Modeling and Simulation in
Science, Engineering and Technology

An Introduction to Continuous-Time Stochastic Processes

Theory, Models, and Applications to Finance, Biology, and Medicine

Vincenzo Capasso
David Bakstein

BIRKHÄUSER

Modeling and Simulation in Science, Engineering and Technology

Vincenzo Capasso
David Bakstein

An Introduction to Continuous-Time Stochastic Processes

Theory, Models, and Applications to Finance, Biology, and Medicine

Birkhäuser
Boston • Basel • Berlin

Vincenzo Capasso
MIRIAM (Milan Research Centre for
Industrial and Applied Mathematics)
University of Milan
Department of Mathematics
20133 Milan
Italy

David Bakstein
MIRIAM (Milan Research Centre for
Industrial and Applied Mathematics)
University of Milan
Department of Mathematics
20133 Milan
Italy

AMS Subject Classifications: 93-xx, 93-02, 93B05, 93B07, 93B10, 93B11, 93B17, 93B25, 93B27, 93B36, 93B40, 93B51, 93B52, 93B55, 93B60, 93C05, 93C15, 93C35, 93C55, 93C70

Library of Congress Cataloging-in-Publication Data
Capasso, V. (Vincenzo), 1945-
An introduction to continuous-time stochastic processes : theory, models, and applications to finance, biology, and medicine / Vincenzo Capasso, David Bakstein.
p. cm — (Modeling and simulation in science, engineering and technology)
Includes bibliographical references and index.
ISBN 0-8176-3234-4 (acid-free paper)
1. Stochastic processes. I. Bakstein, David, 1975- II. Title. III. Modeling and simulation in science, engineering & technology.

QA274.C36 2004
519.2'3–dc22 2003063634

ISBN 0-8176-3234-4 Printed on acid-free paper.

Printed in the United States of America. (SB)

9 8 7 6 5 4 3 2 1 SPIN 10925249

www.birkhauser.com

Preface

This book is a systematic, rigorous, and self-consistent introduction to the theory of continuous-time stochastic processes. But it is neither a tract nor a recipe book as such; rather, it is an account of fundamental concepts as they appear in relevant modern applications and literature. We make no pretense of it being complete. Indeed, we have omitted many results, which we feel are not directly related to the main theme or that are available in easily accessible sources. (Those readers who are interested in the historical development of the subject cannot ignore the volume edited by Wax (1954).)

Proofs are often omitted as technicalities might distract the reader from a conceptual approach. They are produced whenever they may serve as a guide to the introduction of new concepts and methods towards the applications; otherwise, explicit references to standard literature are provided. A mathematically oriented student may find it interesting to consider proofs as exercises.

The scope of the book is profoundly educational, related to modeling real-world problems with stochastic methods. The reader becomes critically aware of the concepts involved in current applied literature, and is moreover provided with a firm foundation of the mathematical techniques. Intuition is always supported by mathematical rigor.

Our book addresses three main groups: first, mathematicians working in a different field; second, other scientists and professionals from a business or academic background; third, graduate or advanced undergraduate students of a quantitative subject related to stochastic theory and/or applications.

As stochastic processes (compared to other branches of mathematics) are relatively new, yet more and more popular in terms of current research output and applications, many pure as well as applied deterministic mathematicians have become interested in learning about the fundamentals of stochastic theory and modern applications. This book is written in a language that both groups will understand, and in its content and structure will allow them to learn the essentials profoundly and in a time-efficient manner. Other scientist-practitioners and academics from fields like finance, biology, or medicine might be very familiar with a less mathematical approach to their specific fields, and thus be interested in learning the mathematical techniques of modeling their applications.

Furthermore, this book would be suitable as a textbook accompanying a graduate or advanced undergraduate course or as a secondary reading for students of mathematical or computational sciences. The book has evolved from course material that has already been tested for many years for various courses in engineering, biomathematics, industrial mathematics, and mathematical finance.

Last but certainly not least, this book should also appeal to anyone who would like to learn about the mathematics of stochastic processes. The reader will see that previous exposure to probability, even though helpful, is not essential and that the fundamentals of measure and integration are provided in a self-consistent way. Only familiarity with calculus and some analysis is required.

The book is divided into three main parts. In part I, comprising chapters 1–4, we introduce the foundations of the mathematical theory of stochastic processes and stochastic calculus, thus providing tools and methods needed in part II (chapters 5 and 6), which is dedicated to major scientific areas of applications. The third part consists of appendices, each of which gives a basic introduction to a particular field of fundamental mathematics (like measure, integration, metric spaces, etc.) and explains certain problems in greater depth (e.g., stability of ODEs) than would be appropriate in the main part of the text.

In chapter 1 the fundamentals of probability are provided following a standard approach based on Lebesgue measure theory due to Kolmogorov. Here the guiding textbook on the subject is the excellent monograph by Métivier (1968). Basic concepts from Lebesgue measure theory are furthermore provided in appendix A.

Chapter 2 gives an introduction to the mathematical theory of stochastic processes in continuous time, including basic definitions and theorems on processes with independent increments, martingales, and Markov processes. The two fundamental classes of processes, namely Poisson and Wiener, are introduced as well as the larger, more general, class of Lévy processes. Further, a significant introduction to marked point processes is also given as a support for the analysis of relevant applications.

Chapter 3 is based on Itô theory. We define the Itô integral, some fundamental results of Itô calculus, and stochastic differentials including Itô's formula, as well as related results like the martingale representation theorem.

Chapter 4 is devoted to the analysis of stochastic differential equations driven by Wiener processes and Itô diffusions, and demonstrates the connections with partial differential equations of second order, via Dynkin and Feynman–Kac formulas.

Chapter 5 is dedicated to financial applications. It covers the core economic concept of arbitrage-free markets and shows the connection with martingales and Girsanov's theorem. It explains the standard Black–Scholes theory and relates it to Kolmogorov's partial differential equations and the Feynman–Kac formula. Furthermore, extensions and variations of the standard theory are discussed as well as interest rate models and insurance mathematics.

Chapter 6 presents fundamental models of population dynamics such as birth and death processes. Furthermore, it deals with an area of important modern research, namely the fundamentals of self-organizing systems, in particular focusing on the social behavior of multiagent systems, with some applications to economics ("price herding"). It also includes a particular application to the neurosciences, illustrating the importance of stochastic differential equations driven by both Poisson and Wiener processes.

Problems and additions are proposed at the end of the volume, listed by chapter. More than being just exercises in a classical way, problems are proposed as a stimulus for discussing further concepts which can be of interest for the reader. Different sources have been used, including a selection of problems submitted to our students over the years. This is the reason why we can provide only selected references.

The core of this monograph, on Itô calculus, was developed during a series of courses that one of the authors VC has been offering at various levels in many universities. That author wishes to acknowledge that the first drafts of the relevant chapters were the outcome of a joint effort by many participating students: Maria Chiarolla, Luigi De Cesare, Marcello De Giosa, Lucia Maddalena, and Rosamaria Mininni, among others. Professor Antonio Fasano is due our thanks for his continuous support, including producing such material as lecture notes within a series that he has coordinated.

It was the success of these lecture notes, and the particular enthusiasm of the coauthor DB, who produced the first English version (indeed, an unexpected Christmas gift), that has led to an extension of the material up to the present status, including in particular a set of relevant and updated applications, which reflect the interests of the two authors.

VC also would like to thank his first advisor and teacher, Professor Grace Yang, who gave him the first rigorous presentation of stochastic processes and mathematical statistics at the University of Maryland at College Park, always referring to real world applications. DB would like to thank the Meregalli and Silvestri families for their kind logistical help while in Milan. He would also like to acknowledge research funding from the EPSRC, ESF, Socrates–Erasmus, and Charterhouse and thank all the people he worked with at OCIAM, University of Oxford, over the years, as this is where he was based when embarking on this project.

The draft of the final volume has been carefully read by Giacomo Aletti, Daniela Morale, Alessandra Micheletti, Matteo Ortisi, and Enea Bongiorno (who also took care of the problems and additions) whom we gratefully acknowledge. Still, we are sure that some odd typos and other, hopefully noncrucial, mistakes remain, for which the authors take all responsibility.

We also wish to thank Professor Nicola Bellomo, editor of the Modeling and Simulation in Science, Engineering, and Technology Series, and Tom Grasso from Birkhäuser for supporting the project. Last but not the least, we cannot forget to thank Rossana VC and Casilda DB for their patience and great tolerance while coping with their “solitude” during the preparation of this monograph.

Vincenzo Capasso and David Bakstein
Milan, November 2003

Contents

Part I

The Theory of Stochastic Processes

1

Fundamentals of Probability

We assume that the reader is already familiar with the basic motivations and notions of probability theory. In this chapter we recall the main mathematical concepts, methods, and theorems according to the Kolmogorov approach (see Kolmogorov (1956)), by using as a main reference the book by Métivier (1968). We shall refer to appendix A of this book for the required theory on measure and integration.

1.1 Probability and Conditional Probability

Definition 1.1. A *probability space* is an ordered triple $(\Omega, \mathcal{F}, P)$, where Ω is any set, $\mathcal{F}$ a σ-algebra of subsets of Ω, and $P : \mathcal{F} \to [0,1]$ a probability measure on $\mathcal{F}$, such that

1. $P(\Omega) = 1$ (and $P(\emptyset) = 0$),
2. for all $A_1, \ldots, A_n, \ldots \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$, $i \neq j$:

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i).$$

The set Ω is called the *sample space*, $\emptyset$ the *empty set*, the elements of $\mathcal{F}$ are called *events*, and every element of Ω is called an *elementary event*.

Definition 1.2. A probability space $(\Omega, \mathcal{F}, P)$ is *finite* if Ω has finitely many elementary events.

Remark 1.3. If Ω is finite, then it suffices to only consider the σ-algebra of all subsets of Ω, i.e., $\mathcal{F} = \mathfrak{P}(\Omega)$.

Definition 1.4. Every finite probability space $(\Omega, \mathcal{F}, P)$ with $\mathcal{F} = \mathfrak{P}(\Omega)$ is an *equiprobable* or *uniform* space, if

$$\forall \omega \in \Omega : \qquad P(\{\omega\}) = k \text{ (constant)};$$

i.e., its elementary events are equiprobable.

Remark 1.5. Following the axioms of a probability space and the definition of a uniform space, if $(\Omega, \mathcal{F}, P)$ is equiprobable, then

$$\forall \omega \in \Omega : \qquad P(\{\omega\}) = \frac{1}{|\Omega|},$$

where $|\cdot|$ denotes the cardinal number of elementary events in Ω, and

$$\forall A \in \mathcal{F} \equiv \mathfrak{P}(\Omega) : \qquad P(A) = \frac{|A|}{|\Omega|}.$$

Intuitively, in this case we may say that $P(A)$ is the ratio of the number of favorable outcomes, divided by the number of all possible outcomes.

Example 1.6. Consider an urn that contains 100 balls, of which 80 are red and 20 are black but that are otherwise identical, from which a player draws a ball. Define the event

R: The first drawn ball is red.

Then

$$P(R) = \frac{|R|}{|\Omega|} = \frac{80}{100} = 0.8.$$

Definition 1.7. We shall call any event $F \in \mathcal{F}$ such that $P(F) = 0$, a *null event*.

Conditional Probability

Definition 1.8. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $A, B \in \mathcal{F}$, $P(B) > 0$. Then the probability of A conditional on B, denoted by $P(A|B)$, is any real number in $[0, 1]$ such that

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

This number is left unspecified whenever $P(B) = 0$.

We must anyway notice that conditioning events of zero probability cannot be ignored. See later a more detailed account of this case in connection with the definition of conditional distributions.

Remark 1.9. Suppose that $P(B) > 0$. Then the mapping $P_B : \mathcal{F} \to [0, 1]$ with

$$\forall A \in \mathcal{F} : \qquad P_B(A) = \frac{P(A \cap B)}{P(B)}$$

defines a probability measure P_B on $\mathcal{F}$. In fact, $0 \leq P_B(A) \leq 1$ and $P_B(\Omega) = \frac{P(B)}{P(B)} = 1$. Moreover, if $A_1, \ldots, A_n, \ldots \in \mathcal{F}$, $A_i \cap A_j = \emptyset$, $i \neq j$, then

$$P_B\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \frac{P(\bigcup_n A_n \cap B)}{P(B)} = \frac{\sum_n P(A_n \cap B)}{P(B)} = \sum_n P_B(A_n).$$

Proposition 1.10. *If $A, B \in \mathcal{F}$, then*

1. $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$;
2. *if* $A_1, \ldots, A_n \in \mathcal{F}$, *then*

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}).$$

Proof: Statement 1 is obvious. Statement 2 is proved by induction. The proposition holds for $n = 2$. Assuming it holds for $n - 1$, we get

$$\begin{aligned}
&P(A_1 \cap \cdots \cap A_n) \\
&= P(A_1 \cap \cdots \cap A_{n-1})P(A_n|A_1 \cap \cdots \cap A_{n-1}) \\
&= P(A_1) \cdots P(A_{n-1}|A_1 \cap \cdots \cap A_{n-2})P(A_n|A_1 \cap \cdots \cap A_{n-1});
\end{aligned}$$

thus it holds for n as well. Since n was arbitrary, the proof is complete. □

Definition 1.11. Two events A and B are *independent* if

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

Thereby A is independent of B, if and only if B is independent of A, and vice versa.

Proposition 1.12. *Let A, B be events and $P(A) > 0$; then the following two statements are equivalent:*

1. *A and B are independent,*
2. $P(B|A) = P(B)$.

If $P(B) > 0$, then the statements hold with interchanged A and B as well.

Example 1.13. Considering the same experiment as in Example 1.6, we define the additional events (x, y) with $x, y \in \{B, R\}$ as, e.g.,

BR: The first drawn ball is black, the second red,
$\cdot R$: The second drawn ball is red.

Now the probability $P(\cdot R|R)$ depends on the rules of the draw.

1. If the draw is with subsequent replacement of the ball, then due to the independence of the draws,

$$P(\cdot R|R) = P(\cdot R) = P(R) = 0.8.$$

2. If the draw is without replacement, then the second draw is dependent on the outcome of the first draw, and we have

$$P(\cdot R|R) = \frac{P(\cdot R \cap R)}{P(R)} = \frac{P(RR)}{P(R)} = \frac{80 \cdot 79 \cdot 100}{100 \cdot 100 \cdot 80} = 0.79.$$

Definition 1.14. Two events A and B are mutually exclusive if $A \cap B = \emptyset$.

Proposition 1.15. *1. Two events cannot be both independent and mutually exclusive, unless one of the two is a null event.*

2. If A and B are independent events, then so are A and $\bar{B}$, $\bar{A}$ and B, as well as $\bar{A}$ and $\bar{B}$, where $\bar{A} := \Omega \setminus A$ is the complementary event.

Definition 1.16. The events A, B, C are independent if

1. $P(A \cap B) = P(A)P(B)$,
2. $P(A \cap C) = P(A)P(C)$,
3. $P(B \cap C) = P(B)P(C)$,
4. $P(A \cap B \cap C) = P(A)P(B)P(C)$.

This definition can be generalized to any number of events.

Remark 1.17. If A, B, C are events that satisfy point 4 of Definition 1.16, then it is not true in general that it satisfies points 1–3 and vice versa.

Example 1.18. Consider a throw of two distinguishable, fair six-sided dice, and the events

A: the roll of the first dice results in 1, 2 or 5,
B: the roll of the first dice results in 4, 5 or 6,
C: the sum of the results of the roll of the dice is 9.

Then $P(A) = P(B) = 1/2$ and $P(A \cap B) = 1/6 \neq 1/4 = P(A)P(B)$. But, since $P(C) = 1/9$ and $P(A \cap B \cap C) = 1/36$, we have that

$$P(A)P(B)P(C) = \frac{1}{36} = P(A \cap B \cap C).$$

On the other hand, consider a uniformly shaped tetrahedron, which has the colors white, green and red on its separate surfaces and all three colors on the fourth. Randomly choosing one side, the events

W: the surface contains white,
G: the surface contains green,
R: the surface contains red,

have the probabilities $P(W) = P(G) = P(R) = 1/2$. Hence $P(W \cap G) = P(W)P(G) = 1/4$, etc., but $P(W)P(G)P(R) = 1/8 \neq 1/4 = P(W \cap G \cap R)$.

Definition 1.19. Let $\mathcal{C}_1, \ldots, \mathcal{C}_k$ be subfamilies of the σ-algebra $\mathcal{F}$. They constitute k *mutually independent* classes of $\mathcal{F}$ if

$$\forall A_1 \in \mathcal{C}_1, \ldots, \forall A_k \in \mathcal{C}_k : \qquad P(A_1 \cap \cdots \cap A_k) = \prod_{i=1}^{k} P(A_i).$$

Definition 1.20. A family of elements $(B_i)_{i\in I}$ of $\mathcal{F}$, with $I \subset \mathbb{N}$, is called a *(countable) partition* of Ω if

1. I is a countable set,
2. $i \neq j \Rightarrow B_i \cap B_j = \emptyset$,
3. $P(B_i) \neq 0$, for all $i \in I$,
4. $\Omega = \bigcup_{i\in I} B_i$.

Theorem 1.21. (Total law of probability). *Let $(B_i)_{i\in I}$ be a partition of Ω and $A \in \mathcal{F}$; then*

$$P(A) = \sum_{i\in I} P(A|B_i)P(B_i).$$

Proof:

$$\begin{aligned}\sum_i P(A|B_i)P(B_i) &= \sum_i \frac{P(A\cap B_i)}{P(B_i)}P(B_i) = \sum_i P(A\cap B_i)\\ &= P\left(\bigcup_i (A\cap B_i)\right) = P\left(A \cap \bigcup_i B_i\right)\\ &= P(A\cap\Omega) = P(A).\end{aligned}$$

□

The following fundamental Bayes theorem provides a formula for the exchange of conditioning between two events; this is why it is also known as the *theorem for probability of causes.*

Theorem 1.22. (Bayes). *Let $(B_i)_{i\in I}$ be a partition of Ω and $A \in \mathcal{F}$, with $P(A) = 0$; then*

$$\forall i \in I: \qquad P(B_i|A) = \frac{P(B_i)}{P(A)}P(A|B_i) = \frac{P(A|B_i)P(B_i)}{\sum_{j\in I} P(A|B_j)P(B_j)}.$$

Proof: Since $A = \bigcup_{j=1}^k (B_j \cap A)$, then

$$P(A) = \sum_{j=1}^k P(B_j)P(A|B_j).$$

Also, because

$$P(B_i\cap A) = P(A)P(B_i|A) = P(B_i)P(A|B_i)$$

and by the total law of probability, we obtain

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^k P(A|B_j)P(B_j)}.$$

□

Example 1.23. Continuing with the experiment of Example 1.6, we further assume that there is a second indistinguishable urn (U_2), containing 40 red balls and 40 black balls. By randomly drawing one ball from one of the two urns, we make a probability estimate about which urn we had chosen:

$$P(U_2|B) = \frac{P(U_2)P(B|U_2)}{\sum_{i=1}^{2} P(U_i)P(B|U_i)} = \frac{1/2 \cdot 1/2}{1/2 \cdot 1/5 + 1/2 \cdot 1/2} = \frac{5}{7};$$

thus $P(U_1|B) = 2/7$.

1.2 Random Variables and Distributions

A random variable is the concept of assigning a numerical magnitude to elementary outcomes of a random experiment, measuring certain of the latter's characteristics. Mathematically, we define it as a function $X : \Omega \to \mathbb{R}$ on the probability space $(\Omega, \mathcal{F}, P)$, such that for every elementary $\omega \in \Omega$ it assigns a numerical value $X(\omega)$. In general, we are then interested in finding the probabilities of events of the type

$$[X \in B] := \{\omega \in \Omega | X(\omega) \in B\} \subset \Omega \tag{1.1}$$

for every $B \subset \mathbb{R}$, i.e., the probability that the random variable will assume values that will lie within a certain range $B \subset \mathbb{R}$. In its simplest case, B can be a possibly unbounded interval or union of intervals of $\mathbb{R}$. More generally, B can be any subset of the Borel σ-algebra $\mathcal{B}_{\mathbb{R}}$, which is generated by the intervals of $\mathbb{R}$. This will require, among other things the results of measure theory and Lebesgue integration in $\mathbb{R}$. Moreover, we will require the events (1.1) to be P-measurable, thus belonging to $\mathcal{F}$. We will later extend the concept of random variables to generic measurable spaces.

Definition 1.24. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then every Borel-measurable mapping $X : \Omega \to \mathbb{R}$, with for all $B \in \mathcal{B}_{\mathbb{R}} : X^{-1}(B) \in \mathcal{F}$, is a *random variable*, denoted by $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. If X takes values in $\bar{\mathbb{R}}$, then it is said to be *extended*.

Definition 1.25. If $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is a random variable, then the mapping $P_X : \mathcal{B}_{\mathbb{R}} \to \mathbb{R}$, where

$$P_X(B) = P(X^{-1}(B)) = P([X \in B]) \qquad \forall B \in \mathcal{B}_{\mathbb{R}},$$

is a probability on $\mathbb{R}$. It is called the *probability law* of X.

If a random variable X has a probability law P_X, we will use the notation $X \sim P_X$.

The following proposition shows that a random variable can be defined in a canonical way in terms of a given probability law on $\mathbb{R}$.

Proposition 1.26. *If* $P : \mathcal{B}_{\mathbb{R}} \to [0,1]$ *is a probability, then there exists a random variable* $X : \mathbb{R} \to \mathbb{R}$ *such that* P *is identical to the probability law* P_X *associated with* X*.*

Proof: We identify $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P)$ as the underlying probability space so that the mapping $X : \mathbb{R} \to \mathbb{R}$, with $X(s) = s$, for all $s \in \mathbb{R}$, is a random variable, and furthermore, denoting its associated probability law by P_X, we obtain

$$P_X(B) = P(X^{-1}(B)) = P(B) \qquad \forall B \in \mathcal{B}_{\mathbb{R}}.$$

□

Definition 1.27. Let X be a random variable. Then the mapping

$$F_X : \mathbb{R} \to [0,1],$$

with

$$F_X(t) = P_X(]-\infty, t]) = P([X \leq t]) \qquad \forall t \in \mathbb{R},$$

is called the *partition function* or *cumulative distribution function* of X.

Proposition 1.28. *1. For all* $a, b \in \mathbb{R}$, $a < b$*:* $F_X(b) - F_X(a) = P_X(]a, b])$*.*
2. F_X *is right-continuous and increasing.*
3. $\lim_{t \to +\infty} F_X(t) = 1$, $\lim_{t \to -\infty} F_X(t) = 0$*.*

Proof: Points 1 and 2 are obvious, given that P_X is a probability. Point 3 can be demonstrated by applying points 2 and 4 of Proposition A.23. In fact, by the former, we obtain

$$\begin{aligned}\lim_{t \to +\infty} F_X(t) &= \lim_{t \to +\infty} P_X(]-\infty, t]) = \lim_{n} P_X(]-\infty, n]) \\ &= P_X\left(\bigcup_n]-\infty, n]\right) = P_X(\mathbb{R}) = 1.\end{aligned}$$

Analogously, by point 4 of Proposition A.23, we get $\lim_{t \to -\infty} F_X(t) = 0$. □

Proposition 1.29. *Conversely, if we assign a function* $F : \mathbb{R} \to [0,1]$ *that satisfies points 2 and 3 of Proposition 1.28, then, by point 1, we can define a probability* $P_X : \mathcal{B}_{\mathbb{R}} \to \mathbb{R}$ *associated with a random variable* X *whose cumulative distribution function is identical to* F*.*

Definition 1.30. If the probability law $P_X : \mathcal{B}_{\mathbb{R}} \to [0,1]$ associated with the random variable X is endowed with a density with respect to Lebesgue measure[1] μ on $\mathbb{R}$, then this density is called the *probability density* of X. If $f : \mathbb{R} \to \mathbb{R}_+$ is the probability density of X, then

$$\forall t \in \mathbb{R} : \qquad F_X(t) = \int_{-\infty}^{t} f d\mu \text{ and } \lim_{t \to +\infty} F_X(t) = \int_{-\infty}^{+\infty} f d\mu = 1,$$

[1] See Definition A.52.

as well as

$$P_X(B) = \int_B f d\mu \qquad \forall B \in \mathcal{B}_{\mathbb{R}}.$$

We may notice that the Lebesgue–Stieltjes measure, canonically associated with F_X as defined in Definition A.50 is identical to P_X.

Definition 1.31. A random variable X is *continuous* if its cumulative distribution function F_X is continuous.

Remark 1.32. X is continuous if and only if $P(X = x) = 0$ for every $x \in \mathbb{R}$.

Definition 1.33. A random variable X is *absolutely continuous* if F_X is absolutely continuous or, equivalently, if P_X is defined through its density.[2]

Proposition 1.34. *Every absolutely continuous random variable is continuous, but not vice versa.*

Example 1.35. Let $F : \mathbb{R} \to [0, 1]$ be an extension to the Cantor function $f : [0, 1] \to [0, 1]$, given by

$$\forall x \in \mathbb{R} : \qquad F(x) = \begin{cases} 1 & \text{if } > 1, \\ f(x) & \text{if } x \in [0,1], \\ 0 & \text{if } x < 0, \end{cases}$$

where f is endowed with the following properties:

1. f is continuous and increasing,
2. $f' = 0$ almost everywhere,
3. f is not absolutely continuous.

Hence X is a random variable with continuous but not absolutely continuous distribution function F.

Remark 1.36. Henceforth we will use "continuous" in the sense of "absolutely continuous".

Remark 1.37. If $f : \mathbb{R} \to \mathbb{R}_+$ is a function that is integrable with respect to Lebesgue measure μ on $\mathbb{R}$ and

$$\int_{\mathbb{R}} f d\mu = 1,$$

then there exists an absolutely continuous random variable with probability density f. Defining

$$F(x) = \int_{-\infty}^{x} f(t) dt \qquad \forall x \in \mathbb{R},$$

then F is a cumulative distribution function.

[2] See Proposition A.56.

Example 1.38. (Continuous probability densities).

1. Uniform (denoted $U(a,b)$):

$$\forall x \in [a,b]: \qquad f(x) = \frac{1}{b-a}, \qquad a,b \in \mathbb{R}, a < b.$$

2. Standard normal or standard Gaussian (denoted $N(0,1)$):

$$\forall x \in \mathbb{R}: \qquad f(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}, \tag{1.2}$$

also denoted $N(0,1)$.

3. Normal or Gaussian (denoted $N(m,\sigma)^2$):

$$\forall x \in \mathbb{R}: \qquad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\}, \qquad \sigma > 0, m \in \mathbb{R}.$$

4. Exponential (denoted $E(\lambda)$):

$$\forall x \in \mathbb{R}_+: \qquad f(x) = \lambda e^{-\lambda x},$$

where $\lambda > 0$.

5. Gamma (denoted $\Gamma(\lambda,\alpha)$):

$$\forall x \in \mathbb{R}_+: \qquad f(x) = \frac{e^{-\lambda x}}{\Gamma(\alpha)} \lambda(\lambda x)^{\alpha-1},$$

where $\lambda, \alpha \in \mathbb{R}_+^*$. Here

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

is the gamma function, which for $n \in \mathbb{N}^*$ is $(n-1)!$, i.e., a generalized factorial.

6. Standard Cauchy (denoted $C(0,1)$):

$$\forall x \in \mathbb{R}: \qquad f(x) = \frac{1}{\pi}\frac{1}{1+x^2}.$$

7. Cauchy (denoted $C(a,h)$):

$$\forall x \in \mathbb{R}: \qquad f(x) = \frac{1}{\pi}\frac{h}{h^2+(x-a)^2}.$$

Definition 1.39. Let X be a random variable and let D denote a countable set of real numbers $D = \{x_1, \ldots, x_n, \ldots\}$. If there exists a function $p: \mathbb{R} \to [0,1]$, with

1. for all $x \in \mathbb{R} \setminus D: p(x) = 0$,

2. for all $B \in \mathcal{B}_{\mathbb{R}}$: $\sum_{x\in B} p(x) < +\infty$,
3. for all $B \in \mathcal{B}_{\mathbb{R}}$: $P_X(B) = \sum_{x\in B} p(x)$,

then X is *discrete* and p is the (discrete) distribution function of X. The set D is called the *support* of the function p.

Remark 1.40. Let p denote the discrete distribution function of the random variable X, having support D. The following properties hold:

1. $\sum_{x\in D} p(x) = 1$.
2. For all $B \in \mathcal{B}_{\mathbb{R}}$ such that $D \cap B = \emptyset$, $P_X(B) = 0$.
3. For all $x \in \mathbb{R}$:
$$P_X(\{x\}) = \begin{cases} 0 & \text{if } x \notin D, \\ p(x) & \text{if } x \in D. \end{cases}$$

Hence P_X corresponds to the discrete measure associated with the "masses" $p(x)$, $x \in D$.

Example 1.41. (Discrete probability distributions).

1. Uniform:
$$\forall i = 1, \ldots, n : \qquad p(x_i) = \frac{1}{n}, \qquad n \in \mathbb{N}.$$
2. Poisson:
$$\forall x \in \mathbb{N} : \qquad p(x) = \exp\{-\lambda\}\frac{\lambda^x}{x!}, \qquad \lambda > 0,$$
also denoted by $P(\lambda)$, where λ is said to be the intensity.
3. Binomial:
$$\forall x = 0, 1, \ldots, n : \qquad p(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}, \qquad n \in \mathbb{N}, p \in [0, 1],$$
also denoted by $B(n, p)$.

Remark 1.42. The cumulative distribution function F_X of a discrete random variable X is an RCLL (right-continuous with left limit) function with finite jumps. If p is the distribution function of X, then

$$p(x) = F_X(x) - F_X(x^-) \qquad \forall x \in D,$$

or, more generally,

$$p(x) = F_X(x) - F_X(x^-) \qquad \forall x \in \mathbb{R}.$$

The concept of random variable can be extended to any function defined on a probability space $(\Omega, \mathcal{F}, P)$ and valued in a measurable space $(E, \mathcal{B})$, i.e., a set E endowed with a σ-algebra $\mathcal{B}$ of its parts.

Definition 1.43. Every measurable function $X : \Omega \to E$, with $X^{-1}(B) \in \mathcal{F}$, for all $B \in \mathcal{B}$, assigned on the probability space $(\Omega, \mathcal{F}, P)$ and valued in $(E, \mathcal{B})$ is a random variable. The probability law P_X associated with X is defined by translating the probability P on $\mathcal{F}$ into a probability on $\mathcal{B}$, through the mapping $P_X : \mathcal{B} \to [0, 1]$, such that

$$\forall B \in \mathcal{B} : \qquad P_X(B) = P(X^{-1}(B)) \equiv P(X \in B).$$

Definition 1.44. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(E, \mathcal{B})$ a measurable space. Further, let E be a normed space of dimension n, and let $\mathcal{B}$ be its Borel σ-algebra. Every vector $\mathbf{X} : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ is called a *random vector*. In particular, we can take $(E, \mathcal{B}) = (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.

Remark 1.45. The Borel σ-algebra on $\mathbb{R}^n$ is identical to the product σ-algebra of the family of n Borel σ-algebras on $\mathbb{R}$: $\mathcal{B}_{\mathbb{R}^n} = \bigotimes_n \mathcal{B}_{\mathbb{R}}$.

Proposition 1.46. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathbf{X} : \Omega \to \mathbb{R}^n$ a mapping. Moreover, let, for all $i = 1, \ldots, n$, $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be the ith projection, and thus $X_i = \pi_i \circ \mathbf{X}$, $i = 1, \ldots, n$, be the ith component of $\mathbf{X}$. Then the following statements are equivalent:*

1. *$\mathbf{X}$ is a random vector of dimension n.*
2. *For all $i \in \{1, \ldots, n\}$, X_i is a random variable.*

Proof: The proposition is an obvious consequence of Proposition A.17. □

Definition 1.47. Under the assumptions of the preceding proposition, the function

$$\forall B_i \in \mathcal{B}_{\mathbb{R}} : \qquad P_{X_i}(B_i) = P(X_i^{-1}(B_i)) : \mathcal{B}_{\mathbb{R}} \to [0, 1], \qquad 1 \le i \le n,$$

is called the *marginal distribution* of the random variable X_i. The probability $P_{\mathbf{X}}$ associated with the random vector $\mathbf{X}$ is called the *joint probability* of the family of random variables $(X_i)_{1 \le i \le n}$.

Remark 1.48. If $\mathbf{X} : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ is a random vector of dimension n and if $X_i = \pi_i \circ \mathbf{X} : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, $1 \le i \le n$, then, knowing the joint probability law $P_{\mathbf{X}}$, it is possible to determine the marginal probability P_{X_i}, for all $i \in \{1, \ldots, n\}$. In fact, if we consider the probability law of X_i, $i \in \{1, \ldots, n\}$, as well as the induced probability $\pi_i(P_{\mathbf{X}})$, for all $i \in \{1, \ldots, n\}$, then we have the relation

$$P_{X_i} = \pi_i(P_{\mathbf{X}}), \qquad 1 \le i \le n.$$

Therefore, for every $B_i \in \mathcal{B}_{\mathbb{R}}$, we obtain

$$\begin{aligned} P_{X_i}(B_i) &= P_{\mathbf{X}}(\pi_i^{-1}(B_i)) = P_{\mathbf{X}}(X_1 \in \mathbb{R}, \ldots, X_i \in B_i, \ldots, X_n \in \mathbb{R}) \\ &= P_{\mathbf{X}}(\mathcal{C}_{B_i}), \end{aligned} \tag{1.3}$$

where $\mathcal{C}_{B_i}$ is the cylinder of base B_i in $\mathbb{R}^n$. This can be further extended by considering, instead of the projection π_i, the projections π_S, where $S \subset \{1, \ldots, n\}$. Then, for every measurable set B_S, we obtain

$$P_{X_S}(B_S) = P_{\mathbf{X}}(\pi_S^{-1}(B_S)).$$

Notice that in general the converse is not true; the knowledge of the marginals does not imply the knowledge of the joint distribution of a random vector $\mathbf{X}$, unless further conditions are imposed.

Definition 1.49. Let $\mathbf{X} : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ be a random vector of dimension n. The mapping $F_{\mathbf{X}} : \mathbb{R}^n \to [0, 1]$, with

$$\mathbf{t} = (t_1, \ldots, t_n) : \qquad F_{\mathbf{X}}(\mathbf{t}) := P(X_1 \leq t_1, \ldots, X_n \leq t_n) \qquad \forall \mathbf{t} \in \mathbb{R}^n,$$

is called the *joint cumulative distribution function* of the random vector $\mathbf{X}$.

Remark 1.50. Analogous to the case of random variables, $F_{\mathbf{X}}$ is increasing and right-continuous on $\mathbb{R}^n$. Further, it is such that

$$\lim_{x_i \to +\infty, \forall i} F(x_1, \ldots, x_n) = 1,$$

and for any $i = 1, \ldots, n$:

$$\lim_{x_i \to -\infty} F(x_1, \ldots, x_n) = 0.$$

Conversely, given a distribution function F satisfying all the above properties, there exists an n-dimensional random vector X with F as its cumulative distribution function. The underlying probability space can be constructed in a canonical way. In the bidimensional case, if $F : \mathbb{R}^2 \to [0, 1]$ satisfies the above conditions, then we can define a probability $P : \mathcal{B}_{\mathbb{R}^2} \to [0, 1]$ in the following way:

$$P(]\mathbf{a}, \mathbf{b}]) = F(b_1, b_2) - F(b_1, a_2) + F(a_1, a_2) - F(a_1, b_2),$$

for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$. Hence there exists a bidimensional random vector $\mathbf{X}$ with P as its probability.

Remark 1.51. Let $\mathbf{X} : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ be a random vector of dimension n, $X_i = \pi_i \circ \mathbf{X}$, $1 \leq i \leq n$, the nth component of $\mathbf{X}$, and let F_{X_i}, $1 \leq i \leq n$, and $F_{\mathbf{X}}$ be the respective cumulative distribution functions of X_i and $\mathbf{X}$. The knowledge of $F_{\mathbf{X}}$ allows one to infer F_{X_i}, $1 \leq i \leq n$, through the relation

$$F_{X_i}(t_i) = P(X_i \leq t_i) = F_{\mathbf{X}}(+\infty, \ldots, t_i, \ldots, +\infty),$$

for every $t_i \in \mathbb{R}$.

Definition 1.52. Let $\mathbf{X} : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ be a random vector of dimension n. If the probability law $P_{\mathbf{X}} : \mathcal{B}_{\mathbb{R}^n} \to [0, 1]$ with respect to $\mathbf{X}$ is endowed with a density with respect to the Lebesgue measure μ_n on $\mathbb{R}^n$ (or product measure of Lebesgue measures μ on $\mathbb{R}$), then this density is called the probability density of $\mathbf{X}$. If $f : \mathbb{R}^n \to \mathbb{R}_+$ is the probability density of $\mathbf{X}$, then

$$F_{\mathbf{X}}(\mathbf{t}) = \int_{-\infty}^{\mathbf{t}} f d\mu_n \qquad \forall \mathbf{t} \in \mathbb{R}^n,$$

and moreover

$$P_{\mathbf{X}}(B) = \int_B f(x_1, \ldots, x_n) d\mu_n \qquad \forall B \in \mathcal{B}_{\mathbb{R}}.$$

Proposition 1.53. *Under the assumptions of the preceding definition, defining* $X_i = \pi_i \circ \mathbf{X}$, $1 \le i \le n$, *then* P_{X_i} *is endowed with density with respect to Lebesgue measure* μ *on* $\mathbb{R}$ *and its density function* $f_i : \mathbb{R} \to \mathbb{R}_+$ *is given by*

$$f_i(x_i) = \int^i f(x_1, \ldots, x_n) d\mu_{n-1},$$

where we have denoted by $\int^i$ *the integration with respect to all variables but the ith one.*

Proof: By (1.3) we have that for all $B_i \in \mathcal{B}_{\mathbb{R}}$:

$$\begin{aligned} P_{X_i}(B_i) = P_{\mathbf{X}}(\mathcal{C}_{B_i}) &= \int_{\mathcal{C}_{B_i}} f(x_1, \ldots, x_n) d\mu_n \\ &= \int_{\mathbb{R}} dx_1 \cdots \int_{B_i} dx_i \cdots \int_{\mathbb{R}} f(x_1, \ldots, x_n) dx_n \\ &= \int_{B_i} dx_i \int^i f(x_1, \ldots, x_n) d\mu_{n-1}. \end{aligned}$$

By putting $f_i(x_i) = \int^i f(x_1, \ldots, x_n) d\mu_{n-1}$, we see that f_i is the density of P_{X_i}. □

Remark 1.54. The definition of a discrete random vector is analogous to Definition 1.39.

1.3 Expectations

Definition 1.55. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ a real-valued random variable. Assume that X is P-integrable, i.e., $X \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$; then

$$E[X] = \int_\Omega X(\omega) dP(\omega)$$

is the *expected value* or *expectation* of the random variable X.

Remark 1.56. By Proposition A.28 it follows that if X is integrable with respect to P, its expected value is given by

$$E(X) = \int_{\mathbb{R}} I_{\mathbb{R}}(x) dP_X(x) := \int x dP_X.$$

Remark 1.57. If X is a continuous real-valued random variable with density function f of P_X, then

$$E[X] = \int x f(x) d\mu.$$

Instead, if f is discrete with probability function p, then

$$E[X] = \sum x p(x).$$

Definition 1.58. A real-valued P-integrable random variable X is *centered*, if it has expectation zero.

Proposition 1.59. *Let $(X_i)_{1 \leq i \leq n}$ be a real, P-integrable family of random variables on the same space $(\Omega, \mathcal{F}, P)$. Then*

$$E[X_1 + \cdots + X_n] = \sum_{i=1}^{n} E[X_i].$$

Moreover, for every $\alpha \in \mathbb{R}$, $E[\alpha X] = \alpha E[X]$, and thus it is linear.

Remark 1.60. If X is a real, P-integrable random variable, then $X - E[X]$ is a centered random variable. This follows directly from the linearity of expectations.

Definition 1.61. Given a real P-integrable random variable X, if $E[(X - E[X])^n] < +\infty$, $n \in \mathbb{N}$, then it is the nth centered moment. The second centered moment is the *variance*, and its square root, the *standard deviation* of a random variable X, denoted by $Var[X]$ and $\sigma = \sqrt{Var[X]}$ respectively.

Proposition 1.62. *Let $(\Omega, \mathcal{F})$ be a probability space and $X : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ a random variable. Then the following two statements are equivalent:*

1. *X is square-integrable with respect to P (see Definition A.61).*
2. *X is P-integrable and $Var[X] < +\infty$.*

Moreover, under these conditions

$$Var[X] = E[X^2] - (E[X])^2. \tag{1.4}$$

Proof: 1⇒2: Because $\mathcal{L}^2(P) \subset \mathcal{L}^1(P)$, $X \in \mathcal{L}^1(P)$. Obviously, the constant $E[X]$ is P-integrable; thus $X - E[X] \in \mathcal{L}^2(P)$ and $Var[X] < +\infty$.

2⇒1: By assumption, $E[X]$ exists and $X - E[X] \in \mathcal{L}^2(P)$; thus $X = X - E[X] + E[X] \in \mathcal{L}^2(P)$. Finally, due to the linearity of expectations,

$$\begin{aligned} Var[X] &= E[(X - E[X])^2] = E[X^2 - 2XE[X] + (E[X])^2] \\ &= E[X^2] - 2(E[X])^2 = E[X^2] - (E[X])^2. \end{aligned}$$

□

Proposition 1.63. *If X is a real-valued P-integrable random variable and $Var[X] = 0$, then $X = E[X]$ almost surely with respect to the measure P.*

Proof: $Var[X] = 0 \Rightarrow \int (X - E[X])^2 dP = 0$. With $(X - E[X])^2$ nonnegative, $X - E[X] = 0$ almost everywhere with respect to P, thus $X = E[X]$ almost surely with respect to P. This is equivalent to

$$P(X \neq E[X]) = P(\{\omega \in \Omega | X(\omega) \neq E[X]\}) = 0.$$

□

Proposition 1.64. (Markov's inequality). *Let X be a nonnegative real P-integrable random variable on a probability space $(\Omega, \mathcal{F}, P)$; then*

$$P(X \geq \lambda E[X]) \leq \frac{1}{\lambda} \qquad \forall \lambda \in \mathbb{R}_+.$$

Proof: If $\lambda \leq 1$, then the inequality is obvious, since $P(X \geq \lambda E[X]) \leq 1$. If $\lambda > 1$, then putting $m = E[X]$ results in

$$m = \int_0^{+\infty} x dP_X \geq \int_{\lambda m}^{+\infty} x dP_X \geq \lambda m P(X \geq \lambda m),$$

thus $P(X \geq \lambda m) \leq 1/\lambda$. □

Proposition 1.65. (Chebyshev's inequality). *If X is a real-valued and P-integrable random variable with variance $Var[X]$ (possibly infinite), then*

$$P(|X - E[X]| \geq \epsilon) \leq \frac{Var[X]}{\epsilon^2}.$$

Proof: Apply Markov's inequality to the random variable $(X - E[X])^2$.

Example 1.66.

1. If X is a P-integrable continuous random variable with density f, where the latter is symmetric around the axis $x = a$, $a \in \mathbb{R}$, then $E[X] = a$.
2. If X is a Gaussian variable, then $E[X] = m$ and $Var[X] = \sigma^2$.
3. If X is a discrete, Poisson distributed random variable, then $E[X] = \lambda$, $Var[X] = \lambda$.
4. If X is binomially distributed, then $E[X] = np$, $Var[X] = np(1-p)$.
5. If X is continuous and uniform with density $f(x) = I_{[a,b]}(x)\frac{1}{b-a}$, $a, b \in \mathbb{R}$, then $E[X] = \frac{a+b}{2}$, $Var[X] = \frac{(b-a)^2}{12}$.
6. If X is a Cauchy variable, then it does not admit an expected value.

Definition 1.67. Let $\mathbf{X} : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ be a vector of random variables with P-integrable components X_i, $1 \leq i \leq n$. The expected value of the vector $\mathbf{X}$ is

$$E[\mathbf{X}] = (E[X_1], \ldots, E[X_2])'.$$

Definition 1.68. If X_1, X_2 and X_1X_2 are P-integrable random variables, then

$$Cov[X_1, X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])]$$

is the *covariance* of X_1 and X_2.

