Show that a predictability property for the process $\phi = (\phi_t)$ is vital in getting the martingale property for a stochastic integral $\int_0^t \phi_s dM_s$.

Hint: Take a rich enough probability space (Ω, F, P) to accommodate two random variables: $\tau \geq 0$ with $P(\tau \leq t) = t \wedge 1$ and a Bernoulli random variable Y such that $P(Y = 1) = P(Y = -1) = \frac{1}{2}$. Define $M_t = Y \cdot I_{\{\tau \leq t\}}$ and filtration $F_t = F_t^M$. In this case, $\int_0^t M_s dM_s$ is not a martingale. ■

Problem 15.36 Let continuous processes $A \in \mathcal{A}_{loc}^+$ and $M \in \mathcal{M}_{loc}$ are defined on a standard stochastic basis $(\Omega, F, (F_t)_{t \geq 0}, P)$. Assume that $b^i = b^i(x)$, $x \in \mathbb{R}^1$, $i = 1, 2$, are bounded continuous functions and X^i , $i = 1, 2$, are (strong) continuous solutions of the stochastic differential equations w.r. to a semimartingale $Y_t = A_t + M_t$, $Y_0 = 0$:

$$dX_t^i = b^i(X_t^i) dA_t + dM_t,$$

where $X_0^i = x \in \mathbb{R}^1$.

Prove that the inequality $b^1(x) < b^2(x)$ for all $x \in \mathbb{R}^1$ implies $X_t^1(\omega) \leq X_t^2(\omega)$ (a.s.) for all $t \geq 0$.

Hint: Apply the method of proof similar to Lemma 9.1. ■

Problem 15.37 Consider a stochastic differential equation as in problem 15.36:

$$dX_t = b(X_t) dA_t + dM_t,$$

where $X_0 = x \in \mathbb{R}^1$, $t \leq T$.

Assume that $b^1 = b(x)$, $x \in \mathbb{R}^1$ satisfies conditions of Theorem 9.3 and $A \ll \langle M, M \rangle$. Prove that this equation admits at least one (strong) solution.

Hint: Adapt the method of proof of Theorem 9.3 for this case. ■

Problem 15.38 Let $dX_t = \mu dt + \sigma dW_t + \nu dN_t$, where W and N be a Wiener process and a Poisson process, respectively, with intensity $\lambda > 0$, $\mu, \sigma, \nu \in \mathbb{R}^1$. Let (τ_i) be the moments of the jumps of N . Prove the Ito formula for $F \in \mathbb{C}^2$:

$$\begin{aligned} F(X_t) = F(X_0) &+ \int_0^t F'(X_s) dX_s + \frac{1}{2} \int_0^t F''(X_{s-}) \sigma^2 ds - \sum_{i=1}^{N_t} F'(X_{\tau_i-}) \Delta X_{\tau_i} \\ &+ \sum_{i=1}^{N_t} (F(X_{\tau_i}) - F(X_{\tau_i-})). \end{aligned}$$

Hint: Adapt the Ito formula for semimartingales in this case. ■

Problem 15.39 Let $N = (N_t)_{t \geq 0}$ be a Poisson process with intensity $\lambda > 0$ and $\alpha = (\alpha_t)$ be a bounded deterministic function.

Define the process

$$L_t = \exp\left\{\int_0^t \alpha_s d(N_s - \lambda s) + \int_0^t (1 + \alpha_s - e^{\alpha_s}) \lambda ds\right\}$$

and prove that $L = (L_t)_{t \geq 0}$ is a martingale satisfying the equation

$$dL_t = L_{t-}(e^{\alpha_t} - 1)d(N_t - \lambda t).$$

Hint: Use the Ito formula. ■

Problem 15.40 Let Z be a normal random variable $N(0, \sigma^2)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Construct a probability $\mathbb{Q}$ such that Z is a normal random variable $N(c, \sigma^2)$, $c \in \mathbb{R}$.