Remark 1.69. Due to the linearity of the $E[\cdot]$ operator, if $E[X_1X_2] < +\infty$, then

$$\begin{aligned} Cov[X_1, X_2] &= E[(X_1 - E[X_1])(X_2 - E[X_2])] \\ &= E[X_1X_2 - X_1E[X_2] - E[X_1]X_2 + E[X_1]E[X_2]] \\ &= E[X_1X_2] - E[X_1]E[X_2]. \end{aligned}$$

Proposition 1.70. *1. If X is a random variable, square-integrable with respect to P, and $a, b \in \mathbb{R}$, then*

$$Var[aX + b] = a^2Var[X].$$

2. If both X_1 and X_2 are in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, then

$$Var[X_1 + X_2] = Var[X_1] + Var[X_2] + 2Cov[X_1, X_2].$$

Proof: 1. Since $Var[X] = E[X^2] - (E[X])^2$, then

$$\begin{aligned} Var[aX + b] &= E[(aX + b)^2] - (E[aX + b])^2 \\ &= a^2E[X^2] + 2abE[X] + b^2 - a^2(E[X])^2 - b^2 - 2abE[X] \\ &= a^2(E[X^2] - (E[X])^2) = a^2Var[X]. \end{aligned}$$

2.

$$\begin{aligned} & Var[X_1] + Var[X_2] + 2Cov[X_1, X_2] \\ = \; & E[X_1^2] - (E[X_1])^2 + E[X_2^2] - (E[X_2])^2 + 2(E[X_1X_2] - E[X_1]E[X_2]) \\ = \; & E[(X_1 + X_2)^2] - 2E[X_1]E[X_2] - (E[X_1])^2 - (E[X_2])^2 \\ = \; & E[(X_1 + X_2)^2] - (E[X_1 + X_2])^2 = Var[X_1 + X_2]. \end{aligned}$$

□

Definition 1.71. If X_1 and X_2 are square-integrable random variables with respect to P, having the respective standard deviations $\sigma_1 > 0$ and $\sigma_2 > 0$, then

$$\rho(X_1, X_2) = \frac{Cov[X_1, X_2]}{\sigma_1\sigma_2}$$

is the *correlation coefficient* of X_1 and X_2.

Remark 1.72. If X_1 and X_2 are $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ random variables, then, by the Cauchy–Schwarz inequality (1.15),

$$|\rho(X_1, X_2)| \leq 1;$$

moreover,

$$|\rho(X_1, X_2)| = 1 \Leftrightarrow \exists a, b \in \mathbb{R} \text{ so that } X_2 = aX_1 + b, \qquad \text{a.s.}$$

Definition 1.73. Let X be a real-valued random variable and $s \in \mathbb{R}$. Then e^{isX} is a complex-valued random variable and its expectation

$$\begin{aligned} \phi_X(s) &= E\left[e^{isX}\right] = \int_{\mathbb{R}} e^{isx} f_X(x) dx \qquad &\text{if } X \text{ is continuous,} \\ \phi_X(s) &= E\left[e^{isX}\right] = \sum_n e^{isx_n} P_X(x_n) \qquad &\text{if } X \text{ is discrete} \end{aligned}$$

is the *characteristic function* of X. It is continuous for all $s \in \mathbb{R}$ and

$$|\phi_X| \leq \phi_X(0) = 1.$$

Example 1.74. The characteristic function of a standard normal random variable X is

$$\begin{aligned} \phi_X(s) = E\left[e^{isX}\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{isX} e^{-\frac{1}{2}x^2} dx \\ &= e^{-\frac{s^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(X-is)^2} dx = e^{-\frac{s^2}{2}}. \end{aligned}$$

Remark 1.75. The characteristic function of a continuous random variable X represents the Fourier transform of its density function f_X. By invoking the inverse Fourier transform

$$f_X(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \phi_X(s) e^{-isx} ds,$$

we recover the probability density function. The probability density of a random variable has a unique characteristic function. Hence if two random variables have identical characteristic functions, they are identical.

1.4 Independence

Definition 1.76. The random variables $X_1, \ldots, X_n$, defined on the same probability space $(\Omega, \mathcal{F}, P)$, are independent if they generate independent classes of σ-algebras. Hence

$$P(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} P(A_i) \qquad \forall A_i \in X_i^{-1}(\mathcal{B}_{\mathbb{R}}).$$

The following is an equivalent definition.

Definition 1.77. The components X_i, $1 \leq i \leq n$, of an n-dimensional random vector $\mathbf{X}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ are independent if

$$P_{\mathbf{X}} = \bigotimes_{i=1}^{n} P_{X_i},$$

where $P_{\mathbf{X}}$ and P_{X_i} are the probability laws of $\mathbf{X}$ and X_i, $1 \leq i \leq n$, respectively. (See Proposition A.43.)

To show that Definitions 1.77 and 1.76 are equivalent, we need to show that the following equivalence holds:

$$P(A_1 \cap \cdots \cap A_n) = \prod_{i=1}^{n} P(A_i) \Leftrightarrow P_{\mathbf{X}} = \bigotimes_{i=1}^{n} P_{X_i} \qquad \forall A_i \in X_i^{-1}(\mathcal{B}_{\mathbb{R}}).$$

We may recall first that $P_{\mathbf{X}} = \bigotimes_{i=1}^{n} P_{X_i}$ is the unique measure on $\mathcal{B}_{\mathbb{R}^n}$ that factorizes on rectangles; i.e., if $B = \prod_{i=1}^{n} B_i$, with $B_i \in \mathcal{B}_{\mathbb{R}}$, we have

$$P_{\mathbf{X}}(B) = \prod_{i=1}^{n} P_{X_i}(B_i).$$

To prove the implication from left to right, we observe that if B is a rectangle in $\mathcal{B}_{\mathbb{R}^n}$ as defined above, then

$$\begin{aligned} P_{\mathbf{X}}(B) = P(\mathbf{X}^{-1}(B)) &= P\left(\mathbf{X}^{-1}\left(\prod_{i=1}^{n} B_i\right)\right) = P\left(\bigcap_{i=1}^{n} X_i^{-1}(B_i)\right) \\ &= \prod_{i=1}^{n} P(X_i^{-1}(B_i)) = \prod_{i=1}^{n} P_{X_i}(B_i). \end{aligned}$$

Vice versa, for all $i = 1, \ldots, n$:

$$A_i \in X_i^{-1}(\mathcal{B}_{\mathbb{R}}) \Rightarrow \exists B_i \in \mathcal{B}_{\mathbb{R}}, \text{ so that } A_i = X_i^{-1}(B_i).$$

Thus, since $A_1 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} X_i^{-1}(B_i)$, we have

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= P\left(\bigcap_{i=1}^{n} X_i^{-1}(B_i)\right) = P(\mathbf{X}^{-1}(B)) = P_{\mathbf{X}}(B) \\ &= \prod_{i=1}^{n} P_{X_i}(B_i) = \prod_{i=1}^{n} P(X_i^{-1}(B_i)) = \prod_{i=1}^{n} P(A_i). \end{aligned}$$

Proposition 1.78. *1. The real-valued random variables $X_1, \ldots, X_n$ are independent if and only if, for every $\mathbf{t} = (t_1, \ldots, t_n)' \in \mathbb{R}^n$,*

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{t}) := P(X_1 \leq t_1 \cap \cdots \cap X_n \leq t_n) &= P(X_1 \leq t_1) \cdots P(X_n \leq t_n) \\ &= F_{X_1}(t_1) \cdots F_{X_n}(t_n). \end{aligned}$$

2. *Let* $\mathbf{X} = (X_1, \ldots, X_n)'$ *be a real-valued random vector with density* f *and probability* $P_{\mathbf{X}}$ *that is absolutely continuous with respect to the measure* μ_n. *The following two statements are equivalent:*
   - $X_1, \ldots, X_n$ *are independent.*
   - $f = f_{X_1} \cdots f_{X_n}$ *almost surely.*

Remark 1.79. From the previous definition it follows that if a random vector $\mathbf{X}$ has independent components, then their marginal distributions determine the joint distribution of $\mathbf{X}$.

Example 1.80. Let $\mathbf{X}$ be a bidimensional random vector with uniform density $f(\mathbf{x}) = c \in \mathbb{R}$, for all $\mathbf{x} = (x_1, x_2)' \in \mathcal{R}$. If $\mathcal{R}$ is, say, a semicircle, then X_1 and X_2 are not independent. But if $\mathcal{R}$ is a rectangle, then X_1 and X_2 are independent.

Proposition 1.81. *Let* $X_1, \ldots, X_n$ *be independent random variables defined on* $(\Omega, \mathcal{F}, P)$ *and valued in* $(E_1, \mathcal{B}_1), \ldots, (E_n, \mathcal{B}_n)$. *If the mappings*

$$g_i : (E_i, \mathcal{B}_i) \to (F_i, \mathcal{U}_i), \qquad 1 \leq i \leq n,$$

are measurable, then the random variables $g_1(X_1), \ldots, g_n(X_n)$ *are independent.*

Proof: Defining $h_i = g_i(X_i)$, $1 \leq i \leq n$, gives

$$h_i^{-1}(U_i) = X_i^{-1}(g_i^{-1}(U_i)) \in X_i^{-1}(\mathcal{B}_i)$$

for every $U_i \in \mathcal{U}_i$. The assertion then follows from Definition 1.76. □

Proposition 1.82. *If* $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ *is a random variable with probability law* P_X *and* $H : (E, \mathcal{B}) \to (F, \mathcal{U})$ *a measurable function, then, defining* $Y = H \circ X = H(X)$, Y *is a random variable. Furthermore, if* $H : (E, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, *then* $Y \in \mathcal{L}^1(P)$ *is equivalent to* $H \in \mathcal{L}^1(P_X)$ *and*

$$E[Y] = \int H(x) P_X(dx).$$

Corollary 1.83. Let $\mathbf{X} = (X_1, X_2)'$ be a random vector defined on $(\Omega, \mathcal{F}, P)$ whose components are valued in $(E_1, \mathcal{B}_1)$ and $(E_2, \mathcal{B}_2)$, respectively. If $h : (E_1 \times E_2, \mathcal{B}_1 \otimes \mathcal{B}_2) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, then $Y = h(\mathbf{X}) \equiv h \circ \mathbf{X}$ is a real-valued random variable. Moreover,

$$E[Y] = \int h(x_1, x_2) dP_{\mathbf{X}}(x_1, x_2),$$

where $P_{\mathbf{X}}$ is the joint probability of X_1 and X_2.

Proposition 1.84. *If X_1 and X_2 are real-valued independent random variables on $(\Omega, \mathcal{F}, P)$ and endowed with finite expectations, then their product $X_1X_2 \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and*

$$E[X_1X_2] = E[X_1]E[X_2].$$

Proof: Given the assumption of independence of X_1 and X_2, it is a a tedious though trivial exercise to show that $X_1, X_2 \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. For the second part, by Corollary 1.83:

$$\begin{aligned} E[X_1X_2] &= \int X_1X_2 dP_{(X_1X_2)} = \int X_1X_2 d(P_{X_1} \otimes P_{X_2}) \\ &= \int X_1 dP_{X_1} \int X_2 dP_{X_2} = E[X_1]E[X_2]. \end{aligned}$$

□

Remark 1.85. From Definition 1.68 and Remark 1.69 it follows that the covariance of two independent variables is zero.

Proposition 1.86. *If two random variables X_1 and X_2 are independent, then the variance operator $Var[\cdot]$ is additive, but not homogeneous. This follows from Proposition 1.70 and Remark 1.85.*

Sums of Two Random Variables

Let X and Y be two real-valued, independent, continuous random variables on $(\Omega, \mathcal{F}, P)$ with densities f and g, respectively. Defining $Z = X + Y$, then Z is a random variable, and let F_Z be its cumulative distribution. It follows that

$$F_Z(t) = P(Z \leq t) = P(X + Y \leq t) = P_{(X,Y)}(\mathcal{R}_t),$$

where $\mathcal{R}_t = \{(x, y) \in \mathbb{R}^2 | x + y \leq t\}$. By Proposition 1.78 (X, Y) is continuous and its density is $f_{(X,Y)} = f(x)g(y)$, for all $(x, y) \in \mathbb{R}^2$. Therefore, for all $t \in \mathbb{R}$:

$$\begin{aligned} F_Z(t) = P_{(X,Y)}(\mathcal{R}_t) &= \int\int_{\mathcal{R}_t} f(x)g(y)dxdy \\ &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{t-x} f(x)g(y)dy = \int_{-\infty}^{+\infty} f(x)dx \int_{-\infty}^{t} g(z-x)dz \\ &= \int_{-\infty}^{t} dz \int_{-\infty}^{+\infty} f(x)g(z-x)dx \qquad \forall z \in \mathbb{R}. \end{aligned}$$

Hence, the function

$$f_Z(z) = \int_{-\infty}^{+\infty} f(x)g(z-x)dx \tag{1.5}$$

is the density of the random variable Z.

Definition 1.87. The function f_Z defined by (1.5) is the convolution of f and g, denoted by $f * g$. Analogously it can be shown that, if f_1, f_2, f_3 are the densities of the independent random variables X_1, X_2, X_3, then the random variable $Z = X_1 + X_2 + X_3$ has density

$$f_1 * f_2 * f_3(z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_1(x) f_2(y - x) f_3(z - y) dx dy$$

for every $z \in \mathbb{R}$. This extends to n independent random variables in an analogous way.

Remark 1.88. Let $Z = X_1 + X_2$ be the sum of two independent random variables. Then the characteristic function of Z is

$$\phi_Z(s) = E\left[e^{isZ}\right] = E\left[e^{isX_1}\right] E\left[e^{isX_2}\right] = \phi_{X_1}(s)\phi_{X_2}(s).$$

By inverting $\phi_Z(s)$ we can recover f_Z as the convolution of f_{X_1} and f_{X_2}. An easier way, whenever applicable, for identifying the probability law of the sum of independent random variables is based on the uniqueness theorem of characteristic functions associated with probability laws.

Example 1.89.

1. The sum of two independent Gaussian random variables distributed as $N(m_1, \sigma_1^2)$ and $N(m_2, \sigma_2^2)$ is distributed as $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ for any $m_1, m_2 \in \mathbb{R}$ and any $\sigma_1^2, \sigma_2^2 \in \mathbb{R}_+^*$. Note that
$$aN(m_1, \sigma_1^2) + b = N(am_1 + b, a^2\sigma_1^2).$$
2. The sum of two independent Poisson variables, distributed as $P(\lambda_1)$ and $P(\lambda_2)$, is distributed as $P(\lambda_1 + \lambda_2)$ for any $\lambda_1, \lambda_2 \in \mathbb{R}_+^*$.
3. The sum of two independent binomial random variables distributed as $B(r_1, p)$ and $B(r_2, p)$ is distributed as $B(r_1 + r_2, p)$ for any $r_1, r_2 \in \mathbb{N}^*$ and any $p \in [0, 1]$.

The Gaussian, Poisson, and binomial distributions are said to *reproduce* themselves.

Definition 1.90. Consider N independent and identically distributed random variables X_i, $i = 1, \ldots, N$, with common probability law P_1, belonging to a family $\mathcal{G}$ of probability laws. Let P_N be the probability law of $X_N = \sum_{i=1}^N X_i$. We say that the family $\mathcal{G}$ is *stable* if $P_1 \in \mathcal{G}$ implies $P_N \in \mathcal{G}$.

Clearly, Gaussian and Poisson laws are stable (see, e.g., Samorodnitsky and Taqqu (1994)).

The Central Limit Theorem for Independent Random Variables

Theorem 1.91. (Central limit theorem for independent and identically distributed random variables). *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of independent identically distributed random variables in* $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ *with* $m = E[X_i]$, $\sigma^2 = Var[X_i]$, *for all* i, *and*

$$S_n = \frac{\frac{1}{n}\sum_{i=1}^n X_i - m}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^n X_i - nm}{\sigma\sqrt{n}}.$$

Then

$$S_n \xrightarrow[n\to\infty]{d} N(0,1);$$

i.e., if we denote by $F_n = P(S_n \leq x)$ *and*

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy, \qquad x \in \mathbb{R},$$

then $\lim_n F_n = \Phi$, *uniformly in* $\mathbb{R}$, *and thus*

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \xrightarrow[n]{} 0.$$

A generalization of the central limit theorem, that does not require the random variables to be identically distributed is the following.

Theorem 1.92. (Lindeberg). *Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of independent random variables in* $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ *with* $E[X_i] = 0$, *for all* i, *and denote by*

$$s_n^2 = \sum_{i=1}^n Var[X_i] = \sum_{i=1}^n E\left[X_i^2\right].$$

If the Lindeberg condition,

$$\forall \epsilon > 0: \qquad \lim_n \frac{1}{s_n^2} \sum_{i=1}^n \int_{|X_i|\geq \epsilon s_n} X_i^2 dP = 0,$$

is satisfied, then

$$S_n \xrightarrow[n\to\infty]{d} N(0,1).$$

Proof: See, e.g., Shiryaev (1995). □

This can further be generalized for noncentered random variables.

Corollary 1.93. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ with $m_i = E[X_i]$, $\sigma_i^2 = Var[X_i]$, and $s_n^2 = \sum_{k=1}^n \sigma_i^2$. If

$$\forall \epsilon: \qquad \lim_n \frac{1}{s_n^2} \sum_{i=1}^n \int_{|X_i - m_i|\geq \epsilon s_n} |X_i - m_i|^2 dP = 0,$$

then

$$\frac{\sum_{i=1}^n X_i - E[S_n]}{\sqrt{Var S_n}} \xrightarrow[n\to\infty]{d} N(0,1).$$

Theorem 1.94. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with $m = E[X_i]$, $\sigma^2 = Var[X_i]$, for all i, and let $(\mathcal{V}_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{N}$-valued random variables such that*

$$\frac{\mathcal{V}_n}{n} \xrightarrow[n]{P} 1.$$

Then

$$\frac{1}{\sqrt{\mathcal{V}_n}} \sum_{i=1}^{n} X_i \xrightarrow[n]{P} N\left(m, \sigma^2\right).$$

Proof: See, e.g., Chung (1974). □

Tail Events

Definition 1.95. Let $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ be a sequence of events and let

$$\sigma(A_n, A_{n+1}, \ldots), \qquad n \in \mathbb{N},$$

as well as

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(A_n, A_{n+1}, \ldots)$$

be σ-algebras. Then $\mathcal{T}$ is the *tail σ-algebra* associated with the sequence $(A_n)_{n\in\mathbb{N}}$ and its elements are called *tail events.*

Example 1.96. The *essential supremum*

$$\limsup_n A_n = \bigcap_{n=1}^{\infty} \bigcup_{i=n}^{\infty} A_i$$

and *essential infimum*

$$\liminf_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{i=n}^{\infty} A_i$$

are both tail events for the sequence $(A_n)_{n\in\mathbb{N}}$. If n is understood to be *time*, then we can write

$$\limsup A_n = \{A_n \text{ i.o.}\};$$

i.e., A_n occurs infinitely often (i.o.). Thus, for infinitely many $n \in \mathbb{N}$, while

$$\liminf A_n = \{A_n \text{ a.a.}\},$$

i.e., A_n occurs almost always (a.a.), thus for all but finitely many $n \in \mathbb{N}$.

Theorem 1.97. (Kolmogorov's zero-one law). *Let $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ be a sequence of independent events. Then for any $A \in \mathcal{T}$: $P(A) = 0$ or $P(A) = 1$.*

Lemma 1.98. (Borel–Cantelli).

1. *Let* $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ *be a sequence of events. If* $\sum_n P(A_n) < +\infty$, *then*
$$P\left(\limsup_n A_n\right) = 0.$$
2. *Let* $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ *be a sequence of independent events. If* $\sum_n P(A_n) = +\infty$, *then*
$$P\left(\limsup_n A_n\right) = 1.$$

Proof: See, e.g., Billingsley (1968). □

1.5 Conditional Expectations

Let $X, Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ be two discrete random variables with joint discrete probability distribution p. There exists an, at most countable, subset $D \subset \mathbb{R}^2$, such that
$$p(x,y) \neq 0 \qquad \forall (x,y) \in D,$$
where $p(x,y) = P(X = x \cap Y = y)$. If, furthermore, D_1 and D_2 are the projections of D along its axes, then the marginal distributions of X and Y are given by
$$p_1(x) = P(X = x) = \sum_y p(x,y) \neq 0 \qquad \forall x \in D_1,$$
$$p_2(y) = P(Y = y) = \sum_x p(x,y) \neq 0 \qquad \forall y \in D_2.$$

Definition 1.99. Given the preceding assumptions and fixing $y \in \mathbb{R}$, then the probability of y *conditional* on $X = x \in D_1$ is
$$p_2(y|x) = \frac{p(x,y)}{p_1(x)} = \frac{P(X = x \cap Y = y)}{P(X = x)} = P(Y = y|X = x).$$
Furthermore,
$$y \to p_2(y|X = x) \in [0,1] \qquad \forall x \in D_1$$
is called the probability function of y conditional on $X = x$.

Definition 1.100. Analogous to the definition of expectation of a discrete random variable, the expectation of Y, conditional on $X = x$, is
$$\begin{aligned} E[Y|X = x] &= \sum_y y p_2(y|x) \qquad \forall x \in D_1, \\ &= \frac{1}{p_1(x)} \sum_y y p(x,y) = \frac{1}{p_1(x)} \sum_{y\in\mathbb{R}} y p(x,y) \\ &= \frac{1}{p_1(x)} \int\int_{\mathcal{R}_x} y dP_{(X,Y)}(x,y) \\ &= \frac{1}{P([X = x])} \int_{[X=x]} Y(\omega) dP(\omega), \end{aligned}$$

with $\mathcal{R}_x = \{x\} \times \mathbb{R}$.

Definition 1.101. Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ be a discrete random variable and $Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ P-integrable. Then the mapping

$$x \to E[Y|X = x] = \frac{1}{P([X = x])} \int_{[X=x]} Y(\omega) dP(\omega) \tag{1.6}$$

is the expected value of Y conditional on X, defined on the set $x \in E$ with $P_X(x) \neq 0$.

Remark 1.102. It is standard to extend the mapping (1.6) to the entire set E by fixing its value arbitrarily at the points $x \in E$ where $P([X = x]) = 0$. Hence there exists an entire equivalence class of functions f defined on E, such that

$$f(x) = E[Y|X = x] \qquad \forall x \in E \text{ such that } P_X(x) \neq 0.$$

An element f of this class is said to be defined on E, almost surely with respect to P_X. A generic element of this class is denoted by either $E[Y|X = \cdot]$, $E[Y|\cdot]$, or $E^X[Y]$. Furthermore, its value at $x \in E$ is denoted by either $E[Y|X = x]$, $E[Y|x]$, or $E^{X=x}[Y]$.

Definition 1.103. Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ be a discrete random variable and $x \in E$ so that $P_X(x) \neq 0$, and let $F \in \mathcal{F}$. The indicator of F, denoted by $I_F : \Omega \to \mathbb{R}$, is a real-valued, P-integrable random variable. The expression

$$P(F|X = x) = E[I_F|X = x] = \frac{P(F \cap [X = x])}{P(X = x)}$$

is the probability of F conditional upon $X = x$.

Remark 1.104. Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ be a discrete random variable. If we define $E_X = \{x \in E | P_X(x) \neq 0\}$, then for every $x \in E_X$ the mapping

$$P(\cdot|X = x) : \mathcal{F} \to [0, 1],$$

so that

$$P(F|X = x) = \frac{P(F \cap [X = x])}{P(X = x)} \qquad \forall F \in \mathcal{F}$$

is a probability measure on $\mathcal{F}$, conditional on $X = x$. Further, if we arbitrarily fix the value of $P(F|X = x)$ at the points $x \in E$ where P_X is zero, then we can extend the mapping

$$x \in E_X \to P(F|X = x)$$

to the whole of E, so that $P(\cdot|X = x) : \mathcal{F} \to [0, 1]$ is again a probability measure on $\mathcal{F}$, defined almost surely with respect to P_X.

Definition 1.105. The family of functions $(P(\cdot|X = x))_{x\in E}$ is called a *regular version of the conditional probability with respect to* X.

Proposition 1.106. *Let* $(P(\cdot|X = x))_{x\in E}$ *be a regular version of the conditional probability with respect to* X*. Then, for any* $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$*:*

$$\int Y(\omega)dP(\omega|X = x) = E[Y|X = x], \qquad P_X\text{-}a.s.$$

Proof: First, we observe that Y, being a random variable, is measurable.[3] Now from (1.6) it follows that

$$E[I_F|X = x] = P(F|X = x) = \int I_F(\omega)P(d\omega|X = x)$$

for every $x \in E$, $P_X(x) \neq 0$. Now let Y be an elementary function so that

$$Y = \sum_{i=1}^{n} \lambda_i I_{F_i}.$$

Then, for every $x \in E_X$:

$$\begin{aligned} E[Y|X = x] &= \sum_{i=1}^{n} \lambda_i E[I_{F_i}|X = x] = \sum_{i=1}^{n} \lambda_i \int I_{F_i}(\omega)P(d\omega|X = x) \\ &= \int \left(\sum_{i=1}^{n} \lambda_i I_{F_i}\right)(\omega)P(d\omega|X = x) = \int Y(\omega)dP(\omega|X = x). \end{aligned}$$

If Y is a positive real-valued random variable, then, by Theorem A.14, there exists an increasing sequence $(Y_n)_{n\in\mathbb{N}}$ of elementary random variables so that

$$Y = \lim_{n\to\infty} Y_n = \sup_{n\in\mathbb{N}} Y_n.$$

Therefore, for every $x \in E$:

$$\begin{aligned} E[Y|X = x] &= \sup_{n\in\mathbb{N}} E[Y_n|X = x] = \sup_{n\in\mathbb{N}} \int Y_n(\omega)dP(\omega|X = x) \\ &= \int \left(\sup_{n\in\mathbb{N}} Y_n\right)(\omega)dP(\omega|X = x) = \int Y(\omega)dP(\omega|X = x), \end{aligned}$$

where the first and third equalities are due to the property of Beppo–Levi (see Proposition A.28). Lastly, if Y is a real-valued, P-integrable random variable, then it satisfies the assumptions, being the difference between two positive integrable functions. □

[3] This only specifies its σ-algebras but not its measure.

Proposition 1.107. *Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ be a discrete random variable and $Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ a P-integrable random variable. Then, for every $B \in \mathcal{B}$ we have that*

$$\int_{[X\in B]} Y(\omega)dP(\omega) = \int_B E[Y|X = x]dP_X(x).$$

Proof: Since X is a discrete random variable and $[X \in B] = \bigcup_{x\in B}[X = x]$, where the elements of the collection $([X = x])_{x\in B}$ are mutually exclusive, we observe that by the additivity of the integral:

$$\begin{aligned}
&\int_{[X\in B]} Y(\omega)dP(\omega) \\
&= \sum_{x\in B}\int_{[X=x]} Y(\omega)dP(\omega) = \sum_{x^*\in B}\int_{[X=x^*]} Y(\omega)dP(\omega) \\
&= \sum_{x^*} E[Y|X = x^*]P(X = x^*) = \sum_{x^*\in B} E[Y|X = x^*]P_X(x^*) \\
&= \sum_{x\in B} E[Y|X = x]P_X(x) = \int_B E[Y|X = x]dP_X(x),
\end{aligned}$$

where the $x^* \in B$ are such that $P_X(x^*) \neq 0$. □

We may generalize the above proposition as follows.

Proposition 1.108. *Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ be a random variable and $Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ a P-integrable random variable. Then there exists a unique class of real-valued $f \in L^1(E, \mathcal{B}, P_X)$, such that*

$$\int_{[X\in B]} Y(\omega)dP(\omega) = \int_B f dP_X \qquad \forall B \in \mathcal{B}.$$

Proof: First we consider Y positive. The mapping $v : \mathcal{B} \to \mathbb{R}_+$ given by

$$v(B) = \int_{[X\in B]} Y(\omega)dP(\omega) \qquad \forall B \in \mathcal{B}$$

is a bounded measure and absolutely continuous with respect to P_X. In fact,

$$P_X(B) = 0 \Leftrightarrow P([X \in B]) = 0 \Rightarrow \int_{[X\in B]} Y(\omega)dP(\omega) = 0 \Leftrightarrow v(B) = 0.$$

Because P_X is bounded, thus σ-finite, then, by the Radon–Nikodym Theorem A.53, there exists a unique $f \in L^1(E, \mathcal{B}, P_X)$ such that

$$v(B) = \int_B f dP_X \qquad \forall B \in \mathcal{B}.$$

The case Y of arbitrary sign can be easily handled by the standard decomposition $Y = Y^+ - Y^-$. □

Definition 1.109. Under the assumptions of the preceding proposition every $f \in L^1(E, \mathcal{B}, P_X)$ such that

$$\int_B f dP_X = \int_{[X \in B]} Y(\omega) dP(\omega) \qquad \forall B \in \mathcal{B}$$

is the expected value of Y conditional on X. We will again resort to the notation of Remark 1.102; any of these f will be denoted by $E[Y|X = \cdot]$. Note that $E[Y|X = \cdot]$ is only defined almost surely with respect to P_X.

Proposition 1.110. *If $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ is a random variable and $f : (E, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ a P-integrable function, then, defining $Y = f \circ X = f(X)$, Y is a P-integrable random variable and*

$$E[Y|X = x] = f(x) \qquad \forall x \in E, P_X(x) \neq 0.$$

Proof: By the definition of a composite function:

$$\int_B f(x) dP_X = \int_{[X \in B]} f \circ X(\omega) dP(\omega) = \int_{[X \in B]} Y(\omega) dP(\omega) \qquad \forall x \in E_X,$$

for every $B \in \mathcal{B}$. By uniqueness in $L^1(E, \mathcal{B}, P_X)$ of the expected value of Y conditional on $X = x$, it follows that

$$E[Y|X = x] = f(x), \qquad P_X\text{-a.s.}$$

□

Proposition 1.111. *If $Y : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ is a positive random variable, then*

$$E[Y|X = x] \geq 0, \qquad P_X - a.s.$$

Proof: Since

$$\int_{[X \in B]} Y dP \geq 0 \qquad \forall B \in \mathcal{B},$$

it follows that

$$\int_B E[Y|X = x] P_X(dx) \geq 0 \qquad \forall B \in \mathcal{B},$$

and therefore $E[Y|X = x] \geq 0$, almost surely with respect to P_X. □

Definition 1.112. Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ be a random variable and Y a real-valued, P-integrable random variable also defined on $(\Omega, \mathcal{F})$. We denote by

$$E[Y|X] = E[Y|X = \cdot] \circ X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$$

such that

$$E[Y|X](\omega) = E[Y|X = X(\omega)] \qquad \forall \omega \in \Omega.$$

Theorem 1.113. *Let* $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ *be a random variable and* Y *a real-valued,* P*-integrable random variable. Then the mapping* $E[Y|X]$ *is a real-valued random variable and*

$$E[E[Y|X]] = E[Y].$$

Proof:

$$E[Y] = \int Y(\omega)dP(\omega) = \int_{[X\in E]} Y(\omega)dP(\omega) = \int_E E[Y|X=x]dP_X(x)$$
$$= E[E[Y|X]].$$

□

Theorem 1.114. (Monotone convergence for conditional expectations). *If* $(Y_n)_{n\in\mathbb{N}}$ *is an increasing sequence of real-valued random variables on* $(\Omega, \mathcal{F}, P)$*, converging almost surely to* $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$*, then the sequence* $E[Y_n|X=x]$ *converges to* $E[Y|X=x]$*,almost surely with respect to* P_X*.*

Proof: If $Y \geq 0$ is a real-valued random variable on $(\Omega, \mathcal{F}, P)$, then

$$E[Y|X=x] \geq 0, \qquad P_X\text{-a.s.},$$

from which it follows that, by monotonicity,

$$\forall n \in \mathbb{N}: \qquad E[Y_{n+1}|X=x] \geq E[Y_n|X=x], \qquad P_X\text{-a.s.}$$

Moreover,

$$\forall B \in \mathcal{B}: \qquad \int_{[X\in B]} Y_n dP = \int_B E[Y_n|X=x]dP_X(x)$$

and

$$\forall B \in \mathcal{B}: \qquad \int_{[X\in B]} Y dP \geq \int_{[X\in B]} Y_n dP.$$

Thus

$$E[Y|X=x] \geq E[Y_n|X=x], \qquad P_X\text{-a.s.}$$

The monotone sequence $(E[Y_n|X=x])_{n\in\mathbb{N}}$ is bounded from above and converges almost surely with respect to P_X. By applying Lebesgue's dominated convergence theorem to the sequences $(E[Y_n|X=x])_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$, we obtain

$$\int_{[X\in B]} Y dP = \lim_n \int_{[X\in B]} Y_n dP = \lim_n \int_B E[Y_n|X=x]dP_X(x)$$
$$= \int_B \lim_n E[Y_n|X=x]dP_X(x), \qquad \text{a.s.}$$

With $\{\lim_n E[Y_n|X = x], x \in E\}$ (defined almost surely with respect to P_X,) being $\mathcal{B}$-measurable, the equality proves that

$$E[Y|X = x] = \lim_n E[Y_n|X = x], \qquad \text{a.s.}$$

□

Theorem 1.115. (dominated convergence for conditional expectations). *If $(Y_n)_{n\in\mathbb{N}}$ is a sequence of real-valued random variables, Z a real random variable belonging to $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, with $E[Z] < \infty$ and*

$$|Y_n| \leq Z \qquad \text{for all } n \in \mathbb{N},$$

then from $Y_n \to Y$, almost surely with respect to P_X, it follows that

$$E[Y_n|X = x] \to E[Y|X = x], \qquad P_X\text{-a.s.}$$

A notable extension of previous results and definitions is the subject of the following presentation.

Expectations Conditional on a σ-Algebra

Again due to the Radon–Nykodim theorem, the following proposition holds.

Proposition 1.116. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{F}'$ a σ-algebra contained in $\mathcal{F}$. For every real-valued random variable $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$, there exists a unique element $g \in L^1(\Omega, \mathcal{F}', P)$ such that $\forall B' \in \mathcal{F}'$:*

$$\int_{B'} Y dP = \int_{B'} g dP.$$

We will call this element the conditional expectation of Y given $\mathcal{F}'$ and will denote it by $E[Y|\mathcal{F}']$ or by $E^{\mathcal{F}'}[Y]$.

Remark 1.117. It is not difficult to identify

$$E[Y|X] = E[Y|\mathcal{F}_X]$$

if $\mathcal{F}_X$ is the σ-algebra generated by X.

Proposition 1.118. (tower laws). *Let $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$. For any two subalgebras $\mathcal{G}$ and $\mathcal{B}$ of $\mathcal{F}$ such that $\mathcal{G} \subset \mathcal{B} \subset \mathcal{F}$, we have*

$$E[E[Y|\mathcal{B}]|\mathcal{G}] = E[Y|\mathcal{G}] = E[E[Y|\mathcal{G}]|\mathcal{B}].$$

Proof: For the first equality, by definition, we have

$$\int_G E[Y|\mathcal{G}]dP = \int_G Y dP = \int_G E[Y|\mathcal{B}]dP = \int_G E[E[Y|\mathcal{B}]|\mathcal{G}]dP$$

for all $G \in \mathcal{G} \subset \mathcal{B}$, where comparing the first and last terms completes the proof. The second equality is proven along the same lines. □

Proposition 1.119. *Let $\mathcal{B}$ be a sub-σ-algebra of $\mathcal{F}$. If Y is a real $\mathcal{B}$-measurable random variable and both Z and YZ are two real-valued random variables in $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, then*

$$E^{\mathcal{B}}[YZ] = YE^{\mathcal{B}}[Z].$$

In particular,

$$E^{\mathcal{B}}[Y] = Y.$$

Proof: See, e.g., Métivier (1968). □

Definition 1.120. Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. We say that a real random variable Y on $(\Omega, \mathcal{F}, P)$ is independent of $\mathcal{G}$ with respect to the probability measure P if

$$\forall B \in \mathcal{B}_{\mathbb{R}},\ \forall G \in \mathcal{G}:\ P(G \cap Y^{-1}(B)) = P(G)P(Y^{-1}(B)).$$

Proposition 1.121. *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$; if $Y \in L^1(\Omega, \mathcal{F}, P)$ is independent of $\mathcal{G}$, then*

$$E[Y|\mathcal{G}] = E[Y],\ a.s.$$

Proof: Let $G \in \mathcal{G}$; then, by independence,

$$\int_G Y dP = \int I_G Y dP = E[I_G Y] = E[I_G]E[Y] = P(G)E[Y] = \int_G E[Y] dP,$$

from which the proposition follows. □

Remark 1.122. Both the monotone and dominated convergence theorems extend in an analogous way to expectations conditional on σ-algebras.

Definition 1.123. A family of random variables $(Y_n)_{n\in\mathbb{N}}$ is *uniformly integrable* if

$$\lim_{m\to\infty} \sup_n \int_{|Y_n|\geq m} |Y_n| dP = 0.$$

Proposition 1.124. *Let $(Y_n)_{n\in\mathbb{N}}$ be a family of random variables in $\mathcal{L}^1$. Then the following two statements are equivalent:*

1. *$(Y_n)_{n\in\mathbb{N}}$ is uniformly integrable,*
2. *$\sup_{n\in\mathbb{N}} E[|Y_n|] < +\infty$ and for all ϵ, there exists $\delta > 0$ such that $A \in \mathcal{F}$, $P(A) \leq \delta \Rightarrow E[|Y_n I_A|] < \epsilon$.*

Proposition 1.125. *Let $(Y_n)_{n\in\mathbb{N}}$ be a family of random variables dominated by a nonnegative $X \in \mathcal{L}^1$ on the same probability space $(\Omega, \mathcal{F}, P)$, so that $|Y_n(\omega)| \leq X(\omega)$ for all $n \in \mathbb{N}$. Then $(Y_n)_{n\in\mathbb{N}}$ is uniformly integrable.*

Theorem 1.126. *Let $Y \in \mathcal{L}^1$ be a random variable on $(\Omega, \mathcal{F}, P)$. Then the class $(E[Y|\mathcal{G}])_{\mathcal{G}\subset\mathcal{F}}$, where $\mathcal{G}$ are sub-σ-algebras, is uniformly integrable.*

Proof: See, e.g., Williams (1991). □

Theorem 1.127. *Let* $(Y_n)_{n\in\mathbb{N}}$ *be a sequence of random variables in* $\mathcal{L}^1$ *and let* $Y \in \mathcal{L}^1$. *Then* $Y_n \xrightarrow{\mathcal{L}^1} Y$ *if and only if*

1. $Y_n \xrightarrow[n]{P} Y$,
2. $(Y_n)_{n\in\mathbb{N}}$ *is uniformly integrable.*

Proposition 1.128. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\mathcal{F}'$ *a sub-* σ*-algebra of* $\mathcal{F}$. *Furthermore, let* Y *and* $(Y_n)_{n\in\mathbb{N}}$ *be real-valued random variables, all belonging to* $\mathcal{L}^1(\Omega, \mathcal{F}, P)$. *Under these assumptions the following properties hold:*

1. $E[E[Y|\mathcal{F}']] = E[Y]$;
2. $E[\alpha Y + \beta|\mathcal{F}'] = \alpha E[Y|\mathcal{F}'] + \beta$ *almost surely* $(\alpha, \beta \in \mathbb{R})$;
3. *if* $Y_n \uparrow Y$, *then* $E[Y_n|\mathcal{F}'] \uparrow E[Y|\mathcal{F}']$ *almost surely;*
4. *if* Y *is* $\mathcal{F}'$*-measurable, then* $E[Y|\mathcal{F}'] = Y$ *almost surely;*
5. *if* $\phi : \mathbb{R} \to \mathbb{R}$ *is convex and* $\phi(Y)$ P*-integrable, then* $\phi(E[Y|\mathcal{F}']) \leq E[\phi(Y)|\mathcal{F}']$ *almost surely* (Jensen's inequality).

Proof:

1. This property follows from Proposition 1.116 with $B' = \Omega$.
2. This is obvious from the linearity of the integral.
3. This point is a consequence of the Beppo–Levi property (see Proposition A.28).
4. This property follows from the fact that for all $B' \in \mathcal{F}' : \int_{B'} Y dP = \int_{B'} Y dP$, with Y $\mathcal{F}'$-measurable and P-integrable.
5. Here we use the fact that every convex function ϕ is of type $\phi(x) = \sup_n(a_n x + b_n)$. Therefore, defining $l_n(x) = a_n x + b_n$, for all n, we have that
$$l_n(E[Y|\mathcal{F}']) = E[l_n(Y)|\mathcal{F}'] \leq E[\phi(Y)|\mathcal{F}']$$
and thus
$$\phi(E[Y|\mathcal{F}']) = \sup_n l_n(E[Y|\mathcal{F}']) \leq E[\phi(Y)|\mathcal{F}'].$$

□

Proposition 1.129. *If* $Y \in L^p(\Omega, \mathcal{F}, P)$, *then* $E[Y|\mathcal{F}']$ *is an element of* $L^p(\Omega, \mathcal{F}', P)$ *and*

$$\|E[Y|\mathcal{F}']\|_p \leq \|Y\|_p \qquad (1 \leq p < \infty). \tag{1.7}$$

Proof: With $\phi(x) = |x|^p$ being convex, we have that $|E[Y|\mathcal{F}']|^p \leq E[|Y|^p|\mathcal{F}']$ and thus $E[Y|\mathcal{F}'] \in L^p(\Omega, \mathcal{F}, P)$, and after integration we obtain (1.7). □

Proposition 1.130. *The conditional expectation* $E[Y|\mathcal{F}']$ *is the unique* $\mathcal{F}'$*-measurable random variable* Z *such that for every* $\mathcal{F}'$*-measurable* $X : \Omega \to \mathbb{R}$, *for which the products* XY *and* XZ *are* P*-integrable, we have*

$$E[XY] = E[XZ]. \tag{1.8}$$

Proof: From the fact that $E[E[XY|\mathcal{F}']] = XY$ (point 1 of Proposition 1.128) and because X is $\mathcal{F}'$-measurable, it follows from Proposition 1.119 that $E[E[XY|\mathcal{F}']] = E[XE[Y|\mathcal{F}']]$. On the other hand, if Z is an $\mathcal{F}'$-measurable random variable, so that for every $\mathcal{F}'$-measurable X, with $XY \in L^1(\Omega, \mathcal{F}, P)$ and $XZ \in L^1(\Omega, \mathcal{F}, P)$, it follows that $E[XY] = E[XZ]$. Taking $X = I_B$, $B \in \mathcal{F}'$ we obtain

$$\int_B Y dP = E[YI_B] = E[ZI_B] = \int_B Z dP$$

and hence, by uniqueness of $E[Y|\mathcal{F}']$, that $Z = E[Y|\mathcal{F}']$ almost surely. □

Theorem 1.131. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\mathcal{F}'$ a sub-σ-algebra of $\mathcal{F}$, and Y be a real-valued random variable on $(\Omega, \mathcal{F}, P)$. If $Y \in L^2(P)$, then $E[Y|\mathcal{F}']$ is the orthogonal projection of Y on $L^2(\Omega, \mathcal{F}', P)$, a closed subspace of the Hilbert space $L^2(\Omega, \mathcal{F}, P)$.*

Proof: By Proposition 1.129, from $Y \in L^2(\Omega, \mathcal{F}, P)$ it follows that

$$E[Y|\mathcal{F}'] \in L^2(\Omega, \mathcal{F}', P)$$

and, by equality (1.8), for all random variables $X \in L^2(\Omega, \mathcal{F}', P)$, it holds that

$$E[XY] = E[XE[Y|\mathcal{F}']],$$

completing the proof, by remembering that $(X, Y) \to E[XY]$ is the scalar product in L^2. □

Remark 1.132. We may interpret the theorem above by stating that $E[Y|\mathcal{F}']$ is the best mean square approximation in $L^2(\Omega, \mathcal{F}', P)$ of $Y \in L^2(\Omega, \mathcal{F}, P)$.

1.6 Conditional and Joint Distributions

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : (\Omega, \mathcal{F}, P) \to (E, \mathcal{B})$ a random variable, and $F \in \mathcal{F}$. Following previous results, a unique element $E[I_F|X = x] \in L^1(E, \mathcal{B}, P_X)$ exists such that for any $B \in \mathcal{B}$

$$P(F \cap [X \in B]) = \int_{[X \in B]} I_F(\omega) dP(\omega) = \int_B E[I_F|X = x] dP_X(x). \quad (1.9)$$

We can write

$$P(F|X = \cdot) = E[I_F|X = \cdot].$$

Remark 1.133. By (1.9) the following properties hold:

1. For all $F \in \mathcal{F} : P(F|X = x) \geq 0$, almost surely with respect to P_X.
2. $P(\emptyset|X = x) = 0$, almost surely with respect to P_X.
3. $P(\Omega|X = x) = 1$, almost surely with respect to P_X.

4. For all $F \in \mathcal{F} : 0 \leq P(F|X = x) \leq 1$, almost surely with respect to P_X.
5. For all $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ collections of mutually exclusive sets:

$$P\left(\bigcup_{n\in\mathbb{N}} A_n|X = x\right) = \sum_{n\in\mathbb{N}} P(A_n|X = x), \qquad P_X\text{-a.s.}$$

If, for a fixed $x \in E$, points 3, 4, and 5 hold simultaneously, then $P(\cdot|X = x)$ is a probability, but in general they do not. For example, it is not in general the case that the set of points $x \in E$, $P_X(x) \neq 0$, for which 4 is satisfied, depends upon $F \in \mathcal{F}$. Even if the set of points for which 4 does not hold has zero measure, their union over $F \in \mathcal{F}$ will not necessarily have measure zero. This is also true for subsets $\mathcal{F}' \subset \mathcal{F}$. Hence, in general, given $x \in E$, $P(\cdot|X = x)$ is not a probability on $\mathcal{F}$, unless $\mathcal{F}$ is a countable family, or countably generated. If it happens that, apart from a set E_0 of P_X-measure zero, $P(\cdot|X = x)$ is a probability, then the collection $(P(\cdot|X = x))_{x\in E-E_0}$ is called a regular version of the conditional probability with respect to X on $\mathcal{F}$.

Definition 1.134. Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ and $Y : (\Omega, \mathcal{F}, P) \to (E_1, \mathcal{B}_1)$ be two random variables. We denote by $\mathcal{F}_Y$ the σ-algebra generated by Y, hence

$$\mathcal{F}_Y = Y^{-1}(\mathcal{B}_1) = \{[Y \in B]|B \in \mathcal{B}_1\}.$$

If there exists a regular version $(P(\cdot|X = x))_{x\in E}$ of the probability conditional on X on the σ-algebra $\mathcal{F}_Y$, denoting by $P_Y(\cdot|X = x)$ the mapping defined on $\mathcal{B}_1$, then

$$P_Y(B|X = x) = P([Y \in B]|X = x) \qquad \forall B \in \mathcal{B}_1, x \in E.$$

This mapping is a probability, called the distribution of Y conditional on X, with $X = x$. $P_Y(\cdot|X = x)$ is also termed the induced measure on Y.

Remark 1.135. From the properties of the induced measure it follows that

$$E[Y|X = x] = \int Y(\omega)dP(\omega|X = x) = \int Y dP_Y(Y|X = x).$$

Existence of Conditional Distributions

The following shows the existence of a regular version of the conditional distribution of a random variable in a very special case.

Proposition 1.136. *Let $X : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ and $Y : (\Omega, \mathcal{F}) \to (E_1, \mathcal{B}_1)$ be two random variables. Then the necessary and sufficient condition for X and Y to be independent is:*

$$\forall A \in \mathcal{B}_1 : \qquad P([Y \in A]|\cdot) = constant(A), \qquad P_X\text{-}a.s.$$

Therefore,

$$P([Y \in A]|\cdot) = P([Y \in A]), \qquad P_X\text{-}a.s.,$$

and if Y is a real-valued integrable random variable, then

$$E[Y|\cdot] = E[Y], \qquad P_X\text{-}a.s.$$

Proof: Independence of X and Y is equivalent to

$$P([X \in B] \cap [Y \in A]) = P([X \in B])P([Y \in A]) \qquad \forall A \in \mathcal{B}_1, B \in \mathcal{B},$$

or

$$\int_{[X \in B]} I_{[Y \in A]}(\omega)P(d\omega) = P([Y \in A]) \int I_B(x)dP_X(x)$$
$$= \int_B P([Y \in A])dP_X(x),$$

and this is equivalent to affirming that

$$P([Y \in A]|\cdot) = P([Y \in A]), \qquad P_X\text{-a.s.} \tag{1.10}$$

which is a constant k for $x \in E$. If we can write

$$P([Y \in A]|\cdot) = k(A), \qquad P_X\text{-a.s.},$$

then

$$\forall B \in \mathcal{B}\colon \int_{[X \in B]} I_{[Y \in A]}(\omega)dP(\omega) = \int_B k(A)dP_X(x) = k(A)P([X \in B]),$$

from which it follows that

$$\forall B \in \mathcal{B}: \qquad P([X \in B] \cap [Y \in A]) = k(A)P([X \in B]).$$

Therefore, for $B = E$, we have that

$$P([Y \in A]) = k(A)P([X \in E]) = k(A).$$

Now, we observe that (1.10) states that there exists a regular version of the probability conditional on X, relative to the σ-algebra $\mathcal{F}'$ generated by Y, where the latter is given by

$$P([Y \in A]|\cdot) = P_Y(A) \qquad \forall x \in E.$$

Hence, by Remark 1.135, it can then be shown that $E[Y|\cdot] = E[Y]$. □

We have already shown that if X is a discrete random variable, then the real random variable Y has a distribution conditional on X. The following theorem provides more general conditions under which this conditional distribution exists.

Theorem 1.137. (Jirina). *Let X and Y be two random variables on $(\Omega, \mathcal{F}, P)$ with values in $(E, \mathcal{B})$ and $(E_1, \mathcal{B}_1)$, respectively. If E and E_1 are complete separable metric spaces with respective Borel σ-algebras $\mathcal{B}$ and $\mathcal{B}_1$, then there exists a distribution of Y conditional on X.*

Definition 1.138. Given the assumptions of Definition 1.134, if $P_Y(\cdot|X = x)$ is defined by density with respect to measure μ_1 on $(E_1, \mathcal{B}_1)$, then this density is said to be conditional on X, written $X = x$, and denoted by $f_Y(\cdot|X = x)$.

Proposition 1.139. *Let $\mathbf{X} = (X_1, \ldots, X_n) : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ be a vector of random variables, whose probability is defined through the density $f_{\mathbf{X}}(x_1, \ldots, x_n)$ with respect to Lebesgue measure μ_n on $\mathbb{R}^n$. Fixing $q = 1, \ldots, n$, we can consider the random vectors*

$$\mathbf{Y} = (X_1, \ldots, X_q) \colon (\Omega, \mathcal{F}) \to \mathbb{R}^q$$

and

$$\mathbf{Z} = (X_{q+1}, \ldots, X_n) \colon (\Omega, \mathcal{F}) \to \mathbb{R}^{n-q}.$$

Then $\mathbf{Z}$ admits a distribution conditional on $\mathbf{Y}$, for almost every $\mathbf{Y} \in \mathbb{R}$, defined through the function

$$f(x_{q+1}, \ldots, x_n | x_1, \ldots, x_q) = \frac{f_{\mathbf{X}}(x_1, \ldots, x_q, x_{q+1}, \ldots, x_n)}{f_{\mathbf{Y}}(x_1, \ldots, x_q)},$$

with respect to Lebesgue measure μ_{n-q} on $\mathbb{R}^{n-q}$. Thereby $f_{\mathbf{Y}}(x_1, \ldots, x_q)$ is the marginal density of $\mathbf{Y}$ at $(x_1, \ldots, x_q)$, given by

$$f_{\mathbf{Y}}(x_1, \ldots, x_q) = \int f_{\mathbf{X}}(x_1, \ldots, x_n) d\mu_{n-q}(x_{q+1}, \ldots, x_n).$$

Proof: Writing $\mathbf{y} = (x_1, \ldots, x_q)$ and $\mathbf{x} = (x_1, \ldots, x_n)$, let $B \in \mathcal{B}_{\mathbb{R}^q}$ and $B_1 \in \mathcal{B}_{\mathbb{R}^{n-q}}$. Then

$$\begin{aligned} P([\mathbf{Y} \in B] \cap [\mathbf{Z} \in B_1]) = P_{\mathbf{X}}((\mathbf{Y}, \mathbf{Z}) = \mathbf{X} \in B \times B_1) &= \int_{B \times B_1} f_{\mathbf{X}}(\mathbf{x}) d\mu_n \\ &= \int_B d\mu_q(x_1, \ldots, x_q) \int_B f_{\mathbf{X}}(\mathbf{x}) d\mu_{n-q}(x_{q+1}, \ldots, x_n) \\ &= \int_B f_{\mathbf{Y}}(\mathbf{x}) d\mu_q \int_{B_1} \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})} d\mu_{n-q} \\ &= \int_B dP_{\mathbf{Y}} \left(\int_{B_1} \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}(\mathbf{y})}} d\mu_{n-q} \right), \end{aligned}$$

where the last equality holds for all points $\mathbf{y}$ for which $f_{\mathbf{Y}}(\mathbf{y}) \neq 0$. By the definition of density, the set of points $\mathbf{y}$ for which $f_{\mathbf{Y}}(\mathbf{y}) = 0$ has zero measure with respect to $P_{\mathbf{Y}}$, and therefore we can write in general:

$$P([\mathbf{Y} \in B] \cap [\mathbf{Z} \in B_1]) = \int_B dP_{\mathbf{Y}}(\mathbf{y}) \int_{B_1} \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})} d\mu_{n-q}.$$

Thus the latter integral is an element of $P([\mathbf{Z} \in B_1]|\mathbf{Y} = \mathbf{y})$. Hence

$$\int_{B_1} \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})} d\mu_{n-q} = P([\mathbf{Z} \in B_1]|\mathbf{Y} = \mathbf{y}) = P_{\mathbf{Z}}(B_1|\mathbf{Y} = \mathbf{y}),$$

from which it follows that $\frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{y})}$ is the density of $P(\cdot|\mathbf{Y} = \mathbf{y})$. □

Example 1.140. Let $f_{X,Y}(x, y)$ be the density of the bivariate Gaussian distribution. Then

$$f_{X,Y}(x, y) = k \exp\left\{-\frac{1}{2}(a(x - m_1)^2 + 2b(x - m_1)(y - m_2) + c(y - m_2)^2)\right\},$$

where

$$k = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}, \qquad a = \frac{1}{(1-\rho^2)\sigma_x^2},$$
$$b = \frac{-\rho}{(1-\rho^2)\sigma_x\sigma_y}, \qquad c = \frac{1}{(1-\rho^2)\sigma_y^2}.$$

The distribution of Y conditional on X is defined through the density

$$f_Y(Y|X = x) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \text{ where } f_X(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x - m}{\sigma_x}\right)^2\right\}.$$

From this it follows that

$$f_Y(Y|X = x) = \frac{1}{\sigma_y\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{y - m_2 - \frac{\sigma_y}{\sigma_x}(x - m_1)}{\sigma_y}\right)^2\right\}.$$

Therefore, the conditional density is normal, but with mean

$$E[Y|X = x] = \int y dP_Y(y|X = x) = \int y f_Y(y|X = x) dy = m_2 + \rho\frac{\sigma_y}{\sigma_x}(x - m_1)$$

and variance $(1-\rho^2)\sigma_y^2$. The conditional expectation in this case is also called the regression line of Y with respect to X.

Remark 1.141. Under the assumptions of Proposition 1.136, two generic random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$ with values in $(E, \mathcal{B})$ and $(E, \mathcal{B}_1)$, respectively, are independent if and only if Y has a conditional distribution with respect to $X = x$, which is independent of x:

$$P_Y(A|X = x) = P_Y(A), \qquad P_X\text{-a.s.}, \tag{1.11}$$

which can be rewritten to hold for every $x \in E$. If X and Y are independent, then their joint probability is given by

$$P_{(X,Y)} = P_X \otimes P_Y.$$

Integrating a function $f(x, y)$ with respect to $P(X, Y)$, by Fubini's theorem, results in

$$\int f(x,y)P_{(X,Y)}(dx, dy) = \int dP_X(x) \int f(x, y)dP_Y(y). \tag{1.12}$$

Using (1.11), then (1.12) can be rewritten in the form

$$\int f(x,y)P_{(X,Y)}(dx, dy) = \int dP_X(x) \int f(x, y)dP_Y(y|X = x).$$

The following proposition asserts that this relation holds in general.

Proposition 1.142. (generalization of Fubini's theorem). *Let X and Y be two generic random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$ with values in $(E, \mathcal{B})$ and $(E, \mathcal{B}_1)$, respectively. Moreover, let P_X be the probability of X and $P_Y(\cdot|X = x)$ the probability of Y conditional on $X = x$, for every $x \in E$. Then, for all $M \in \mathcal{B} \otimes \mathcal{B}_1$, the function*

$$h : x \in E \to \int I_M(x, y)P_Y(dy|x)$$

is $\mathcal{B}$-measurable and positive, resulting in

$$P_{(X,Y)}(M) = \int P_X(dx) \left(\int I_M(x, y)P_Y(dy|x) \right). \tag{1.13}$$

In general, if $f : E \times E_1 \to \mathbb{R}$ is $P_{(X,Y)}$-integrable, then the function

$$h' : x \in E \to \int f(x, y)P_Y(dy|x)$$

is defined almost surely with respect to P_X, and is P_X-integrable. Thus we obtain

$$\int f(x, y)P_{(X,Y)}(dx, dy) = \int h'(x)P_X(dx). \tag{1.14}$$

Proof: We observe that if $M = B \times B_1$, $B \in \mathcal{B}$, and $B_1 \in \mathcal{B}_1$, then

$$P_{(X,Y)}(B \times B_1) = P([X \in B] \cap [Y \in B_1]) = \int_B P([Y \in B_1]|X = x)dP_X(x),$$

and by the definition of conditional probability

$$\begin{aligned} P_{(X,Y)}(B \times B_1) &= \int I_B(x) P_Y(B_1|x) dP_X(x) \\ &= \int dP_X(x) \int P_Y(dy|x) I_B(x) I_{B_1}(y). \end{aligned}$$

This shows that (1.13) holds for $M = B \times B_1$. It is then easy to show that (1.13) holds for every elementary function on $\mathcal{B} \otimes \mathcal{B}_1$. With the usual limiting procedure, we can show that for every $\mathcal{B} \otimes \mathcal{B}_1$-measurable positive f we get

$$\int^* f(x,y) dP_{(X,Y)}(x,y) = \int^* dP_X(x) \int^* f(x,y) P_Y(dy|x).$$

As usual we have denoted by $\int^*$ the integral of a nonnegative measurable function, independently of its finiteness. If, then, f is measurable as well as both $P_{(X,Y)}$-integrable and positive, then

$$\int^* dP_X(x) \int^* f(x,y) P_Y(dy|x) < \infty,$$

where

$$\int^* f(x,y) P_Y(dy|x) < \infty, \qquad P_X\text{-a.s., } x \in E.$$

Thus h' is defined almost surely with respect to P_X and (1.14) holds. Finally, if f is $P_{(X,Y)}$-integrable and of arbitrary sign, applying the preceding results to f^+ and f^-, we obtain that

$$\int f(x,y) P_Y(dy|x) = \int f^+(x,y) P_Y(dy|x) - \int f^-(x,y) P_Y(dy|x)$$

is defined almost surely with respect to P_X, and again (1.14) holds. □

1.7 Convergence of Random Variables

Convergence in Mean of Order p

Definition 1.143. Let X be a real-valued random variable on the probability space $(\Omega, \mathcal{F}, P)$. X is *integrable to the pth exponent* ($p \geq 1$) if the random variable $|X|^p$ is P-integrable; thus $|X|^p \in \mathcal{L}^1(P)$. By $\mathcal{L}^p(P)$ we denote the whole of the real-valued random variables on $(\Omega, \mathcal{F}, P)$ that are integrable to the pth exponent. Then, by definition,

$$X \in \mathcal{L}^p(P) \Leftrightarrow |X|^p \in \mathcal{L}^1(P).$$

The following results are easy to show:

Theorem 1.144.

$$X, Y \in \mathcal{L}^p(P) \Rightarrow \begin{cases} \alpha X \in \mathcal{L}^p(P) & (\alpha \in \mathbb{R}), \\ X + Y \in \mathcal{L}^p(P), \\ \sup\{X, Y\} \in \mathcal{L}^p(P), \\ \inf\{X, Y\} \in \mathcal{L}^p(P). \end{cases}$$

Theorem 1.145. *If $X, Y \in \mathcal{L}^p(P)$ with $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, then $XY \in \mathcal{L}^p(P)$.*

Corollary 1.146. If $1 \leq p' \leq p$, then $\mathcal{L}^p(P) \subset \mathcal{L}^{p'}(P)$.

Proposition 1.147. Putting $N_p(X) = (\int |X|^p dP)^{\frac{1}{p}}$ for $X \in \mathcal{L}^p(P)$ $(p \geq 1)$, we get the following results.

1. Hölder's inequality: If $X \in \mathcal{L}^p(P)$, $Y \in \mathcal{L}^q(P)$ with $p, q > 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, then $N_1(XY) \leq N_p(X)N_q(Y)$.
2. Cauchy–Schwarz inequality:

$$\left| \int XY dP \right| \leq N_2(X)N_2(Y), \qquad X, Y \in \mathcal{L}^2(P). \tag{1.15}$$

3. Minkowski's inequality:

$$N_p(X + Y) \leq N_p(X) + N_p(Y) \text{ for } X, Y \in \mathcal{L}^p(P), \ (p \geq 1).$$

Proposition 1.148. The mapping $N_p : \mathcal{L}^p(P) \to \mathbb{R}_+$ $(p \geq 1)$ has the following properties:

1. $N_p(\alpha X) = |\alpha| N_p(X)$ for $X \in \mathcal{L}^p(P)$, $\alpha \in \mathbb{R}$;
2. $X = 0 \Rightarrow N_p(X) = 0$.

By 1 and 2 of Proposition 1.148 as well as 3 of Proposition 1.147, we can assert that N_p is a *seminorm on* $\mathcal{L}^p(P)$, but not a norm.

Definition 1.149. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of elements of $L^p(P)$ and let X be another element of $L^p(P)$. Then the sequence $(X_n)_{n\in\mathbb{N}}$ *converges to* X *in mean of order* p, if $\lim_{n\to\infty} \|X_n - X\|_p = 0$.

Convergence in Distribution

Now we will define a different type of convergence of random variables, which is associated with its partition function (see Loève (1963) for further references). We consider a sequence of probabilities $(P_n)_{n\in\mathbb{N}}$ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and define the following.

Definition 1.150. The sequence of probabilities $(P_n)_{n\in\mathbb{N}}$ *converges weakly* to a probability P if the following conditions are satisfied:

$$\forall f : \mathbb{R} \to \mathbb{R} \text{ continuous and bounded: } \lim_{n\to\infty} \int f dP_n = \int f dP.$$

We write

$$P_n \xrightarrow[n\to\infty]{\mathcal{W}} P.$$

Definition 1.151. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables on the probability space $(\Omega, \mathcal{F}, P)$ and X a further random variable defined on the same space. $(X_n)_{n\in\mathbb{N}}$ *converges in distribution* to X if the sequence $(P_{X_n})_{n\in\mathbb{N}}$ converges weakly to P_X. We write

$$X_n \xrightarrow[n\to\infty]{d} X.$$

Theorem 1.152. *Denoting by F the partition function associated with X, then, for every $n \in \mathbb{N}$, with F_{X_n} being the partition function associated with X_n, the following two conditions are equivalent:*

1. *for all $f : \mathbb{R} \to \mathbb{R}$ continuous and bounded:* $\lim_{n\to\infty} \int f dP_{X_n} = \int f dP_X$;
2. *for all $t \in \mathbb{R}$ such that F is continuous in t:* $\lim_{n\to\infty} F_{X_n}(t) = F(t)$.

We will henceforth denote the characteristic functions associated with the random variables X and X_n, by ϕ_X and ϕ_{X_n}, for all $n \in \mathbb{N}$, respectively.

Theorem 1.153. (Lévy's continuity theorem). *If $(P_{X_n})_{n\in\mathbb{N}}$ converges weakly to P_X, then for all $t \in \mathbb{R}$: $\phi_{X_n}(t) \xrightarrow[n]{} \phi_X(t)$. If there exists $\phi : \mathbb{R} \to \mathbb{C}$ such that for all $t \in \mathbb{R}$: $\phi_{X_n}(t) \xrightarrow[n]{} \phi_X(t)$ and, moreover, if ϕ is continuous in zero, then ϕ is the characteristic function of a probability P on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, such that $(P_{X_n})_{n\in\mathbb{N}}$ converges weakly to P.*

Theorem 1.154. (Polya). *If F_X is continuous and for all $t \in \mathbb{R}$:*

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t),$$

then $(\phi_{X_n})_{n\in\mathbb{N}}$ converges pointwise to ϕ_X and the convergence is uniform on all the bounded intervals $[-T, T]$.

Almost Sure Convergence and Convergence in Probability

Definition 1.155. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables on the probability space $(\Omega, \mathcal{F}, P)$ and X a further random variable defined on the same space. $(X_n)_{n\in\mathbb{N}}$ *converges almost surely* to X, denoted by $X_n \xrightarrow[n]{a.s.} X$ or, equivalently, $\lim_{n\to\infty} X_n = X$ almost surely, if

$$\exists S_0 \subset \Omega \text{ such that } P(S_0) = 0 \text{ and } \forall \omega \in \Omega \setminus S_0 \colon \lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Definition 1.156. $(X_n)_{n\in\mathbb{N}}$ *converges in probability (or stochastically)* to X, denoted by $X_n \xrightarrow[n]{P} X$ or, equivalently, $P-\lim_{n\to\infty} X_n = X$, if

$$\forall\epsilon > 0 : \lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

Theorem 1.157. *The following relationships hold:*

1. *almost sure convergence ⇒ convergence in probability ⇒ convergence in distribution;*
2. *convergence in mean ⇒ convergence in probability;*
3. *if the limit is a degenerate random variable (i.e., a deterministic quantity) then convergence in probability ⇔ convergence in distribution.*

Skorohod Representation Theorem

A fundamental result was obtained by Skorohod, relating convergence in law and almost sure convergence (see, e.g., Billingsley (1968)).

Theorem 1.158. (Skorohod representation theorem). *Consider a sequence* $(P_n)_{n\in\mathbb{N}}$ *of probability measures and a probability measure* P *on* $(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k})$, *such that* $P_n \xrightarrow[n\to\infty]{\mathcal{W}} P$. *Let* F_n *be the distribution function corresponding to* P_n, *and* F *the distribution function corresponding to* P. *Then there exists a sequence of random variables* $(Y_n)_{n\in\mathbb{N}}$ *and a random variable* Y *defined on a common probability space* $(\Omega, \mathcal{F}, P)$, *with values in* $(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k})$, *such that* Y_n *has distribution function* F_n, Y *has distribution function* F, *and*

$$Y_n \xrightarrow[n\to\infty]{a.s.} Y.$$

Note

For proofs of the various results, see, e.g., Ash (1972), Bauer (1981), or Métivier (1968).

1.8 Exercises and Additions

1.1. Prove Proposition 1.15.

1.2. Prove all the points of Example 1.66.

1.3. Show that the statement of Example 1.80 is true.

1.4. Prove all points of Example 1.89 and, in addition, the following: Let X be a Cauchy distributed random variable, i.e., $X \sim C(0,1)$; then $Y = a + hX \sim C(a,h)$.

1.5. Give an example of two random variables that are uncorrelated but not independent.

1.6. If X has an absolutely continuous distribution with pdf $f(x)$, its *entropy* is defined as

$$H(X) = -\int_D f(x) \ln f(x) dx,$$

where $D = \{x \in \mathbb{R} | f(x) > 0\}$.

1. Show that the maximal value of entropy within the set of nonnegative random variables with a given expected value μ is attained by the exponential $E(\mu^{-1})$.
2. Show that the maximal value of entropy within the set of real random variables with fixed mean μ and variance σ^2 is attained by the Gaussian $N(\mu, \sigma^2)$.

1.7. We say that X is a *compound Poisson* random variable if it can be expressed as

$$X = \sum_{k=1}^{N} Y_k$$

for $N \in \mathbb{N}^*$, and $X = 0$ for $N = 0$, where N is a Poisson random variable with some parameter $\lambda \in \mathbb{R}_+^*$, and $(Y_k)_{k \in \mathbb{N}^*}$ is a family of independent and identically distributed random variables, independent of N. Determine the characteristic function of X.

1.8. Let X be a random variable with characteristic function ϕ. We say that ϕ is *infinitely divisible* (i.d.), if for any $n \in \mathbb{N}^*$, there exists a characteristic function ϕ_n such that

$$\phi(s) = (\phi_n(s))^n \text{ for any } s \in \mathbb{R}.$$

1. Show that the characteristic function of a Gaussian random variable $X \sim N(\mu, \sigma^2)$ is i.d.
2. Show that the characteristic function of a Poisson random variable $X \sim P(\lambda)$ is i.d.
3. Show that the characteristic function of a compound Poisson random variable is i.d.
4. Show that the exponential distribution $E(\lambda)$ is i.d.
5. Show that the characteristic function of a Gamma random variable $X \sim \Gamma(\alpha, \beta)$ is i.d.
6. Show that an i.d. characteristic function never vanishes.
7. Show that the characteristic function of a uniform random variable $X \sim U(0, 1)$ is not i.d.

1.9. (*Kolmogorov*) Show that a function ϕ is an i.d. characteristic function with finite variance if and only if

$$\ln \phi(s) = ias + \int_{\mathbb{R}} \frac{e^{isx} - 1 - isx}{x^2} G(dx) \text{ for any } s \in \mathbb{R},$$

where $a \in \mathbb{R}$ and G is an increasing function of bounded variation (the reader may refer to Gnedenko (1963)).

1.10. (*Lévy–Khintchine*) Show that a function ϕ is an infinitely divisible characteristic function if and only if

$$\ln \phi(s) = ias - \frac{\sigma^2 s^2}{2} + \int_{\mathbb{R}-\{0\}} (e^{isx} - 1 - is\chi(x))\lambda_L(dx) \text{ for any } s \in \mathbb{R},$$

where $a \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+^*$,

$$\chi(x) = -I_{]-\infty,1]}(x) + xI_{]-1,1[}(x) + I_{[1,+\infty[},$$

and λ_L is a Lévy measure, i.e. a measure defined on $\mathbb{R}^*$ such that

$$\int_{\mathbb{R}^*} \min\{x^2, 1\}\lambda_L(dx) < +\infty.$$

The triplet (a, σ^2, λ_L) is called the *generating triplet* of the infinitely divisible characteristic function ϕ. (The reader may refer to Fristedt and Gray (1997) or Sato (1999).)

1.11. A distribution is infinitely divisible if and only if it is the weak limit of a sequence of distributions, each of which is compound Poisson (the reader may refer to Breiman (1968)).

1.12. We will say that two distribution functions F and G on $\mathbb{R}$ are of the *same type* if there exist two constants $a \in \mathbb{R}_+^*$ and $b \in \mathbb{R}$ such that

$$F(ax + b) = G(x) \text{ for any } x \in \mathbb{R}.$$

It is easy to see that this is an equivalence relation.

We may then introduce the definition of *stable law* as follows: F is stable, if the convolution of any two distributions of the same type as F is again of the same type.

Show that ϕ is the characteristic function of a stable law if and only if for any a_1 and a_2 in $\mathbb{R}_+^*$, there exist two constants $a \in \mathbb{R}_+^*$ and $b \in \mathbb{R}$ such that

$$\phi(a_1 s)\phi(a_2 s) = e^{ibs}\phi(as).$$

1.13. Show that any Cauchy distribution is stable.

1.14. Show that every stable law is infinitely divisible. What about the converse?

1.15. Show that if ϕ is the characteristic function of a stable law which is symmetric about the origin, then there exist $c \in \mathbb{R}_+^*$ and $\alpha \in]0, 2]$ such that

$$\phi(s) = e^{-c|s|^\alpha} \text{ for any } x \in \mathbb{R}.$$

1.16. If $\phi_1(t) = \sin t$, $\phi_2(t) = \cos t$ are characteristic functions, then give an example of random variables associated with ϕ_1, ϕ_2, respectively.

Let $\phi(t)$ be a characteristic function, and describe the random variable with characteristic function $|\phi(t)|^2$.

1.17. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with common density f, and

$$Y_j = j\text{th smallest of the } X_1, X_2, \ldots, X_n, \qquad j = 1, \ldots, n.$$

It follows that $Y_1 \leq \cdots \leq Y_j \leq \cdots \leq Y_n$. Show that

$$f_{Y_1,\ldots,Y_n} = \begin{cases} n! \prod_{i=1}^n f(y_i), & \text{if } y_1 < y_2 < \cdots < y_n, \\ 0, & \text{otherwise.} \end{cases}$$

1.18. Let X and $(Y_n)_{n\in\mathbb{N}}$ be random variables such that

$$X \sim E(1), \qquad Y_n(\omega) = \begin{cases} n, & \text{if } X(\omega) \leq \frac{1}{n}, \\ 0, & \text{otherwise.} \end{cases}$$

Give, if it exists, the limit $\lim_{n\to\infty} Y_n$:

- in distribution,
- in probability,
- almost surely,
- in mean of order $p \geq 1$.

1.19. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of uncorrelated random variables with common expected value $E[X_i] = \mu$ and such that $\sup Var[X_i] < +\infty$.

Show that $\sum_{i=1}^n \dfrac{X_i}{n}$ converges to μ in mean of order $p = 2$.

1.20. Give an example of random variables $X, X_1, X_2, \ldots$ such that $(X_n)_{n\in\mathbb{N}}$ converges to X

- in probability but not almost surely,
- in probability but not in mean,
- almost surely but not in mean and vice versa,
- in mean of order 1 but not in mean of order $p = 2$ (generally $p > 1$).

1.21. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables such that $X_i \sim B(p)$ for all i. Let Y be uniformly distributed on $[0,1]$ and independent of X_i, for all i. If $S_n = \frac{1}{n}\sum_{k=1}^{n}(X_k - Y)^2$, show that $(S_n)_{n\in\mathbb{N}}$ converges almost surely and determine its limit.

1.22. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables; determine the limit almost surely of

$$\frac{1}{n}\sum_{k=1}^{n}\sin\left(\frac{X_i}{X_{i+1}}\right)$$

in the following case:

- $X_i = \pm 1$ with probability $1/2$,
- X_i is a continuous random variable and its density function f_{X_i} is an even function.

(*Hint:* Consider the sum on the natural even numbers.)

1.23. (*Large deviations*). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables and suppose that their moment generating function $M(t) = E[e^{tX_1}]$ exists and is finite in $[0,a]$, $a \in \mathbb{R}_+^*$. Prove that for any $t \in [0,a]$

$$P(\bar{X} > E[X_1] + \epsilon) \leq (e^{-t(E[X_1]+\epsilon)}M(t))^n < 1,$$

where $\bar{X}$ denotes the arithmetic mean of $X_1, \ldots, X_n$, $n \in \mathbb{N}$.

Apply the above result to the cases $X_1 \sim B(1,p)$ and $X_1 \sim N(0,1)$.

1.24. (*Chernoff*). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed simple (finite range) random variables, satisfying $E[X_n] < 0$ and $P(X_n > 0) > 0$ for any $n \in \mathbb{N}$, and suppose that their moment generating function $M(t) = E[e^{tX_1}]$ exists and is finite in $[0,a]$, $a \in \mathbb{R}_+^*$. Show that

$$\lim_{n\to\infty}\frac{1}{n}\ln P(X_1 + \cdots + X_n \geq 0) = \ln\inf_t M(t).$$

1.25. (*Law of iterated logarithms*). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed simple (finite range) random variables with mean zero and variance 1. Show that

$$P\left(\limsup_n \frac{S_n}{\sqrt{2n\ln\ln n}} = 1\right) = 1.$$

1.26. Let X be a d-dimensional Gaussian vector. Prove that for every Lipschitz function f on $\mathbb{R}^d$, with $\|f\|_{Lip} \leq\leq 1$, the following inequality holds for any $\lambda \geq 0$:

$$P(f(X) - E[f(X)] \geq \lambda) \leq e^{-\frac{\lambda^2}{2}}.$$

1.27. Let X be an n-dimensional centered Gaussian vector. Show that

$$\lim_{r\to+\infty} \frac{1}{r^2} \ln P\left(\max_{1\le i\le n} X_i \ge r\right) = -\frac{1}{2\sigma^2}.$$

1.28. Let $(Y_n)_{n\in\mathbb{N}}$ be a family of random variables in $\mathcal{L}^1$, then the following two statements are equivalent:

1. $(Y_n)_{n\in\mathbb{N}}$ is uniformly integrable,
2. $\sup_{n\in\mathbb{N}} E[|Y_n|] < +\infty$ and for all ϵ, there exists a $\delta > 0$ such that $A \in \mathcal{F}$, $P(A) \le \delta \Rightarrow E[|Y_n I_A|] < \epsilon$.

(*Hint:* $\int_A |Y_n| \le rP(A) + \int_{|Y_n|>r} Y_n$ for $r > 0$.)

1.29. Show that the random variables $(Y_n)_{n\in\mathbb{N}}$ are uniformly integrable if and only if $\sup_n E[f(|Y_n|)] < \infty$ for some increasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(x)/x \to \infty$ as $n \to \infty$.

1.30. Show that for any $Y \in \mathcal{L}^1$, the family of conditional expectations $\{E[Y|\mathcal{G}], \mathcal{G} \subset F\}$ is uniformly integrable.

The following exercises are extending the concept of sequences of random variables and are introducing (discrete) processes and martingales. The latter's continuous equivalents will be the subject of the following chapters.

1.31. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\mathcal{F}_n)_{n\ge 0}$ be a *filtration*, that is, an increasing family of sub-σ-algebras of $\mathcal{F}$:

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}.$$

We define $\mathcal{F}_\infty := \sigma(\bigcup_n \mathcal{F}_n) \subseteq \mathcal{F}$. A process $X = (X_n)_{n\ge 0}$ is called *adapted* (to the filtration $(\mathcal{F}_n)_{n\ge 0}$) if for each n, X_n is $\mathcal{F}_n$-measurable.

A process X is called a *martingale* (relative to $(\mathcal{F}_n, P)$) if

- X is adapted,
- $E[|X_n|] < \infty$ for all n ($\Leftrightarrow X_n \in \mathcal{L}^1$),
- $E[X_n|\mathcal{F}_n] = X_{n-1}$ almost surely ($n \ge 1$).

1. Show that if $(X_n)_{n\in\mathbb{N}}$ is a sequence of independent random variables with $E[X_n] = 0$ for all $n \in \mathbb{N}$, then $S_n = X_1 + X_2 + \cdots + X_n$ is a martingale with respect to $(\mathcal{F}_n = \sigma(X_1, \ldots, X_n), P)$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
2. Show that if $(X_n)_{n\in\mathbb{N}}$ is a sequence of independent random variables with $E[X_n] = 1$ for all $n \in \mathbb{N}$, then $M_n = X_1 \cdot X_2 \cdot \cdots \cdot X_n$ is a martingale with respect to $(\mathcal{F}_n = \sigma(X_1, \ldots, X_n), P)$ and $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
3. Show that if $\{\mathcal{F}_n : n \ge 0\}$ is a filtration in $\mathcal{F}$ and $\xi \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$, then $M_n \equiv E[\xi|\mathcal{F}_n]$ is a martingale.
4. An urn contains white and black balls; we draw a ball and replace it with two balls of the same color; the process is repeated many times. Let X_n be the proportion of white balls in the urn before the nth draw. Show that the process $(X_n)_{n\ge 0}$ is a martingale.

1.32. A process $C = (C_n)_{n\geq 1}$ is called *predictable* if

$$C_n \text{ is } \mathcal{F}_{n-1}\text{-measurable } (n \geq 1).$$

We define

$$(C \bullet X)_n := \sum_{k=1}^{n} C_k(X_k - X_{k-1}).$$

Prove that if C is a bounded predictable process and X is a martingale, then $(C \bullet X)$ is a martingale null at $n = 0$ (*stochastic integration theorem*).

1.33. Let $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$. A random variable $T : (\Omega, \mathcal{F}) \to (\overline{\mathbb{N}}, \mathcal{B}_{\overline{\mathbb{N}}})$ is a stopping time if and only if

$$\forall n \in \overline{\mathbb{N}}: \ \{T \leq n\} \in \mathcal{F}_n.$$

Let X be a martingale with respect to the natural filtration $(\mathcal{F}_t)_{n\in\mathbb{R}_+}$ and let T be a stopping time with respect to the same filtration. Show that the stopped process $X_T; = (X_{n\wedge T(\omega)})_{n\geq 0}$ is a martingale with the same expected value of X.
(*Hint:* Consider the predictable process $C_n = I_{(T\geq n)}$ and apply the result of problem 1.32 to the process $(X_T - X_0)_n = (C^T \bullet X)_n$.)

1.34. Let $(X_n)_{n\geq 0}$ be an adapted process with $X_n \in \mathcal{L}^1$ for all n. Prove that X admits a *Doob decomposition*

$$X = X_0 + M + A,$$

where M is a martingale null at $n = 0$ and A is a predictable process null at $n = 0$. Moreover, this decomposition is unique in the sense that if $X = X_0 + \widetilde{M} + \widetilde{A}$ is another such decomposition, then

$$P(M_n = \widetilde{M}_n, A_n = \widetilde{A}_n, \forall n) = 1.$$

1.35. Consider the model

$$\Delta X_n = X_{n+1} - X_n = pX_n + \Delta M_n,$$

where M_n is a zero-mean-martingale. Prove that

$$\hat{p} = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{X_j} \Delta X_j$$

is an unbiased estimator of p (i.e., $E[\hat{p}] = p$). (*Hint:* Use the stochastic integration theorem.)

2

Stochastic Processes

2.1 Definition

We commence along the lines of the founding work of Kolmogorov by regarding stochastic processes as a family of random variables defined on a probability space and thereby define a probability law on the set of trajectories of the process. More specifically, stochastic processes generalize the notion of (finite-dimensional) vectors of random variables to the case of any family of random variables indexed in a general set T. Typically, the latter represents "time" and is an interval of $\mathbb{R}$ (in the continuous case) or $\mathbb{N}$ (in the discrete case).

Definition 2.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space, T an index set, and (E, B) a measurable space. An $(E, \mathcal{B})$-valued *stochastic process* on $(\Omega, \mathcal{F}, P)$ is a family $(X_t)_{t\in T}$ of random variables $X_t : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ for $t \in T$.

$(\Omega, \mathcal{F}, P)$ is called the underlying *probability space* of the process $(X_t)_{t\in T}$, while $(E, \mathcal{B})$ is the *state space* or *phase space*. Fixing $t \in T$, the random variable X_t is the *state of the process at "time" t*. Moreover, for all $\omega \in \Omega$, the mapping $X(\cdot, \omega) : t \in T \to X_t(\omega) \in E$ is called the *trajectory* or *path of the process* corresponding to ω. Any trajectory $X(\cdot, \omega)$ of the process belongs to the space E^T of functions defined in T and valued in E. Our aim is to introduce a suitable σ-algebra $\mathcal{B}^T$ on E^T that makes the family of trajectories of our stochastic process a random function $X : (\Omega, \mathcal{F}) \to (E^T, \mathcal{B}^T)$.

More generally, let us consider the family of measurable spaces $(E_t, \mathcal{B}_t)_{t\in T}$ (as a special case, all E_t may coincide with a unique E) and define $W^T = \prod_{t\in T} E_t$. If $S \in \mathcal{S}$, where $\mathcal{S} = \{S \subset T |\ \text{S is finite}\}$, the product σ-algebra $\mathcal{B}^S = \bigotimes_{t\in S} \mathcal{B}_t$ is well defined as the σ-algebra generated by the family of rectangles with sides in $\mathcal{B}_t$, $t \in S$.

Definition 2.2. If $A \in \mathcal{B}^S$, $S \in \mathcal{S}$, then the subset $\pi_{ST}^{-1}(A)$ is a *cylinder* in W^T with base A, where π_{ST} is the canonical projection of W^T on W^S.

It is easy to show that if C_A and $C_{A'}$ are cylinders with bases $A \in \mathcal{B}^S$ and $A' \in \mathcal{B}^{S'}$, $S, S' \in \mathcal{S}$, respectively, then $C_A \cap C_{A'}$, $C_A \cup C_{A'}$, and $C_A \setminus C_{A'}$ are cylinders with base in $W^{S \cup S'}$. From this it follows that the set of cylinders with finite-dimensional base is a *ring* of subsets of W^T (or, better, an *algebra*). We denote by $\mathcal{B}^T$ the σ-algebra generated by it. (See, e.g., Métivier (1968).)

Definition 2.3. The measurable space $(W^T, \mathcal{B}^T)$ is called the *product space of the measurable spaces* $(E_t, \mathcal{B}_t)_{t \in T}$.

From the definition of $\mathcal{B}^T$ we have the following result.

Theorem 2.4. *$\mathcal{B}^T$ is the smallest σ-algebra of the subsets of W^T that makes all canonical projections π_{ST} measurable.*

Furthermore the following is true.

Lemma 2.5. *The canonical projections π_{ST} are measurable if and only if $\pi_{\{t\}T}$ for all $t \in T$, are measurable as well.*

Moreover, from a well-known result of measure theory, we have the following proposition.

Proposition 2.6. *A function $f : (\Omega, \mathcal{F}) \to (W^T, \mathcal{B}^T)$ is measurable if and only if for all $t \in T$, the composite mapping $\pi_{\{t\}} \circ f : (\Omega, \mathcal{F}) \to (E_t, \mathcal{B}_t)$ is measurable.*

For proofs of Theorem 2.4, Lemma 2.5, and Proposition 2.6, see, e.g., Métivier (1968).

Remark 2.7. Let $(\Omega, \mathcal{F}, P, (X_t)_{t \in T})$ be a stochastic process with state space $(E, \mathcal{B})$. Since the function space $E^T = \prod_{t \in T} E$, the mapping $f : \Omega \to E^T$, which associates every $\omega \in \Omega$ with its corresponding trajectory of the process, is $(\mathcal{F} - \mathcal{B}^T)$-measurable, and, in fact, we have that

$$\forall t \in T: \qquad \pi_{\{t\}} \circ f(\omega) = \pi_{\{t\}}(X(\cdot, \omega)) = X_t(\omega),$$

$\pi_{\{t\}} \circ f = X_t$, which is a random variable, is obviously measurable.

Definition 2.8. A function $f : \Omega \to E^T$ defined on a probability space $(\Omega, \mathcal{F}, P)$ and valued in a measurable space $(E^T, \mathcal{G})$ is called a *random function* if it is $(\mathcal{F}\text{-}\mathcal{G})$-measurable.

How can we define a probability law P^T on $(E^T, \mathcal{B}^T)$ for the stochastic process $(X_t)_{t \in T}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ in a coherent way? We may observe that from a *physical* point of view, it is natural to assume that in principle we are able, from experiments, to define all possible *finite-dimensional* joint probabilities

$$P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n)$$

for any $n \in \mathbb{N}$, for any $\{t_1, \ldots, t_n\} \subset T$, and for any $B_1, \ldots, B_n \in \mathcal{B}$; i.e., the joint probability laws P^S of all finite-dimensional random vectors $(X_{t_1}, \ldots, X_{t_n})$, for any choice of $S = \{t_1, \ldots, t_n\} \subset \mathcal{S}$, such that

$$P^S(B_1 \times \cdots \times B_n) = P(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n).$$

Accordingly, we require that, for any $S \subset \mathcal{S}$,

$$P^T(\pi_{ST}^{-1}(B_1 \times \cdots \times B_n) = P^S(B_1 \times \cdots \times B_n) = P(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n).$$

A general answer comes from the following theorem. After having constructed the σ-algebra $\mathcal{B}^T$ on E^T, we now define a measure μ_T on $(W^T, \mathcal{B}^T)$, supposing that, for all $S \in \mathcal{S}$, a measure μ_S is assigned on $(W^S, \mathcal{B}^S)$. If $S \in \mathcal{S}$, $S' \in \mathcal{S}'$, and $S \subset S'$, we denote the canonical projection of W^S on $W^{S'}$ by $\pi_{SS'}$, which is certainly $(\mathcal{B}^{S'}$-$\mathcal{B}^S)$-measurable.

Definition 2.9. If, for all $(S, S') \in \mathcal{S} \times \mathcal{S}'$, with $S \subset S'$, we have that $\pi_{SS'}(\mu_{S'}) = \mu_S$. Moreover,

$$(W^S, \mathcal{B}^S, \mu_S, \pi_{SS'})_{S,S' \in \mathcal{S}; S \subset S'}$$

is called a *projective system* of measurable spaces and $(\mu_S)_{S \in \mathcal{S}}$ is called a *compatible system* of measures on the finite products $(W^S, \mathcal{B}^S)_{S \in \mathcal{S}}$.