Solution Note that under probability $\mathbb{Q}$, the random variable Z must have a density

$\frac{1}{\sqrt{2\pi}\sigma} \left\{ -\frac{(z-c)^2}{2\sigma^2} \right\}$. So, we have

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(z-c)^2}{2\sigma^2}\right\} = d\mathbb{Q} = \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P}.$$

Let us note that

$$\exp\left\{-\frac{(z-c)^2}{2\sigma^2}\right\} = \exp\left\{-\frac{z^2 - 2cz + c^2}{2\sigma^2}\right\} = \exp\left\{-\frac{z^2}{2\sigma^2}\right\} \cdot \exp\left\{\frac{cz - \frac{1}{2}c^2}{\sigma^2}\right\},$$

and we arrive at the conclusion that $\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left\{\frac{cz - \frac{1}{2}c^2}{\sigma^2}\right\}$. ■

Problem 15.41 Let $(w_t^i)_{t \geq 0}$, $i = 1, 2$, be two independent Wiener processes. For the given constant $-1 < \rho < 1$, construct the process $X_t = \rho w_t^1 + \sqrt{1 - \rho^2} w_t^2$. Is it a Wiener process?

Hint: Check the definition of Wiener process. ■

Problem 15.42 Let $(w_t^i)_{t \geq 0}$ be a Wiener process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\tau_a = \inf\{t : w_t = a\}$, $a > 0$. Prove a specific form of the so-called *reflection principle* for a Wiener process w_t :

$$p\{w_t \geq a | \tau_a \leq t\} = \frac{1}{2}.$$

Hint: We note that

$$p(w_t \geq a) = p(w_t \geq a | \tau_a \leq t) \cdot p(\tau_a \leq t) + p(w_t \geq a | \tau_a > t) \cdot p(\tau_a > t)$$

and find

$$p(w_t \geq a) = p(w_t \geq a | \tau_a) \cdot p(\tau_a \leq t).$$

Therefore,

$$p(w_t \geq a | \tau_a \leq t) = p(w_t \geq a | w_{\tau_a} = a, \tau_a \leq t) = p(w_t - w_{\tau_a} \geq 0 | \tau_a \leq t) = \frac{1}{2}.$$
■

Problem 15.43 Let $(\varepsilon_n)_{n=1,2,\dots}$ be a sequence of independent standard random variables, $\varepsilon_n \sim N(0, 1)$. Define a new sequence $(Y_n)_{n=1,2,\dots}$ as follows: $Y_n = \sum_{i=1}^n \frac{\varepsilon_i}{\sqrt{n \log \log n}}$. Prove that $\lim_{n \rightarrow \infty} Y_n = 0$ in probability, but Y_n does not converge (a.s.).

Hint: First, we note that in the sense of distribution

$$Y_n \stackrel{d}{=} \frac{\sqrt{n} \varepsilon_1}{\sqrt{n \log \log n}} = \frac{\varepsilon_1}{\sqrt{\log \log n}},$$

and, hence, $Y_n \xrightarrow[n \rightarrow \infty]{p} 0$. Second, using the iterated logarithm law, we have (a.s.)

$$\limsup_{n \rightarrow \infty} Y_n = \sqrt{2} \text{ and } \liminf_{n \rightarrow \infty} Y_n = -\sqrt{2}.$$

■

Problem 15.44 Let $(\varepsilon_n)_{n=1,2,\dots}$ be a sequence of independent random variables taking values 0 or 1 on $(\Omega, \mathcal{F}, \mathbb{P})$. Prove the following relations:

$$P - \lim_n \varepsilon_n = 0 \Leftrightarrow P(\varepsilon_n = 1) \rightarrow 0 \quad (n \rightarrow \infty),$$

$$\varepsilon_n \rightarrow 0 \quad (n \rightarrow \infty) \quad (a.s.) \Leftrightarrow \sum_{n=1}^{\infty} P(\varepsilon_n = 1) < \infty,$$

$$P(\varepsilon_n \xrightarrow[n \rightarrow \infty]{} 0) = 0 \Leftrightarrow \sum_{n=1}^{\infty} P(\varepsilon_n = 1) = \infty.$$

Hint: Use the Borel-Cantelli lemma. We also note in the context of this problem that the sequence $(\varepsilon_n)_{n=1,2,\dots}$ which chosen as follows $P(\varepsilon_n = 1) = \frac{1}{n}$ will converge in probability, but not almost surely.

■

Problem 15.45 Let $(X_t^n)_{t \geq 0, n = 1, 2, \dots}$ be strong solutions of the stochastic differential equations

$$dX_t^n = -n^3 X_t^n dt + n dW_t, \quad X_0^n = 0,$$

where $(W_t)_{t \geq 0}$ is a Wiener process. Show that (X_t^n) converges to zero in space L_2 as $n \rightarrow \infty$.