Theorem 2.10. (Kolmogorov–Bochner). *Let $(E_t, \mathcal{B}_t)_{t \in T}$ be a family of Polish spaces (i.e., metric, complete, separable) endowed with their respective Borel σ-algebras, and let $\mathcal{S}$ be the collection of finite subsets of T and, for all $S \in \mathcal{S}$ with $W^S = \prod_{t \in S} E_t$ and $\mathcal{B}^S = \bigotimes_{t \in S} \mathcal{B}_t$, let μ_S be a finite measure on $(W^S, \mathcal{B}^S)$. Under these assumptions the following two statements are equivalent:*

1. *there exists a μ_T measure on $(W^T, \mathcal{B}^T)$ such that for all $S \in \mathcal{S}:\ \mu_S = \pi_{ST}(\mu_T)$;*
2. *the system $(W^S, \mathcal{B}^S, \mu_S, \pi_{SS'})_{S,S' \in \mathcal{S}; S \subset S'}$ is projective.*

Moreover, in both cases, μ_T, as defined in 1, is unique.

Proof: See, e.g., Métivier (1968). □

Definition 2.11. The unique measure μ_T of Theorem 2.10 is called the *projective limit* of the projective system $(W^S, \mathcal{B}^S, \mu_S, \pi_{SS'})_{S,S' \in \mathcal{S}; S \subset S'}$.

As a special case we consider a family of probability spaces $(E_t, \mathcal{B}_t, P_t)_{t \in T}$. If, for all $S \in \mathcal{S}$, we define $P_S = \bigotimes_{t \in S} P_t$, then $(W^S, \mathcal{B}^S, P_S, \pi_{SS'})_{S,S' \in \mathcal{S}; S \subset S'}$ is a projective system and the projective probability limit $\bigotimes_{t \in T} P_t$ is called the *probability product* of the family of probabilities $(P_t)_{t \in T}$.

With respect to the projective system of finite-dimensional probability laws $P_S = \bigotimes_{t \in S} P_{X_t}$ of a stochastic process $(X_t)_{t \in \mathbb{R}_+}$, the projective limit will be the required *probability law of the process*.

Theorem 2.12. *Two real-valued stochastic processes* $(X_t)_{t\in\mathbb{R}_+}$ *and* $(Y_t)_{t\in\mathbb{R}_+}$ *that have the same finite-dimensional probability laws have the same probability law.*

Definition 2.13. Two stochastic processes are *equivalent* if and only if they have the same projective system of finite-dimensional joint distributions.

A more stringent notion is the following.

Definition 2.14. Two real-valued stochastic processes $(X_t)_{t\in\mathbb{R}_+}$ and $(Y_t)_{t\in\mathbb{R}_+}$ on the probability space $(\Omega, \mathcal{F}, P)$ are called *modifications* or *versions* of one another if,

$$\text{for any } t \in T,\ P(X_t = Y_t) = 1.$$

Remark 2.15. It is obvious that two processes that are modifications of one another are also equivalent.

An even more stringent requirement comes from the following definition.

Definition 2.16. Two processes are *indistinguishable* if

$$P(X_t = Y_t, \forall t \in \mathbb{R}_+) = 1.$$

Remark 2.17. It is obvious that two indistinguishable processes are modifications of each other.

Example 2.18. Let $(X_t)_{t\in T}$ be a family of independent random variables defined on $(\Omega, \mathcal{F}, P)$ and valued in $(E, \mathcal{B})$. (In fact, in this case, it is sufficient to assume that only finite families of $(X_t)_{t\in T}$ are independent.) We know that for all $t \in T$ the probability $P_t = X_t(P)$ is defined on $(E, \mathcal{B})$. Then

$$\forall S = \{t_1, \ldots, t_r\} \in \mathcal{S}: \qquad P_S = \bigotimes_{k=1}^{r} P_{t_k},\ \text{for some } r \in \mathbb{N}^*,$$

and the system $(P_S)_{S\in\mathcal{S}}$ is compatible with its finite products $(E^S, \mathcal{B}^S)_{S\in\mathcal{S}}$. In fact, if B is a rectangle of $\mathcal{B}^S$, i.e., $B = B_{t_1} \times \cdots \times B_{t_r}$, and if $S \subset S'$, where $S, S' \in \mathcal{S}$, then

$$\begin{aligned} P_S(B) &= P_S(B_{t_1} \times \cdots \times B_{t_r}) = P_{t_1}(B_{t_1}) \cdot \cdots \cdot P_{t_r}(B_{t_r}) \\ &= P_{t_1}(B_{t_1}) \cdot \cdots \cdot P_{t_r}(B_{t_r}) P_{t_{r+1}}(E) \cdot \cdots \cdot P_{t_{r'}}(E) \\ &= P_{S'}(\pi_{SS'}^{-1}(B)). \end{aligned}$$

By the extension theorem we obtain that $P_S = \pi_{SS'}(P_{S'})$.

Remark 2.19. The compatibility condition $P_S = \pi_{SS'}(P_{S'})$, for all $S, S' \in \mathcal{S}$ and $S \subset S'$, can be expressed in an equivalent way by either the distribution function F_S of the probability P_S or its density f_S. Respectively, we obtain:

1. for all $S, S' \in \mathcal{S}, S \subset S'$, for all $(x_{t_1}, \ldots, x_{t_r}) \in E^S$:
   $F_S(x_{t_1}, \ldots, x_{t_r}) = F_{S'}(x_{t_1}, \ldots, x_{t_r}, +\infty, \ldots, +\infty)$;
2. for all $S, S' \in \mathcal{S}, S \subset S'$, for all $(x_{t_1}, \ldots, x_{t_r}) \in \mathbb{R}^S$:
   $f_S(x_{t_1}, \ldots, x_{t_r}) = \int \cdots \int dx_{t_{r+1}} \cdots dx_{t_{r'}} f_{S'}(x_{t_1}, \ldots, x_{t_r}, x_{t_{r+1}}, \ldots, x_{t_{r'}})$.

Definition 2.20. A real-valued stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is *continuous in probability* if

$$P - \lim_{s\to t} X_s = X_t, \qquad s, t \in \mathbb{R}_+.$$

Definition 2.21. A function $f : \mathbb{R}_+ \to \mathbb{R}$ is *right-continuous* if for any $t \in \mathbb{R}_+$, with $s > t$,

$$\lim_{s\downarrow t} f(s) = f(t).$$

Instead, the function is *left-continuous* if for any $t \in \mathbb{R}_+$, with $s < t$,

$$\lim_{s\uparrow t} f(s) = f(t).$$

Definition 2.22. A stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is *right-(left-)continuous* if its trajectories are right-(left-)continuous almost surely.

Definition 2.23. A stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *right-continuous with left limits* (RCLL) or *continu à droite avec limite à gauche* (càdlàg) if, almost surely, it has trajectories that are RCLL. The latter is denoted $X_{t^-} = \lim_{s\uparrow t} X_s$.

Theorem 2.24. *Let $(X_t)_{t\in\mathbb{R}_+}$ and $(Y_t)_{t\in\mathbb{R}_+}$ be two RCLL processes. X_t and Y_t are modifications of each other if and only if they are indistinguishable.*

Definition 2.25. A real-valued stochastic process $(X_t)_{t\in\mathbb{R}_+}$ on the probability space $(\Omega, \mathcal{F}, P)$ is called *separable* if

- there exists a $T_0 \subset \mathbb{R}_+$, countable and dense everywhere in $\mathbb{R}_+$;
- there exists an $A \in \mathcal{F}, P(A) = 0$ (negligible),

such that

- for all $t \in \mathbb{R}_+$: there exists $(t_n)_{n\in\mathbb{N}} \in T_0^{\mathbb{N}}$, such that $\lim_{n\to\infty} t_n = t$;
- for all $\omega \in \Omega \setminus A$: $\lim_{n\to\infty} X_{t_n}(\omega) = X_t(\omega)$.

The subset T_0 of $\mathbb{R}_+$, as defined above, is called the *separating set.*

Theorem 2.26. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a separable process, having T_0 and A as its separating and negligible sets, respectively. If $\omega \notin A$, $t_0 \in \mathbb{R}_+$, and $\lim_{t\to t_0} X_t(\omega)$ for $t \in T_0$ exists, then so does the limit $\lim_{t\to t_0} X_t(\omega)$ for $t \in \mathbb{R}_+$, and they coincide.*

Proof: See, e.g., Ash and Gardner (1975). □

Theorem 2.27. *Every real stochastic process* $(X_t)_{t\in\mathbb{R}_+}$ *admits a separable modification.*

Proof: See, e.g., Ash and Gardner (1975). □

Remark 2.28. By virtue of Theorem 2.27, we may henceforth only consider separable processes.

Definition 2.29. Let $(X_t)_{t\in\mathbb{R}_+}$ be a stochastic process defined on the probability space $(\Omega,\mathcal{F},P)$ and valued in $(E,\mathcal{B}_E)$. The process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *measurable* if it is measurable as a function defined on $\mathbb{R}_+\times\Omega$ (with the σ-algebra $\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{F}$) and valued in E.

Proposition 2.30. *If the process* $(X_t)_{t\in\mathbb{R}_+}$ *is measurable, then the trajectory* $X(\cdot,\omega):\mathbb{R}_+\to E$ *is measurable for all* $\omega\in\Omega$*.*

Proof: Let $\omega\in\Omega$ and $B\in\mathcal{B}_E$. We want to show that $(X(\cdot,\omega))^{-1}(B)$ is an element of $\mathcal{B}_{\mathbb{R}_+}$. In fact,

$$(X(\cdot,\omega))^{-1}(B)=\{t\in\mathbb{R}_+|X(t,\omega)\in B\}=\{t\in\mathbb{R}_+|(t,\omega)\in X^{-1}(B)\},$$

meaning that $(X(\cdot,\omega))^{-1}(B)$ is the path ω of X^{-1}, which is certainly measurable, because $X^{-1}(B)\in\mathcal{B}_{\mathbb{R}_+}\otimes\mathcal{F}$ (as follows from the properties of the product σ-algebra). □

If the process is measurable, it makes sense to consider the integral $\int_a^b X(t,\omega)dt$ along a trajectory. By Fubini's theorem, we have

$$\int_\Omega d\omega\int_a^b dtX(t,\omega)=\int_a^b dt\int_\Omega d\omega X(t,\omega).$$

However, in general, it is not true that a function $f(\omega_1,\omega_2)$ is jointly measurable in both variables, even if it is separately measurable in each of them. It is therefore required to impose conditions that guarantee the joint measurability of f in both variables. Evidently, if $(X_t)_{t\in\mathbb{R}_+}$ is a stochastic process, then for all $t\in\mathbb{R}_+: X(t,\cdot)$ is measurable.

Definition 2.31. The process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *progressively measurable* with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, which is an increasing family of subalgebras of $\mathcal{F}$, if, for all $t\in\mathbb{R}_+$, the mapping $(s,\omega)\in[0,t]\times\Omega\to X(s,\omega)\in E$ is $(\mathcal{B}_{[0,t]}\otimes\mathcal{F}_t)$-measurable. Furthermore, we henceforth suppose that $\mathcal{F}_t=\sigma(X(s),0\le s\le t)$, $t\in\mathbb{R}_+$, which is called the *generated* or *natural* filtration of the process X_t.

Proposition 2.32. *If the process* $(X_t)_{t\in\mathbb{R}_+}$ *is progressively measurable, then it is also measurable.*

Proof: Let $B \in \mathcal{B}_E$. Then

$$X^{-1}(B) = \{(s,\omega) \in \mathbb{R}_+ \times \Omega | X(s,\omega) \in B\}$$
$$= \bigcup_{n=0}^{\infty} \{(s,\omega) \in [0,n] \times \Omega | X(s,\omega) \in B\}.$$

Since

$$\forall n: \qquad \{(s,\omega) \in [0,n] \times \Omega | X(s,\omega) \in B\} \in \mathcal{B}_{[0,n]} \otimes \mathcal{F}_n,$$

we obtain that $X^{-1}(B) \in \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{F}$. □

Theorem 2.33. *If the process $(X_t)_{t\in\mathbb{R}_+}$ is continuous in probability, then it admits a separable and progressively measurable modification.*

Proof: See, e.g., Ash and Gardner (1975). □

Definition 2.34. A filtered complete probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in\mathbb{R}_+})$ is said to satisfy the *usual hypotheses* if

1. $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$,
2. $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$, for all $t \in \mathbb{R}_+$; i.e., the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is right-continuous.

Henceforth we will always assume that the usual hypotheses hold, unless specified otherwise.

Definition 2.35. Let $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in\mathbb{R}_+})$ be a filtered probability space. The σ-algebra on $\mathbb{R}_+ \times \Omega$ generated by all sets of the form $\{0\} \times A$, $A \in \mathcal{F}_0$, and $]a,b] \times A$, $0 \leq a < b < +\infty$, $A \in \mathcal{F}_a$, is said to be the *predictable σ-algebra* for the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$.

Definition 2.36. A real-valued process $(X_t)_{t\in\mathbb{R}_+}$ is called *predictable* with respect to a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, or *$\mathcal{F}_t$-predictable*, if as a mapping from $\mathbb{R}_+ \times \Omega \to \mathbb{R}$ it is measurable with respect to the predictable σ-algebra generated by this filtration.

Definition 2.37. A *simple predictable process* is of the form

$$X = k_0 I_{\{0\}\times A} + \sum_{i=1}^{n} k_i I_{]a_i,b_i]\times A_i},$$

where $A_0 \in \mathcal{F}_0$, $A_i \in \mathcal{F}_{a_i}$, $i = 1, \ldots, n$, and $k_0, \ldots, k_n$ are real constants.

Proposition 2.38. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a process that is $\mathcal{F}_t$-predictable. Then, for any $t > 0$, X_t is $\mathcal{F}_{t-}$-measurable.*

Lemma 2.39. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a left-continuous real-valued process adapted to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. Then X_t is predictable.*

Lemma 2.40. *A process is predictable if and only if it is measurable with respect to the smallest σ-algebra on $\mathbb{R}_+ \times \Omega$ generated by the adapted left-continuous processes.*

Proposition 2.41. *Every predictable process is progressively measurable.*

Proposition 2.42. *If the process $(X_t)_{t\in\mathbb{R}_+}$ is right-(left-)continuous, then it is progressively measurable.*

Proof: See, e.g., Métivier (1968). □

2.2 Stopping Times

In what follows we are given a probability space $(\Omega, \mathcal{F}, P)$ and a filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ on $\mathcal{F}$.

Definition 2.43. A random variable T defined on Ω (endowed with the σ-algebra $\mathcal{F}$) and valued in $\bar{\mathbb{R}}_+$ is called a *stopping time* (or *Markov time*) with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, or simply an $\mathcal{F}_t$*-stopping time*, if

$$\forall t \in \mathbb{R}_+ : \qquad \{\omega | T(\omega) \leq t\} \in \mathcal{F}_t.$$

The stopping time is said to be finite if $P(T = \infty) = 0$.

Remark 2.44. If $T(\omega) \equiv k$ (constant), then T is always a stopping time. If T is a stopping time with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ generated by the stochastic process $(X_t)_{t\in\mathbb{R}_+}$, $t \in \mathbb{R}_+$, then T is called the *stopping time of the process*.

Definition 2.45. Let T be an $\mathcal{F}_t$-stopping time. $A \in \mathcal{F}$ is said to *precede* T if, for all $t \in \mathbb{R}_+ :\ A \cap \{T \leq t\} \in \mathcal{F}_t$.

Proposition 2.46. *Let T be an $\mathcal{F}_t$-stopping time, and let $\mathcal{F}_T = \{A \in \mathcal{F} | A$ precedes $T\}$; then $\mathcal{F}_T$ is a σ-algebra of the subsets of Ω. It is called the σ-algebra* of T-preceding events.

Proof: See, e.g., Métivier (1968).

Theorem 2.47. *The following relationships hold:*

1. *If both S and T are stopping times, then so are $S \wedge T = \inf\{S, T\}$ and $S \vee T = \sup\{S, T\}$.*
2. *If T is a stopping time and $a \in [0, +\infty[$, then $T \wedge a$ is a stopping time.*
3. *If T is a finite stopping time, then it is $\mathcal{F}_T$-measurable.*
4. *If both S and T are stopping times and $A \in \mathcal{F}_S$, then $A \cap \{S \leq T\} \in \mathcal{F}_T$.*
5. *If both S and T are stopping times and $S \leq T$, then $\mathcal{F}_S \subset \mathcal{F}_T$.*

Proof: See, e.g., Métivier (1968).

Theorem 2.48. *Let* $(X_t)_{t\in\mathbb{R}_+}$ *be a progressively measurable stochastic process valued in* $(S, \mathcal{B}_S)$*. If* T *is a finite stopping time, then the function*

$$X(T) : \omega \in \Omega \to X(T(\omega), \omega) \in E$$

is $\mathcal{F}_T$*-measurable (and hence a random variable).*

Proof: We need to show that

$$\forall B \in \mathcal{B}_E: \qquad \{\omega | X(T(\omega)) \in B\} \in \mathcal{F}_T,$$

hence

$$\forall B \in \mathcal{B}_E, \forall t \in \mathbb{R}_+: \qquad \{\omega | X(T(\omega)) \in B\} \cap \{T \leq t\} \in \mathcal{F}_t.$$

Fixing $B \in \mathcal{B}_E$ we have

$$\forall t \in \mathbb{R}_+ : \qquad \{\omega | X(T(\omega)) \in B\} \cap \{T \leq t\} = \{X(T \wedge t) \in B\} \cap \{T \leq t\},$$

where $\{T \leq t\} \in \mathcal{F}_t$, since T is a stopping time. We now show that $\{X(T \wedge t) \in B\} \in \mathcal{F}_t$. In fact, $T \wedge t$ is a stopping time (by point 2 of Theorem 2.47) and is $\mathcal{F}_{T\wedge t}$-measurable (by point 3 of Theorem 2.47). But $\mathcal{F}_{T\wedge t} \subset \mathcal{F}_t$ and thus $T \wedge t$ is $\mathcal{F}_t$-measurable. Now $X(T \wedge t)$ is obtained as a composite of the mapping

$$\omega \in \Omega \to (T \wedge t(\omega), \omega) \in [0, t] \times \Omega \tag{2.1}$$

with

$$(s, \omega) \in [0, t] \times \Omega \to X(s, \omega) \in E. \tag{2.2}$$

The mapping (2.1) is $(\mathcal{F}_t - \mathcal{B}_{[0,t]} \otimes \mathcal{F}_t)$-measurable (because $T \wedge t$ is $\mathcal{F}_t$-measurable) and the mapping (2.2) is $(\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t - \mathcal{B}_E)$-measurable, since X is progressively measurable. Therefore, $X(T \wedge t)$ is $\mathcal{F}_t$-measurable, completing the proof. □

2.3 Canonical Form of a Process

Let $(\Omega, \mathcal{F}, P, (X_t)_{t\in T})$ be a stochastic process valued in $(E, \mathcal{B})$ and, for every $S \in \mathcal{S}$, let P_S be the joint probability law for the random variables $(X_t)_{t\in S}$ that is the probability on $(E^S, \mathcal{B}^S)$ induced by P through the function

$$X^S : \omega \in \Omega \to (X_t(\omega))_{t\in S} \in E^S = \prod_{t\in S} E.$$

Evidently, if

$$S \subset S' (S, S' \in \mathcal{S}), \qquad X^S = \pi_{SS'} \circ X^{S'},$$

then it follows that

$$P_S = X^S(P) = (\pi_{SS'} \circ X^{S'})(P) = \pi_{SS'}(P_{S'}),$$

and therefore $(E^S, \mathcal{B}^S, P_S, \pi_{SS'})_{S,S' \in \mathcal{S}, S \subset S'}$ is a projective system of probabilities.

On the other hand, the random function $f : \Omega \to E^T$ that associates every $\omega \in \Omega$ with a trajectory of the process in ω is measurable (following Proposition 2.6). Hence we can consider the induced probability P_T on $\mathcal{B}^T$, $P_T = f(P)$; P_T is the projective limit of $(P_S)_{S \in \mathcal{S}}$. From this it follows that $(E^T, \mathcal{B}^T, P_T, (\pi_t)_{t \in T})$ is a stochastic process with the property that, for all $S \in \mathcal{S}$, the random vectors $(\pi_t)_{t \in S}$ and $(X_t)_{t \in S}$ have the same joint distribution.

Definition 2.49. The stochastic process $(E^T, \mathcal{B}^T, P_T, (\pi_t)_{t \in T})$ is called the *canonical form* of the process $(\Omega, \mathcal{F}, P, (X_t)_{t \in T})$.

Remark 2.50. From this it follows that two stochastic processes are equivalent if they admit the same canonical process.

2.4 Gaussian Processes

Definition 2.51. The real-valued stochastic process $(\Omega, \mathcal{F}, P, (X_t)_{t \in \mathbb{R}_+})$ is called a *Gaussian process* if, for all $n \in \mathbb{N}^*$ and for all $(t_1, \ldots, t_n) \in \mathbb{R}_+^n$, the n-dimensional random vector $\mathbf{X} = (X_{t_1}, \ldots, X_{t_n})'$ has multivariate Gaussian distribution, with probability density

$$f_{t_1,\ldots,t_n}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det K}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{m})' K^{-1}(\mathbf{x}-\mathbf{m})\right\}, \qquad (2.3)$$

where

$$\begin{cases} m_i = E[X_{t_i}], & i = 1, \ldots, n, \\ K = (\sigma_{ij}) = Cov[X_{t_i}, X_{t_j}], & i,j = 1, \ldots, n. \end{cases}$$

Remark 2.52. Vice versa, by assigning n numbers $m_i, i = 1, \ldots, n$, and a positive definite $n \times n$ matrix $K = (\sigma_{ij})$ (i.e., such that for all $\mathbf{a} \in \mathbb{R}^n$: $\sum_{i,j=1}^n a_i \sigma_{ij} a_j > 0$), which in particular is nonsingular, we uniquely determine a Gaussian distribution with density given by (2.3). Now $P_{t_1,\ldots,t_n}$ represents the marginal probability law of $P_{t_1,\ldots,t_n,t_{n+1},\ldots,t_m}$, $m > n$, if and only if their densities satisfy the condition

$$\begin{aligned} &f_{t_1,\ldots,t_n}(x_{t_1}, \ldots, x_{t_n}) \\ &= \int \cdots \int dx_{t_{n+1}} \cdots dx_{t_m} f_{t_1,\ldots,t_n,t_{n+1},\ldots,t_m}(x_{t_1}, \ldots, x_{t_n}, x_{t_{n+1}}, \ldots, x_{t_m}). \end{aligned}$$

Hence, assigning a projective system of Gaussian laws $(P_S)_{S \in \mathcal{S}}$ (where $\mathcal{S}$ is the set of finite intervals of $\mathbb{R}_+$) is equivalent to assigning two vectors/matrices of functions

$$\begin{cases} m_i : \mathbb{R}_+ \to \mathbb{R}, \\ \\ K : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}, \\ \text{where } \forall n \in \mathbb{N}^*, \forall (t_1, \ldots, t_n) \in \mathbb{R}_+^n, \forall \mathbf{a} \in \mathbb{R}^n : \sum_{i,j=1}^n K(t_i, t_j) a_i a_j > 0. \end{cases}$$

Then the projective system $(P_S)_{S \in \mathcal{S}}$ is given by the density (2.3). Because $\mathbb{R}$ is a Polish space, by the Kolmogorov–Bochner Theorem 2.10, we can now assert that:
there exists a Gaussian process $(X_t)_{t \in \mathbb{R}_+}$, such that

$$\begin{cases} \forall t \in \mathbb{R}_+ : \ m(t) = E[X_t] \\ \forall (t, r) \in \mathbb{R}_+ \times \mathbb{R}_+ : \ K(t, r) = Cov[X_t, X_r]. \end{cases}$$

2.5 Processes with Independent Increments

Definition 2.53. The stochastic process $(\Omega, \mathcal{F}, P, (X_t)_{t \in \mathbb{R}_+})$, with state space $(E, \mathcal{B})$, is called a *process with independent increments* if, for all $n \in \mathbb{N}$ and for all $(t_1, \ldots, t_n) \in \mathbb{R}_+^n$, where $t_1 < \cdots < t_n$, the random variables $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

Theorem 2.54. *If* $(\Omega, \mathcal{F}, P, (X_t)_{t \in \mathbb{R}_t})$ *is a process with independent increments, then it is possible to construct a compatible system of probability laws* $(P_S)_{S \in \mathcal{S}}$*, where again* $\mathcal{S}$ *is a collection of finite subsets of the index set.*

Proof: To do this, we need to assign a joint distribution to every random vector $(X_{t_1}, \ldots, X_{t_n})$ for all $(t_1, \ldots, t_n)$ in $\mathbb{R}_+^n$ with $t_1 < \cdots < t_n$. Thus, let $(t_1, \ldots, t_n) \in \mathbb{R}_+^n$, with $t_1 < \cdots < t_n$, and μ_0, $\mu_{s,t}$ be the distributions of X_0 and $X_t - X_s$, for every $(s, t) \in \mathbb{R}_+ \times \mathbb{R}_+$, with $s < t$, respectively. We define

$$\begin{aligned} Y_0 &= X_0, \\ Y_1 &= X_{t_1} - X_0, \\ &\cdots \\ Y_n &= X_{t_n} - X_{t_{n-1}}, \end{aligned}$$

where $Y_0, Y_1, \ldots, Y_n$ have the distributions $\mu_0, \mu_{0,t_1}, \ldots, \mu_{t_{n-1},t_n}$, respectively. Moreover, since the Y_i are independent, $(Y_0, \ldots, Y_n)$ have joint distribution $\mu_0 \otimes \mu_{0,t_1} \otimes \cdots \otimes \mu_{t_n,t_{n-1}}$. Let f be a real-valued, $\bigotimes^n \mathcal{B}$-measurable function and consider the random variable $f(X_{t_1}, \ldots, X_{t_n})$, then

$$\begin{aligned} &E[f(X_{t_1}, \ldots, X_{t_n})] \\ &= E[f(Y_0 + Y_1, \ldots, Y_0 + \cdots + Y_n)] \\ &= \int f(y_0 + y_1, \ldots, y_0 + \cdots + y_n) d(\mu_0 \otimes \mu_{0,t_1} \otimes \cdots \otimes \mu_{t_{n-1},t_n})(y_0, \ldots, y_n). \end{aligned}$$

In particular, if $f = I_B$, with $B \in \bigotimes^n \mathcal{B}$, we obtain the joint distribution of $X_{t_1}, \ldots, X_{t_n}$:

$$P((X_{t_1},\ldots,X_{t_n})\in B)=E[I_B(X_{t_1},\ldots,X_{t_n})] \tag{2.4}$$
$$=\int I_B(y_0+y_1,\ldots,y_0+\cdots+y_n)d(\mu_0\otimes\mu_{0,t_1}\otimes\cdots\otimes\mu_{t_{n-1},t_n})(y_0,\ldots,y_n).$$

Thus having obtained P_S, where $S=\{t_1,\ldots,t_n\}$, with $t_1<\cdots<t_n$, we show that $(P_S)_{S\in\mathcal{S}}$ is a compatible system. Let $S,S'\in\mathcal{S};S\subset S'$, $S=\{t_1,\ldots,t_n\}$, with $t_1<\cdots<t_n$ and $S'=\{t_1,\ldots,t_j,s,t_{j+1},\ldots,t_n\}$, with $t_1<\cdots<t_j<s<t_{j+1}<\cdots<t_n$. For $B\in\mathcal{B}^S$ and $B'=\pi_{SS'}^{-1}(B)$, we will show that $P_S(B)=P_{S'}(B')$.

We can observe, by the definition of B', that

$$I_{B'}(x_{t_1},\ldots,x_{t_j},x_s,x_{t_{j+1}},\ldots,x_{t_n})$$

does not depend on x_s and is therefore identical to $I_B(x_{t_1},\ldots,x_{t_n})$. Thus putting $U=X_s-X_{t_j}$ and $V=X_{t_{j+1}}-X_s$, we obtain

$$\begin{aligned}P_{S'}(B')=\int I_{B'}(&y_0+y_1,\ldots,y_0+\cdots+y_j,y_0+\cdots+y_j+u,y_0+\cdots\\&+y_j+u+v,\ldots,y_0+\cdots+y_n)d(\mu_0\otimes\mu_{0,t_1}\otimes\cdots\\&\otimes\mu_{t_j,s}\otimes\mu_{s,t_{j+1}}\otimes\cdots\otimes\mu_{t_{n-1},t_n})(y_0,\ldots,y_j,u,v,y_{j+2},\ldots,y_n)\\=\int I_B(&y_0+y_1,\ldots,y_0+\cdots+y_j,y_0+\cdots+y_j+u+v,y_0+\cdots\\&+u+v+y_{j+2},\ldots,y_0+\cdots+y_n)d(\mu_0\otimes\mu_{0,t_1}\otimes\cdots\\&\otimes\mu_{t_j,s}\otimes\mu_{s,t_{j+1}}\otimes\cdots\otimes\mu_{t_{n-1},t_n})(y_0,\ldots,y_j,u,v,y_{j+2},\ldots,y_n).\end{aligned}$$

Integrating with respect to all the variables except u and v, after applying Fubini's theorem, we obtain

$$P_{S'}(B')=\int h(u+v)d(\mu_{t_j,s}\otimes\mu_{s,t_{j+1}})(u,v).$$

Letting $y_{j+1}=u+v$ we have

$$P_{S'}(B')=\int h(y_{j+1})d(\mu_{t_j,s}*\mu_{s,t_{j+1}})(y_{j+1}).$$

Moreover, we observe that the definition of $y_{j+1}=u+v$ is compatible with the above notation $Y_{j+1}=X_{t_{j+1}}-X_{t_j}$. In fact, we have

$$u+v=x_s-x_{t_j}+x_{t_{j+1}}-x_s=x_{t_{j+1}}-x_{t_j}.$$

Furthermore, for the independence of $(X_{t_{j+1}}-X_s)$ and $(X_s-X_{t_j})$, the sum of random variables

$$X_{t_{j+1}}-X_s+X_s-X_{t_j}=X_{t_{j+1}}-X_{t_j}$$

has to have the distribution $\mu_{t_j,s} * \mu_{s,t_{j+1}}$, where $*$ denotes the convolution product. Therefore, having denoted the distribution of $X_{t_{j+1}} - X_{t_j}$ with $\mu_{t_j,t_{j+1}}$, we obtain

$$\mu_{t_j,s} * \mu_{s,t_{j+1}} = \mu_{t_j,t_{j+1}}.$$

As a consequence we have

$$P_{S'}(B') = \int h(y_{j+1}) d\mu_{t_j,t_{j+1}}(y_{j+1}).$$

This integral coincides with the one in (2.4), and thus

$$P_S(B') = P((X_{t_1}, \dots, X_{t_n}) \in B) = P_S(B).$$

If now $S' = S \cup \{s_1, \dots, s_k\}$, the proof is completed by induction. □

Definition 2.55. A process with independent increments is called *time-homogeneous* if

$$\mu_{s,t} = \mu_{s+h,t+h} \qquad \forall s,t,h \in \mathbb{R}_+, s < t.$$

If $(\Omega, \mathcal{F}, P, (X_t)_{t\in\mathbb{R}_+})$ is a homogeneous process with independent increments, then as a particular case we have

$$\mu_{s,t} = \mu_{0,t-s} \qquad \forall s,t \in \mathbb{R}_+, s < t.$$

Definition 2.56. A family of measures $(\mu_t)_{t\in\mathbb{R}_+}$ that satisfy the condition

$$\mu_{t_1+t_2} = \mu_{t_1} * \mu_{t_2}$$

is called a *convolution semigroup.*

Remark 2.57. A time-homogeneous process with independent increments is completely defined by assigning it a convolution semigroup.

2.6 Martingales

Definition 2.58. Let $(X_t)_{t\in\mathbb{R}_+}$ be a real-valued family of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ and let $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtration. The stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *adapted* to the family $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if, for all $t \in \mathbb{R}_+$, X_t is $\mathcal{F}_t$-measurable.

Definition 2.59. The stochastic process $(X_t)_{t\in\mathbb{R}_+}$, adapted to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, is a *martingale* with respect to this filtration, provided the following conditions hold:

1. X_t is P-integrable, for all $t \in \mathbb{R}_+$;
2. for all $(s,t) \in \mathbb{R}_+ \times \mathbb{R}_+, s < t : E[X_t|\mathcal{F}_s] = X_s$ almost surely.

$(X_t)_{t\in\mathbb{R}_+}$ is said to be a *submartingale* (*supermartingale*) with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if, in addition to condition 1 and instead of condition 2, we have:

2′. for all $(s,t) \in \mathbb{R}_+ \times \mathbb{R}_+, s < t : E[X_t|\mathcal{F}_s] \geq X_s$ $(E[X_t|\mathcal{F}_s] \leq X_s)$ almost surely.

Remark 2.60. When the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is not specified, it is understood to be the increasing σ-algebra generated by the random variables of the process $(\sigma(X_s, 0 \leq s \leq t))_{t\in\mathbb{R}_+}$. In this case we can write $E[X_t|X_r, 0 \leq r \leq s]$, instead of $E[X_t|\mathcal{F}_s]$.

Example 2.61. The evolution of a gambler's wealth in a game of chance, the latter specified by the sequence of real-valued random variables $(X_n)_{n\in\mathbb{N}}$, will serve as a descriptive example of the above definitions. Suppose that two players flip a coin and the loser pays the winner (who guessed head or tail correctly) the amount α after every round. If $(X_n)_{n\in\mathbb{N}}$ represents the cumulative fortune of player 1, then after n throws he holds

$$X_n = \sum_{i=0}^{n} \Delta_i.$$

The random variables Δ_i (just like every flip of the coin) are independent and take values α and $-\alpha$ with probabilities p and q, respectively. Therefore, we see that

$$\begin{aligned} E[X_{n+1}|X_0,\dots,X_n] &= E[\Delta_{n+1} + X_n|X_0,\dots,X_n] \\ &= X_n + E[\Delta_{n+1}|X_0,\dots,X_n]. \end{aligned}$$

Since Δ_{n+1} is independent of every $\sum_{i=0}^{k} \Delta_i, k = 0,\dots,n$, we obtain

$$E[X_{n+1}|X_0,\dots,X_n] = X_n + E[\Delta_{n+1}] = X_n + \alpha(p-q).$$

- If the game is fair, then $p = q$ and $(X_n)_{n\in\mathbb{N}}$ is a martingale.
- If the game is in player 1's favor, then $p > q$ and $(X_n)_{n\in\mathbb{N}}$ is a submartingale.
- If the game is to the disadvantage of player 1, then $p < q$ and $(X_n)_{n\in\mathbb{N}}$ is a supermartingale.

Example 2.62. Let $(X_t)_{t\in\mathbb{R}_+}$ be (for all $t \in \mathbb{R}_+$) a P-integrable stochastic process on $(\Omega, \mathcal{F}, P)$ with independent increments. Then $(X_t - E[X_t])_{t\in\mathbb{R}_+}$ is a martingale. In fact:[4]

$$E[X_t|\mathcal{F}_s] = E[X_t - X_s|\mathcal{F}_s] + E[X_s|\mathcal{F}_s], \qquad s < t,$$

[4] For simplicity, but without loss of generality, we will assume that $E[X_t] = 0$, for all t. In the case where $E[X_t] \neq 0$, we can always define a variable $Y_t = X_t - E[X_t]$, so that $E[Y_t] = 0$. In that case $(Y_t)_{t\in\mathbb{R}_+}$ will again be a process with independent increments, so that the analysis is analogous.

and recalling that both X_s is $\mathcal{F}_s$-measurable and $(X_t - X_s)$ is independent of $\mathcal{F}_s$, we obtain that

$$E[X_t|\mathcal{F}_s] = E[X_t - X_s] + X_s = X_s, \qquad s < t.$$

Proposition 2.63. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a real-valued martingale. If the function $\phi : \mathbb{R} \to \mathbb{R}$ is both convex and measurable and such that*

$$\forall t \in \mathbb{R}_+ : \qquad E[\phi(X_t)] < +\infty,$$

then $(\phi(X_t))_{t\in\mathbb{R}_+}$ is a submartingale.

Proof: Let $(s,t) \in \mathbb{R}_+ \times \mathbb{R}_+, s < t$. Following Jensen's inequality and the properties of the martingale $(X_t)_{t\in\mathbb{R}_+}$, we have that

$$\phi(X_s) = \phi(E[X_t|\mathcal{F}_s]) \leq E[\phi(X_t)|\mathcal{F}_s].$$

Letting

$$\mathcal{V}_s = \sigma(\phi(X_r), 0 \leq r \leq s) \qquad \forall s \in \mathbb{R}_+$$

and with the measurability of ϕ, it is easy to verify that $\mathcal{V}_s \subset \mathcal{F}_s$ for all $s \in \mathbb{R}_+$ and therefore

$$\phi(X_s) = E[\phi(X_s)|\mathcal{V}_s] \leq E[E[\phi(X_t)|\mathcal{F}_s]|\mathcal{V}_s] = E[\phi(X_t)|\mathcal{V}_s].$$

□

Lemma 2.64. *Let X and Y be two positive real random variables defined on $(\Omega, \mathcal{F}, P)$. If $X \in L^p(P)$ $(p > 1)$ and if, for all $\alpha > 0$,*

$$\alpha P(Y \geq \alpha) \leq \int_{\{Y\geq\alpha\}} X dP, \tag{2.5}$$

then $Y \in L^p(P)$ and $\|Y\|_p \leq q\|X\|_p$, where $\frac{1}{p} + \frac{1}{q} = 1$.

Proof: We have

$$\begin{aligned}
E[Y^p] &= \int_\Omega Y^p(\omega)dP(\omega) = \int_\Omega dP(\omega)p\int_0^{Y(\omega)} \lambda^{p-1}d\lambda \\
&= p\int_\Omega dP(\omega)\int_0^\infty \lambda^{p-1}I_{\{\lambda\leq Y(\omega)\}}(\lambda)d\lambda \\
&= p\int_0^\infty d\lambda\lambda^{p-1}\int_\Omega dP(\omega)I_{\{\lambda\leq Y(\omega)\}}(\lambda) \\
&= p\int_0^\infty d\lambda\lambda^{p-1}P(\lambda \leq Y) = p\int_0^\infty d\lambda\lambda^{p-2}\lambda P(Y \geq \lambda) \\
&\leq p\int_0^\infty d\lambda\lambda^{p-2}\int_{\{Y\geq\lambda\}} XdP
\end{aligned}$$

$$
\begin{aligned}
&= p \int_{\Omega} dP(\omega) X(\omega) \int_0^{\infty} d\lambda \lambda^{p-2} I_{\{Y(\omega) \geq \lambda\}}(\lambda) \\
&= p \int_{\Omega} dP(\omega) X(\omega) \int_0^{Y(\omega)} d\lambda \lambda^{p-2} = \frac{p}{p-1} \int_{\Omega} dP(\omega) X(\omega) Y^{p-1}(\omega) \\
&= \frac{p}{p-1} E[Y^{p-1} X],
\end{aligned}
$$

where, throughout, λ denotes Lebesgue measure, and when changing the order of integration we invoke Fubini's theorem. By Hölder's inequality, we obtain

$$
E[Y^p] \leq \frac{p}{p-1} E[Y^{p-1} X] \leq \frac{p}{p-1} E[X^p]^{\frac{1}{p}} E[Y^p]^{\frac{p-1}{p}},
$$

which, after substitution and rearrangement, gives

$$
E[Y^p]^{\frac{1}{p}} \leq q E[X^p]^{\frac{1}{p}},
$$

as long as $E[Y^p] < +\infty$ (in such case we may, in fact, divide the left- and right-hand sides by $E[Y^p]^{\frac{p-1}{p}}$). But in any case we can consider the sequence of random variables $(Y \wedge n)_{n \in \mathbb{N}}$ ($Y \wedge n$ is the random variable defined letting, for all $\omega \in \Omega$, $Y \wedge n(\omega) = \inf\{Y(\omega), n\}$); since, for all n, $Y \wedge n$ satisfies condition (2.5), then we obtain

$$
\|Y \wedge n\|_p \leq q \|X\|_p,
$$

and in the limit

$$
\|Y\|_p = \lim_{n \to \infty} \|Y \wedge n\|_p \leq q \|X\|_p.
$$

□

Proposition 2.65. *Let $(X_n)_{n \in \mathbb{N}^*}$ be a sequence of real random variables defined on the probability space $(\Omega, \mathcal{F}, P)$, and X_n^+ the positive part of X_n.*

1. If $(X_n)_{n \in \mathbb{N}^*}$ is a submartingale, then

$$
P\left(\max_{1 \leq k \leq n} X_k > \lambda\right) \leq \frac{1}{\lambda} E[X_n^+], \qquad \lambda > 0, n \in \mathbb{N}^*.
$$

2. If $(X_n)_{n \in \mathbb{N}^*}$ is a martingale and if, for all $n \in \mathbb{N}^*$, $X \in L^p(P)$, $p > 1$, then

$$
E\left[\left(\max_{1 \leq k \leq n} |X_k|\right)^p\right] \leq \left(\frac{p}{p-1}\right)^p E[|X_n|^p], \qquad n \in \mathbb{N}^*.
$$

(Points 1 and 2 are called Doob's inequalities.*)*

Proof: 1. For all $k \in \mathbb{N}^*$ we put $A_k = \bigcap_{j=1}^{k-1} \{X_j \leq \lambda\} \cap \{X_k > \lambda\}$ ($\lambda > 0$), where all A_k are pairwise disjoint and $A = \{\max_{1 \leq k \leq n} X_k > \lambda\}$. Thus it is obvious that $A = \bigcup_{k=1}^n A_k$. Because in A_k, X_k is greater than λ, we have

$$\int_{A_k} X_k dP \geq \lambda \int_{A_k} dP.$$

Therefore,

$$\forall k \in \mathbb{N}^*, \qquad \lambda P(A_k) \leq \int_{A_k} X_k dP,$$

resulting in

$$\begin{aligned}\lambda P(A) &= \lambda P\left(\bigcup_{k=1}^{n} A_k\right) = \lambda \sum_{k=1}^{n} P(A_k) \\ &\leq \sum_{k=1}^{n} \int_{A_k} X_k dP = \sum_{k=1}^{n} \int_{\Omega} X_k I_{A_k} dP = \sum_{k=1}^{n} E[X_k I_{A_k}]. \qquad (2.6)\end{aligned}$$

Now, we have

$$\begin{aligned}E[X_n^+] &= \int_{\Omega} X_n^+ dP \\ &\geq \int_A X_n^+ dP = \sum_{k=1}^{n} \int_{A_k} X_n^+ dP = \sum_{k=1}^{n} \int_{\Omega} X_n^+ I_{A_k} dP \\ &= \sum_{k=1}^{n} E[X_n^+ I_{A_k}] = \sum_{k=1}^{n} E[E[X_n^+ I_{A_k} | X_1, \ldots, X_k]] \\ &= \sum_{k=1}^{n} E[I_{A_k} E[X_n^+ | X_1, \ldots, X_k]] \geq \sum_{k=1}^{n} E[I_{A_k} E[X_n | X_1, \ldots, X_k]],\end{aligned}$$

where the last row follows from the fact that I_{A_k} is $\sigma(X_1, \ldots, X_k)$-measurable. Moreover, since $(X_n)_{n \in \mathbb{N}^*}$ is a submartingale we have

$$E[X_n^+] \geq \sum_{k=1}^{n} E[I_{A_k} X_k]. \qquad (2.7)$$

By (2.6) and (2.7), $E[X_n^+] \geq \lambda P(A)$, and this completes the proof of 1. We can also observe that

$$\begin{aligned}\sum_{k=1}^{n} E[I_{A_k} X_n^+] &= \sum_{k=1}^{n} E[E[X_n^+ I_{A_k} | X_1, \ldots, X_k]] \\ &\geq \sum_{k=1}^{n} E[I_{A_k} E[X_n | X_1, \ldots, X_k]] \geq \sum_{k=1}^{n} E[I_{A_k} X_k] \geq \lambda P(A)\end{aligned}$$

and therefore

$$\lambda P\left(\max_{1 \leq k \leq n} X_k > \lambda\right) \leq \sum_{k=1}^{n} E[I_{A_k} X_n^+]. \qquad (2.8)$$

2. Let $(X_n)_{n\in\mathbb{N}^*}$ be a martingale such that $X_n \in L^p(P)$ for all $n \in \mathbb{N}^*$. Since $\phi = |x|$ is a convex function, it follows from Proposition 2.63 that $(|X_n|)_{n\in\mathbb{N}^*}$ is a submartingale. Thus from (2.8) we have

$$\begin{aligned}\lambda P\left(\max_{1\le k\le n}|X_k|>\lambda\right) &\le \sum_{k=1}^{n} E[I_{A_k}|X_n^+|] = \sum_{k=1}^{n} E[I_{A_k}|X_n|]\\ &= \sum_{k=1}^{n}\int_{A_k}|X_n|dP = \int_A |X_n|dP \qquad (\lambda>0, n\in\mathbb{N}^*).\end{aligned}$$

Putting $X = \max_{1\le k\le n}|X_k|$ and $Y = |X_n|$, we obtain

$$\lambda P(X>\lambda) \le \int_A Y dP = \int_{\{X>\lambda\}} Y dP,$$

and from Lemma 2.64 it follows that $\|X\|_p \le q\|Y\|_p$. Thus $E[X^p] \le q^p E[Y^p]$, proving 2. □

Remark 2.66. Because

$$\max_{1\le k\le n}|X_k|^p = \left(\max_{1\le k\le n}|X_k|\right)^p,$$

by point 2 of Proposition 2.65 it is also true that

$$E\left[\max_{1\le k\le n}|X_k|^p\right] \le \left(\frac{p}{p-1}\right)^p E[|X_n|^p].$$

Corollary 2.67. If $(X_n)_{n\in\mathbb{N}^*}$ is a martingale such that $X_n \in L^p(P)$ for all $n \in \mathbb{N}^*$, then

$$P\left(\max_{1\le k\le n}|X_k|>\lambda\right) \le \frac{1}{\lambda^p}E[|X_n|^p], \qquad \lambda>0.$$

Proof: From Proposition 2.63 we can assert that $(|X_n|^p)_{n\in\mathbb{N}^*}$ is a submartingale. In fact, $\phi(x) = |x|^p, p > 1$, is convex. By point 1 of Proposition 2.65, it follows that

$$P\left(\max_{1\le k\le n}|X_k|^p>\lambda^p\right) \le \frac{1}{\lambda^p}E[|X_n|^p],$$

which is equivalent to

$$P\left(\max_{1\le k\le n}|X_k|>\lambda\right) \le \frac{1}{\lambda^p}E[|X_n|^p].$$

□

Lemma 2.68. *The following are true:*

1. *If $(X_t)_{t\in\mathbb{R}_+}$ is a martingale, then so is $(X_t)_{t\in I}$ for all $I \subset \mathbb{R}_+$.*

2. *If, for all* $I \subset \mathbb{R}_+$ *and* I *finite,* $(X_t)_{t\in I}$ *is a (discrete) martingale, then so is* $(X_t)_{t\in\mathbb{R}_+}$.

Proof: 1. Let $I \subset \mathbb{R}_+, (s,t) \in I^2, s < t$. Because $(X_r)_{r\in\mathbb{R}_+}$ is a martingale,

$$X_s = E[X_t|X_r, 0 \le r \le s, r \in \mathbb{R}_+].$$

Observing that

$$\sigma(X_r, 0 \le r \le s, r \in I) \subset \sigma(X_r, 0 \le r \le s, r \in \mathbb{R}_+)$$

and remembering that in general

$$E[X|B_1] = E[E[X|B_2]|B_1], \qquad B_1 \subset B_2 \subset \mathcal{F},$$

we obtain:

$$\begin{aligned} &E[X_t|X_r, 0 \le r \le s, r \in I] \\ &= E[E[X_t|X_r, 0 \le r \le s, r \in \mathbb{R}_+]|X_r, 0 \le r \le s, r \in I] \\ &= E[X_s|X_r, 0 \le r \le s, r \in I] \\ &= X_s. \end{aligned}$$

The last equality holds because X_s is measurable with respect to $\sigma(X_r, 0 \le r \le s, r \in I)$.

2. See, e.g., Doob (1953). □

Proposition 2.69. *Let* $(X_t)_{t\in\mathbb{R}_+}$ *be a stochastic process on* $(\Omega, \mathcal{F}, P)$ *valued in* $\mathbb{R}$.

1. *If* $(X_t)_{t\in\mathbb{R}_+}$ *is a submartingale, then*

$$P\left(\sup_{0\le s\le t} X_s > \lambda\right) \le \frac{1}{\lambda} E[X_t^+], \qquad \lambda > 0, t \ge 0.$$

2. *If* $(X_t)_{t\in\mathbb{R}_+}$ *is a martingale such that, for all* $t \ge 0$, $X_t \in L^p(P)$, $p > 1$, *then*

$$E\left[\sup_{0\le s\le t} |X_s|^p\right] \le \left(\frac{p}{p-1}\right)^p E[|X_t|^p].$$

Proof: See, e.g., Doob (1953). □

Definition 2.70. A subset H of $L^1(\Omega, \mathcal{F}, P)$ is *uniformly integrable* if

$$\lim_{c\to\infty} \sup_{Y\in H} \int_{\{|Y|>c\}} |Y| dP = 0.$$

Theorem 2.71. *A martingale is uniformly integrable if and only if it is of the form* $M_n = E[Y|\mathcal{F}_n]$, *where* $Y \in L^1(\Omega, \mathcal{F}, P)$. *Under these conditions* $\{M_n\}_n$ *converges almost surely and in* L^1.

Proof: See, e.g., Baldi (1984). □

The subsequent proposition specifies the limit of a uniformly integrable martingale.

Proposition 2.72. *Let* $Y \in L^1(\Omega, \mathcal{F}, P)$, $\{\mathcal{F}_n\}_n$ *be a filtration and* $\mathcal{F}_\infty = \bigcup_n \mathcal{F}_n$ *be the* σ*-algebra generated by* $\{\mathcal{F}_n\}_n$*. Then*

$$\lim_{n\to\infty} E[Y|\mathcal{F}_n] = E[Y|\mathcal{F}_\infty] \text{ almost surely and in } L^1.$$

Proof: See, e.g., Baldi (1984). □

Doob–Meyer Decomposition

Proposition 2.73. *Every martingale has a right-continuous version.*

Theorem 2.74. *Let* X_t *be a supermartingale. Then the mapping* $t \to E[X_t]$ *is right-continuous if and only if there exists an RCLL modification of* X_t*. This modification is unique.*

Proof: See, e.g., Protter (1990). □

Definition 2.75. Consider the set $\mathcal{S}$ of stopping times T, with $P(T < \infty) = 1$, of the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. The right-continuous adapted process $(X_t)_{t\in\mathbb{R}_+}$ is said to be of *class D* if the family $(X_T)_{T\in\mathcal{S}}$ is uniformly integrable. Instead, if $\mathcal{S}_a$ is the set of stopping times with $P(T \leq a) = 1$, for a finite $a > 0$, and the family $(X_T)_{T\in\mathcal{S}_a}$ is uniformly integrable, then it is said to be of *class DL*.

Proposition 2.76. *Let* $(X_t)_{t\in\mathbb{R}_+}$ *be a right-continuous submartingale. Then* X_t *is of class DL under either of the following two conditions:*

1. $X_t \geq 0$ *almost surely for every* $t \geq 0$;
2. X_t *has the form*
$$X_t = M_t + A_t, \qquad t \in \mathbb{R}_+,$$
where $(M_t)_{t\in\mathbb{R}_+}$ *is a martingale and* $(A_t)_{t\in\mathbb{R}_+}$ *an adapted increasing process.*

Lemma 2.77. *If* $(X_t)_{t\in\mathbb{R}_+}$ *is a uniformly integrable martingale, then it is of class D.*

Definition 2.78. Let $(X_t)_{t\in\mathbb{R}_+}$ be an adapted stochastic process with RCLL trajectories. It is said to be *decomposable* if it can be written as

$$X_t = X_0 + M_t + Z_t,$$

where $M_0 = Z_0 = 0$, M_t is a locally square-integrable martingale, and Z_t has RCLL-adapted trajectories of bounded variation.

Theorem 2.79. (Doob–Meyer). *Let $(X_t)_{t\in\mathbb{R}_+}$ be an adapted right-continuous process. It is a submartingale of class D, with $X_0 = 0$ almost surely if and only if it can be decomposed as*

$$\forall t \in \mathbb{R}_+, \qquad X_t = M_t + A_t \text{ a.s.},$$

where M_t is a uniformly integrable martingale with $M_0 = 0$ and $A_t \in L^1(P)$ is an increasing predictable process with $A_0 = 0$. The decomposition is unique and if, in addition, X_t is bounded, then M_t is uniformly integrable and A_t integrable.

Definition 2.80. Resorting to the notation of Theorem 2.79, the process $(A_t)_{t\in\mathbb{R}_+}$ is called the *compensator* of X_t.

Proposition 2.81. Under the assumptions of Theorem 2.79, the compensator A_t of X_t is continuous if and only if X_t is regular in the sense that for every predictable finite stopping time T we have that $E[X_T] = E[X_{T-}]$.

Definition 2.82. A stochastic process $(M_t)_{t\in\mathbb{R}_+}$ is a *local martingale* with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if there exists a "localizing" sequence $(T_n)_{n\in\mathbb{N}}$ such that for each $n \in \mathbb{N}$, $(M_{t\wedge T_n})_{t\in\mathbb{R}_+}$ is an $\mathcal{F}_t$-martingale.

Definition 2.83. Let $(X_t)_{t\in\mathbb{R}_+}$ be a stochastic process. *Property* $\mathcal{P}$ is said to hold locally if

1. there exists $(T_n)_{n\in\mathbb{N}}$, a sequence of stopping times, with $T_n < T_{n+1}$,
2. $\lim_n T_n = +\infty$ almost surely,

such that $X_{T_n} I_{\{T_n>0\}}$ has property $\mathcal{P}$ for $n \in \mathbb{N}^*$.

Theorem 2.84. *Let $(M_t)_{t\in\mathbb{R}_+}$ be an adapted and RCLL stochastic process and let $(T_n)_{n\in\mathbb{N}}$ be as in the preceding definition. If $M_{T_n} I_{\{T_n>0\}}$ is a martingale for each $n \in \mathbb{N}^*$, then M_t is a local martingale.*

Remark 2.85. Any RCLL martingale is a local martingale. (Choose $T_n = n$ for all $n \in \mathbb{N}^*$.)

Lemma 2.86. *Any martingale is a local martingale.*

Proof: Simply take $T_n = t$. □

Theorem 2.87. (local form Doob–Meyer). *Let $(X_t)_{t\in\mathbb{R}_+}$ be a nonnegative right-continuous $\mathcal{F}_t$-local submartingale with localizing sequence $(T_n)_{n\in\mathbb{N}}$ and $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ a right-continuous filtration. Then there exists a unique increasing right-continuous predictable process $(A_t)_{t\in\mathbb{R}_+}$ such that $A_0 = 0$ almost surely and $P(A_t < \infty) = 1$ for all $t > 0$, so that $X_t - A_t$ is a right-continuous local martingale.*

Theorem 2.88. *Let $(M_t)_{t\in\mathbb{R}_+}$ be a martingale with $M_0 = 0$ almost surely and bounded in L^2. Then M_t^2 is a submartingale and by Doob's L^2 inequality (Proposition 2.69)*

$$E\left[\sup_t M_t^2\right] \leq 4E[M_\infty^2] < \infty.$$

Therefore, M_t^2 is a submartingale of class D and there exists a unique integrable predictable increasing process $\langle M_t\rangle$, with $\langle M_0\rangle = 0$, such that $M_t^2 - \langle M_t\rangle$ is a uniformly integrable martingale.

Definition 2.89. For two L^2 martingales M and N, the process

$$\langle M, N\rangle = \frac{1}{4}(\langle M+N\rangle - \langle M-N\rangle)$$

is called the *predictable covariation* of M and N. Furthermore, if $\langle M, N\rangle = 0$, then the two martingales are said to be *orthogonal*.

Remark 2.90. Evidently $\langle M, M\rangle = \langle M\rangle$, and by the latter's martingale assumption we have that $\langle M, N\rangle$ is the unique finite variation predictable RCLL process such that $\langle M, N\rangle_0 = 0$ and $MN - \langle M, N\rangle$ is a martingale. Thus M and N are orthogonal if and only if MN is a martingale.

2.7 Markov Processes

Definition 2.91. Let $(X_t)_{t\in\mathbb{R}_+}$ be a stochastic process on a probability space, valued in $(E, \mathcal{B})$ and adapted to the increasing family $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ of σ-algebras of subsets of $\mathcal{F}$. $(X_t)_{t\in\mathbb{R}_+}$ is a *Markov process* with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if the following condition is satisfied:

$$\forall B \in \mathcal{B}, \forall (s,t) \in \mathbb{R}_+ \times \mathbb{R}_+, s < t: \qquad P(X_t \in B|\mathcal{F}_s) = P(X_t \in B|X_s) \text{ a.s.} \tag{2.9}$$

Remark 2.92. If, for all $t \in \mathbb{R}_+$, $\mathcal{F}_t = \sigma(X_r, 0 \leq r \leq t)$, then the condition (2.9) becomes

$$P(X_t \in B|X_r, 0 \leq r \leq s) = P(X_t \in B|X_s) \text{ a.s.}$$

for all $B \in \mathcal{B}$, for all $(s,t) \in \mathbb{R}_+ \times \mathbb{R}_+$, and $s < t$.

Proposition 2.93. *Under the assumptions of Definition 2.91, the following two statements are equivalent:*

1. *for all $B \in \mathcal{B}$ and all $(s,t) \in \mathbb{R}_+ \times \mathbb{R}_+, s < t$: $P(X_t \in B|\mathcal{F}_s) = P(X_t \in B|X_s)$ almost surely;*
2. *for all $g : E \to \mathbb{R}$, $\mathcal{B}$-$\mathcal{B}_{\mathbb{R}}$-measurable such that $g(X_t) \in L^1(P)$ for all t, for all $(s,t) \in \mathbb{R}_+^2, s < t$: $E[g(X_t)|\mathcal{F}_s] = E[g(X_t)|X_s]$ almost surely.*

Proof: The proof is left to the reader as an exercise. □

Lemma 2.94. *If $(Y_k)_{k\in\mathbb{N}^*}$ is a sequence of real, independent random variables, then, putting*

$$X_n = \sum_{k=1}^{n} Y_k \qquad \forall n \in \mathbb{N}^*,$$

the new sequence $(X_n)_{n\in\mathbb{N}^*}$ is Markovian with respect to the family of σ-algebras $(\sigma(Y_1,\ldots,Y_n))_{n\in\mathbb{N}^*}$.

Proof: From the definition of X_k it is obvious that

$$\sigma(Y_1,\ldots,Y_n) = \sigma(X_1,\ldots,X_n) \qquad \forall n \in \mathbb{N}^*.$$

We thus first prove that, for all $C, D \in \mathcal{B}_{\mathbb{R}}$, for all $n \in \mathbb{N}^*$:

$$\begin{aligned} &P(X_{n-1} \in C, Y_n \in D|Y_1,\ldots,Y_{n-1}) \\ &= P(X_{n-1} \in C, Y_n \in D|X_{n-1}) \qquad \text{a.s.} \end{aligned} \tag{2.10}$$

To do this we fix $C, D \in \mathcal{B}_{\mathbb{R}}$ and $n \in \mathbb{N}^*$ and separately look at the left- and right-hand sides of (2.10). We get

$$\begin{aligned} &P(X_{n-1} \in C, Y_n \in D|Y_1,\ldots,Y_{n-1}) = E[I_C(X_{n-1})I_D(Y_n)|Y_1,\ldots,Y_{n-1}] \\ &= I_C(X_{n-1})E[I_D(Y_n)|Y_1,\ldots,Y_{n-1}] = I_C(X_{n-1})E[I_D(Y_n)] \text{ a.s.}, \end{aligned} \tag{2.11}$$

where the second equality of (2.11) holds because $I_C(X_{n-1})$ is $\sigma(Y_1,\ldots,Y_{n-1})$-measurable, and for the last one we use the fact that $I_D(Y_n)$ is independent of $Y_1,\ldots,Y_{n-1}$. On the other hand, we obtain that

$$\begin{aligned} P(X_{n-1} \in C, Y_n \in D|X_{n-1}) &= E[I_C(X_{n-1})I_D(Y_n)|X_{n-1}] \\ &= I_C(X_{n-1})E[I_D(Y_n)] \text{ a.s.} \end{aligned} \tag{2.12}$$

In fact, $I_C(X_{n-1})$ is $\sigma(X_{n-1})$-measurable and $I_D(Y_n)$ is independent of $X_{n-1} = \sum_{k=1}^{n-1} Y_k$. For (2.11) and (2.12), equation (2.10) follows and hence

$$\begin{aligned} &P((X_{n-1}, Y_n) \in C \times D|Y_1,\ldots,Y_{n-1}) \\ &= P((X_{n-1}, Y_n) \in C \times D|X_{n-1}) \qquad \text{a.s.} \end{aligned} \tag{2.13}$$

As (2.13) holds for the rectangles of $\mathcal{B}_{\mathbb{R}^2}$ $(= \mathcal{B}_{\mathbb{R}} \otimes \mathcal{B}_{\mathbb{R}})$, by the measure extension theorem (see, e.g., Bauer (1981)), it follows that (2.13) is also true for every $B \in \mathcal{B}_{\mathbb{R}^2}$. If now $A \in \mathcal{B}_{\mathbb{R}}$, then the two events

$$\{X_{n-1} + Y_n \in A\} = \{(X_{n-1}, Y_n) \in B\},$$

where $B \in \mathcal{B}_{\mathbb{R}^2}$ is the inverse image of A for a generic mapping $+ : \mathbb{R}^2 \to \mathbb{R}$ (which is continuous and hence measurable), are identical. Applying (2.13) to B, we obtain

$$P(X_{n-1} + Y_n \in A|Y_1, \ldots, Y_{n-1}) = P(X_{n-1} + Y_n \in A|X_{n-1}) \text{ a.s.},$$

and thus

$$P(X_{n-1} + Y_n \in A|X_1, \ldots, X_{n-1}) = P(X_{n-1} + Y_n \in A|X_{n-1}) \text{ a.s.},$$

and then

$$P(X_n \in A|X_1, \ldots, X_{n-1}) = P(X_n \in A|X_{n-1}) \text{ a.s.}$$

Therefore, $(X_n)_{n\in\mathbb{N}^*}$ is Markovian with respect to $(\sigma(X_1, \ldots, X_n))_{n\in\mathbb{N}^*}$ or, equivalently, with respect to $(\sigma(Y_1, \ldots, Y_n))_{n\in\mathbb{N}^*}$. □

Proposition 2.95. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a real stochastic process defined on the probability space $(\Omega, \mathcal{F}, P)$. The following two statements are true:*

1. *If $(X_t)_{t\in\mathbb{R}_+}$ is a Markov process, then so is $(X_t)_{t\in J}$ for all $J \subset \mathbb{R}_+$.*
2. *If for all $J \subset \mathbb{R}_+$, J finite: $(X_t)_{t\in J}$ is a Markov process, and then so is $(X_t)_{t\in\mathbb{R}_+}$.*

Proof: See, e.g., Ash and Gardner (1975). □

Theorem 2.96. *Every real stochastic process $(X_t)_{t\in\mathbb{R}_+}$ with independent increments is a Markov process.*

Proof: We define $(t_1, \ldots, t_n) \in \mathbb{R}_+^n$ such that $0 < t_1 < \cdots < t_n$ and $t_0 = 0$. If, for simplicity, we further suppose that $X_0 = 0$, then $X_{t_n} = \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})$. Putting $Y_k = X_{t_k} - X_{t_{k-1}}$, then, for all $k = 1, \ldots, n$, the Y_k are independent (because the process $(X_t)_{t\in\mathbb{R}_+}$ has independent increments) and we have that

$$X_{t_n} = \sum_{k=1}^n Y_k.$$

From Lemma 2.94 we can assert that

$$\forall B \in \mathcal{B}_\mathbb{R}: \qquad P(X_{t_n} \in B|X_{t_1}, \ldots, X_{t_{n-1}}) = P(X_{t_n} \in B|X_{t_{n-1}}) \text{ a.s.}$$

Thus $\forall J \subset \mathbb{R}_+, J$ finite, $(X_t)_{t\in J}$ is Markovian. The theorem then follows by point 2 of Proposition 2.95. □

Definition 2.97. Let $p(s, x, t, A)$ be a nonnegative function defined for $0 \leq s < t < \infty, x \in \mathbb{R}, A \in \mathcal{B}_\mathbb{R}$. Then p is a *Markov transition probability function* if

1. for all $0 \leq s < t < \infty$, for all $A \in \mathcal{B}_\mathbb{R}$, $p(s, \cdot, t, A)$ is $\mathcal{B}_\mathbb{R}$-measurable;
2. for all $0 \leq s < t < \infty$, for all $x \in \mathbb{R}$, $p(s, x, t, \cdot)$ is a probability measure on $\mathcal{B}_\mathbb{R}$;
3. p satisfies the Chapman–Kolmogorov equation:

$$p(s, x, t, A) = \int_\mathbb{R} p(s, x, r, dy)p(r, y, t, A) \qquad \forall x \in \mathbb{R}, s < r < t.$$

Definition 2.98. If $(X_t)_{t\in[t_0,T]}$ is a Markov process, then the distribution P_0 of $X(t_0)$ is the *initial distribution* of the process.

Theorem 2.99. *Let E be a Polish space endowed with the σ-algebra $\mathcal{B}_E$ of its Borel sets, P_0 a probability measure on $\mathcal{B}_E$, and $p(r,x,s,A), t_0 \leq r < s \leq T, x \in E, A \in \mathcal{B}_E$ a Markov transition probability function. Then there exists a unique (in the sense of equivalence) Markov process $(X_t)_{t\in[t_0,T]}$ valued in E, with P_0 as its initial distribution and p as its transition probability.*

Proof: See, e.g., Ash and Gardner (1975) or Dynkin (1965). □

Remark 2.100. From Theorem 2.99 we can deduce that

$$p(s,x,t,A) = P(X_t \in A|X_s = x), \qquad 0 \leq s < t < \infty, x \in \mathbb{R}, A \in \mathcal{B}_{\mathbb{R}}.$$

Definition 2.101. A Markov process $(X_t)_{t\in[t_0,T]}$ is said to be *homogeneous* if the transition probability functions $p(s,x,t,A)$ depend on t and s only through their difference $t-s$. Therefore, for all $(s,t) \in [t_0,T]^2$, $s<t$, for all $u \in [0, T-t]$, for all $A \in \mathcal{B}_{\mathbb{R}}$, and for all $x \in \mathbb{R}$:

$$p(s,x,t,A) = p(s+u,x,t+u,A) \qquad \text{a.s.}$$

Remark 2.102. If $(X_t)_{t\in[t_0,T]}$ is a homogeneous Markov process with transition probability function p, then, for all $(s,t) \in [t_0,T]^2, s<t$, for all $A \in \mathcal{B}_{\mathbb{R}}$, and for all $x \in \mathbb{R}$, we obtain

$$p(t_0,x,t_0+t-s,A) = p(s,x,t,A) \text{ a.s.},$$

where $p(t_0,x,t_0+t-s,A)$ is denoted by $p(\bar{t},x,A)$, with $\bar{t} = (t-s) \in [0, T - t_0], x \in \mathbb{R}, A \in \mathcal{B}_{\mathbb{R}}$.

Semigroups Associated with Markov Transition Probability Functions

Let $BC(\mathbb{R})$ be the space of all continuous and bounded functions on $\mathbb{R}$, endowed with the norm $\|f\| = \sup_{x\in\mathbb{R}} |f(x)| (< \infty)$, and let $p(s,x,t,A)$ be a transition probability function ($0 \leq s < t \leq T, x \in \mathbb{R}, A \in \mathcal{B}_{\mathbb{R}}$). We consider the operator

$$T_{s,t} : BC(\mathbb{R}) \to BC(\mathbb{R}), \qquad 0 \leq s < t \leq T,$$

defined by assigning, for all $f \in BC(\mathbb{R})$,

$$(T_{s,t}f)(x) = \int_{\mathbb{R}} f(y)p(s,x,t,dy).$$

If $s = t$, then

$$p(s, x, s, A) = \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ if } x \notin A. \end{cases}$$

Therefore,

$$T_{t,t} = I \text{ (identity).} \tag{2.14}$$

Moreover, we have that

$$T_{s,t}T_{t,u} = T_{s,u}, \qquad 0 \leq s < t < u. \tag{2.15}$$

In fact, if $f \in BC(\mathbb{R})$ and $x \in \mathbb{R}$,

$$\begin{aligned}
&(T_{s,t}(T_{t,u}f))(x) \\
&= \int_{\mathbb{R}} (T_{t,u}f)(y)p(s, x, t, dy) \\
&= \int\int_{\mathbb{R}^2} f(z)p(t, y, u, dz)p(s, x, t, dy) \\
&= \int_{\mathbb{R}} f(z) \int_{\mathbb{R}} p(t, y, u, dz)p(s, x, t, dy) \text{ (by Fubini's theorem)} \\
&= \int_{\mathbb{R}} f(z)p(s, x, u, dz) \text{ (by the Chapman–Kolmogorov equation)} \\
&= (T_{s,u}f)(x).
\end{aligned}$$

Definition 2.103. The family $\{T_{s,t}\}_{0 \leq s \leq t \leq T}$ is a *semigroup associated with the transition probability function* $p(s, x, t, A)$ (or with its corresponding Markov process).

Definition 2.104. If $(X_t)_{t \in \mathbb{R}_+}$ is a Markov process with transition probability function p and associated semigroup $\{T_{s,t}\}$, then the operator

$$\mathcal{A}_s f = \lim_{h \downarrow 0} \frac{T_{s,s+h}f - f}{h}, \qquad s \geq 0, f \in BC(\mathbb{R})$$

is called the *infinitesimal generator of the Markov process* $(X_t)_{t \geq 0}$. Its domain $\mathcal{D}_{\mathcal{A}_s}$ consists of all $f \in BC(\mathbb{R})$ for which the above limit exists uniformly (and therefore in the norm of $BC(\mathbb{R})$) (see, e.g., Feller (1971)).

Remark 2.105. From the preceding definition we observe that

$$(\mathcal{A}_s f)(x) = \lim_{h \downarrow 0} \frac{1}{h} \int_{\mathbb{R}} [f(y) - f(x)]p(s, x, s + h, dy).$$

Definition 2.106. Let $(X_t)_{t \in \mathbb{R}_+}$ be a Markov process with transition probability function $p(s, x, t, A)$, and $\{T_{s,t}\}$ $(s, t \in \mathbb{R}_+, s \leq t)$ its associated semigroup. If, for all $f \in BC(\mathbb{R})$, the function

$$(t, x) \in \mathbb{R}_+ \times \mathbb{R} \to (T_{t,t+\lambda}f)(x) = \int_{\mathbb{R}} p(t, x, t + \lambda, dy)f(y) \in \mathbb{R}$$

is continuous for all $\lambda > 0$, then we say that the process satisfies the *Feller property*.

Theorem 2.107. *If $(X_t)_{t\in\mathbb{R}_+}$ is a Markov process with right-continuous trajectories satisfying the Feller property, then, for all $t \in \mathbb{R}_+$, $\mathcal{F}_t = \mathcal{F}_{t^+}$, where $\mathcal{F}_{t^+} = \bigcap_{t'>t} \sigma(X(s), 0 \leq s \leq t')$, and the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is right-continuous.*

Proof: See, e.g., Friedman (1975). □

Remark 2.108. It can be shown that $\mathcal{F}_{t^+}$ is a σ-algebra.

Example 2.109. Examples of processes with the Feller property, or simply *Feller processes*, include Wiener processes (Brownian motions), Poisson processes, and all Lévy processes (see later sections).

Examples of Stopping Times

Let $(\mathbf{X}_t)_{t\in\mathbb{R}_+}$ be a continuous Markov process taking values in $\mathbb{R}^v$ and suppose that the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, generated by the process, is right-continuous. Let $B \in \mathcal{B}_{\mathbb{R}^v} \setminus \{\emptyset\}$, and we define $T : \Omega \to \bar{\mathbb{R}}_+$ as:

$$\forall \omega \in \Omega, \qquad T(\omega) = \begin{cases} \inf\{t \geq 0 | \mathbf{X}(t,\omega) \in B\} & \text{if the set is} \neq \emptyset, \\ +\infty & \text{if the set is} = \emptyset. \end{cases}$$

This gives rise to the following theorem.

Theorem 2.110. *If B is an open or closed subset of $\mathbb{R}^v$, then T is a stopping time.*

Proof: For B open, let $t \in \mathbb{R}_+$. In this case it can be shown that

$$\{T < t\} = \bigcup_{r<t, r\in\mathbb{Q}^+} \{\omega | \mathbf{X}(r,\omega) \in B\}.$$

Since $\mathbf{X}(r)$ is $\mathcal{F}$-measurable,

$$\{\omega | \mathbf{X}(r,\omega) \in B\} \in \mathcal{F}_r \subset \mathcal{F}_t \qquad \forall r < t, r \in \mathbb{Q}^+,$$

and therefore the (countable) union of such events will be an element of $\mathcal{F}_t$ as well, and thus $\{T < t\} \in \mathcal{F}_t$. Now, fixing $\delta > 0$ and $N \in \mathbb{N}$ such that $\delta > \frac{1}{N}$, we have that

$$\forall n \in \mathbb{N}, n \geq N\colon \qquad \left\{T < t + \frac{1}{n}\right\} \in \mathcal{F}_{t+\delta}.$$

Hence

$$\{T \leq t\} = \bigcap_{n=N}^{\infty} \left\{T < t + \frac{1}{n}\right\} \in \mathcal{F}_{t+\delta}$$

and, due to the arbitrary choice of δ, this results in

$$\{T \leq t\} \in \bigcap_{\delta>0} \mathcal{F}_{t+\delta} = \mathcal{F}_t^+ = \mathcal{F}_t.$$

For B closed, for all $n \in \mathbb{N}$, we define $V_n = \{\mathbf{x} \in \mathbb{R}^v | d(\mathbf{x}, B) < \frac{1}{n}\}$ and

$$T_n = \begin{cases} \inf\{t \geq 0 | \mathbf{X}(t,\omega) \in V_n\} & \text{if the set is } \neq \emptyset, \\ +\infty & \text{if the set is } = \emptyset. \end{cases}$$

It can be shown that $B = \bigcap_{n\in\mathbb{N}} V_n$, $\{T \leq t\} = \bigcap_{n\in\mathbb{N}}\{T_n < t\}$, and, since (with V_n open) $\{T_n < t\} \in \mathcal{F}_{t^+}$, we finally get that $\{T \leq t\} \in \mathcal{F}_{t^+} = \mathcal{F}_t$. □

Definition 2.111. The stopping time T is the *first hitting time* of B or, equivalently, the *first exit time* from $\mathbb{R}^v \setminus B$.

Definition 2.112. A Markov process $(X_t)_{t\in\mathbb{R}_+}$ with transition probability function $p(s,x,t,A)$ is said to have the *strong Markov property* if, for all $A \in \mathcal{B}_\mathbb{R}$,

$$P(X(T+t) \in A|\mathcal{F}_t) = p(T, X(T), T+t, A) \qquad \text{a.s.} \tag{2.16}$$

Remark 2.113. Equation (2.16) is formally analogous to the Markov property

$$P(X(t) \in A|\mathcal{F}_s) = p(s, X(s), t, A) \text{ for } s < t,$$

with which it coincides when $T = k$ (constant).

Proposition 2.114. *Equation (2.16) is equivalent to the assertion that for all* $f : \mathbb{R} \to \mathbb{R}$, *measurable, bounded,*

$$E[f(X(T+t))|\mathcal{F}_T] = E[f(X(T+t))|X(T)] \qquad a.s. \tag{2.17}$$

Proof: See, e.g., Ash and Gardner (1975). □

Remark 2.115. By Proposition 2.42 and Theorem 2.48, if $(X_t)_{t\in\mathbb{R}_+}$ is right-continuous and if T is a finite stopping time of the process, then $X(T)$ is $\mathcal{F}_T$-measurable.

Lemma 2.116. *Every Markov process* $(X_t)_{t\in\mathbb{R}_+}$ *that satisfies the Feller property has the strong Markov property, at least for a discrete stopping time* T.

Proof: Let T be a discrete stopping time of the process $(X_t)_{t\in\mathbb{R}_+}$ and $\{t_j\}_{j\in\mathbb{N}}$ its codomain. Fixing a $j \in \mathbb{N}$ we have $\{T \leq t_j\} \in \mathcal{F}_{t_j}$ and $\{T < t_j\} = \bigcup_{t_l<t_j}\{T \leq t_l\} \in \mathcal{F}_{t_j}$. Therefore,

$$G_j \equiv \{T = t_j\} = \{T \leq t_j\} \setminus \{T < t_j\} \in \mathcal{F}_{t_j}$$

and

$$\forall t \in \mathbb{R}_+, \qquad G_j \cap \{T \leq t\} = \begin{cases} \emptyset & \text{for } t_j > t, \\ G_j & \text{for } t \geq t_j. \end{cases}$$

From this we obtain, for all $t \in \mathbb{R}_+$, $G_j \cap \{T \leq t\} \in \mathcal{F}_t$, that is, $G_j \in \mathcal{F}_T$. Proving equation (2.16) is equivalent to showing that if $t \in \mathbb{R}_+, A \in \mathcal{B}_\mathbb{R}$, then:

1. $p(T, X(T), T+t, A)$ is $\mathcal{F}_T$-measurable;
2. for all $E \in \mathcal{F}_T$, $P((X(T+t) \in A) \cap E) = \int_E p(T, X(T), T+t, A) dP$.

Before proving point 1, we will show that 2 holds. Let $E \in \mathcal{F}_T$; then, by $\Omega = \bigcup_{j \in \mathbb{N}} G_j$, it follows that

$$\begin{aligned} P((X(T+t) \in A) \cap E) &= \sum_{j \in \mathbb{N}} P((X(T+t) \in A) \cap E \cap G_j) \\ &= \sum_{j \in \mathbb{N}} P((X(T+t) \in A) \cap E \cap (T = t_j)) \\ &= \sum_{j \in \mathbb{N}} P((X(t+t_j) \in A) \cap E \cap (T = t_j)) \\ &= \sum_{j \in \mathbb{N}} P((X(t+t_j) \in A) \cap E \cap G_j). \end{aligned} \tag{2.18}$$

But

$$E \cap G_j = E \cap (\{T \leq t_j\} \setminus \{T < t_j\}) \in \mathcal{F}_{t_j}$$

(in fact, $E \cap \{T \leq t_j\} \in \mathcal{F}_{t_j}$ following point 4 of Theorem 2.47), and therefore

$$P((X(t+t_j) \in A) \cap E \cap G_j) = \int_{E \cap G_j} P(X(t+t_j) \in A | \mathcal{F}_{t_j}) dP.$$

Moreover, by the Markov property,

$$P(X(t+t_j) \in A | \mathcal{F}_{t_j}) = p(t_j, X(t_j), t_j + t, A) \text{ a.s.} \tag{2.19}$$

Using (2.18) and (2.19), we obtain

$$\begin{aligned} P((X(T+t) \in A) \cap E) &= \bigcup_{j \in \mathbb{N}} \int_{E \cap G_j} p(t_j, X(t_j), t_j + t, A) dP \\ &= \bigcup_{j \in \mathbb{N}} \int_{E \cap \{T = t_j\}} p(t_j, X(t_j), t_j + t, A) dP \\ &= \bigcup_{j \in \mathbb{N}} \int_{E \cap \{T = t_j\}} p(T, X(T), T + t, A) dP \\ &= \int_E p(T, X(T), T + t, A) dP. \end{aligned}$$

For the proof of 1, we now observe that, by the Feller property, the mapping

$$(r, z) \in \mathbb{R}_+ \times \mathbb{R} \to \int_{\mathbb{R}} p(r, z, r+t, dy) f(y) \in \mathbb{R}$$

is continuous (for $f \in BC(\mathbb{R})$). Furthermore, T and $X(T)$ are $\mathcal{F}_T$-measurable, and therefore the mapping

$$\omega \in \Omega \to (T(\omega), X(T(\omega), \omega))$$

is $\mathcal{F}_T$-measurable. Hence the composite of the two mappings

$$\omega \in \Omega \to \int_{\mathbb{R}} p(T, X(T), T+t, dy) f(y) \in \mathbb{R}$$

is $\mathcal{F}_T$-measurable (for $f \in BC(\mathbb{R})$). Now let $(f_m)_{m\in\mathbb{N}} \in (BC(\mathbb{R}))^{\mathbb{N}}$ be a sequence of uniformly bounded functions such that $\lim_{m\to\infty} f_m = I_A$. Then, from our previous observations,

$$\forall m \in \mathbb{N}, \qquad \int_{\mathbb{R}} p(T, X(T), T+t, dy) f_m(y)$$

is $\mathcal{F}_T$-measurable and, following Lebesgue's theorem on integral limits, we get

$$\int_{\mathbb{R}} p(T, X(T), T+t, dy) I_A(y) = \lim_{m\to\infty} \int_{\mathbb{R}} p(T, X(T), T+t, dy) f_m(y),$$

and thus

$$p(T, X(T), T+t, A) = \int_{\mathbb{R}} p(T, X(T), T+t, dy) I_A(y)$$

is $\mathcal{F}_T$-measurable. □

Before generalizing Lemma 2.116, we assert the following.

Lemma 2.117. *If T is a stopping time of the stochastic process $(X_t)_{t\in\mathbb{R}_+}$, then there exists a sequence of stopping times $(T_n)_{n\in\mathbb{N}}$ such that:*

1. *for all $n \in \mathbb{N}, T_n$ has a codomain that is at most countable;*
2. *for all $n \in \mathbb{N}, T_n \geq T$;*
3. *$T_n \downarrow T$ almost surely for $n \to \infty$.*

Moreover, $\{T_n = \infty\} = \{T = \infty\}$ for every n.

Proof: See, e.g., Friedman (1975). □

Theorem 2.118. *If $(X_t)_{t\in\mathbb{R}_+}$ is a right-continuous Markov process that satisfies the Feller property, then it satisfies the strong Markov property.*

Proof: Let T be a finite stopping time of the process $(X_t)_{t\in\mathbb{R}_+}$ and $(T_n)_{n\in\mathbb{N}}$ a sequence of stopping times satisfying properties 1, 2, and 3 of Lemma 2.117 with respect to T. We observe that, for all $n \in \mathbb{N}$, $\mathcal{F}_T \subset \mathcal{F}_{T_n}$. In fact, if $A \in \mathcal{F}_T$, then

$$\forall t \in \mathbb{R}_+, \qquad A \cap \{T_n \leq t\} = (A \cap \{T \leq t\}) \cap \{T_n \leq t\} \in \mathcal{F}_t,$$

provided that $A \cap \{T \leq t\} \in \mathcal{F}_t, \{T_n \leq t\} \in \mathcal{F}_t$. Just like for Lemma 2.116, we will need to show that the points 1 and 2 of its proof hold in this present case. Following Proposition 2.114, point 2 is equivalent to asserting that for all $E \in \mathcal{F}_T$ and all $f \in BC(\mathbb{R})$:

$$\int_E f(X(T+t))dP = \int_E dP \int_{\mathbb{R}} p(T, X(T), T+t, dy) f(y). \tag{2.20}$$

Then, by Proposition 2.42, for all $n \in \mathbb{N}$, we have that for all $E \in \mathcal{F}_{T_n}$ and all $f \in BC(\mathbb{R})$

$$\int_E f(X(T_n+t))dP = \int_E dP \int_{\mathbb{R}} p(T_n, X(T_n), T_n+t, dy) f(y).$$

Moreover, since $T_n \downarrow T$ for $n \to \infty$ and by the right-continuity of the process X, it follows that

$$X(T_n) \to X(T) \text{ for } n \to \infty.$$

From the continuity[5] of the mapping

$$(\lambda, x) \in \mathbb{R}_+ \times \mathbb{R} \to \int_{\mathbb{R}} p(\lambda, x, \lambda+t, \lambda y) f(y) \text{ for } f \in BC(\mathbb{R}),$$

we have that, for $n \to \infty$:

$$\int_{\mathbb{R}} p(T_n, X(T_n), T_n+t, dy) f(y) \to \int_{\mathbb{R}} p(T, X(T), T+t, dy) f(y). \tag{2.21}$$

On the other hand, if f is continuous, then we also get

$$f(X(T_n+t)) \to f(X(T+t)) \text{ for } n \to \infty. \tag{2.22}$$

Therefore, if $E \in \mathcal{F}_T$ and $f \in BC(\mathbb{R})$, then $E \in \mathcal{F}_{T_n}$ for all n, and we have

$$\lim_{n\to\infty} \int_E f(X(T_n+t))dP = \lim_{n\to\infty} \int_E dP \int_{\mathbb{R}} p(T_n, X(T_n), T_n+t, dy) f(y).$$

Since f and p are bounded, following Lebesgue's theorem, we can take the limit of the integral and then (2.20) follows from (2.21) and (2.22). The proof of point 1 is entirely analogous to the proof of Lemma 2.117. □

The above results may be extended to more general, possibly uncountable state spaces. In particular, we will assume that E is a subset of $\mathbb{R}^d$ for $d \in \mathbb{N}^*$. If we consider the time-homogeneous case, a Markov process $(X_t)_{t\in\mathbb{R}_+}$ on $(E, \mathcal{B}_E)$, will be defined in terms of a transition kernel $p(t, x, B)$ for $t \in \mathbb{R}_+$, $x \in E$, $B \in \mathcal{B}_E$, such that

$$p(h, X_t, B) = P(X_{t+h} \in B | \mathcal{F}_t) \qquad \forall t, h \in \mathbb{R}_+, B \in \mathcal{B}_E,$$

given that $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is the natural filtration of the process. Equivalently, if we denote by $BC(E)$ the space of all continuous and bounded functions on E, endowed with the *sup norm*,

[5] By the Feller property.

$$E[g(X_{t+h})|\mathcal{F}_t] = \int_E g(y)p(h, X_t, dy) \qquad \forall t, h \in \mathbb{R}_+, g \in BC(E).$$

In this case the transition semigroup of the process is a one-parameter contraction semigroup $(T(t), t \in \mathbb{R}_+)$ on $BC(E)$ defined by

$$T(t)g(x) := \int_E g(y)p(t, x, dy) = E[g(X_t)|X_0 = x], \qquad x \in E,$$

for any $g \in BC(E)$. The infinitesimal generator will be time independent. It is defined as

$$\mathcal{A}g = \lim_{t \to 0+} \frac{1}{t}(T(t)g - g)$$

for $g \in \mathcal{D}(\mathcal{A})$, the subset of $BC(E)$ for which the above limit exists, in $BC(E)$, with respect to the sup norm. Given the above definitions, it is obvious that for all $g \in \mathcal{D}(\mathcal{A})$,

$$\mathcal{A}g(x) = \lim_{t \to 0+} \frac{1}{t} E[g(X_t)|X_0 = x], \qquad x \in E.$$

If $(T(t), t \in \mathbb{R}_+)$ is the contraction semigroup associated with a Markov process, it is not difficult to show that the mapping $t \to T(t)g$ is right-continuous in $t \in \mathbb{R}_+$ provided that $g \in BC(E)$ is such that the mapping $t \to T(t)g$ is right continuous in $t = 0$. Then, for all $g \in \mathcal{D}(\mathcal{A})$ and $t \in \mathbb{R}_+$,

$$\int_0^t T(s)gds \in \mathcal{D}(\mathcal{A})$$

and

$$T(t)g - g = \mathcal{A} \int_0^t T(s)gds = \int_0^t \mathcal{A}T(s)gds = \int_0^t T(s)\mathcal{A}gds$$

by considering Riemann integrals. The following, so-called *Dynkin's formula*, establishes a fundamental link between Markov processes and martingales (see Rogers and Williams (1994), page 253).

Theorem 2.119. *Assume $(X_t)_{t \in \mathbb{R}_+}$ is a Markov process on $(E, \mathcal{B}_E)$, with transition kernel $p(t, x, B)$, $t \in \mathbb{R}_+$, $x \in E$, $B \in \mathcal{B}_E$. Let $(T(t), t \in \mathbb{R}_+)$ denote its transition semigroup and $\mathcal{A}$ its infinitesimal generator. Then, for any $g \in \mathcal{D}(\mathcal{A})$, the stochastic process*

$$M(t) := g(X_t) - g(X_0) - \int_0^t \mathcal{A}g(X_s)ds$$

is an $\mathcal{F}_t$-martingale.

Proof: The following equations hold:

$$
\begin{aligned}
&E\left[M(t+h)|\mathcal{F}_t\right]+g(X_0)\\
&=E\left[g(X_{t+h})-\int_t^{t+h}\mathcal{A}g(X_s)ds\middle|\mathcal{F}_t\right]-\int_0^t\mathcal{A}g(X_s)ds\\
&=\int_E g(y)p(h,X_t,dy)-\int_t^{t+h}\int_E\mathcal{A}g(y)p(s-t,X_t,dy)-\int_0^t\mathcal{A}g(X_s)ds\\
&=T(h)g(X_t)-\int_0^h T(s)\mathcal{A}g(X_t)ds-\int_0^t\mathcal{A}g(X_s)ds\\
&=g(X_t)-\int_0^t\mathcal{A}g(X_s)ds=M(t)+g(X_0).
\end{aligned}
$$

□

The next proposition shows that a Markov process is indeed characterized by its infinitesimal generator via a martingale problem (see, e.g., Rogers and Williams (1994), page 253).