Hint: Using a general formula for the Orstein-Uhlenbeck process, we obtain that

$$X_t^n = n \cdot e^{-n^3 t} \int_0^t e^{n^3 s} dW_s,$$

and, hence, $E X_t^n = 0$. Further, applying the isometric property of the Ito stochastic integral, we get

$$E(X_t^n)^2 = n^2 e^{-2n^3 t} E\left(\int_0^t e^{n^3 s} dW_s\right)^2 = n^2 \cdot e^{-2n^3 t} \frac{1}{2n^3} \cdot e^{2n^3 t} = \frac{1}{2n} \rightarrow 0 \quad (n \rightarrow \infty).$$

■

Problem 15.46 Show the way how to construct a predictable stopping time (s.t.) on the stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ without *usual conditions*.

Hint: Let σ be a wide sense stopping time and τ be a $\mathcal{F}_{\sigma+}$ -measurable r.v. such that $\sigma < \tau$ on $\{\sigma < \infty\}$. Then τ is a desirable stopping time since the sequence $\tau_n = n \wedge (\frac{1}{n}\sigma + \frac{n-1}{n}\tau)$ foretells τ .

■

Problem 15.47 Let τ be a stopping time. Show that $\mathcal{F}_{\tau-} \leq \mathcal{F}_{\tau}$ and σ is $F_{\sigma-}$ -measurable.

Hint: Take a set B such that $B \cap \{\sigma \leq t\} \in \mathcal{F}_t$ for all t . So, we have either $B \in \mathcal{F}_0$ or $B = A \cap \{\sigma > u\}$, $u \geq 0$, $A \in \mathcal{F}_u$.

■

Problem 15.48 Let $\xi, \xi_1, \dots, \xi_n$ be nonnegative random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\xi_1 + \xi_2 + \dots + \xi_n \geq \xi$. Prove the following useful inequality: for $\eta \geq 0$,

$$P(\xi > \eta) \leq \sum_{i=1}^n P(\xi_i > \frac{\eta}{n}).$$

Hint: Use an induction method or the following inclusion $\{\xi > \eta\} \subseteq \bigcup_{i=1}^n \{\xi_i > \frac{\eta}{n}\}$. This exercise can be used to bound the probability that a sum of random variables exceeds an amount in terms of the individual components of the sum.

■

Problem 15.49 In the framework of Problem 15.42, prove that the density of τ_a is equal to $\frac{a}{\sqrt{2\pi}} t^{-\frac{3}{2}} \exp(-\frac{a^2}{2t})$ and, hence,

$$P(\max_{0 \leq s \leq t} W_s \geq a) = P(\tau_a \leq t) = \frac{2}{\sqrt{2\pi}} \int_{\frac{a}{\sqrt{t}}}^{\infty} \exp(-\frac{y^2}{2}) dy.$$

Hint: Use the reflection principle for (W_t) :

$$P(W_t \geq a) = \frac{1}{2} P(\tau_a \leq t).$$

■

Problem 15.50 Let ξ be a standard normal random variable. Prove that $E\xi^4 = 3$, and, hence, skewness and kurtosis of ξ are 0 and 3, correspondingly.

Hint: Use the direct calculations.