Theorem 2.120. (martingale problem for Markov processes). *If an RCLL Markov process $(X_t)_{t\in\mathbb{R}_+}$ is such that*

$$
g(X_t)-g(X_0)-\int_0^t\mathcal{A}g(X_s)ds
$$

is an $\mathcal{F}_t$-martingale for any function $g\in\mathcal{D}(\mathcal{A})$, where $\mathcal{A}$ is the infinitesimal generator of a contraction semigroup on E, then X_t is equivalent to a Markov process having $\mathcal{A}$ as its infinitesimal generator.

Example 2.121. A Poisson process (see the following section for more details) is an integer-valued Markov process $(N_t)_{t\in\mathbb{R}_+}$. If its intensity parameter is $\lambda>0$, the process $(X_t)_{t\in\mathbb{R}_+}$, defined by $X_t=N_t-\lambda t$, is a stationary Markov process with independent increments. The transition kernel of X_t is

$$
p(h,x,B)=\sum_{k=0}^{\infty}\frac{(\lambda h)^k}{k!}e^{-\lambda h}I_{\{x+h-\lambda h\in B\}}\text{ for }x\in\mathbb{N},h\in\mathbb{R}_+,B\subset\mathbb{N}.
$$

Its transition semigroup is then

$$
T(h)g(x)=\sum_{k=0}^{\infty}\frac{(\lambda h)^k}{k!}e^{-\lambda h}g(x+h-\lambda h)\quad\text{for }x\in\mathbb{N},g\in BC(\mathbb{R}).
$$

The infinitesimal generator is then

$$
\mathcal{A}g(x)=\lambda(g(x+1)-g(x))-\lambda g'(x+).
$$

According to previous theorems,

$$
M(t)=g(X_t)-\int_0^t ds(\lambda(g(X_s+1)-g(X_s))-\lambda g'(X_s+)
$$

is a martingale for any $g\in BC(\mathbb{R})$.

Markov Diffusion Processes

Definition 2.122. A Markov process on $\mathbb{R}$ with transition probability function $p(s, x, t, A)$ is called a *diffusion process* if

1. for all $\epsilon > 0$, for all $t \geq 0$, and for all $x \in \mathbb{R}$: $\lim_{h\downarrow 0} \frac{1}{h} \int_{|x-y|>\epsilon} p(t, x, t + h, dy) = 0$;
2. there exist $a(t, x)$ and $b(t, x)$ such that, for all $\epsilon > 0$, for all $t \geq 0$, and for all $x \in \mathbb{R}$,

$$\lim_{h\downarrow 0} \frac{1}{h} \int_{|x-y|<\epsilon} (y - x)p(t, x, t + h, dy) = a(t, x),$$
$$\lim_{h\downarrow 0} \frac{1}{h} \int_{|x-y|<\epsilon} (y - x)^2 p(t, x, t + h, dy) = b(t, x).$$

$a(t, x)$ is the *drift coefficient* and $b(t, x)$ the *diffusion coefficient* of the process.

Lemma 2.123. *Conditions 1 and 2 of Definition 2.122 are satisfied if*

1.* *there exists a* $\delta > 0$ *such that, for all* $t \geq 0$ *and for all* $x \in \mathbb{R}$, $\lim_{h\downarrow 0} \frac{1}{h} \int_{\mathbb{R}} |x - y|^{2+\delta} p(t, x, t + h, dy) = 0$;
2.* *for all* $t \geq 0$ *and for all* $x \in \mathbb{R}$,

$$\lim_{h\downarrow 0} \frac{1}{h} \int_{\mathbb{R}} (y - x)p(t, x, t + h, dy) = a(t, x),$$
$$\lim_{h\downarrow 0} \frac{1}{h} \int_{\mathbb{R}} (y - x)^2 p(t, x, t + h, dy) = b(t, x).$$

Proof: We fix $\epsilon > 0$, $x \in \mathbb{R}$, $|x - y| > \epsilon \Rightarrow \dfrac{|y - x|^{2+\delta}}{\epsilon^{2+\delta}} \geq 1$, and hence

$$\frac{1}{h} \int_{|x-y|>\epsilon} p(t, x, t + h, dy) \leq \frac{1}{h\epsilon^{2+\delta}} \int_{|x-y|>\epsilon} |y - x|^{2+\delta} p(t, x, t + h, dy)$$
$$\leq \frac{1}{h\epsilon^{2+\delta}} \int_{\mathbb{R}} |y - x|^{2+\delta} p(t, x, t + h, dy).$$

From this, due to 1^*, follows 1 of Definition 2.122. Analogously, for $j = 1, 2$,

$$\frac{1}{h} \int_{|x-y|>\epsilon} |y - x|^j p(t, x, t + h, dy) \leq \frac{1}{h\epsilon^{2+\delta-j}} \int_{\mathbb{R}} |y - x|^{2+\delta} p(t, x, t + h, dy),$$

and again from 1^*, we obtain

$$\lim_{h\downarrow 0} \frac{1}{h} \int_{|x-y|>\epsilon} |y - x|^j p(t, x, t + h, dy) = 0.$$

Moreover,

$$\lim_{h\downarrow 0}\frac{1}{h}\int_{\mathbb{R}}|y-x|^j p(t,x,t+h,dy) = \lim_{h\downarrow 0}\frac{1}{h}\left(\int_{|x-y|>\epsilon}|y-x|^j p(t,x,t+h,dy) + \int_{|x-y|<\epsilon}|y-x|^j p(t,x,t+h,dy)\right),$$

which, along with 2^*, gives point 2 of Definition 2.122. □

Proposition 2.124. *If $(X_t)_{t\in\mathbb{R}_+}$ is a diffusion process with transition probability function p and drift and diffusion coefficients $a(x,t)$ and $b(x,t)$, respectively, and if $\mathcal{A}_s$ is the infinitesimal generator associated with p, then we have that*

$$(\mathcal{A}_s f)(x) = \frac{\partial f}{\partial x}a(s,x) + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}b(s,x), \tag{2.23}$$

provided that f is bounded and twice continuously differentiable.

Proof: Let $f \in BC(\mathbb{R}) \cap C^2(\mathbb{R})$. From Taylor's formula we obtain

$$f(y) - f(x) = f'(x)(y-x) + \frac{1}{2}f''(x)(y-x)^2 + o(|y-x|^2) \tag{2.24}$$

for $|y-x| < \delta$ (which is in a suitable neighbourhood of x), and thus

$$\begin{aligned}(\mathcal{A}_s f)(x) &= \lim_{h\downarrow 0}\frac{1}{h}\int_{\mathbb{R}}[f(y)-f(x)]p(s,x,s+h,dy)\\ &= \lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|<\delta} f'(x)(y-x)p(s,x,s+h,dy)\\ &\quad +\frac{1}{2}\lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|<\delta} f''(x)(y-x)^2 p(s,x,s+h,dy)\\ &\quad +\lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|<\delta} o(|y-x|^2)p(s,x,s+h,dy)\\ &\quad +\lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|\geq\delta}[f(y)-f(x)]p(s,x,s+h,dy).\end{aligned}$$

Because $f \in BC(\mathbb{R})$

$$\begin{aligned}&\lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|\geq\delta}[f(y)-f(x)]p(s,x,s+h,dy)\\ &\leq \lim_{h\downarrow 0}\frac{1}{h}c\int_{|y-x|\geq\delta}p(s,x,s+h,dy) = 0,\end{aligned}$$

by point 1 of Definition 2.122, where c is a constant. By point 2 of the same definition:

$$\begin{aligned}
&\lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|<\delta} f'(x)(y-x)p(s,x,s+h,dy)\\
&= f'(x)\lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|<\delta}(y-x)p(s,x,s+h,dy)\\
&= f'(x)a(t,x),
\end{aligned}$$

as well as

$$\frac{1}{2}\lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|<\delta} f''(x)(y-x)^2p(s,x,s+h,dy) = \frac{1}{2}f''(x)b(x,t).$$

Fixing $\epsilon > 0$, we finally observe that if we choose δ such that Taylor's formula (2.24) holds, so that

$$|y-x| < \delta \Rightarrow \frac{o(|y-x|^2)}{|y-x|^2} < \epsilon,$$

we get

$$\begin{aligned}
&\lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|<\delta} o(|y-x|^2)p(s,x,s+h,dy)\\
&\leq \lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|<\delta}\epsilon|y-x|^2p(s,x,s+h,dy)\\
&= \epsilon b(t,x)
\end{aligned}$$

and, from the fact that ϵ is arbitrary, we conclude that

$$\lim_{h\downarrow 0}\frac{1}{h}\int_{|y-x|<\delta} o(|y-x|^2)p(s,x,s+h,dy) = 0.$$

□

Markov Jump Processes

Consider a Markov process $(X_t)_{t\in\mathbb{R}_+}$ valued in a countable set E (say, $\mathbb{N}$ or $\mathbb{Z}$). In such a case it is sufficient (with respect to Theorem 2.99) to provide the so-called *one-point* transition probability function

$$p_{ij}(s,t) := p(s,i,t,j) := P(X_t = j|X_s = i)$$

for $t_0 \leq s < t$, $i,j \in E$. It follows from the general structure of Markov processes that the one-point transition probabilities satisfy the following relations:

(a) $p_{ij}(s,t) \geq 0$,
(b) $\sum_{j\in E} p_{ij}(s,t) = 1$,
(c) $p_{ij}(s,t) = \sum_{k\in E} p_{ik}(s,r)p_{kj}(r,t)$,

provided $t_0 \le s \le r \le t$, in $\mathbb{R}_+$, and $i, j \in E$. To these three conditions we need to add
(d)

$$\lim_{t \to s+} p_{ij}(s,t) = p_{ij}(s,s) = \delta_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \ne j. \end{cases}$$

The time-homogeneous case gives transition probabilities $(\tilde{p}_{ij}(t))_{t \in \mathbb{R}_+}$, such that

$$p_{ij}(s,t) = \tilde{p}_{ij}(t-s), \qquad s \le t.$$

From now on we shall limit our analysis to the time-homogeneous case, whose transition probabilities will be denoted $(p_{ij}(t))_{t \in \mathbb{R}_+}$. The following theorems hold (Gihman and Skorohod (1974), pp. 304–306).

Theorem 2.125. *The transition probabilities $(p_{ij}(t))_{t \in \mathbb{R}_+}$ of a homogeneous Markov process on a countable state space E are uniformly continuous in $t \in \mathbb{R}_+$ for any fixed $i, j \in E$.*

Theorem 2.126. *The limit*

$$q_i = \lim_{h \to 0+} \frac{1 - p_{ii}(h)}{h} \le +\infty$$

always exists (finite or not), and for arbitrary $t > 0$:

$$\frac{1 - p_{ii}(t)}{t} \le q_i.$$

If $q_i < +\infty$, then for all $t > 0$ the derivatives $p'_{ij}(t)$ exist for any $i, j \in E$ and are continuous. They satisfy the following relations:

1. $p'_{ij}(t+s) = \sum_{k \in E} p'_{ik}(t) p_{kj}(s)$,
2. $\sum_{j \in E} p'_{ij}(t) = 0$,
3. $\sum_{j \in E} |p'_{ij}(t)| \le 2q_i$.

In the following theorem the condition $q_i < +\infty$ is not required.

Theorem 2.127. *The limits*

$$\lim_{t \to 0+} \frac{p_{ij}(t)}{t} = p'_{ij}(0) =: q_{ij} < +\infty$$

always exist (finite) for any $i \ne j$.

As a consequence of Theorems 2.126 and 2.127, provided $q_i < +\infty$, we obtain evolution equations for $p_{ij}(t)$:

$$p'_{ij}(t) = \sum_{k \in E} q_{ik} p_{kj}(t),$$

with $q_{ii} = -q_i$. These equations are known as *Kolmogorov backward equations.* Consider the family of matrices $(P(t))_{t \in \mathbb{R}_+}$, with entries $(p_{ij}(t))_{t \in \mathbb{R}_+}$, for $i, j \in E$. We may rewrite conditions (c) and (d) in matrix form as follows:

(c′) $P(s+t) = P(s)P(t)$ for any $s, t \geq 0$;
(d′) $\lim_{h \to 0+} P(h) = P(0) = I$.

A family of stochastic matrices fulfilling conditions (c′) and (d′) is called a *matrix transition function*. If a matrix transition function satisfies the condition

$$\sum_{j \neq i} q_{ij} = -q_{ii} \equiv q_i < +\infty$$

for any $i \in E$, it is called *conservative*. The matrix $Q = (q_{ij})_{i,j \in E}$ is called the *intensity matrix*. The Kolmogorov backward equations can be rewritten in matrix form as

$$P'(t) = QP(t), \qquad t > 0,$$

subject to

$$P(0) = I.$$

If Q is a finite-dimensional matrix, the function $exp\{tQ\}$ for $t > 0$ is well defined.

Theorem 2.128. *(see Karlin and Taylor (1975), page 152). If E is finite, the matrix transition function can be represented in terms of its intensity matrix Q via*

$$P(t) = e^{tQ}, \qquad t \geq 0.$$

Given an intensity matrix Q of a conservative Markov jump process with stationary (time-homogeneous) transition probabilities, we have that (see Doob (1953))

$$P(X_u = i \ \forall u \in]s, s+t] | X_s = i) = e^{-q_i t}$$

for every $s, t \in \mathbb{R}_+$, and state $i \in E$. This shows that the sojourn time in state i is exponentially distributed with parameter q_i. This is independent of the initial time $s \geq 0$.

Furthermore, let π_{ij}, $i \neq j$, be the conditional probability of a jump to state j, given that a jump from state i has occurred. It can be shown (Doob (1953)) that

$$\pi_{ij} = \frac{q_{ij}}{q_i},$$

provided that $q_i > 0$. For $q_i = 0$, state i is *absorbing*, which obviously means that once state i is entered, the process remains there permanently. Indeed,

$$P(X_u = i, \text{ for all } u \in]s, s+t] | X_s = i) = e^{-q_i t} = 1,$$

for all $t \geq 0$. A state i for which $q_i = +\infty$ is called an *instantaneous state*. The expected sojourn time in such a state is zero. A state i for which $0 \leq q_i < +\infty$ is called a *stable state*.

Example 2.129. If $(X_t)_{t \in \mathbb{R}_+}$ is a homogeneous Poisson process with intensity $\lambda > 0$, then

$$p_{ij}(t) = \begin{cases} e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!} & \text{for } j > i, \\ 0 & \text{otherwise.} \end{cases}$$

This implies that

$$q_{ij} = p'_{ij}(0) \begin{cases} \lambda & \text{for } j = i+1, \\ -\lambda & \text{for } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

For the following result we refer again to Doob (1953).

Theorem 2.130. *For any $x \in E$, there exists a unique RCLL Markov process associated with a given intensity matrix Q and such that $P(X(0) = x) = 1$.*

Consider a time-homogeneous Markov jump process on a countable state space E with intensity matrix $Q = (q_{ij})_{i,j \in E}$. The matrix Q can be seen as a functional operator on E as follows: For any $f : E \to \mathbb{R}_+$ define

$$Q : f \to Q(f) = \sum_{j \in E} q_{ij} f(j) = \sum_{j \neq i} q_{ij}(f(j) - f(i)).$$

For f bounded in E we may define, for any $x \in E$,

$$\begin{aligned} & E_x[f(X(t+s))] - E_x[f(X(t))] \\ &= E_x[E_{X(t)}[f(X(s)) - f(X(0))]] \\ &= \sum_{j \neq i} (f(j) - f(i)) P(X(s) = j | X(0) = i) P_x(X(t) = i). \end{aligned}$$

Assume we may interchange the derivative and sum of the series:

$$\frac{d}{dt} E_x[f(X(t))] = \sum_{j \neq i} q_{ij}(f(j) - f(i)) P_x(X(s) = i),$$

which can be written as

$$\frac{d}{dt} E_x[f(X(t))] = E_x[Q(f)(X(t))].$$

By returning to the integral formulation

$$E_x[f(X(t))] - E_x[f(X(0))] = \int_0^t E_x[Q(f)(X(s))] ds, \tag{2.25}$$

the above can be seen as a Dynkin's formula for Markov jump processes in terms of the intensity matrix Q. Indeed, from Rogers and Williams (1994) (pp. 30–37), we obtain the following theorem.

Theorem 2.131. *For any function $g \in C^{1,0}(\mathbb{R}_+ \times E)$ such that the mapping*

$$t \to \frac{\partial}{\partial t} g(t,x)$$

is continuous for all $x \in E$, the process

$$\left(g(t, X(t)) - g(0, X(0)) - \int_0^t \left(\frac{\partial g}{\partial t} + Q(g(s,\cdot)) \right) (s, X(s)) ds \right)_{t \in \mathbb{R}_+}$$

is a local martingale.

Corollary 2.132. For any real function f defined on E, the process

$$\left(f(X(t)) - f(X(0)) - \int_0^t Q(f) X(s) ds \right)_{t \in \mathbb{R}_+} \qquad (2.26)$$

is a local martingale. Whenever the local martingale is a martingale, we recover equation (2.25).

Proposition 2.133. (martingale problem for Markov jump processes). *Given an intensity matrix Q, if an RCLL Markov process $X \equiv (X(t))_{t \in \mathbb{R}_+}$ on E is such that the process (2.26) is a local martingale, then Q is the intensity matrix of the Markov process X.*

Further discussions on this topic may be found in Doob (1953) and Karlin and Taylor (1981) (an additional and updated source regarding discrete-space continuous-time Markov chains is Anderson (1991)). For applications, see, for example, Robert (2003).

2.8 Brownian Motion and the Wiener Process

A small particle (e.g., a pollen corn) suspended in a liquid is subject to infinitely many collisions with atoms, and therefore it is impossible to observe its exact trajectory. With the help of a microscope it is only possible to confirm that the movement of the particle is entirely chaotic. This type of movement, discovered under similar circumstances by the botanist Robert Brown, is called Brownian motion. As its mathematical inventor Einstein already observed, it is necessary to make approximations, in order to describe the process. The formalized mathematical model defined on the basis of these is called a Wiener process. Henceforth, we will limit ourselves to the study of the one-dimensional Wiener process in $\mathbb{R}$, under the assumption that the three components determining its motion in space are independent.

Definition 2.134. The real-valued process $(W_t)_{t \in \mathbb{R}_+}$ is a *Wiener process* if it satisfies the following conditions:

1. $W_0 = 0$ almost surely;
2. $(W_t)_{t\in\mathbb{R}_+}$ is a process with independent increments;
3. $W_t - W_s$ is normally distributed with $N(0, t-s)$, $(0 \leq s < t)$.

Remark 2.135. From point 3 of Definition 2.134 it becomes obvious that every Wiener process is homogeneous.

Proposition 2.136. *If $(W_t)_{t\in\mathbb{R}_+}$ is a Wiener process, then*

1. $E[W_t] = 0 \qquad$ *for all* $t \in \mathbb{R}_+$,
2. $K(s,t) = Cov[W_t, W_s] = \min\{s,t\}, \qquad s,t \in \mathbb{R}_+$.

Proof: 1. Fixing $t \in \mathbb{R}$, we observe that $W_t = W_0 + (W_t - W_0)$ and thus $E[W_t] = E[W_0] + E[W_t - W_0] = 0$. The latter is given by the fact that $E[W_0] = 0$ (by 1 of Definition 2.134) and $E[W_t - W_0] = 0$ (by 3 of Definition 2.134).

2. Let $s,t \in \mathbb{R}_+$ and $Cov[W_t, W_s] = E[W_tW_s] - E[W_t]E[W_s]$, which (by point 1) gives $Cov[W_t, W_s] = E[W_tW_s]$. For simplicity, if we suppose that $s < t$, then

$$E[W_tW_s] = E[W_s(W_s + (W_t - W_s))] = E[W_s^2] + E[W_s(W_t - W_s)].$$

Since $(W_t)_{t\in\mathbb{R}_+}$ has independent increments, we obtain

$$E[W_s(W_t - W_s)] = E[W_s]E[W_t - W_s]$$

and by point 1 of Proposition 2.136 (or 3 of Definition 2.134) it follows that this is equal to zero, thus

$$Cov[W_t, W_s] = E[W_s^2] = Var[W_s].$$

If we now observe that $W_s = W_0 + (W_s - W_0)$ and hence $Var[W_s] = Var[W_0 + (W_s - W_0)]$, then, by the independence of the increments of the process, we get

$$Var[W_0 + (W_s - W_0)] = Var[W_0] + Var[W_s - W_0].$$

Therefore, by points 1 and 3 of Definition 2.134 it follows that

$$Var[W_s] = s = \inf\{s,t\},$$

which completes the proof. □

Remark 2.137. By 1 of Definition 2.134, it follows, for all $t \in \mathbb{R}_+$, $W_t = W_t - W_0$ almost surely and by 3 of the same definition, that W_t is distributed as $N(0,t)$. Thus

$$P(a \leq W_t \leq b) = \frac{1}{\sqrt{2\pi t}} \int_a^b e^{-\frac{x^2}{2t}}\, dx, \qquad a \leq b.$$

Remark 2.138. The Wiener process is a Gaussian process. In fact, if $n \in \mathbb{N}^*, (t_1, \ldots, t_n) \in \mathbb{R}_+^n$ with $0 = t_0 < t_1 < \ldots < t_n$ and $(a_1, \ldots, a_n) \in \mathbb{R}^n, (b_1, \ldots, b_n) \in \mathbb{R}^n$, such that $a_i \leq b_i, i = 1, 2, \ldots, n$, it can be shown that

$$P(a_1 \leq W_{t_1} \leq b_1, \ldots, a_n \leq W_{t_n} \leq b_n) \tag{2.27}$$
$$= \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} g(0|x_1, t_1) g(x_1|x_2, t_2 - t_1) \cdots g(x_{n-1}|x_n, t_n - t_{n-1}) dx_n \cdots dx_1,$$

where

$$g(x|y,t) = \frac{e^{-\frac{|x-y|^2}{2t}}}{\sqrt{2\pi t}}.$$

In order to prove that the density of $(W_{t_1}, \ldots, W_{t_n})$ is given by the integrand of (2.27), by uniqueness of the characteristic function, it is sufficient to show that the characteristic function ϕ' of the n-dimensional real-valued random vector, whose density is given by the integrand of (2.27), is identical to the characteristic function ϕ of $(W_{t_1}, \ldots, W_{t_n})$. Thus, let $\lambda = (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$. Then

$$\begin{aligned}
\phi(\lambda) &= E\left[e^{i(\lambda_1 W_{t_1} + \cdots + \lambda_n W_{t_n})}\right] \\
&= E\left[e^{i(\lambda_n (W_{t_n} - W_{t_{n-1}}) + (\lambda_n + \lambda_{n-1})(W_{t_{n-1}} - W_{t_{n-2}}) + \cdots + (\lambda_1 + \cdots + \lambda_n) W_{t_1})}\right] \\
&= E\left[e^{i\lambda_n (W_{t_n} - W_{t_{n-1}})}\right] E\left[e^{i(\lambda_n + \lambda_{n-1})(W_{t_{n-1}} - W_{t_{n-2}})}\right] \cdots \\
&\quad \cdots E\left[e^{i(\lambda_1 + \cdots + \lambda_n) W_{t_1}}\right],
\end{aligned}$$

where we exploit the independence of the random variables $W_{t_i} - W_{t_{i-1}}$, $i = 1, \ldots, n$. Furthermore, because $(W_{t_i} - W_{t_{i-1}})$ is $N(0, t_i - t_{i-1}), i = 1, \ldots, n$, we get

$$\phi(\lambda) = e^{\frac{-\lambda_n^2}{2}(t_n - t_{n-1})} e^{\frac{-(\lambda_n + \lambda_{n-1})^2}{2}(t_{n-1} - t_{n-2})} \cdots e^{\frac{-(\lambda_1 + \cdots + \lambda_n)^2}{2} t_1}.$$

We continue by calculating the characteristic function ϕ':

$$\begin{aligned}
\phi'(\lambda) &= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{i\lambda \cdot \mathbf{x}} g(0|x_1, t_1) \cdots g(x_{n-1}|x_n, t_n - t_{n-1}) dx_n \cdots dx_1 \\
&= \int_{-\infty}^{+\infty} \cdots \left(\int_{-\infty}^{+\infty} e^{i\lambda_n x_n} g(x_{n-1}|x_n, t_n - t_{n-1}) dx_n \right) \cdots dx_1.
\end{aligned}$$

Because

$$\int_{-\infty}^{+\infty} e^{i\lambda x} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-|x-m|^2}{2\sigma^2}} dx = e^{im\lambda - \frac{\lambda^2 \sigma^2}{2}}, \tag{2.28}$$

we obtain

$$\phi'(\lambda) = \int_{-\infty}^{+\infty} \cdots \left(e^{i\lambda_n x_{n-1} - \frac{\lambda_n^2}{2}(t_n - t_{n-1})} \right) \cdots dx_1$$
$$= e^{-\frac{\lambda_n^2}{2}(t_n - t_{n-1})} \int_{-\infty}^{+\infty} \cdots$$
$$\left(\int_{-\infty}^{+\infty} e^{i(\lambda_n + \lambda_{n-1})x_{n-1}} g(x_{n-2}|x_{n-1}, t_{n-1} - t_{n-2}) dx_{n-1} \right) \cdots dx_1.$$

By recalling (2.28) and applying it to each variable, we obtain

$$\phi'(\lambda) = e^{\frac{-\lambda_n^2}{2}(t_n - t_{n-1})} e^{\frac{-(\lambda_n + \lambda_{n-1})^2}{2}(t_{n-1} - t_{n-2})} \cdots e^{\frac{-(\lambda_1 + \ldots + \lambda_n)^2}{2} t_1}$$

and hence $\phi'(\lambda) = \phi(\lambda)$. We now show that $g(0|x_1, t_1) \cdots g(x_{n-1}|x_n, t_n - t_{n-1})$ is of the form

$$\frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det K}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})' K^{-1} (\mathbf{x}-\mathbf{m})}.$$

We will only show it for the case where $n = 2$; then

$$g(0|x_1, t_1) g(x_1|x_2, t_2 - t_1) = \frac{1}{2\pi \sqrt{t_1(t_2 - t_1)}} e^{-\frac{1}{2}\left[\frac{x_1^2}{t_1} + \frac{(x_2 - x_1)^2}{t_2 - t_1}\right]}$$
$$= \frac{1}{2\pi \sqrt{t_1(t_2 - t_1)}} e^{-\frac{1}{2}\left[\frac{x_1^2(t_2 - t_1) + (x_2 - x_1)^2 t_1}{t_1(t_2 - t_1)}\right]}.$$

If we put

$$K = \begin{pmatrix} t_1 & t_1 \\ t_1 & t_2 \end{pmatrix} \qquad (\text{where } K_{ij} = Cov[W_{t_i}, W_{t_j}]; i, j = 1, 2),$$

then

$$K^{-1} = \begin{pmatrix} \frac{t_2}{t_1(t_2 - t_1)} & -\frac{1}{t_2 - t_1} \\ -\frac{1}{t_2 - t_1} & \frac{1}{t_2 - t_1} \end{pmatrix},$$

resulting in

$$g(0|x_1, t_1) g(x_1|x_2, t_2 - t_1) = \frac{1}{2\pi \sqrt{\det K}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})' K^{-1} (\mathbf{x}-\mathbf{m})},$$

where $m_1 = E[W_{t_1}] = 0, m_2 = E[W_{t_2}] = 0$. Thus

$$g(0|x_1, t_1) g(x_1|x_2, t_2 - t_1) = \frac{1}{2\pi \sqrt{\det K}} e^{-\frac{1}{2}\mathbf{x}' K^{-1} \mathbf{x}},$$

completing the proof. □

Proposition 2.139. *If $(W_t)_{t \in \mathbb{R}_+}$ is a Wiener process, then it is also a martingale.*

Proof: The proposition follows from Example 2.62, because $(W_t)_{t\in\mathbb{R}_+}$ is a centered process with independent increments. □

Theorem 2.140. *Every Wiener process $(W_t)_{t\in\mathbb{R}_+}$ is a Markov process.*

Proof: The theorem follows directly by Theorem 2.96. □

Theorem 2.141. (Kolmogorov's continuity theorem). *Let $(X_t)_{t\in\mathbb{R}_+}$ be a separable real-valued stochastic process. If there exist positive real numbers r, c, ϵ, δ such that*

$$\forall h < \delta, \forall t \in \mathbb{R}_+, \qquad E[|X_{t+h} - X_t|^r] \leq ch^{1+\epsilon}, \tag{2.29}$$

then, for almost every $\omega \in \Omega$, the trajectories are continuous in $\mathbb{R}_+$.

Proof: For simplicity, we will only consider the interval $I =]0,1[$, instead of $\mathbb{R}_+$, so that $(X_t)_{t\in]0,1[}$. Let $t \in]0,1[$ and $0 < h < \delta$ such that $t + h \in]0,1[$. Then by the Markov inequality and by (2.29) we obtain

$$P(|X_{t+h} - X_t| > h^k) \leq h^{-rk} E[|X_{t+h} - X_t|^r] \leq ch^{1+\epsilon-rk} \tag{2.30}$$

for $k > 0$ and $\epsilon - rk > 0$. Therefore

$$\lim_{h\to 0} P\left(|X_{t+h} - X_t| > h^k\right) = 0;$$

namely, the process is continuous in probability and, by hypothesis, separable. Under these two conditions it can be shown that any arbitrary countable dense subset T_0 of $]0,1[$ can be regarded as a separating set. Thus we define

$$T_0 = \left\{ \frac{j}{2^n} \,\middle|\, j = 1, \ldots, 2^n - 1; n \in \mathbb{N}^* \right\}$$

and observe that, by (2.30),

$$\begin{aligned} P\left(\max_{1\leq j\leq 2^n-2} \left|X_{\frac{j+1}{2^n}} - X_{\frac{j}{2^n}}\right| \geq \frac{1}{2^{nk}}\right) &\leq \sum_{j=1}^{2^n-2} P\left(\left|X_{\frac{j+1}{2^n}} - X_{\frac{j}{2^n}}\right| \geq \frac{1}{2^{nk}}\right) \\ &\leq 2^n c 2^{-n(1+\epsilon-rk)} = c2^{-n(\epsilon-rk)}. \end{aligned}$$

Because $(\epsilon - rk) > 0$ and $\sum_n 2^{-n(\epsilon-rk)} < \infty$, we can apply the Borel–Cantelli Lemma 1.98 to the sets

$$F_n = \left\{ \max_{0\leq j\leq 2^n-1} \left|X_{\frac{j+1}{2^n}} - X_{\frac{j}{2^n}}\right| \geq \frac{1}{2^{nk}} \right\},$$

yielding $P(B) = 0$, where $B = \limsup_n F_n = \bigcap_n \bigcup_{k\leq n} F_k$. As a consequence, if $\omega \notin B$, then $\omega \in \Omega \setminus (\bigcap_n \bigcup_{k\geq n} F_k)$; i.e., there exists an $N = N(\omega) \in \mathbb{N}^*$, such that, for all $n \geq N$,

$$\left|X_{\frac{j+1}{2^n}}(\omega) - X_{\frac{j}{2^n}}(\omega)\right| < \frac{1}{2^{nk}}, \qquad j = 0, \ldots, 2^n - 1. \tag{2.31}$$

Now, let $\omega \notin B$ and s be a rational number, such that

$$s = j2^{-n} + a_1 2^{-(n+1)} + \cdots + a_m 2^{-(n+m)}, \qquad s \in [j2^{-n}, (j+1)2^{-n}[,$$

where either $a_j = 0$ or $a_j = 1$ and $m \in \mathbb{N}^*$. If we put

$$b_r = j2^{-n} + a_1 2^{-(n+1)} + \cdots + a_r 2^{-(n+r)},$$

with $b_0 = j2^{-n}$ and $b_m = s$ for $r = 0, \ldots, m$, then

$$|X_s(\omega) - X_{j2^{-n}}(\omega)| \leq \sum_{r=0}^{m-1} |X_{b_{r+1}}(\omega) - X_{b_r}(\omega)|.$$

If $a_{r+1} = 0$, then $[b_r, b_{r+1}[= \emptyset$; if $a_{r+1} = 1$, then $[b_r, b_{r+1}[$ is of the form $[l2^{-(n+r+1)}, (l+1)2^{-(n+r+1)}[$. Hence from (2.31), it follows that

$$|X_s(\omega) - X_{j2^{-n}}(\omega)| \leq \sum_{r=0}^{m-1} 2^{-(n+r+1)k} \leq 2^{-nk} \sum_{r=0}^{\infty} 2^{-(r+1)k} \leq M2^{-nk}, \quad (2.32)$$

with $M \geq 1$. Fixing $\epsilon > 0$, there exists an $N_1 > 0$ such that, for all $n \geq N_1$, $M2^{-nk} < \frac{\epsilon}{3}$, and from the fact that $M \geq 1$, it also follows that, for all $n \geq N_1$, $2^{-nk} < \frac{\epsilon}{3}$. Let t_1, t_2 be elements of T_0 (separating set) such that $|t_1 - t_2| < \min\{2^{-N_1}, 2^{-N(\omega)}\}$. If $n = \max\{N_1, N(\omega)\}$, then there is at most one rational number of the form $\frac{j+1}{2^n}$ $(j = 1, \ldots, 2^n - 1)$ between t_1 and t_2. Therefore, by (2.31) and (2.32) it follows that

$$\begin{aligned} &| X_{t_1}(\omega) - X_{t_2}(\omega)| \\ &\leq \left|X_{t_1}(\omega) - X_{\frac{j}{2^n}}(\omega)\right| + \left|X_{\frac{j+1}{2^n}}(\omega) - X_{\frac{j}{2^n}}(\omega)\right| + \left|X_{t_2}(\omega) - X_{\frac{j+1}{2^n}}(\omega)\right| \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Hence the trajectory is uniformly continuous almost everywhere in T_0 and has a continuous extension in $[0, 1]$. By Theorem 2.26, the extension coincides with the original trajectory. Therefore, the trajectory is continuous almost everywhere in $]0, 1[$. □

Theorem 2.142. *If* $(W_t)_{t\in\mathbb{R}_+}$ *is a real-valued Wiener process, then it has continuous trajectories almost surely.*

Proof: Let $t \in \mathbb{R}_+$ and $h > 0$. Because $W_{t+h} - W_t$ is normally distributed as $N(0, h)$, putting $Z_{t,h} = \frac{W_{t+h}-W_t}{\sqrt{h}}$, $Z_{t,h}$ has standard normal distribution. Therefore, it is clear that there exists an $r > 2$ such that $E[|Z_{t,h}|^r] > 0$, and thus $E[|W_{t+h} - W_t|^r] = E[|Z_{t,h}|^r]h^{\frac{r}{2}}$. If we write $r = 2(1 + \epsilon)$, we obtain $E[|W_{t+h} - W_t|^r] = ch^{1+\epsilon}$, with $c = E[|Z_{t,h}|^r]$. The assertion then follows by Kolmogorov's continuity theorem. □

Remark 2.143. Since Brownian motion is continuous in probability, it admits a separable and progressively measurable modification.

Lemma 2.144. *Let $(W_t)_{t\in\mathbb{R}_+}$ be a real-valued Wiener process. If $a > 0$, then*

$$P\left(\max_{0\leq s\leq t} W_s > a\right) = 2P(W_t > a).$$

Proof: We employ the *reflection principle* by defining the process $(\tilde{W}_t)_{t\in\mathbb{R}}$ as

$$\begin{cases} \tilde{W}_t = W_t & \text{if } W_s < a, \forall s < t, \\ \tilde{W}_t = 2a - W_t & \text{if } \exists s < t \text{ such that } W_s = a. \end{cases}$$

The name arises because once $W_s = a$, then $\tilde{W}_s$ becomes a reflection of W_s about the *barrier* a. It is obvious that $(\tilde{W}_t)_{t\in\mathbb{R}}$ is a Wiener process as well. Moreover, we can observe that

$$\max_{0\leq s\leq t} W_s > a \tag{2.33}$$

if and only if either $W_t > a$ or $\tilde{W}_t > a$. These two events are mutually exclusive and thus their probabilities are additive. As they are both Wiener processes it is obvious that the two events have the same probability and thus

$$P\left(\max_{0\leq s\leq t} W_s > a\right) = P(W_t > a) + P(\tilde{W}_t > a) = 2P(W_t > a),$$

completing the proof. For a more general case, see (B.11). □

Theorem 2.145. *If $(W_t)_{t\in\mathbb{R}_+}$ is a real-valued Wiener process, then*

1. $P(\sup_{t\in\mathbb{R}_+} W_t = +\infty) = 1,$
2. $P(\inf_{t\in\mathbb{R}_+} W_t = -\infty) = 1.$

Proof: For $a > 0$,

$$P\left(\sup_{t\in\mathbb{R}_+} W_t > a\right) \geq P\left(\sup_{0\leq s\leq t} W_s > a\right) = P\left(\max_{0\leq s\leq t} W_s > a\right),$$

where the last equality follows by continuity of trajectories. By Lemma 2.144:

$$P\left(\sup_{t\in\mathbb{R}_+} W_t > a\right) \geq 2P\left(W_t > a\right) = 2P\left(\frac{W_t}{\sqrt{t}} > \frac{a}{\sqrt{t}}\right), \text{ for } t > 0.$$

Because W_t is normally distributed as $N(0,t)$, $\frac{W_t}{\sqrt{t}}$ is standard normal and, denoting by Φ its cumulative distribution, we get

$$2P\left(\frac{W_t}{\sqrt{t}} > \frac{a}{\sqrt{t}}\right) = 2\left(1 - \Phi\left(\frac{a}{\sqrt{t}}\right)\right).$$

By $\lim_{t\to\infty} \Phi(\frac{a}{\sqrt{t}}) = \frac{1}{2}$, it follows that

$$\lim_{t\to\infty} 2P\left(\frac{W_t}{\sqrt{t}} > \frac{a}{\sqrt{t}}\right) = 1,$$

and because

$$\left\{\sup_{t\in\mathbb{R}_+} W_t = +\infty\right\} = \bigcap_{a=1}^{\infty} \left\{\sup_{t\in\mathbb{R}_+} W_t > a\right\},$$

we obtain 1.

Point 2 follows directly from 1, by observing that if $(W_t)_{t\in\mathbb{R}_+}$ is a real-valued Wiener process, then so is $(-W_t)_{t\in\mathbb{R}_+}$. □

Theorem 2.146. *If $(W_t)_{t\in\mathbb{R}_+}$ is a real-valued Wiener process, then,*

$$\forall h > 0, \qquad P\left(\max_{0\le s\le h} W_s > 0\right) = P\left(\min_{0\le s\le h} W_s < 0\right) = 1.$$

Moreover, for almost every $\omega \in \Omega$ the process $(W_t)_{t\in\mathbb{R}_+}$ has a zero (i.e., crosses the spatial axis) in $]0,h]$, for all $h > 0$.

Proof: If $h > 0$ and $a > 0$, it is obvious that

$$P\left(\max_{0\le s\le h} W_s > 0\right) \ge P\left(\max_{0\le s\le h} W_s > a\right).$$

Then, by Lemma 2.144,

$$P\left(\max_{0\le s\le h} W_s > a\right) = 2P(W_h > a) = 2P\left(\frac{W_h}{\sqrt{h}} > \frac{a}{\sqrt{h}}\right) = 2\left(1 - \Phi\left(\frac{a}{\sqrt{h}}\right)\right).$$

For $a \to 0$, $2(1 - \Phi(\frac{a}{\sqrt{h}})) \to 1$ and thus $P(\max_{0\le s\le h} W_s > 0) = 1$. Furthermore,

$$P\left(\min_{0\le s\le h} W_s < 0\right) = P\left(\max_{0\le s\le h} (-W_s) > 0\right) = 1.$$

Now we can observe that

$$P\left(\max_{0\le s\le h} W_s > 0, \forall h > 0\right) \ge P\left(\bigcap_{n=1}^{\infty}\left(\max_{0<s\le\frac{1}{n}} W_s > 0\right)\right) = 1.$$

Hence

$$P\left(\max_{0\le s\le h} W_s > 0, \forall h > 0\right) = 1$$

and, analogously,

$$P\left(\min_{0\le s\le h} W_s < 0, \forall h > 0\right) = 1.$$

From this it can be deduced that for almost every $\omega \in \Omega$ the process $(W_t)_{t\in\mathbb{R}_+}$ becomes zero in $]0,h]$, for all $h > 0$. On the other hand, since $(W_t)_{t\in\mathbb{R}_+}$ is a time-homogeneous Markov process with independent increments, it has the same behavior in $]h,2h]$ as in $]0,h]$, and thus it has zeros in every interval. □

Theorem 2.147. *Almost every trajectory of the Wiener process* $(W_t)_{t\in\mathbb{R}_+}$ *is differentiable almost nowhere.*

Proof: Let $D = \{\omega \in \Omega | W_t(\omega)$ is differentiable for at least one $t \in \mathbb{R}_+\}$. We will show that $D \subset G$, with $P(G) = 0$ (obviously, if P is complete, then $D \in \mathcal{F}$). Let $k > 0$ and

$$A_k = \left\{\omega \left| \limsup_{h\downarrow 0} \frac{|W_{t+h}(\omega) - W_t(\omega)|}{h} < k \text{ for at least one } t \in [0,1[\right.\right\}.$$

Then, if $\omega \in A_k$, we can choose $m \in \mathbb{N}$ sufficiently large, such that $\frac{j-1}{m} \leq t < \frac{j}{m}$ for $j \in \{1,\ldots,m\}$, and for $t \leq s \leq \frac{j+3}{m}$, $W(s,\omega)$ is enveloped by the cone with slope k. Then, for an integer $j \in \{1,\ldots,m\}$, we get

$$\begin{aligned}\left|W_{\frac{j+1}{m}}(\omega) - W_{\frac{j}{m}}(\omega)\right| &\leq \left|W_{\frac{j+1}{m}}(\omega) - W_t(\omega)\right| + \left|-W_t(\omega) + W_{\frac{j}{m}}(\omega)\right| \\ &< \left(\frac{j+1}{m} - \frac{j-1}{m}\right)k + \left(\frac{j}{m} - \frac{j-1}{m}\right)k \\ &= \frac{2k}{m} + \frac{k}{m} = \frac{3k}{m}. \end{aligned} \tag{2.34}$$

Analogously, we obtain that

$$\left|W_{\frac{j+2}{m}}(\omega) - W_{\frac{j+1}{m}}(\omega)\right| \leq \frac{5k}{m} \tag{2.35}$$

and

$$\left|W_{\frac{j+3}{m}}(\omega) - W_{\frac{j+2}{m}}(\omega)\right| \leq \frac{7k}{m}. \tag{2.36}$$

Because $\frac{W_{t+h}-W_t}{\sqrt{h}}$ is distributed as $N(0,1)$, it follows that

$$\begin{aligned} P(|W_{t+h} - W_t| < a) &= P\left(\frac{|W_{t+h} - W_t|}{\sqrt{h}} < \frac{a}{\sqrt{h}}\right) \\ &= \int_{-\frac{a}{\sqrt{h}}}^{\frac{a}{\sqrt{h}}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx \\ &\leq \frac{1}{\sqrt{2\pi}} 2\frac{a}{\sqrt{h}} = \frac{2a}{\sqrt{2\pi h}}. \end{aligned}$$

Putting $A_{m,j} = \{\omega | (2.34),(2.35),(2.36) \text{ are true}\}$, because the process has independent increments, we obtain

$$\begin{aligned} P(A_{m,j}) &= P(\{\omega|(2.34) \text{ is true}\})P(\{\omega|(2.35) \text{ is true}\})P(\{\omega|(2.36) \text{ is true}\}) \\ &\leq 8\left(\frac{2\pi}{m}\right)^{-\frac{3}{2}} \frac{3k}{m}\frac{5k}{m}\frac{7k}{m}, \end{aligned}$$

and thus $P(A_{m,j}) \leq cm^{-\frac{3}{2}}$, $j = 1,\ldots,m$. Putting $A_m = \bigcup_{j=1}^m A_{m,j}$, then

$$P(A_m) \leq \sum_{j=1}^{m} P(A_{m,j}) \leq cm^{-\frac{1}{2}}. \tag{2.37}$$

Now let $m = n^4$ $(n \in \mathbb{N}^*)$; we obtain $P(A_{n^4}) \leq cn^{-2} = \frac{c}{n^2}$ and thus

$$\sum_n P(A_{n^4}) \leq c \sum_n \frac{1}{n^2} < \infty.$$

Therefore, by the Borel–Cantelli Lemma 1.98,

$$P\left(\limsup_n A_{n^4}\right) = 0.$$

It can now be shown that

$$A_k \subset \liminf_m A_m \equiv \bigcup_m \bigcap_{i \geq m} A_i \subset \liminf_n A_{n^4} \subset \limsup_n A_{n^4},$$

hence $A_k \subset A''_{n^4}$ and $P(A''_{n^4}) = 0$. Let

$$D_0 = \{\omega | W(\cdot, \omega) \text{ is differentiable in at least one } t \in [0, 1[\}.$$

Then $D_0 \subset \bigcup_{k=1}^{\infty} A_k = G_0$, which means that D_0 is contained in a set of probability zero, namely $D_0 \subset G_0$ and $P(G_0) = 0$. Decomposing $\mathbb{R}_+ = \bigcup_n [n, n+1[$, since the motion is Brownian and of independent increments,

$$D_n = \{\omega | W(\cdot, \omega) \text{ is differentiable in at least one } t \in [n, n+1[\},$$

analogously to D_0, will be contained in a set of probability zero, i.e., $D_n \subset G_n$ and $P(G_n) = 0$. But $D \subset \bigcup_n D_n \subset \bigcup_n G_n$, thus completing the proof. □

Proposition 2.148. (scaling property). *Let $(W_t)_{t \in \mathbb{R}_+}$ be a Wiener process. Then the time-scaled process $(\tilde{W}_t)_{t \in \mathbb{R}_+}$ defined by*

$$\tilde{W}_t = tW_{1/t},\ t > 0, \qquad \tilde{W}_0 = 0$$

is also a Wiener process.