References

1. Abdelghani, M., Melnikov, A.V.: Optional Processes: Theory and Applications. Chapman & Hall/CRC (2020)
2. Abdelghani, M., Melnikov, A.V.: On linear stochastic equations of optional semimartingales and their applications. *Stat. Probab. Lett.* **125**, 207–214 (2017)
3. Abdelghani, M., Melnikov, A.V.: Optional decomposition of optional supermartingales and applications to filtering and finance. *Stochastics* **91**(6), 797–816 (2019)
4. Abdelghani, M.N., Melnikov, A.V.: Existence and uniqueness of stochastic equations of optional semimartingales under monotonicity condition. *Stochastics* **92**(1), 67–89 (2020)
5. Abdelghani, M., Melnikov, A.V., Pak, A.: On statistical estimation and inferences in optional regression models. *Stat. A J. Theor. Appl. Stat.* **55**(2), 445–457 (2021)
6. Asmussen, S., Albrecher, H.: *Ruin Probabilities*. World Scientific, Singapore (2010)
7. Baldi, P.: *An Introduction Through Theory and Exercises. Stochastic Calculus*. Universitext (2017)
8. Beiglboeck, M., Schachermayer, W., Veliyev, B.: A short proof of the Doob–Meyer theorem. *Stoch. Process. Appl.* **122**(4), 1204–1209 (2012)
9. Bishwal, J.P.N.: *Parameter Estimation in Stochastic Differential Equations*. Springer, Berlin/Heidelberg (2008)
10. Borkar, V.S.: *Stochastic Approximation: A Dynamical Systems Viewpoint*. Cambridge University Press, Cambridge (2008)
11. Borodin, A.N.: *Stochastic Processes*. Birkhäuser, Basel (2018)
12. Bulinski, A.V., Shirayev, A.N.: *Theory of Stochastic Processes*. Fizmatlit, Moscow (2005)
13. Çinlar, E.: *Probability and Stochastics*, vol. 261. Springer, Berlin (2011)
14. Dellacherie, C., Meyer, P.-A.: Un nouveau théorème de projection et de section. *Séminaire de Probabilités IX*, pp. 239–245. Springer, Berlin (1975)
15. Cohen, S.N., Elliott, R.J.: *Stochastic Calculus and Applications*, 2nd edn. Springer, Berlin (2015)
16. Doléans-Dade, C.: *Stochastic Processes and Stochastic Differential Equations*, pp. 7–73. Springer, Berlin (2010)
17. Durrett, R.: *Essentials of Stochastic Processes*, 3rd edn. Springer, Berlin (2018)
18. Eberlein, E., Kallsen, J.: *Mathematical Finance*. Springer, Berlin (2019)
19. Edgar, G.A., Sucheston, L.: Amarts: a class of asymptotic martingales. *J. Multivariate Anal.* **6**, 193–221 (1976)
20. Etheridge, A.: *A Course in Financial Calculus*. Cambridge University Press, Cambridge (2002)
21. Falkowski, A.: SDE's with two reflecting barriers driven by optional processes with regulated trajectories. *Stoch. Process. Appl.* **173**, 225 (2025)

22. Gal'chuk, L.I., Davis, M.H.A.: A comparison theorem for equations with different diffusions. *Stochastics* **6**(2), 147–149 (1982)
23. Gal'chuk, L.I.: A comparison theorem for stochastic equations with integrals with respect to martingales and random measures. *Theory Probab. Appl.* **27**(1), 93–108 (1983)
24. Gal'chuk, L.I.: Optional martingales with respect to optional semimartingales. *Theory Probab. Appl.* **29**(1), 483–521 (1980)
25. Gasparyan, K.: Stochastic equations with respect to optional semimartingales. *Investigacia Vysshikh Uchebnykh Zavedenii* **12**, 57–60 (1985)
26. Horowitz, J.: Optional supermartingales and the Andersen–Jessen theorem. *Probab Theory Related Fields* **43**(3), 263–272 (1978)
27. Ikeda, N., Watanabe, S.: *Stochastic Differential Equations and Diffusion Processes*, 2nd edn. North-Holland, Amsterdam (1989)
28. Jacod, J., Protter, P.: *Probability Essentials*, 2nd edn. Springer, Berlin (2003)
29. Jarni, L., Ouknine, Y.: On reflection with two-sided jumps. *J. Theor. Probab.* **34**(4), 1811–1830 (2021)
30. Kallianpur, G., Karandikar, R.L.: *Introduction to Option Pricing Theory*. Springer, Berlin (2012)
31. Karatzas, I., Shreve, S.: *Brownian Motion and Stochastic Calculus*. Springer, New York (1998)
32. Klebaner, F.C.: *Introduction to Stochastic Calculus with Applications*. World Scientific Publishing Company, Singapore (2012)
33. Kolmogorov, A.N.: *Foundations of the Theory of Probability*, 2nd edn. Chelsea, New York (1956)
34. Krasin, V.Y., Smirnov, I., Melnikov, A.V.: Approximate option pricing and hedging in the CEV model via path-wise comparison of stochastic processes. *Ann. Financ.* **14**(2), 195–209 (2018)
35. Kruglov, V.M.: *Stochastic Processes*. Academy, Moscow (2013)
36. Krylov, N.V.: *Controlled Diffusion Processes*. Springer, Berlin (1980)
37. Krylov, N.V.: *Introduction to the Theory of Random Processes*. AMS, Providence (2002)
38. Kühn, C., Stroh, M., et al.: A note on stochastic integration with respect to optional semimartingales. *Electron. Commun. Probab.* **14**, 192–201 (2009)
39. Lamberton, D., Lapeyre, B.: *Introduction to Stochastic Calculus Applied to Finance*. Chapman & Hall/CRC (1996)
40. Le Gall, J.-F.: *Brownian Motion, Martingales, and Stochastic Calculus*. Springer, Berlin (2016)
41. Lenglart, E.: Tribus de Meyer et théorie des processus. *Séminaire de Probabilités XIV 1978/79*, pp. 500–546. Springer, Berlin (1980)
42. Lepingle, D.: Sur la représentation des sauts des martingales. *Séminaire de Probabilités XI*, pp. 418–434. Springer, Berlin (1977)
43. Liptser, R. Sh., Shiryaev, A.N.: *Theory of Martingales*. Kluwer Academic Publishers, Dordrecht (1989)
44. Liptser, R. Sh., Shiryaev, A.N.: *Statistics of Random Processes*, 2nd edn. Springer, Berlin (2001)
45. Melnikov, A.V.: On strong solutions of stochastic differential equations with nonsmooth coefficients. *Theory Probab. Appl.* **24**(1), 146–149 (1979)
46. Melnikov, A.V.: On the theory of stochastic equations in components of semimartingales. *Sb. Math.* **38**(3), 381–394 (1981)
47. Melnikov, A.V.: Stochastic differential equations: singularity of coefficients, regression models and stochastic approximation. *Russian Math. Surv.* **51**(5), 43–136 (1996)
48. Meyer, P.-A.: *Probability and Potential*. Blaisdell Publishing Company (1966)
49. Nevel'son, M. B., Has'minskii, R.Z.: *Stochastic Approximation and Recursive Estimation*. AMS, Providence (1976)
50. Øksendal, B.: *Stochastic Differential Equations*, 5th edn. Springer, Berlin (2000)
51. Protter, P.E.: *Stochastic Integration and Differential Equations*, 2nd edn. Springer, Berlin (2005)
52. Revuz, D., Yor, M.: *Continuous Martingales and Brownian Motion*, 2nd edn. Springer, Berlin (1999)

53. Shahrokhbadi, M.A., Melnikov, A.V., Pak, A.: An optional semimartingales approach to risk theory. *Risks* **13**(61), 1–27 (2025)
54. Schachermayer, W., Teichmann, J.: How close are the option pricing formulas of Bachelier and Black–Merton–Scholes? *Math. Financ* **18**(1), 155–170 (2008)
55. Shiryaev, A. N.: *Probability*, 2nd edn. Springer, Berlin (1996)
56. Shiryaev, A.N.: *Essentials of Stochastic Finance*. World Scientific, Singapore (1999)
57. Skorokhod, A.V.: *Lectures on the Theory of Stochastic Processes*. VSP, Utrecht (1996)
58. Sorensen, M.: A semimartingale approach to some problems in risk theory. *ASTIN Bull.* **26**(1), 15–23 (1996)
59. Tikhonov, A.N., Vasileva, A.B., Sveshnikov, A.G.: *Differential Equations*. Springer, Berlin (1985)
60. Valkeila, E., Melnikov, A.V.: Martingale models of stochastic approximation and their convergence. *Theory Probab. Appl.* **44**(2), 333–360 (2000)
61. Wentzell, A.D.: *A Course in the Theory of Stochastic Processes*. McGraw-Hill, New York (1981)
62. Williams, D.: *Probability and Martingales*. Cambridge University Press, Cambridge (1991)

Index

A

Absolute continuity, 43
 local, 79
Absolute continuity of measures, 43
Accessible stopping time, 157
Algebra, 3
Atom, 48
Autoregression model, 72, 209

B

Bachelier
 discrete model, 81
 formula, 81
Bank account, 81
Bernoulli distribution, 8
Binomial market model, 201
Black-Scholes
 formula, 144
 model, 143
Borel-Cantelli lemma, 98
Borel function, 160
Borel space, 5, 13
Brownian motion, 90