Proof: See, e.g., Karlin and Taylor (1975). □

Proposition 2.149. (strong law of large numbers). *Let $(W_t)_{t \in \mathbb{R}_+}$ be a Wiener process. Then*

$$\frac{W_t}{t} \to 0, \qquad \textit{as} \qquad t \to +\infty, \qquad \textit{a.s.}$$

Proof: See, e.g., Karlin & Taylor (1975). □

Proposition 2.150. (law of iterated logarithms). *Let $(W_t)_{t\in\mathbb{R}_+}$ be a Wiener process. Then*

$$\limsup_{t\to+\infty} \frac{W_t}{\sqrt{2t\ln\ln t}} = 1, \qquad a.s.,$$
$$\liminf_{t\to+\infty} \frac{W_t}{\sqrt{2t\ln\ln t}} = -1, \qquad a.s.$$

As a consequence, for any $\epsilon > 0$, there exists a $t_0 > 0$, such that for any $t > t_0$ we have

$$-(1+\epsilon)\sqrt{2t\ln\ln t} \le W_t \le (1+\epsilon)\sqrt{2t\ln\ln t}, \qquad a.s.$$

Proof: See, e.g., Breiman (1968). □

Brownian Motion After a Stopping Time

Let $(W(t))_{t\in\mathbb{R}_+}$ be a Wiener process with a finite stopping time T and $\mathcal{F}_T$ the σ-algebra of events preceding T. By Remark 2.143 and Theorem 2.48, $W(T)$ is $\mathcal{F}_T$-measurable and hence measurable.

Remark 2.151. Brownian motion is endowed with the Feller property and therefore also with the strong Markov property. (This can be shown by using the representation of the semigroup associated with $(W(t))_{t\in\mathbb{R}_+}$.)

Theorem 2.152. *Resorting to the previous notation, we have that*

1. *the process $y(t) = W(T+t) - W(T), t \ge 0$, is again a Brownian motion;*
2. *$\sigma(y(t), t \ge 0)$ is independent of $\mathcal{F}_T$*

(thus a Brownian motion remains a Brownian motion after a stopping time).

Proof: If $T = s$ (s constant), the assertion is obvious. We now suppose that T has a countable codomain $(s_j)_{j\in\mathbb{N}}$ and that $B \in \mathcal{F}_T$. If we consider further that $0 \le t_1 < \cdots < t_n$ and that $A_1, \ldots, A_n$ are Borel sets of $\mathbb{R}$, then

$$\begin{aligned}
&P(y(t_1) \in A_1, \ldots, y(t_n) \in A_n, B)\\
&= \sum_{j\in\mathbb{N}} P(y(t_1) \in A_1, \ldots, y(t_n) \in A_n, B, T = s_j)\\
&= \sum_{j\in\mathbb{N}} P((W(t_1 + s_j) - W(s_j)) \in A_1, \ldots\\
&\qquad\qquad \ldots, (W(t_n + s_j) - W(s_j)) \in A_n, B, T = s_j).
\end{aligned}$$

Moreover, $(T = s_j) \cap B = (B \cap (T \le s_j)) \cap (T = s_j) \in \mathcal{F}_{s_j}$ (as observed in the proof of Theorem 2.48), and since a Wiener process has independent increments, the events $((W(t_1+s_j)-W(s_j)) \in A_1, \ldots, (W(t_n+s_j)-W(s_j)) \in A_n)$ and $(B, T = s_j)$ are independent; therefore,

$$\begin{aligned}
&P(y(t_1) \in A_1, \ldots, y(t_n) \in A_n, B) \\
&= \sum_{j \in \mathbb{N}} P((W(t_1 + s_j) - W(s_j)) \in A_1, \ldots \\
&\qquad \ldots, (W(t_n + s_j) - W(s_j)) \in A_n) P(B, T = s_j) \\
&= \sum_{j \in \mathbb{N}} P(W(t_1) \in A_1, \ldots, W(t_n) \in A_n) P(B, T = s_j) \\
&= P(W(t_1) \in A_1, \ldots, W(t_n) \in A_n) P(B),
\end{aligned}$$

where we note that $W(t_k + s_j) - W(s_j)$ has the same distribution as $W(t_k)$. From these equations (having factorized) follows point 2. Furthermore, if we take $B = \Omega$, we obtain

$$P(y(t_1) \in A_1, \ldots, y(t_n) \in A_n) = P(W(t_1) \in A_1, \ldots, W(t_n) \in A_n).$$

This shows that the finite-dimensional distributions of the process $(y(t))_{t \geq 0}$ coincide with those of W. Therefore, by the Kolmogorov–Bochner theorem, the proof of 1 is complete.

Let T be a generic finite stopping time of the Wiener process $(W_t)_{t \geq 0}$ and (as in Lemma 2.117) $(T_n)_{n \in \mathbb{N}}$ a sequence of stopping times, such that $T_n \geq T, T_n \downarrow T$ as $n \to \infty$ and T_n has an at most countable codomain. We put, for all $n \in \mathbb{N}$, $y_n(t) = W(T_n + t) - W(T_n)$ and let $B \in \mathcal{F}_T, 0 \leq t_1 \leq \cdots \leq t_k$. Then, because for all $n \in \mathbb{N}$, $\mathcal{F}_T \subset \mathcal{F}_{T_n}$ (see the proof of Theorem 2.118) and for all $n \in \mathbb{N}$, the theorem holds for T_n (as already shown above), we have

$$P(y_n(t_1) \leq x_1, \ldots, y_n(t_k) \leq x_k, B) = P(W(t_1) \leq x_1, \ldots, W(t_k) \leq x_k) P(B).$$

Moreover, since W is continuous, from $T_n \downarrow T$ as $n \to \infty$, it follows that $y_n(t) \to y(t)$ a.s., for all $t \geq 0$. Thus, if $(x_1, \ldots, x_k)$ is a point of continuity of the k-dimensional distribution F_k of $(W(t_1), \ldots, W(t_k))$, we get by Lévy's continuity Theorem 1.153

$$\begin{aligned}
&P(y(t_1) \leq x_1, \ldots, y(t_k) \leq x_k, B) \\
&= P(W(t_1) \leq x_1, \ldots, W(t_k) \leq x_k) P(B).
\end{aligned} \tag{2.38}$$

Since F_k is continuous almost everywhere (given that Gaussian distributions are absolutely continuous with respect to Lebesgue measure and thus have density), (2.38) holds for every $x_1, \ldots, x_k$. Therefore, for every Borel set $A_1, \ldots, A_k$ of $\mathbb{R}$, we have that

$$P(y(t_1) \in A_1, \ldots, y(t_k) \in A_k, B) = P(W(t_1) \in A_1, \ldots, W(t_n) \in A_k) P(B),$$

completing the proof. □

Definition 2.153. The real-valued process $(W_1(t), \ldots, W_n(t))'_{t \geq 0}$ is said to be an *n-dimensional Wiener process (or Brownian motion)* if:

1. for all $i \in \{1, \ldots, n\}$, $(W_i(t))_{t \geq 0}$ is a Wiener process,
2. the processes $(W_i(t))_{t \geq 0}$, $i = 1, \ldots, n$, are independent

(thus the σ-algebras $\sigma(W_i(t), t \geq 0)$, $i = 1, \ldots, n$, are independent).

Proposition 2.154. If $(W_1(t), \ldots, W_n(t))'_{t \geq 0}$ is an n-dimensional Brownian motion, then it can be shown that:

1. $(W_1(0), \ldots, W_n(0)) = (0, \ldots, 0)$ almost surely;
2. $(W_1(t), \ldots, W_n(t))'_{t \geq 0}$ has independent increments;
3. $(W_1(t), \ldots, W_n(t))' - (W_1(s), \ldots, W_n(s))', 0 \leq s < t$, has multivariate normal distribution $N(\mathbf{0}, (t-s)I)$ (where $\mathbf{0}$ is the null-vector of order n and I is the $n \times n$ identity matrix).

Proof: The proof follows from Definition 2.153. □

2.9 Counting, Poisson, and Lévy Processes

Whereas Brownian motion and the Wiener process are continuous in space and time, there exists a family of processes that are continuous in time, but discontinuous in space, admitting jumps. The simplest of these is a counting process, of which the Poisson process is a special case. The latter also allows many explicit results. The most general process admitting both continuous and discontinuous movements is the Lévy process, which contains both Brownian motion and the Poisson process. Finally, a stable process is a particular type of Lévy process, which reproduces itself under addition.

Definition 2.155. Let $(\tau_i)_{i \in \mathbb{N}^*}$ be a strictly increasing sequence of positive random variables on the space $(\Omega, \mathcal{F}, P)$, with $\tau_0 \equiv 0$. Then the process $(N_t)_{t \in \bar{\mathbb{R}}_+}$ given by

$$N_t = \sum_{i \in \mathbb{N}^*} I_{[\tau_i, +\infty]}(t), \qquad t \in \bar{\mathbb{R}}_+,$$

valued in $\bar{\mathbb{N}}$, is called a *counting process* associated with the sequence $(\tau_i)_{i \in \mathbb{N}^*}$. Moreover, the random variable $\tau = \sup_i \tau_i$ is the *explosion time* of the process. If $\tau = \infty$ almost surely, then N_t is *nonexplosive*.

Theorem 2.156. *Let $(\mathcal{F}_t)_{t \in \bar{\mathbb{R}}_+}$ be a filtration that satisfies the usual hypotheses (see Definition 2.34). A counting process $(N_t)_{t \in \bar{\mathbb{R}}_+}$ is adapted to $(\mathcal{F}_t)_{t \in \bar{\mathbb{R}}_+}$ if and only if its associated random variables $(\tau_i)_{i \in \mathbb{N}^*}$ are stopping times.*

Proof: See, e.g., Protter (1990). □

Proposition 2.157. An RCLL process may admit at most jump discontinuities. If $P(X_t \neq X_{t^-}) > 0$ such a process $(X_t)_{t \in \mathbb{R}_+}$ has a *fixed jump* at a time t.

Remark 2.158. A nonexplosive counting process is RCLL. Its trajectories are right-continuous step functions with upward jumps of magnitude 1 and $N_0 = 0$ almost surely.

Theorem 2.159. *Let $(N_t)_{t\in\mathbb{R}_+}$ be a counting process. Then its natural filtration is right-continuous.*

Poisson Process

Definition 2.160. Let $(T_i)_{i\in\mathbb{N}^*}$ be a sequence of independent and identically distributed random variables with common exponential probability law (i.e. $E[T_i] = 1/\lambda$ for a given parameter $\lambda > 0$), and let $\tau_n = \sum_{i=1}^n T_i$. Then the process $(N_t)_{t\in\mathbb{R}_+}$, given by

$$N_t = \sum_{n\in\mathbb{N}^*} I_{[\tau_n,+\infty]}(t) = \sum_{n\in\mathbb{N}^*} nI_{[\tau_n,\tau_{n+1}[}(t), \qquad t \in \mathbb{R}_+,$$

is a *Poisson process* of *intensity* λ.

Remark 2.161. Following the definition of the Poisson process we have that

$$\tau_n = \inf\{t \in \mathbb{R}_+ | N_t = n\}, \qquad n \in \mathbb{N}^*.$$

Remark 2.162. The sequence of random variables $(T_i)_{i\in\mathbb{N}^*}$ underlying a Poisson process is usually referred to as its interarrival times.

Theorem 2.163. *A Poisson process is an adapted counting process without explosions that has both independent and stationary increments.*

The subsequent theorem specifies the distribution of the random variable N_t, $t \in \mathbb{R}_+$.

Theorem 2.164. *Let $(N_t)_{t\in\mathbb{R}_+}$ be a Poisson process of intensity $\lambda > 0$. Then for any $t \in \mathbb{R}_+$ the random variable N_t has the Poisson probability distribution*

$$P(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \qquad n \in \mathbb{N},$$

with parameter λt. Furthermore, $E[N_t] = \lambda t$, $Var[N_t] = \lambda t$, its characteristic function is

$$\phi_{N_t}(u) = E\left[e^{iuN_t}\right] = e^{-\lambda t(1-\exp\{iu\})},$$

and its probability generating function is

$$g_{N_t}(u) = E\left[u^{N_t}\right] = e^{\lambda t(u-1)}, \qquad u \in \mathbb{R}_+^*.$$

Proof: The characteristic function of T_1 is

$$\phi_{T_1}(u) = E\left[e^{iuT_1}\right] = \int_0^\infty e^{iuT_1}\lambda e^{-\lambda T_1}dT_1 = \frac{\lambda}{\lambda - iu}, \qquad u \in \mathbb{R}.$$

From the independence and identical distribution property of $T_1, \ldots, T_n$, it follows that

$$\phi_{\tau_n}(u) = E\left[e^{iu\tau_n}\right] = \left(E\left[e^{iuT_1}\right]\right)^n = \left(\frac{\lambda}{\lambda - iu}\right)^n, \qquad u \in \mathbb{R},$$

which can be shown to be the characteristic function of the Gamma distribution $\Gamma(\lambda, n)$. Hence for any $n \in \mathbb{N}$, assuming that $\tau_0 = 0$, we have that τ_n are Gamma distributed and thus

$$\begin{aligned} P(N_t = n) &= P(\tau_n \le t) - P(\tau_{n+1} \le t) \\ &= \int_0^t \lambda e^{-\lambda x}\frac{(\lambda x)^{n-1}}{(n-1)!}dx - \int_0^t \lambda e^{-\lambda x}\frac{(\lambda x)^n}{n!}dx = \frac{(\lambda t)^n}{n!}e^{-\lambda t}. \end{aligned}$$

The other quantities can be calculated by solving

$$\begin{aligned} E[N_t] &= \sum_{n=0}^\infty n\frac{(\lambda t)^n}{n!}e^{-\lambda t} = \lambda t\sum_{n=0}^\infty \frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t} = \lambda t, \\ E\left[N_t^2\right] &= \sum_{n=0}^\infty n^2\frac{(\lambda t)^n}{n!}e^{-\lambda t} = \lambda t\sum_{n=0}^\infty((n-1)+1)\frac{(\lambda t)^{n-1}}{(n-1)!}e^{-\lambda t} \\ &= (\lambda t)^2 + \lambda t, \\ Var[N_t] &= E\left[N_t^2\right] - (E[N_t])^2, \\ E\left[e^{iuN_t}\right] &= \sum_{n=0}^\infty e^{iun}\frac{(\lambda t)^n}{n!}e^{-\lambda t} = e^{-\lambda t(1-\exp\{iu\})}\sum_{n=0}^\infty \frac{\left(\lambda t e^{iu}\right)^n}{n!}e^{-\lambda t\exp\{iu\}} \\ &= e^{-\lambda t(1-\exp\{iu\})}, \\ E\left[u^{N_t}\right] &= \sum_{n=0}^\infty u^n\frac{(\lambda t)^n}{n!}e^{-\lambda t} = e^{\lambda t(u-1)}\sum_{n=0}^\infty \frac{(u\lambda t)^n}{n!}e^{-u\lambda t} = e^{\lambda t(u-1)}. \end{aligned}$$

□

Theorem 2.165. *Let $(N_t)_{t\in\mathbb{R}_+}$ be a Poisson process of intensity $\lambda > 0$ and $\mathcal{F}_t = \sigma(N_s, s \le t)$. Then*

1. *$(N_t)_{t\in\mathbb{R}_+}$ is a process with independent increments; i.e., for $s, t \in \mathbb{R}_+$, $N_{t+s} - N_t$ is independent of $\mathcal{F}_t$.*
2. *$(N_t)_{t\in\mathbb{R}_+}$ is stationary; i.e., for $s, t \in \mathbb{R}_+$, $N_{t+s} - N_t$ has the same probability law as*

$$N_s - N_0 \equiv N_s.$$

Remark 2.166. The jump times τ_n, for any $n \in \mathbb{N}^*$, are $\mathcal{F}_t$ stopping times:

$$\{\tau_n \leq t\} = \{N_t = n\} \in \mathcal{F}_t, \qquad t \in \mathbb{R}_+.$$

Definition 2.167. P^0: Let $(N_t)_{t\in\mathbb{R}_+}$ be a stochastic process. For each $\omega \in \Omega$ we have that

1. $N_t(\omega) \in \mathbb{N}$, $t \in \mathbb{R}_+$;
2. $N_0(\omega) = 0$ almost surely and $\lim_{t\to+\infty} P(N_t(\omega) = 0) = 0$;
3. $N_t(\omega)$ is nondecreasing and right-continuous, and at points of discontinuity, the jump
$$N_t(\omega) - \sup_{s<t} N_s(\omega) = 1.$$

Theorem 2.168. *Assumption* (P^0) *of the preceding definition is equivalent to the following:*

$(P^0)'$ $(N_t)_{t\in\mathbb{R}_+}$ *is an RCLL Markov process with probability law*

$$P(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \qquad t \in \mathbb{R}_+, n \in \mathbb{N}.$$

Proposition 2.169. Under assumption P^0 let $\tau_n' = \inf\{t \in \mathbb{R}_+ : N_t \geq n\}$ and $T_n = \tau_n - \tau_{n-1}$, for all $n \in \mathbb{N}$. Then the following statements are all equivalent:

P^1 : T_n are independent exponentially distributed random variables with parameter λ.

P^2 : For any $0 < t_1 < \cdots < t_k$ the increments $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_k} - N_{t_{k-1}}$ are independent and each of them is identically Poisson distributed:

$$N_{t_i} - N_{t_{i-1}} \sim P(\lambda(t_i - t_{i-1})).$$

P^3 : For any $0 < t_1 < \cdots < t_k$ the increments $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_k} - N_{t_{k-1}}$ are independent and the distribution of $N_{t_i} - N_{t_{i-1}}$ depends only on the difference $t_i - t_{i-1}$.

P^4 : For any $0 < t_1 < \cdots < t_k$ in $\mathbb{R}_+$ and $n_1, \ldots, n_k$ in $\mathbb{N}$ we have that

$$\lim_{h\downarrow 0} P\left(N_{t_{k+h}} - N_{t_k} = 1 \,\middle|\, N_{t_j} = n_j, j \leq k\right) = \lambda h + o(h),$$
$$\lim_{h\downarrow 0} P\left(N_{t_{k+h}} - N_{t_k} \geq 2 \,\middle|\, N_{t_j} = n_j, j \leq k\right) = o(h).$$

Furthermore, $(N_t)_{t\in\mathbb{R}_+}$ has no fixed discontinuities.

Theorem 2.170. *A process* $(N_t)_{t\in\mathbb{R}_+}$ *with stationary increments has a version in which it is constant on all sample paths except for upward jumps of magnitude* 1*, if and only if there exists a parameter* $\lambda > 0$ *so that its characteristic function*

$$\phi_{N_t}(u) = E\left[e^{iuN_t}\right] = e^{-\lambda t(1-\exp\{iu\})}$$

or, equivalently, $N_t \sim P(\lambda t)$*.*

Proof: See, e.g., Breiman (1968). □

Theorem 2.171. *Let $(N_t)_{t\in\mathbb{R}_+}$ be a Poisson process of intensity λ. Then $(N_t - \lambda t)_{t\in\mathbb{R}_+}$ and $((N_t - \lambda t)^2 - \lambda t)_{t\in\mathbb{R}_+}$ are martingales.*

Remark 2.172. Because $M_t = (N_t - \lambda t)^2 - \lambda t$ is a martingale, by uniqueness, the process $(\lambda t)_{t\in\mathbb{R}_+}$ is the predictable compensator of $(N_t - \lambda t)^2$, i.e., $\langle (N_t - \lambda t)^2 \rangle = \lambda t$, for all $t \in \mathbb{R}_+$, as well as the compensator of the Poisson process $(N_t)_{t\in\mathbb{R}_+}$ directly.

Theorem 2.173. *Let $(N_t)_{t\in\bar{\mathbb{R}}_+}$ be a simple counting process on $\mathbb{R}_+$ adapted to $\mathcal{F}_t$. If the $\mathcal{F}_t$-compensator A_t of N_t is continuous and $\mathcal{F}_0$-measurable, then N_t is a doubly stochastic Poisson process (with stochastic intensity), directed by A_t, also known as a Cox process.*

Proof: For $u \in \mathbb{R}$ let

$$M_t(u) = e^{iuN_t - (\exp\{iu\}-1)A_t}.$$

Then, by using the properties of stochastic integrals, it can be shown that

$$E[M_t(u)|\mathcal{F}_0] = E\left[e^{iuN_t - (\exp\{iu\}-1)A_t} \middle| \mathcal{F}_0 \right] = 1.$$

Because A_t is assumed to be $\mathcal{F}_0$-measurable

$$E\left[e^{iuN_t}|\mathcal{F}_0\right] = e^{(\exp\{iu\}-1)A_t},$$

representing the characteristic function of a Poisson distribution with (stochastic) intensity A_t. □

Lévy Process

Definition 2.174. Let $(X_t)_{t\in\mathbb{R}_+}$ be an adapted process with $X_0 = 0$ almost surely. If X_t

1. has independent increments,
2. has stationary increments,
3. is continuous in probability so that $X_s \xrightarrow[s\to t]{P} X_t$,

then it is a *Lévy process.*

Proposition 2.175. Both the Wiener and the Poisson processes are Lévy processes.

Theorem 2.176. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a Lévy process. Then it has an RCLL version $(Y_t)_{t\in\mathbb{R}_+}$, which is also a Lévy process.*

Proof: See, e.g., Kallenberg (1997). □

For Lévy processes we can invoke examples of filtrations that satisfy the usual hypotheses.

Theorem 2.177. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a Lévy process and $\mathcal{G}_t = \sigma(\mathcal{F}_t, \mathcal{N})$, where $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is the natural filtration of X_t and $\mathcal{N}$ the family of P-null sets of $\mathcal{F}_t$. Then $(\mathcal{G}_t)_{t\in\mathbb{R}_+}$ is right-continuous.*

Theorem 2.178. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a Lévy process and T a stopping time. Then the process $(Y_t)_{t\in\mathbb{R}_+}$, given by*

$$Y_t = X_{T+t} - X_T,$$

is a Lévy process on the set $]T,\infty[$, adapted to $\mathcal{F}_{T+t}$. Furthermore, Y_t is independent of $\mathcal{F}_T$ and has the same distribution as X_t.

Remark 2.179. Because, by Theorem 2.176, every Lévy process has an RCLL version, by Proposition 2.157, the only type of discontinuity it may admit are jumps.

Definition 2.180. Taking the left limit $X_{t^-} = \lim_{s\to t} X_s$, $s < t$, we define

$$\Delta X_t = X_t - X_{t^-}$$

as the jump at t. If $\sup_t |\Delta X_t| \leq c$ almost surely, $c \in \mathbb{R}_+$, constant and nonrandom, then X_t is said to have *bounded jumps.*

Theorem 2.181. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a Lévy process with bounded jumps. Then*

$$E\left[|X_t|^p\right] < \infty, \qquad \textit{i.e., } X_t \in \mathcal{L}^p \qquad \textit{for any } p \in \mathbb{N}^*.$$

Theorem 2.182. *Let $(X_t)_{t\in\mathbb{R}_+}$ be a Lévy process. Then it has an RCLL version without fixed jumps (see Proposition 2.157).*

Proof: See, e.g., Kallenberg (1997). □

We proceed with the general representation theorem of a Lévy process, commencing with the analysis of the structure of its jumps. Along the lines of the definition of counting and Poisson processes, let $\Lambda \in \mathcal{B}_{\mathbb{R}}$, such that 0 is not in $\bar{\Lambda}$, the closure of Λ. For a Lévy process $(X_t)_{t\in\mathbb{R}_+}$ we also, as before, define the random variables

$$\tau^{\Lambda}_{i+1} = \inf\left\{t > \tau^{\Lambda}_i | \Delta X_t \in \Lambda\right\}, \qquad i = 0,\ldots,n; \tau^{\Lambda}_0 \equiv 0.$$

Because $(X_t)_{t\in\mathbb{R}_+}$ has RCLL paths and $0 \notin \Lambda$ it is easy to demonstrate that

$$\left\{\tau^{\Lambda}_n \geq t\right\} \in \mathcal{F}_{t+} = \mathcal{F}_t,$$

thus $(\tau^{\Lambda}_i)_{i\in\mathbb{N}^*}$ are stopping times, and moreover $\tau^{\Lambda}_i > 0$ almost surely as well as $\lim_{n\to\infty} \tau^{\Lambda}_n = +\infty$ almost surely. If we now define

$$N_t^{\Lambda} = \sum_{0<s\leq t} I_{\{\Lambda\}}(\Delta X_s) \equiv \sum_{i=1}^{\infty} I_{\left[\tau_i^{\Lambda}\leq t\right]}(t),$$

then $(N_t^{\Lambda})_{t\in\mathbb{R}_+}$ is a nonexplosive counting process and, more specifically, we have the following theorem.

Theorem 2.183. *Let $\Lambda \in \mathcal{B}_{\mathbb{R}}$, with $0 \notin \bar{\Lambda}$. Then $(N_t^{\Lambda})_{t\in\mathbb{R}_+}$ is a time-homogeneous Poisson process with intensity*

$$\nu(\Lambda) = E\left[N_1^{\Lambda}\right].$$

Remark 2.184. If the Lévy process $(X_t)_{t\in\mathbb{R}_+}$ has bounded jumps, then $\nu(\Lambda) < +\infty$.

Theorem 2.185. *For any $t \in \mathbb{R}_+$ the mapping*

$$\Lambda \to N_t(\Lambda) \equiv N_t^{\Lambda}, \qquad \Lambda \in \mathcal{B}_{\mathbb{R}}; 0 \notin \bar{\Lambda},$$

is a random (counting) measure. Furthermore the mapping

$$\Lambda \to \nu(\Lambda), \qquad \Lambda \in \mathcal{B}_{\mathbb{R}}; 0 \notin \bar{\Lambda},$$

is a σ-finite measure.

Proof: See, e.g., Protter (1990). □

Definition 2.186. The measure ν given by

$$\nu(\Lambda) = E\left[\sum_{0<s\leq 1} I_{\{\Lambda\}}(\Delta X_s)\right], \qquad \Lambda \in \mathcal{B}_{\mathbb{R}\setminus\{0\}},$$

is called the *Lévy measure* of the Lévy process $(X_t)_{t\in\mathbb{R}_+}$.

Theorem 2.187. *Under the assumptions of Theorem 2.185, let f be a measurable function, finite on Λ. Then*

$$\int_{\Lambda} f(x) N_t(dx) = \sum_{0<s\leq t} f(\Delta X_s) I_{\{\Lambda\}}(\Delta X_s).$$

Because by Theorem 2.183 $(N_t^{\Lambda})_{t\in\mathbb{R}_+}$ is a time-homogeneous Poisson process, we also have the following proposition.

Proposition 2.188. Under the assumptions of Theorem 2.187, the process $(\int_{\Lambda} f(x) N_t(dx))_{t\in\mathbb{R}_+}$ is a Lévy process. In particular, if $f(x) = x$, then the process is nonexplosive almost surely for any $t \in \mathbb{R}_+$.

Theorem 2.189. *Let* $\Lambda \in \mathcal{B}_{\mathbb{R}}$, $0 \notin \bar{\Lambda}$. *Then the process*

$$\left(X_t - \int_{\Lambda} f(x) N_t(dx)\right)_{t \in \mathbb{R}_+}$$

is a Lévy process.

Now, if we define

$$J_t = \int_{\{|x| \geq 1\}} x N_t(dx) = \sum_{0 < s \leq t} \Delta X_s I_{\{|\Delta X_s| \geq 1\}}(|\Delta X_s|),$$

then because $(X_t)_{t \in \mathbb{R}_+}$ has RCLL paths for each $\omega \in \Omega$, its trajectory has only finitely many jumps bigger than 1 during the interval $[0, t]$. Therefore $(J_t)_{t \in \mathbb{R}_+}$ has paths of finite variation on compacts.

Both $(J_t)_{t \in \mathbb{R}_+}$ (by Proposition 2.188) and $V_t = X_t - J_t$ (by Theorem 2.189) are Lévy processes, where in particular the latter has jumps bounded by 1. Hence all moments of $(V_t)_{t \in \mathbb{R}_+}$ exist and are finite. Because $E[V_1] = \mu$ (and $E[V_0] = 0$), we have $E[V_t] = \mu t$, by the stationarity of the increments. If we define $Y_t = V_t - E[V_t]$, for all $t \in \mathbb{R}_+$, then $(Y_t)_{t \in \mathbb{R}_+}$ has independent increments and mean zero. Hence it is a martingale. If we further define $Z_t = J_t + \mu t$, then the following decomposition theorem holds.

Theorem 2.190. *Let* $(X_t)_{t \in \mathbb{R}_+}$ *be a Lévy process. Then it can be decomposed as*

$$X_t = Y_t + Z_t,$$

where Y_t *and* Z_t *are both Lévy processes and, furthermore,* Y_t *is a martingale with bounded jumps and* $Y_t \in L^p$, *for all* $p \geq 1$, *whereas* Z_t *has trajectories of finite variation on compacts.*

Proposition 2.191. Let $(X_t)_{t \in \mathbb{R}_+}$ be a Lévy process and ν its Lévy measure. Then for any $a \in \mathbb{R}_+^*$

$$Z_t = \int_{\{|x| < a\}} x[N_t(dx) - t\nu(dx)] \tag{2.39}$$

is a zero mean martingale.

By Theorem 2.171 the process $(N_t - \lambda t)_{t \in \mathbb{R}_+}$ is also a zero mean martingale and $(Z_t)_{t \in \mathbb{R}_+}$ can be interpreted as a mixture of compensated Poisson processes.

Theorem 2.192. *Let* $(X_t)_{t \in \mathbb{R}_+}$ *be a Lévy process with jumps bounded by* $a \in \mathbb{R}_+^*$ *and let*

$$V_t = X_t - E[X_t] \qquad \forall t \in \mathbb{R}_+.$$

Then $(V_t)_{t\in\mathbb{R}_+}$ *is a zero mean martingale that can be decomposed as*

$$V_t = Z_t^c + Z_t \qquad \forall t \in \mathbb{R}_+,$$

where Z_t^c *is a martingale with continuous paths and* Z_t *as defined in (2.39). In fact,* $Z_t^c = W_t$ *is Brownian motion.*

Theorem 2.192 can be interpreted by saying that a Lévy process with bounded jumps can be decomposed as the sum of a continuous martingale (Brownian motion) and another martingale that is a mixture of compensated Poisson processes. More generally, a third component would be due to the presence of unbounded jumps.

Theorem 2.193. (Lévy decomposition). *Let* $(X_t)_{t\in\mathbb{R}_+}$ *be a Lévy process and* $\mu \in \mathbb{R}$. *Then*

$$X_t = W_t + \int_{\{|x|<1\}} x[N_t(dx) - t\nu(dx)] + \mu t + \sum_{0<s\leq t} \Delta X_s I_{\{|\Delta X_s|\geq 1\}}(|\Delta X_s|),$$

where

1. W_t *is Brownian motion,*
2. *for any set* $\Lambda \in \mathcal{B}_{\mathbb{R}\setminus\{0\}}$, $0 \notin \bar{\Lambda}$:
   - $N_t^\Lambda \equiv \int_\Lambda N_t(dx)$ *is a Poisson process independent of* W_t,
   - N_t^Λ *is independent of* $N_t^{\Lambda'}$ *if* $\Lambda \cap \Lambda' = \emptyset$,
   - N_t^Λ *has intensity* $\nu(\Lambda)$,
   - $\nu(dx)$ *is a measure on* $\mathbb{R}\setminus\{0\}$ *such that* $\int \min\{1, x^2\}\nu(dx) < \infty$.

Proof: See, e.g., Jacod and Shiryaev (1987). □

Theorem 2.194. (Lévy–Khintchine formula). *Under the assumptions of Theorem 2.193 let* $u \in \mathbb{R}$ *and*

$$\phi_{X_t}(u) = E\left[e^{iuX_t}\right] = \exp\{-t\psi(u)\}, \tag{2.40}$$

where

$$\psi(u) = \frac{1}{2}\sigma^2 u^2 - i\mu u + \int_{\{|x|<1\}} (1 - \exp\{iux\} + iux)\nu(dx) + \int_{\{|x|\geq 1\}} (1 - \exp\{iux\})\nu(dx).$$

For given ν, σ^2, μ *there exists a unique probability measure on the probability space of* X_t *under which (2.40) is the characteristic function of a Lévy process* $(X_t)_{t\in\mathbb{R}_+}$.

Proof: See, e.g., Bertoin (1996). □

Stable Lévy Process

As a corollary we will briefly mention stable processes. They are a particular family of Lévy processes that reproduce themselves under linear combinations. Hence a number of processes already discussed fall into this category.

Definition 2.195. A random variable X has a *stable distribution* if there exists a number $n \geq 2$ of independently identically distributed random variables Y_i, $i = 1, \ldots, n$, as well as numbers $a_n \in \mathbb{R}$, $b_n \in \mathbb{R}_+$, such that

$$\sum_{i=1}^{n} Y_i \sim a_n + b_n X.$$

In fact, it can be demonstrated (see, e.g., Samorodnitsky and Taqqu (1994)) that this is equivalent to what follows.

Definition 2.196. A random variable X has a stable distribution, if its characteristic function is of the form

$$E\left[e^{iuX}\right] = \begin{cases} e^{-\sigma^\alpha |u|^\alpha \left(1 - i\beta \operatorname{sgn}\{u\} \tan \frac{\pi\alpha}{2}\right) + i\mu u} & \text{for } \alpha \in]0,2] \setminus \{1\}, \\ e^{-\sigma|u|\left(1 + i\beta \frac{2}{\pi} \operatorname{sgn}\{u\} \ln |u|\right) + i\mu u} & \text{for } \alpha = 1, \end{cases}$$

where $\beta \in [-1, 1]$, $\sigma^2 \in \mathbb{R}_+$, and $\mu \in \mathbb{R}$ are unique. We also say that X is α-stable.

Remark 2.197. For $\alpha = 2$ we obtain the characteristic function of a Gaussian distribution implying that the latter is also stable. Further well-known distributions can be obtained for $\alpha = 1$, $\beta = 0$ (Cauchy), $\alpha = 0.5$, $\beta = 1$ (Lévy), and $\alpha = \beta = \sigma = 0$ (constant). For other parameter ranges no closed-form formulae for the densities exist.

Definition 2.198. A stochastic process $(X_t)_{t \in T}$ is stable if for $t_i \in T$, $i = 1, \ldots, n$, the densities of $(X_{t_1}, \ldots, X_{t_n})$ are stable.

Remark 2.199. A stable process is a Lévy process.

Corollary 2.200. If $\alpha \in]0, 2] \setminus \{1\}$ and $\beta = 0$, then a Lévy stable process $(X_t)_{t \in T}$ has the *scaling property*, i.e., the rescaled process $(t^{\frac{1}{\alpha}} X_1)_{t \in T}$ has the same probability law as $(X_t)_{t \in T}$. This is a generalization of the specific case for the Wiener process of Proposition 2.148.

2.10 Marked Point Processes

We will now generalize the notion of a compensator (see Definition 2.80) to a larger class of counting processes, including the so-called marked point processes. For this we will commence with a point process on $\mathbb{R}_+$,

$$N = \sum_{n \in \mathbb{N}^*} \epsilon_{\tau_n},$$

defined by the sequence of random times $(\tau_n)_{n \in \mathbb{N}^*}$ on the underlying probability space $(\Omega, \mathcal{F}, P)$. Here ϵ_t is the Dirac measure (also called point mass) on $\mathbb{R}_+$, i.e.,

$$\forall A \in \mathcal{B}_{\mathbb{R}_+}: \qquad \epsilon_t(A) = \begin{cases} 1 & \text{if } t \in A, \\ 0 & \text{if } t \notin A. \end{cases}$$

The corresponding definition of the same process as a counting process was given in Definition 2.155.

Definition 2.201. (A^*): Let $\mathcal{F}_t = \sigma(N_s, 0 \leq s \leq t)$, $t \in \mathbb{R}_+$, be the natural filtration of the counting process $(N_t)_{t \in \mathbb{R}_+}$. We assume that

1. the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}_+}, P)$ satisfies the usual hypotheses (see Definition 2.34);
2. $E[N_t] < \infty$, for all $t \in \mathbb{R}_+$, i.e., avoiding the problem of exploding martingales in the Doob–Meyer decomposition (see Theorem 2.87).

Proposition 2.202. *Under assumption (A^*) of Definition 2.201 there exists a unique increasing right-continuous predictable process $(A_t)_{t \in \mathbb{R}_+}$, such that*

1. $A_0 = 0$,
2. $P(A_t < \infty) = 1$ *for any* $t > 0$,
3. *the process $(M_t)_{t \in \mathbb{R}_+}$ defined as $M_t = N_t - A_t$ is a right-continuous zero mean martingale.*

The process $(A_t)_{t \in \mathbb{R}_+}$ is called the compensator of the process $(N_t)_{t \in \mathbb{R}_+}$.

Proposition 2.203. (See Bremaud (1981), Karr (1986).) *For every nonnegative $\mathcal{F}_t$-predictable process $(C_t)_{t \in \mathbb{R}_+}$, by Proposition 2.202, we have that*

$$E\left[\int_t^\infty C_t dN_t\right] = E\left[\int_0^\infty C_t dA_t\right]. \tag{2.41}$$

Theorem 2.204. *Given a point (or counting) process $(N_t)_{t \in \mathbb{R}_+}$ satisfying assumption (A^*) of Definition 2.201 and a predictable random process $(A_t)_{t \in \mathbb{R}_+}$, the following two statements are equivalent:*

1. *$(A_t)_{t \in \mathbb{R}_+}$ is the compensator of $(N_t)_{t \in \mathbb{R}_+}$.*
2. *The process $M_t = N_t - A_t$ is a zero mean martingale.*

Remark 2.205. In infinitesimal form (2.41) provides the heuristic expression

$$dA_t = E[dN_t | \mathcal{F}_{t-}],$$

giving a dynamical interpretation to the compensator. In fact, the increment $dM_t = dN_t - dA_t$ is the unpredictable part of dN_t over $[0, t[$, also therefore known as the *innovation martingale* of $(N_t)_{t \in \mathbb{R}_+}$.

In the case where the innovation martingale M_t is bounded in L^2 we may apply Theorem 2.88 and introduce the predictable variation process $\langle M \rangle_t$, with $\langle M \rangle_0 = 0$ and $M_t^2 - \langle M \rangle_t$ being a uniformly integrable martingale. Then the variation process can be compensated in terms of A_t by the following.

Theorem 2.206. (See Karr (1986), page 64.) *Let $(N_t)_{t\in\mathbb{R}_+}$ be a point process on $\mathbb{R}_+$ with compensator $(A_t)_{t\in\mathbb{R}_+}$ and let the innovation process $M_t = N_t - A_t$ be an L^2-martingale. Defining $\Delta A_t = A_t - A_{t-}$, then*

$$\langle M \rangle_t = \int_0^t (1 - \Delta A_s) dA_s.$$

Remark 2.207. In particular, if A_t is continuous in t, then $\Delta A_t = 0$, so that $\langle M \rangle_t = A_t$. Formally in this case we have

$$E\left[(dN_t - E[dN_t|\mathcal{F}_{t-}])^2 |\mathcal{F}_{t-}\right] = dA_t = E[dN_t|\mathcal{F}_{t-}],$$

so that the counting process has locally and conditionally the typical behavior of a Poisson process.

Stochastic Intensities

Let N be a simple point process on $\mathbb{R}_+$ with a compensator A, satisfying the assumptions of Proposition 2.202.

Definition 2.208. We say that N admits an $\mathcal{F}_t$-stochastic intensity if a (non-trivial) nonnegative, predictable process $\lambda = (\lambda_t)_{t\in\mathbb{R}_+}$ exists, such that

$$A_t = \int_0^t \lambda_s ds, \qquad t \in \mathbb{R}_+.$$

Remark 2.209. Due to the uniqueness of the compensator, the stochastic intensity, whenever it exists, is unique.

Formally, from

$$dA_t = E[dN_t|\mathcal{F}_{t-}]$$

it follows that

$$\lambda_t dt = E[dN_t|\mathcal{F}_{t-}],$$

i.e.,

$$\lambda_t dt = \lim_{\Delta t \to 0+} \frac{1}{\Delta t} E[\Delta N_t|\mathcal{F}_{t-}]$$

and, because of the simplicity of the process, we also have

$$\lambda_t dt = \lim_{\Delta t \to 0+} \frac{1}{\Delta t} P(\Delta N_t = 1|\mathcal{F}_{t-}),$$

meaning that $\lambda_t dt$ is the conditional probability of a new *event* during $[t, t+dt]$, given the history of the process over $[0, t]$.

Example 2.210. (Poisson process). A stochastic intensity does exist for a Poisson process with intensity $(\lambda_t)_{t\in\mathbb{R}_+}$ and, in fact, is identically equal to the latter (hence deterministic).

A direct consequence of Theorem 2.206 and of the previous definitions is the following theorem.

Theorem 2.211. (Karr (1986), page 64). *Let $(N_t)_{t\in\mathbb{R}_+}$ be a point process satisfying assumption (A^*) of Definition 2.201 and admitting stochastic intensity $(\lambda_t)_{t\in\mathbb{R}_+}$. Assume further that the innovation martingale*

$$M_t = N_t - \int_0^t \lambda_s ds, \qquad t \in \mathbb{R}_+,$$

is an L^2-martingale. Then for any $t \in \mathbb{R}_+$:

$$\langle M \rangle_t = \int_0^t \lambda_s ds.$$

An important theorem that further explains the role of the stochastic intensity for counting processes is the following (see Karr (1986), page 71).

Theorem 2.212. *Let $(\Omega, \mathcal{F}, P)$ be a probability space over which a simple point process with an $\mathcal{F}_t$-stochastic intensity $(\lambda_t)_{t\in\mathbb{R}_+}$ is defined. Suppose that P_0 is another probability measure on $(\Omega, \mathcal{F})$ with respect to which $(N_t)_{t\in\mathbb{R}_+}$ is a stationary Poisson process with rate 1. Then $P << P_0$ and for any $t \in \mathbb{R}_+$ we have*

$$\frac{dP}{dP_0}|_{\mathcal{F}_t} = \exp\left\{\int_0^t (1-\lambda_s)ds + \int_0^t \ln \lambda_s dN_s\right\}. \tag{2.42}$$

Conversely, if P_0 is as above and P a probability measure on $(\Omega, \mathcal{F})$, absolutely continuous with respect to P_0, then there exists a predictable process λ, such that N has stochastic intensity λ with respect to P (and equation (2.42) holds).

Marked Point Processes

We will now consider a generic Polish space endowed with its σ-algebra $(E, \mathcal{E})$ and introduce a sequence of $(E, \mathcal{E})$-valued random variables $(Z_n)_{n\in\mathbb{N}^*}$ in addition to the sequence of random times $(\tau_n)_{n\in\mathbb{N}^*}$, which are $\bar{\mathbb{R}}_+$-valued random variables.

Definition 2.213. The random measure on $\bar{\mathbb{R}}_+ \times E$,

$$N = \sum_{n\in\mathbb{N}^*} \epsilon_{(\tau_n, z_n)},$$

is called a *marked point process* with *mark space* $(E, \mathcal{B})$. z_n is called the *mark* of the event occurring at time τ_n. The process

$$N_t = N([0,t] \times E), \qquad t \in \mathbb{R}_+,$$

is called the *underlying counting process* of the process N. As usual, we assume that the process $(N_t)_{t\in\mathbb{R}_+}$ is simple.