C

Cadlag process, 158
Call option, 140
Canonical decomposition, 189
Cantor function, 9
Capital of strategy, 196
Cauchy-Schwartz inequality, 30
Central limit theorem (CLT), 33
CEV market model, 247

Change of measure, 55
Change of time, 52
Change of variables formula, 20, 79
Characteristic function, 39
Chebyshev inequality, 28
Class D, 155
Comparison theorem, 126, 247
Compensator, 55, 163, 164, 215
Complete probability space, 91
Conditional expectation, 45
Consumption process, 242
Controlled diffusion process, 147
Convergence
 weak, 41
Convergence of random variables, 31
Cox-Ross-Rubinstein model, 201
Cumulative process, 259
Cylinder, 6

D

Debut, 158
Diffusion coefficient, 133
Diffusion process, 133
Dirichlet function, 19
Discrete distribution, 8
Discrete stochastic integral, 56
Distribution
 Bernoulli, 8
 binomial, 8
 density function, 11
 discrete, 8
 finite-dimensional, 131
 function, 7
 normal, 9

Poisson, 8
 uniform, 8, 9
 Doleans exponent, 188
 Doob decomposition, 54
 Doob inequalities, 161
 Doob-Meyer decomposition, 163
 Downcrossing, 61
 Drift coefficient, 127
 Dual projection, 216

E

Elementary event, outcome, 1
 Equivalence of measures, 115
 Existence of a solution, 122
 Expectation, 18
 Extended random variable, 13

F

Fair price, 140
 Fatou lemma, 24
 Feynman-Kac representation, 137
 Filtration, 52
 Financial market, 81, 143
 Finite-additive measure, 3
 Finite-dimensional
 process, 107
 Finite-dimensional distribution, 131
 First variation, 94
 Fisk decomposition, 176
 Fixed accuracy property, 195
 Fokker-Plank equation, 135
 Formula
 Bachelier, 81
 Ito, 109
 Merton, 201
 Function
 Borel, 15
 Cantor, 9
 characteristic, 39
 Dirichlet, 19
 Functional space, 89
 Fundamental sequence, 104

G

Galchuk integral, 218
 Gaussian process, 100
 Generalized martingale, 86
 General theory of stochastic processes,
 155
 Generator, 133

Geometric Brownian motion (GBM), 124, 143
 Girsanov theorem, 80
 Graph, 157
 Gronwall lemma, 121, 229

H

Haar system, 97
 Hamilton-Jacobi-Bellman (HJB) equation, 153
 Hamilton-Jacobi-Bellman (HJB) principle, 147
 Helly principle, 34

I

Incomplete market, 197
 Independent random variables, 16
 Indicator, 14
 Indistinguishable processes, 157
 Inequality
 Cauchy-Schwartz, 29
 Chebyshev, 28
 Jensen, 29
 Kolmogorov-Doob, 59
 Kunita-Watanabe, 178
 Information flow, 52
 Integrable, 19
 Integral
 Lebesgue, 19
 Interval of non-arbitrage prices, 149
 Inverse image, 13
 Isometric property, 102
 Ito formula, 109
 Ito process, 109
 Ito stochastic integral, 107

J

Jensen inequality, 29
 Joint quadratic characteristic, 55
 Jump-diffusion model, 201

K

Kolmogorov
 backward, forward equation, 134
 consistency theorem, 6
 variance condition, 72
 Kolmogorov-Chapman equation, 131
 Kolmogorov-Doob inequality, 59
 Kunita-Watanabe
 decomposition, 56
 inequality, 178
 Kurtosis, 21

L

Ladlag process, 212
 Large numbers law (LNL), 31
 Least-squares (LS) estimate, 72
 Lebesgue
 dominated convergence theorem, 25
 integral, 19
 Lebesgue measure, 6
 Levy theorem, 68
 Lindeberg condition, 42
 Lipschitz condition, 120
 Localization, 106
 Localizing sequence, 86
 Local Lipschitz conditions, 120
 Locally square-integrable martingale, 170
 Local martingale, 86, 170
 Local martingale deflator, 245
 Lower option price, 149

M

Markov process, 130
 Martingale, 54
 generalized, 86
 local, 86, 170
 locally square-integrable, 170
 square integrable, 55
 Martingale difference, 54
 Martingale measure, 141
 Martingale representation, 116
 Mathematical finance, 124, 140
 Maturity time, 140
 Measurable
 mapping, 16
 Measurable space, 4
 Measure, 1
 finite-additive, 3
 Lebesgue, 6
 Martingale, 141
 Wiener, 12
 Merton formula, 201
 Method of monotonic approximations, 126
 Modification
 continuous, 93
 right-continuous, 96
 Monotonic class, 5
 Multiplication rule, 203

N

Normal distribution, 9
 Novikov condition, 81

O

Optimal control, 147
 Option
 call, 81
 Optional
 local martingales, 231
 martingales, 239
 semimartingales, 239
 stochastic analysis, 211
 stochastic calculus, 217
 stochastic differential equations, 225
 stochastic exponent, 231
 stochastic logarithm, 231
 stochastic process, 211
 Optional decomposition, 250
 Optional processes, 212
 Optional regression analysis, 250
 Optional risk theory, 257
 Optional sampling theorem, 58
 Optional sigma-algebra, 159
 Ornstein-Uhlenbeck process, 125

P

Parseval identity, 100
 Partial differential equation (PDE), 137
 Poisson distribution, 8
 Polynomial, 112
 Predictable measure, 214
 Predictable process, 159
 Predictable sequence, 55
 Predictable sigma-algebra, 159
 Predictable stopping time, 157
 Principle of optimality, 147
 Probability measure, 3
 Probability space, 3, 6
 Process with finite variation, 160
 Progressively measurable, 109
 Prokhorov theorem, 36
 Purely discontinuous martingale, 167
 Put-call parity, 142

Q

Quadratic bracket, 168
 Quadratic characteristic, 55, 169
 Quasimartingale, 85

R

Radon-Nikodym
 density/derivative, 44
 theorem, 44
 Random measure, 213

Random set, 159
 Random variable, 13
 extended, 13
 uniformly integrable, 25
 Regression
 analysis, 72
 function, 77
 Regression model, 72
 Relative compactness, 36
 Replicating strategy, 196
 Risk process, 257
 Ruin probability, 257

S

Sample path, 51
 Scalar product, 97
 Schauder system, 97
 Semimartingale, 173
 Sequential estimate, 146
 Sigma-algebra, 4
 optional, 159
 predictable, 159
 Singular distribution, measure, 9
 Skewness, 21
 Small perturbations method, 151
 Step function, 101
 Stochastically continuous process, 96
 Stochastic analysis, 52
 Stochastic approximation, 77, 203
 Stochastic basis, 57
 Stochastic calculus, 178
 Stochastic differential equation (SDE), 119
 Stochastic exponential, 56, 188
 multiplication rule, 188
 Stochastic interval, 159
 Stochastic Kronecker's Lemma, 70
 Stochastic regression analysis, 203
 Stochastic volatility, 149
 Stock price, 49, 124
 Stopped process, 172
 Stopping time, 52
 accessible, 157
 graph, 157
 predictable, 157
 totally inaccessible, 157
 Stopping time in wide sense, 212
 Strategy, 196
 Stratonovich integral, 108

Strongly predictable process, 211
 Strong solution, 125
 Structural condition, 227
 Structural LS-estimator, 255
 Subdivision, 92
 Submartingale, 54
 Supermartingale, 54

T

Theorem
 Caratheodory, 6
 CLT, 41
 comparison, 126
 Girsanov, 79
 Kolmogorov consistency, 7
 Lebesgue dominated convergence, 30
 Levy, 68
 optional sampling, 58
 Prokhorov, 36
 Radon-Nikodym, 44
 Theory of probability, 1
 Tightness, 36
 Totally inaccessible stopping time, 157
 Transition probability function, 131

U

Uniform distribution, 9
 Uniform Doob-Meyer decomposition, 259
 Uniformly integrable random variable, 25
 Uniqueness of solutions, 125, 147
 Unusual stochastic basis, 211
 Upcrossing, 60
 Upper option price, 149
 Usual conditions, 155, 161

V

Value function, 148
 Variance, 21

W

Wald identity, 105
 Weak convergence, 41
 Weak solution, 125
 Wiener measure, 12