For $B \in \mathcal{E}$ the process

$$N_t(B) := N([0,t] \times B) = \sum_{n\in\mathbb{N}^*} I_{[\tau_n \le t, Z_n \in B]}(t), \qquad t \in \mathbb{R}_+,$$

represents the counting process of events occurring up to time t with marks in $B \in \mathcal{E}$. The *history* of the process up to time t is denoted as

$$\mathcal{F}_t := \sigma(N_s(B)|0 \le s \le t, B \in \mathcal{E}).$$

We will assume throughout that the filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+}, P)$ satisfies the usual hypotheses (see Definition 2.34).

Remark 2.214. Note that, for any $n \in \mathbb{N}^*$, while τ_n is $\mathcal{F}_{\tau_n -}$-measurable, Z_n is $\mathcal{F}_{\tau_n}$-measurable but not $\mathcal{F}_{\tau_n -}$-measurable; i.e.,

$$\mathcal{F}_{\tau_n} = \sigma\left((\tau_1, Z_1), \ldots, (\tau_n, Z_n)\right),$$

whereas

$$\mathcal{F}_{\tau_n -} = \sigma\left((\tau_1, Z_1), \ldots, (\tau_{n-1}, Z_{n-1}), \tau_n\right).$$

Hence τ_n is an $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ stopping time.

By a reasoning similar to the one employed for regular conditional probabilities in chapter 1, the following theorem can be proved, which provides an extension of Theorem 2.204 to marked point processes.

Theorem 2.215. (Bremaud (1981), Karr (1986), Last and Brandt (1995).) *Let N be a marked point process such that the underlying counting process $(N_t)_{t\in\mathbb{R}_+}$ satisfies the assumptions of Proposition 2.202. Then there exists a unique random measure ν on $\mathbb{R}_+ \times E$ such that*

1. *for any $B \in \mathcal{E}$, the process $\nu([0,t] \times B)$ is $\mathcal{F}_t$-predictable;*
2. *for any nonnegative $\mathcal{F}_t$-predictable process C on $\mathbb{R}_+ \times E$:*

$$E\left[\int C(t,z) N(dt \times dz)\right] = E\left[\int C(t,z) \nu(dt \times dz)\right].$$

The random measure ν introduced in the preceding theorem is called the $\mathcal{F}_t$-compensator of the process N. The above theorem again suggests that formally the following holds:

$$\nu(dt \times dz) = E\left[N(dt \times dz)|\mathcal{F}_{t-}\right].$$

The following propositions mimic the corresponding results for the unmarked point processes.

Proposition 2.216. *For any $B \in \mathcal{E}$, the process*

$$M_t(B) := N_t(B) - \nu([0,t] \times B), \qquad t \in \mathbb{R}_+,$$

is a zero-mean martingale.

We will call the process $M = (M_t(B))_{t\in\mathbb{R}_+, B\in\mathcal{E}}$ the innovation process of N. From now on let us denote $A_t(B) := \nu([0,t] \times B)$.

Proposition 2.217. (Karr (1986), page 65.) *Let N be a marked point process on $\mathbb{R}_+ \times E$, with compensator ν, and let B_1 and B_2 be two disjoint sets in $\mathcal{E}$ for which $M_t(B_1)$ and $M_t(B_2)$ are L^2- martingales. Then*

$$\langle M_t(B_1), M_t(B_2)\rangle_t = -\int_0^t \Delta A_s(B_1)\Delta A_s(B_2) ds.$$

Hence, if $(A_t(B))_{t\in\mathbb{R}_+}$ is continuous in t for any $B \in \mathcal{E}$, the two martingales $M_t(B_1)$ and $M_t(B_2)$ are orthogonal.

Definition 2.218. Let N be a marked point process on $\mathbb{R}_+ \times E$. We say that $(\lambda_t(B))_{t\in\mathbb{R}_+, B\in\mathcal{E}}$ is the $\mathcal{F}_t$-stochastic intensity of N provided that,

1. for any $t \in \mathbb{R}_+$, the map
$$B \in \mathcal{E} \to \lambda_t(B) \in \mathbb{R}_+$$
is a random measure on $\mathcal{E}$;
2. for any $B \in \mathcal{E}$ the process $(\lambda_t(B))_{t\in\mathbb{R}_+}$ is the stochastic intensity of the counting process
$$N_t(B) = \sum_{n\in\mathbb{N}^*} I_{[\tau_n \le t, Z_n \in B]}(t);$$

i.e., for any $t \in \mathbb{R}_+, B \in \mathcal{E}$:

$$A_t(B) = \int_0^t \lambda_s(B) ds,$$

in which case the process $(A_t(B))_{t\in\mathbb{R}_+, B\in\mathcal{E}}$ is known as the *cumulative stochastic intensity* of N.

In the presence of the absolute continuity (hence the continuity) of the process $A_t(B)$ as a function of t, the following is an obvious consequence of Proposition 2.217.

Proposition 2.219. *Let N be a marked point process on $\mathbb{R}_+$ with mark space $(E, \mathcal{E})$ and stochastic intensity $(\lambda_t(B))_{t\in\mathbb{R}_+, B\in\mathcal{E}}$. Let B_1 and B_2 be two disjoint sets in $\mathcal{E}$, such that the corresponding innovation martingales are bounded in L^2. Then $M(B_1)$ and $M(B_2)$ are orthogonal; i.e.,*

$$\langle M_t(B_1), M_t(B_2)\rangle_t = 0 \qquad \text{for any } t \in \mathbb{R}_+.$$

Representation of Point Process Martingales

Let N be a point process on $\mathbb{R}_+$ with $\mathcal{F}$-compensator A. From the section on martingales, we know that, if $M = N - A$ is the innovation martingale of N and H is a bounded predictable process, then

$$\tilde{M}_t = \int_0^t H(s) dM_s, \qquad t \in \mathbb{R}_+,$$

is also a martingale. In fact, the converse also holds, as stated by the following theorem, which extends an analogous result for Wiener processes to marked point processes.

Theorem 2.220. (Martingale representation.) *Let N be a marked point process on $\mathbb{R}_+$ with mark space $(E, \mathcal{E})$, and let M be its innovation process with respect to the internal history $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. Suppose the assumptions of Proposition 2.202 are satisfied and let $(\tilde{M}_t)_{t\in\mathbb{R}_+}$ be a right-continuous and uniformly integrable $\mathcal{F}_t$-martingale. Then there exists a process $(H(t,x))_{t\in\mathbb{R}_+, x\in E}$, such that*

$$\tilde{M}_t = \tilde{M}_0 + \int_{[0,t]\times E} H(s,x) M_s(dx).$$

Proof: See Last and Brandt (1995), page 342. □

The Marked Poisson Process

A marked Poisson process is a marked point process, such that any univariate point process counting its points with a mark in a fixed Borel set is Poisson. It turns out that these processes are necessarily independent whenever the corresponding mark sets are disjoint. Consider a marked point process N on $\mathbb{R}_+ \times E$ and let Λ be a σ-finite deterministic measure on $\mathbb{R}_+ \times E$. Then, formally, we have the following definition.

Definition 2.221. N is a marked Poisson process if, for any $s, t \in \mathbb{R}_+, s < t$ and any $B \in \mathcal{E}$,

$$P(N(]s,t] \times B) = k | \mathcal{F}_s) = \frac{(\Lambda(]s,t] \times B))^k}{k!} \exp\{-\Lambda(]s,t] \times B)\},$$

for $k \in \mathbb{N}$, almost surely with respect to P.

In the preceding case the intensity measure Λ is such that

$$\Lambda(]s,t] \times B) = E[N(]s,t] \times B)]$$

for any $s, t \in \mathbb{R}_+, s < t$, and any $B \in \mathcal{E}$. It is the (deterministic) compensator of the marked Poisson process, formally:

$$\Lambda(dt \times dx) = E[N(dt \times dx) | \mathcal{F}_{t-}] = E[N(dt \times dx)],$$

thus confirming the independence of increments for the marked Poisson process. Now the following theorem is a consequence of the definitions.

Theorem 2.222. *Let N be a marked Poisson process and $B_1, \ldots, B_m \in \mathcal{E}$, for $m \in \mathbb{N}^*$ mutually disjoint sets. Then $N(\cdot \times B_1), \ldots, N(\cdot \times B_m)$ are independent Poisson processes with intensity measures $\Lambda(\cdot \times B_1), \ldots, \Lambda(\cdot \times B_m)$, respectively.*

Proof: See Last and Brandt (1995), page 182. □

The underlying counting process of a marked Poisson process $N(]0, t] \times E)$ is itself a univariate Poisson process with intensity measure $\bar{\Lambda}(]s, t]) = \Lambda(]s, t] \times E)$ for any $s, t \in \mathbb{R}_+, s < t$. The intensity measure may be chosen to be continuous, in which case $\bar{\Lambda}(\{t\}) = 0$, or even absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}_+$, so that

$$\bar{\Lambda}([0, t]) = \int_0^t \lambda(s) ds,$$

where $\lambda \in \mathcal{L}^1(\mathbb{R}_+)$.

2.11 Exercises and Additions

2.1. Let $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ be a filtration on the measurable space $(\Omega, \mathcal{F})$. Show that $\mathcal{F}_{t^+} = \bigcap_{u>t} \mathcal{F}_u$ is a σ-algebra (see Theorem 2.107 and Remark 2.108).

2.2. Prove that two processes that are modifications of each other are equivalent.

2.3. A real-valued stochastic process, indexed in $\mathbb{R}$, is *strictly stationary*, if and only if all its joint finite-dimensional distributions are invariant under a parallel time shift; i.e.,

$$F_{X_{t_1}, \ldots, X_{t_n}}(x_1, \ldots, x_n) = F_{X_{t_1+h}, \ldots, X_{t_n+h}}(x_1, \ldots, x_n)$$

for any $n \in \mathbb{N}$, any choice of $t_1, \ldots, t_n \in \mathbb{R}$ and $h \in \mathbb{R}$, and any $x_1, \ldots, x_n \in \mathbb{R}$.

1. Prove that a process of independent and identically distributed random variables is strictly stationary.
2. Prove that a time-homogeneous process with independent increments is strictly stationary.
3. Prove that a Gaussian process $(X_t)_{t \in \mathbb{R}}$ is strictly stationary if and only if the following two conditions hold:
   (a) $E[X_t] = constant$ for any $t \in \mathbb{R}$;
   (b) $Cov[s, t] = K(t - s)$ for any $s, t \in \mathbb{R},\ s < t$.

2.4. An L^2 real-valued stochastic process indexed in $\mathbb{R}$ is *weakly stationary* if and only if the following two conditions hold:

(a) $E[X_t] = constant$ for any $t \in \mathbb{R}$;

(b) $Cov[s,t] = K(t-s)$ for any $s, t \in \mathbb{R},\ s < t$.

1. Prove that an L^2 strictly stationary process is also weakly stationary.
2. Prove that a weakly stationary Gaussian process is also strictly stationary.

2.5. Show that Brownian motion is not stationary.

2.6. (*prediction*) Let $(X_{r-j}, \ldots, X_r)$ be a family of random variables representing a sample of a (weakly) stationary stochastic process in L^2. We know that the best approximation in L^2 of an additional random variable X_{r+s}, for any $s \in \mathbb{N}^*$, in terms of $(X_{r-j}, \ldots, X_r)$ is given by $E[Y|X_{r-j}, \ldots, X_r]$. To evaluate this quantity is generally a hard task. On the other hand, the problem of the best linear approximation can be handled in terms of the covariances of the random variables $X_{r-j}, \ldots, X_r, X_{r+s}$ as follows.

Prove that the best approximation of X_{r+s} in terms of a linear function of $(X_{r-j}, \ldots, X_r)$, is given by

$$\widehat{X}_{r+s} = \sum_{k=0}^{j} a_k X_{r-k},$$

where the a_k satisfy the linear system

$$\sum_{k=0}^{j} a_k c(|k-i|) = c(s+i) \text{ for } 0 \leq i \leq j.$$

Here we have denoted $c(m) = Cov[X_i, X_{i+m}]$.

2.7. Refer to Proposition 2.46. Prove that $\mathcal{F}_T$ is a σ-algebra of the subsets of Ω.

2.8. Prove all the statements of Theorem 2.47.

2.9. Prove Lemma 2.117 by considering the sequence

$$T_n = \sum_{k=1}^{\infty} k 2^{-n} I_{(k-1)2^{-n} \leq T \leq k 2^{-n}}.$$

2.10. Let $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$ be a filtration and prove that T is a stopping time, if and only if the process $X_t = I_{\{T \leq t\}}$ is adapted to $(\mathcal{F}_t)_{t \in \mathbb{R}_+}$. Show that if T and S are stopping times, then so is $T + S$.

2.11. Show that any (sub- or super-) martingale remains a (sub- or super-) martingale with respect to the induced filtration.

2.12. Let $(X_t)_{t \in \mathbb{R}_+}$ be a martingale in L^2. Show that its increments on nonoverlapping intervals are orthogonal.

2.13. Prove Proposition 2.93. (*Hint:* To prove that 1⇒2 it suffices to use the indicator function on B; to prove that 2⇒1 it should first be shown for simple measurable functions, and then the theorem of approximation of measurable functions through elementary functions is invoked.)

2.14. Prove Remark 2.105.

2.15. Verify Example 2.129.

2.16. Determine the infinitesimal generator of a time-homogeneous Poisson process.

2.17. We say that $(Z_t)_{t\in\mathbb{R}_+}$ is a *compound Poisson process* if it can be expressed as

$$Z_0 = 0$$

and

$$Z_t = \sum_{k=1}^{N_t} Y_k \text{ for } t > 0,$$

where N_t is a Poisson process with intensity parameter $\lambda \in \mathbb{R}_+^*$ and $(Y_k)_{k\in\mathbb{N}^*}$ is a family of independent and identically distributed random variables, independent of N_t. Show that the compound Poisson process $(Z_t)_{t\in\mathbb{R}_+}$ is a stochastic process with time-homogeneous (stationary) independent increments.

2.18. Show that

1. The Brownian motion and the compound Poisson process are both almost surely continuous at any $t \geq 0$.
2. The Brownian motion is sample continuous, but the compound Poisson process is not sample continuous.

Hence almost sure continuity does not imply sample continuity.

2.19. In the compound Poisson process, assume that the random variables Y_n are independent and identically distributed with common distribution

$$P(Y_n = a) = P(Y_n = -a) = \frac{1}{2},$$

where $a \in \mathbb{R}_+^*$.

1. Find the characteristic function ϕ of the process $(Z_t)_{t\in\mathbb{R}_+}$.
2. Discuss the limiting behavior of the characteristic function ϕ when $\lambda \to +\infty$ and $a \to +\infty$ in such a way that the product λa^2 is constant.

2.20. An integer-valued stochastic process $(N_t)_{t\in\mathbb{R}_+}$ with stationary (time-homogeneous) independent increments is called a *generalized Poisson process*.

1. Show that the characteristic function of a generalized Poisson process necessarily has the form

$$\phi_{N_t}(u) = e^{\lambda t[\phi(u)-1]}$$

for some $\lambda \in \mathbb{R}_+^*$ and some characteristic function ϕ of a nonnegative integer valued random variable. The Poisson process corresponds to the degenerate case $\phi(u) = e^{iu}$.

2. Let $(N_t^{(k)})_{t\in\mathbb{R}_+}$ be a sequence of independent Poisson processes with respective parameters λ_k. Assume that $\lambda = \sum_{k=1}^{+\infty} \lambda_k < +\infty$. Show that the process

$$N_t^{(k)} = \sum_{k=1}^{+\infty} kN_t^{(k)}, \ t \in \mathbb{R}_+,$$

is a generalized Poisson process, with characteristic function

$$\phi(u) = \sum_{k=1}^{+\infty} \frac{\lambda_k}{\lambda} e^{iku}.$$

3. Show that any generalized Poisson process can be represented as a compound Poisson process. Vice versa, if the random variables Y_k in the compound Poisson process are integer valued, then the process is a generalized Poisson process.

2.21. Let $(X_n)_{n\in\mathbb{N}} \subset E$ be a *Markov chain*, i.e., a discrete-time Markov jump process, where E is a countable set. Let $i, j \in E$ be *states* of the process; j is said to be *accessible* from state i if for some integer $n \geq 0$, $p_{ij}(n) > 0$: i.e., state j is accessible from state i if there is positive probability that in a finite number of transition states j can be reached starting from state i. Two states i and j, each accessible to the other, are said to *communicate*, and we write $i \leftrightarrow j$. If two states i and j do not communicate, then either

$$p_{ij}(n) = 0 \quad \forall n \geq 0$$

or

$$p_{ji}(n) = 0 \quad \forall n \geq 0,$$

or both relations are true.

We define the period of state i, written $d(i)$, as the greatest common divisor of all integers $n \geq 1$ for which $p_{ii}(n) > 0$ (if $p_{ii}(n) = 0$ for all $n \geq 1$, define $d(i) = 0$).

1. Show that the concept of communication is an equivalence relationship.
2. Show that, if $i \leftrightarrow j$, then $d(i) = d(j)$.

3. Show that if state i has period $d(i)$, then there exists an integer N depending on i such that for all integers $n \geq N$

$$p_{ii}(nd(i)) > 0.$$

2.22. 1. Consider two urns A and B containing a total of N balls. A ball is selected at random (all selections are equally likely) at time $t = 1, 2, \ldots$ from among the N balls. The drawn ball is placed with probability p in urn A and with probability $q = 1 - p$ in urn B. The state of the system at each trial is represented by the number of balls in A. Determine the transition matrix for this Markov chain.

2. Assume that at each time t there are exactly k balls in A. At time $t+1$ an urn is selected at random proportionally to its content (i.e., A is chosen with probability k/N and B with probability $(N-k)/N$). Then a ball is selected either from A with probability p or from B with probability $1-p$ and placed in the previously chosen urn. Determine the transition matrix for this Markov chain.
3. Now assume that at time $t+1$ a ball and an urn are chosen with probability depending on the contents of the urn (i.e., a ball is chosen from A with probability $p = k/N$ or from B with probability q. Urn A is chosen with probability p and B with probability q). Determine the transition matrix of the Markov chain.
4. Determine the equivalence classes in parts 1, 2, and 3.

2.23. Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain whose transition probabilities are $p_{ij} = 1/[e(j-i)!]$ for $i = 0, 1, \ldots$ and $j = i, i+1, \ldots$. Verify the martingale property for

- $Y_n = X_n - n$,
- $U_n = Y_n^2 - n$,
- $V_n = \exp\{X_n - n(e-1)\}$.

2.24. Let $(X_t)_{t\in\mathbb{R}_+}$ be a process with the following property:

- $X_0 = 0$;
- for any $0 \leq t_0 < t_1 < \cdots < t_n$, the random variables $X_{t_k} - X_{t_{k-1}}$ $(1 \leq k \leq n)$ are independent;
- if $0 \leq s < t$, $X_t - X_s$ is normally distributed with

$$E(X_t - X_s) = (t-s)\mu, \qquad E\left[(X_t - X_s)^2\right] = (t-s)\sigma^2$$

where μ, σ are real constants $(\sigma \neq 0)$.

The process $(X_t)_{t\in\mathbb{R}_+}$ is called Brownian motion with *drift* μ and *variance* σ^2 (Note that if $\mu = 0$ and $\sigma = 1$, then X_t is the so-called standard Brownian motion). Show that $Cov(X_t, X_s) = \sigma^2 \min\{s, t\}$ and $(X_t - \mu t)/\sigma$ is a standard Brownian motion.

2.25. Show that if $(X_t)_{t\in\mathbb{R}_+}$ is a Brownian motion, then the processes

$$Y_t = cX_{t/c^2} \quad \text{for fixed} \quad c > 0,$$

$$U_t = \begin{cases} tX_{1/t} & \text{for } \ t > 0, \\ 0 & \text{for } \ t = 0, \end{cases}$$

and

$$V_t = X_{t+h} - X_h \quad \text{for fixed} \quad h > 0$$

are each Brownian motions.

2.26. Let $(X_t)_{t\in\mathbb{R}_+}$ be a Brownian motion and let $M_t = \max_{0\leq s\leq t} X_s$. Prove that $Y_t = M_t - X_t$ is a continuous-time Markov process. (*Hint:* Note that for $t' < t$,

$$Y(t) = \max\left\{\max_{t'\leq s\leq t}\{(X_s - X_{t'})\}, Y_{t'}\right\} - (X_t - X_{t'}).)$$

2.27. Let T be a stopping time for a Brownian motion $(X_t)_{t\in\mathbb{R}_+}$. Then the process

$$Y_t = X_{t+T} - X_T, \qquad t \geq 0,$$

is a Brownian motion, and $\sigma(Y_t, t \geq 0)$ is independent of $\sigma(X_t, 0 \leq t \leq T)$.

(*Hint:* At first consider T constant. Then suppose that the range of T is a countable set and finally approximate T by a sequence of stopping times such as in Lemma 2.117.)

2.28. Let $(X_t)_{t\in\mathbb{R}_+}$ be an n-dimensional Brownian motion starting at 0 and let $U \in \mathbb{R}^{n\times n}$ be a (constant) orthogonal matrix, i.e., $UU^T = I$. Prove that

$$\widetilde{X}_t \equiv UX_t$$

is also a Brownian motion.

2.29. Let $(X_t)_{t\in\mathbb{R}_+}$ be a Lévy process:

1. Show that the characteristic function of X_t is infinitely divisible.
2. Suppose that the law of X_1 is $P_{X_1} = \mu$. Then, for any $t > 0$ the law of X_t is $P_{X_1} = \mu^t$.
3. Given two Lévy processes $(X_t)_{t\in\mathbb{R}_+}$ and $(X'_t)_{t\in\mathbb{R}_+}$, if $P_{X_1} = P_{X'_1}$, then the two processes are identical in law.

We call $\mu = P_{X_1}$ the infinitely divisible distribution of the Lévy process $(X_t)_{t\in\mathbb{R}_+}$.

2.30. A Lévy process $(X_t)_{t\in\mathbb{R}_+}$ is a *subordinator* if it is also a real and nonnegative process.

1. Show that sample paths of a subordinator are increasing.
2. Show that a Lévy process $(X_t)_{t\in\mathbb{R}_+}$ is a subordinator if and only if $X_1 \geq 0$ almost surely.

2.31. Show that the Brownian motions with drift, i.e.,

$$X_t = \sigma W_t + \alpha t \text{ for } \alpha, \sigma \in \mathbb{R},$$

are the only Lévy processes with continuous paths.

2.32. Consider two sequences of real numbers $(\alpha_k)_{k\in\mathbb{N}}$ and $(\beta_k)_{k\in\mathbb{N}}$ such that $\sum_{k\in\mathbb{N}} \beta_k^2 \alpha_k < +\infty$. Let N_t^k be a sequence of Poisson processes with intensities α_k, $k \in \mathbb{N}$, respectively.

Then the process

$$X_t = \sum_{k\in\mathbb{N}} \beta_k (N_t^k - \alpha_k t), \quad t \in \mathbb{R}_+,$$

is a Lévy process having ν as its Lévy measure.

2.33. Show that

1. any Lévy process is a Markov process;
2. conversely, any stochastically continuous and temporarily homogeneous Markov process on $\mathbb{R}$ is a Lévy process.

2.34. According to, e.g., Grigoriu (2002), we define as a classical semimartingale any adapted, RCLL process X_t that admits the following decomposition:

$$X_t = X_0 + M_t + A_t,$$

where M_t is a local martingale and A_t is a finite variation (on compacts) RCLL process such that $M_0 = A_0 = 0$.

1. Show that any Lévy process is a semimartingale.
2. Show that the Poisson process is a semimartingale.
3. Show that the square of a Wiener process is a semimartingale.

2.35. (*Poisson process and order statistics*). Let $X_1, \ldots, X_n$ denote a *sample*, i.e. a family of nondegenerate independent and identically distributed random variables with common cumulative distribution function F. We define the *ordered sample* as the family

$$X_{n,n} \leq \cdots \leq X_{1,n},$$

so that $X_{n,n} = \min\{X_1, \ldots, X_n\}$ and $X_{1,n} = \max\{X_1, \ldots, X_n\}$. The random variable $X_{k,n}$ is called the *k-order statistic.*

Let $N = (N_t)_{t\in\mathbb{R}_+}$ be a homogeneous Poisson process with intensity $\lambda > 0$. Prove that the arrival times T_i of N in $]0, t]$, conditionally upon $\{N_t = n\}$, have the same distribution as the order statistics of a uniform sample on $]0, t[$ of size n; i.e., for all Borel sets A in $\mathbb{R}_+$ and any $n \in \mathbb{N}$, we have

$$P((T_1, T_2, \ldots, T_{N_t}) \in A | N_t = n) = P((U_{n,n}, \ldots, U_{1,n}) \in A).$$

2.36. (*self-similarity*). A real-valued stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *self-similar* with *index* $H > 0$ (H-ss) if its finite-dimensional distributions satisfy the relation

$$(X_{at_1}, \ldots, X_{at_n}) \stackrel{d}{=} a^H (X_{t_1}, \ldots, X_{t_n})$$

for any choice of $a > 0$ and $t_1, \ldots, t_n \in \mathbb{R}_+$. Show that a Gaussian process with mean function $m_t = E[X_t]$ and covariance function $K(s,t) = Cov(X_s, X_t)$ is H-ss for some $H > 0$ if and only if

$$m_t = ct^H, \text{ and } K(s,t) = s^{2H} C(t/s, 1)$$

for some constant $c \in \mathbb{R}$ and some nonnegative definite function C. As a consequence, show that the standard Brownian motion is 1/2-ss. Also, show that any α-stable process is $1/\alpha$-ss.

2.37. (*affine processes*). Let $\Phi = (\Phi_t)_{t\in\mathbb{R}_+}$ be a process on a given probability space $(\Omega, \mathcal{F}, P)$, such that $E[\|\Phi_t\|] < +\infty$ for each $t \in \mathbb{R}_+$. The past-future filtration associated with Φ is defined as the family

$$\mathcal{F}_{s,T} = \sigma\{\Phi_u | u \in [0,s] \cup [T, +\infty[\}.$$

We shall call Φ an *affine process* if it satisfies

$$E[\Phi_t | \mathcal{F}_{s,T}] = \frac{T-t}{T-s}\Phi_s + \frac{t-s}{T-s}\Phi_T, \ s < t < T.$$

Show that the condition above is equivalent to the property that for $s \leq t < t' \leq u$, the quantity

$$E\left[\left.\frac{\Phi_t - \Phi_{t'}}{t - t'} \right| \mathcal{F}_{s,u}\right] = \frac{\Phi_u - \Phi_s}{u - s},$$

and hence does not depend on the pair (t, t').

2.38. Prove that the Brownian motion is an affine process.

2.39. Let $X = (X_t)_{t\in\mathbb{R}_+}$ be a Lévy process, such that $E[\|X_t\|] < +\infty$ for each $t \in \mathbb{R}_+$. Show that X is an affine process.

2.40. Consider a process $M = (M_t)_{t\in\mathbb{R}_+}$ that is adapted to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ on a probability space $(\Omega, \mathcal{F}, P)$ and satisfies

$$E[\|M_t\|] < +\infty \text{ and } E\left[\int_0^t du |M_u|\right] < +\infty \text{ for any } t > 0.$$

Prove that the following two conditions are equivalent:

1. M is an $\mathcal{F}_t$-martingale;

2. for every $t > s$,

$$E\left[\frac{1}{t-s}\int_s^t duM_u \middle| \mathcal{F}_s\right] = M_s.$$

2.41. (*empirical process and Brownian bridge*). Let $U_1, \ldots, U_n, \ldots,$ be a sequence of independent and identically distributed random variables uniformly distributed on $[0,1]$. Define the stochastic process $b^{(n)}$ on the interval $[0,1]$ as follows:

$$b^{(n)}(t) = \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} I_{[0,t]}(U_k) - t\right), \quad t \in [0,1].$$

1. For any s and t in $[0,1]$, compute $E[b^{(n)}(t)]$ and $Cov[b^{(n)}(s), b^{(n)}(t)]$.
2. Prove that, as $n \to \infty$, the finite-dimensional distributions of the process $(b^{(n)}(t))_{t\in[0,1]}$ converge weakly towards those of a Gaussian process on $[0,1]$ whose mean and covariance functions are the same as those of $b^{(n)}$.

3

The Itô Integral

3.1 Definition and Properties

The remaining chapters on the theory of stochastic processes will primarily focus on Brownian motion, as it is by far the most useful and applicable model, which allows for many explicit calculations and, as has been demonstrated in the pollen grain example, arises naturally. Continuing the formal analysis of this example, suppose that a small amount of liquid flows with the macroscopic velocity $a(t, u(t))$ (where $u(t)$ is its position at time t). Then a microscopic particle that is suspended in this liquid will, as mentioned, display evidence of Brownian motion. The change in the particle's position $u(t + dt) - u(t)$ over the time interval $[t, t + dt[$ is due to, first, the macroscopic flow of the liquid, with the latter's contribution given by $a(t, u(t))dt$. But, second, there is the additional molecular bombardment of the particle, which contributes to its dynamics with the term $b(t, u(t))[W_{t+dt} - W_t]$, where $(W_t)_{t\geq 0}$ is Brownian motion. Summming the terms results in the equation

$$du(t) = a(t, u(t))dt + b(t, u(t))dW_t,$$

which, however, in the current form does not make sense, because the trajectories of $(W_t)_{t\geq 0}$ are not differentiable. Instead, we will try to interpret it in the form

$$\forall \omega \in \Omega: \qquad u(t) - u(0) = \int_0^t a(s, u(s))ds + \int_0^t b(s, u(s))dW_s,$$

which requires us to give meaning to an integral $\int_a^b f(t)dW_t$ that, as will be demonstrated, is not of Lebesgue–Stieltjes[6] hence neither of Riemann–Stieltjes type.

[6] For a revision, see the appendix or, in addition, e.g., Kolmogorov and Fomin (1961).

Definition 3.1. Let $F : [a,b] \to \mathbb{R}$ be a function and Π the set of the partitions $\pi : a = x_0 < x_1 < \cdots < x_n = b$ of the interval $[a,b]$. Putting

$$\forall \pi \in \Pi: \qquad V_F(\pi) = \sum_{i=1}^{n} |F(x_i) - F(x_{i-1})|,$$

then F is of *bounded variation*, if

$$\sup_{\pi \in \Pi} V_F(\pi) < \infty.$$

Also, $V_F(a,b) = \sup_{\pi \in \Pi} V_F(\pi)$ is called the *total variation* of F in the interval $[a,b]$.

Remark 3.2. If $F : [a,b] \to \mathbb{R}$ is monotonic, then F is of bounded variation and

$$V_F(a,b) = |F(b) - F(a)|.$$

Lemma 3.3. *Let $F : [a,b] \to \mathbb{R}$. Then the following two statements are equivalent*

1. *F is of bounded variation;*
2. *there exists an $F_1 : [a,b] \to \mathbb{R}$, and there exists an $F_2 : [a,b] \to \mathbb{R}$ monotonically increasing, such that $F = F_1 - F_2$.*

Lemma 3.4. *If $F : [a,b] \to \mathbb{R}$ is monotonically increasing, then F is λ-almost everywhere differentiable in $[a,b]$ (where λ is the Lebesgue measure).*

Corollary 3.5. If $F : [a,b] \to \mathbb{R}$ is of bounded variation, then F is differentiable almost everywhere.

Definition 3.6. Let $f : [a,b] \to \mathbb{R}$ be continuous and $F : [a,b] \to \mathbb{R}$ of bounded variation, for all $\pi \in \Pi, \pi : a = x_0 < x_1 < \cdots < x_n = b$. We will fix points ξ_i arbitrarily in $[x_{i-1}, x_i[$, $i = 1, \ldots, n$ and construct the sum

$$S_n = \sum_{i=1}^{n} f(\xi_i)[F(x_i) - F(x_{i-1})].$$

If for $\max_{i \in \{1,\ldots,n\}}(x_i - x_{i-1}) \to 0$ the sum S_n tends to a limit (that depends neither on the choice of the partition nor on the selection of the points ξ_i within the partial intervals of the partition), then this limit is the *Riemann–Stieltjes integral of f with respect to the function F* over $[a,b]$ and is denoted by the symbol $\int_a^b f(x)dF(x)$.

Remark 3.7. By Theorem 2.147 and by Corollary 3.5, it can be shown that a Wiener process is not of bounded variation and hence $\int_a^b f(t)dW_t$ cannot be interpreted in the sense of Riemann–Stieltjes.

Definition 3.8. Let $(W_t)_{t\geq 0}$ be a Wiener process defined on the probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{C}$ the set of functions $f(t,\omega) : [a,b] \times \Omega \to \mathbb{R}$ satisfying the following conditions:

1. f is $\mathcal{B}_{[a,b]} \otimes \mathcal{F}$-measurable;
2. for all $t \in [a,b]$, $f(t,\cdot) : \Omega \to \mathbb{R}$ is $\mathcal{F}_t$-measurable, where $\mathcal{F}_t = \sigma(W_s, 0 \leq s \leq t)$;
3. for all $t \in [a,b]$, $f(t,\cdot) \in L^2(\Omega, \mathcal{F}, P)$ and $\int_a^b E[|f(t)|^2]dt < \infty$.

Remark 3.9. Condition 2 of Definition 3.8 stresses the nonanticipatory nature of f through the fact that it only depends on the present and the past history of the Brownian motion, but not on the future.

Definition 3.10. Let $f \in \mathcal{C}$. If there exist both a partition π of $[a,b]$, $\pi : a = t_0 < t_1 < \cdots < t_n = b$ and some real-valued random variables $f_0, \ldots, f_{n-1}$ defined on $(\Omega, \mathcal{F}, P)$, such that

$$f(t,\omega) = \sum_{i=0}^{n-1} f_i(\omega) I_{[t_i, t_{i+1}[}(t)$$

(with the convention that $[t_{n-1}, t_n[= [t_{n-1}, b])$, then f is a *piecewise function.*

Remark 3.11. By condition 2 of Definition 3.8 it follows that, for all $i \in \{0, \ldots, n\}$, f_i is $\mathcal{F}_{t_i}$-measurable.

Definition 3.12. If $f \in \mathcal{C}$, with $f(t,\omega) = \sum_{i=0}^{n-1} f_i(\omega) I_{[t_i, t_{i+1}[}(t)$, is a piecewise function, then the real random variable $\Phi(f)$ is a *(stochastic) Itô integral of the process f*, where

$$\forall \omega \in \Omega : \qquad \Phi(f)(\omega) = \sum_{i=0}^{n-1} f_i(\omega)(W_{t_{i+1}}(\omega) - W_{t_i}(\omega)).$$

$\Phi(f)$ is denoted by the symbol $\int_a^b f(t)dW_t$, henceforth suppressing the explicit dependence on the trajectory ω wherever obvious.

Lemma 3.13. *Let $f, g \in \mathcal{C}$ be piecewise functions. Then they have the properties that*

1. $E[\int_a^b f(t)dW_t] = 0,$
2. $E[\int_a^b f(t)dW_t \int_a^b g(t)dW_t] = \int_a^b E[f(t)g(t)]dt.$

Proof: 1. Let $f(t,\omega) = \sum_{i=0}^{n-1} f_i(\omega) I_{[t_i, t_{i+1}[}(t)$. Then

$$E\left[\int_a^b f(t)dW_t\right] = E\left[\sum_{i=0}^{n-1} f_i(W_{t_{i+1}} - W_{t_i})\right]$$
$$= E\left[\sum_{i=0}^{n-1} E[f_i(W_{t_{i+1}} - W_{t_i})|\mathcal{F}_{t_i}]\right]$$
$$= E\left[\sum_{i=0}^{n-1} f_i E[W_{t_{i+1}} - W_{t_i}|\mathcal{F}_{t_i}]\right],$$

where the last step follows from Remark 3.11. Now, because $(W_t)_{t\geq 0}$ has independent increments, $(W_{t_{i+1}} - W_{t_i})$ is independent of $\mathcal{F}_{t_i}$. Hence $E[W_{t_{i+1}} - W_{t_i}|\mathcal{F}_{t_i}] = E[W_{t_{i+1}} - W_{t_i}]$ and the completion of the proof follows from the fact that the Wiener process has mean zero.

2. The piecewise functions f and g can be represented by means of the same partition $a = t_0 < t_1 < \cdots < t_n = b$ of the interval $[a, b]$. For this purpose it suffices to choose the union of the partitions associated with f and g, respectively. Thus let $f(t,\omega) = \sum_{i=0}^{n-1} f_i(\omega) I_{[t_i,t_{i+1}[}(t)$ and $g(t,\omega) = \sum_{i=0}^{n-1} g_i(\omega) I_{[t_i,t_{i+1}[}(t)$. Then

$$E\left[\int_a^b f(t)dW_t \int_a^b g(t)dW_t\right]$$
$$= E\left[\sum_{i=0}^{n-1} f_i(W_{t_{i+1}} - W_{t_i}) \sum_{j=0}^{n-1} g_j(W_{t_{j+1}} - W_{t_j})\right]$$
$$= E\left[\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} f_i g_i (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j})\right]$$
$$= E\left[\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} E[f_i g_i (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_i \vee t_j}]\right],$$

where $t_i \vee t_j = \max\{t_i, t_j\}$. If $i < j$, then $t_i < t_j$ and therefore $\mathcal{F}_{t_i} \subset \mathcal{F}_{t_j}$, resulting in f_i being $\mathcal{F}_{t_j}$-measurable (already being $\mathcal{F}_{t_i}$-measurable) and $(W_{t_{i+1}} - W_{t_i})$ being $\mathcal{F}_{t_j}$-measurable (already being $\mathcal{F}_{t_{i+1}}$-measurable with $t_{i+1} \leq t_j$). Finally, by Remark 3.11, g_j is $\mathcal{F}_{t_j}$-measurable. Thus

$$E[f_i g_j (W_{t_{i+1}} - W_{t_i})(W_{t_{j+1}} - W_{t_j})|\mathcal{F}_{t_j}]$$
$$= f_i g_j (W_{t_{i+1}} - W_{t_i}) E[W_{t_{j+1}} - W_{t_j}|\mathcal{F}_{t_j}]$$
$$= 0,$$

given that $(W_{t_{j+1}} - W_{t_j})$ is independent of $\mathcal{F}_{t_j}$ ($(W_t)_{t\geq 0}$ having independent increments) and $E[W_{t_{j+1}} - W_{t_j}] = 0$.

Instead if $i = j$, then

$$E[f_i g_i (W_{t_{i+1}} - W_{t_i})(W_{t_{i+1}} - W_{t_i})|\mathcal{F}_{t_i}] = f_i g_i E[(W_{t_{i+1}} - W_{t_i})^2|\mathcal{F}_{t_i}]$$
$$= f_i g_i E[(W_{t_{i+1}} - W_{t_i})^2].$$

But since $(W_{t_{i+1}} - W_{t_i})$ is normally distributed as $N(0, t_{i+1} - t_i)$,

$$E[(W_{t_{i+1}} - W_{t_i})^2] = t_{i+1} - t_i$$

and therefore

$$E[f_i g_i (W_{t_{i+1}} - W_{t_i})^2|\mathcal{F}_{t_i}] = f_i g_i (t_{i+1} - t_i).$$

Putting parts together, we obtain

$$E\left[\int_a^b f(t)dW_t \int_a^b g(t)dW_t\right] = E\left[\sum_{i=0}^{n-1} f_i g_i (t_{i+1} - t_i)\right]$$
$$= \sum_{i=0}^{n-1} E[f_i g_i](t_{i+1} - t_i) = \int_a^b E[f(t)g(t)]dt.$$

□

Corollary 3.14. If $f \in \mathcal{C}$ is a piecewise function, then

$$E\left[\left(\int_a^b f(t)dW_t\right)^2\right] = \int_a^b E\left[(f(t))^2\right] dt < \infty.$$

Lemma 3.15. *If $\mathcal{S}$ denotes the space of piecewise functions belonging to the class $\mathcal{C}$, then $\mathcal{S} \subset L^2([a,b] \times \Omega)$ and $\Phi : \mathcal{S} \to L^2(\Omega)$ is linearly continuous.*

Proof: By point 3 of the characterization of the class $\mathcal{C}$, it follows that $\mathcal{S} \subset L^2([a,b] \times \Omega)$, whereas by Corollary 3.14, it follows that Φ takes values in $L^2(\Omega)$. The linearity and continuity of Φ can be inferred from Definition 3.12 and, again, from Corollary 3.14, respectively, the latter by observing that if $f \in \mathcal{S}$, then

$$\|f\|^2_{L^2([a,b]\times\Omega)} = \int_a^b E\left[(f(t))^2\right] dt,$$
$$\|\Phi(f)\|^2_{L^2(\Omega)} = E\left[(\Phi(f))^2\right] = E\left[\left(\int_a^b f(t)dW_t\right)^2\right].$$

Thus $\|\Phi(f)\|^2_{L^2(\Omega)} = \|f\|^2_{L^2([a,b]\times\Omega)}$, which guarantees the continuity of the linear mapping Φ.[7] □

[7] For this classical result of analysis, see, e.g., Kolmogorov and Fomin (1961).

Lemma 3.16. *$\mathcal{C}$ is a closed subspace of the Hilbert space $L^2([a,b] \times \Omega)$ and is therefore a Hilbert space as well. The scalar product is defined as*

$$\langle f, g \rangle = \int_a^b \int_\Omega f(t,\omega) g(t,\omega) dP(\omega) dt = \int_a^b E[f(t)g(t)]dt.$$

Hence Φ has a unique linear continuous extension in the closure of $\mathcal{S}$ in $\mathcal{C}$ (which we will continue to denote by Φ), i.e., $\Phi : \bar{\mathcal{S}} \to L^2(\Omega)$.

Lemma 3.17. *$\mathcal{S}$ is dense in $\mathcal{C}$.*

Proof: See, e.g., Dieudonné (1960). □

Theorem 3.18. *The (stochastic) Itô integral $\Phi : \mathcal{S} \to L^2(\Omega)$ has a unique linear continuous extension in $\mathcal{C}$. If $f \in \mathcal{C}$, we denote $\Phi(f)$ by $\int_a^b f(t)dW_t$.*

Proposition 3.19. *If $f, g \in \mathcal{C}$, then*

1. *$E[\int_a^b f(t)dW_t] = 0$,*
2. *$E[\int_a^b f(t)dW_t \int_a^b g(t)dW_t] = \int_a^b E[f(t)g(t)]dt$,*
3. *$E[(\int_a^b f(t)dW_t)^2] = \int_a^b E[(f(t))^2]dt$ (Itô isometry).*

Proof: 1. Let $f \in \mathcal{C}$, then, because the closure of $\mathcal{S}$ in $\mathcal{C}$ coincides with $\mathcal{C}$, there exists $(f_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ such that $\lim_{n\to\infty} \int_a^b E[(f(t) - f_n(t))^2]dt = 0$, and hence $\lim_{n\to\infty} f_n = f$ in $L^2([a,b] \times \Omega)$. Because Φ is linearly continuous, we also have that

$$\lim_{n\to\infty} \Phi(f_n) = \Phi(f) \text{ (in } L^2(\Omega)),$$

thus

$$\lim_{n\to\infty} E\left[\left(\int_a^b (f(t) - f_n(t))dW_t\right)^2\right] = 0.$$

Because P is a probability on $(\Omega, \mathcal{F})$ (hence a finite measure), the convergence of $\Phi(f_n)$ to $\Phi(f)$ in $L^2(\Omega)$ implies the convergence in $L^1(\Omega)$. Therefore, $\lim_{n\to\infty} E[|\int_a^b (f(t) - f_n(t))dW_t|] = 0$, from which it follows that

$$\lim_{n\to\infty} E\left[\int_a^b (f(t) - f_n(t))dW_t\right] = 0,$$

and from the linearity of both the stochastic integral and its expectation, we obtain

$$\lim_{n\to\infty} E\left[\int_a^b f_n(t)dW_t\right] = E\left[\int_a^b f(t)dW_t\right].$$

Now 1 follows by point 1 of Lemma 3.13.

2. Let $f, g \in \mathcal{C}$. Then

$$\exists (f_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}} \text{ such that } f_n \xrightarrow{n} f \text{ in } L^2([a,b]\times\Omega);$$
$$\exists (g_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}} \text{ such that } g_n \xrightarrow{n} g \text{ in } L^2([a,b]\times\Omega).$$

By the continuity of the scalar product (in $L^2([a,b]\times\Omega)$):

$$\langle f_n, g_n\rangle \xrightarrow{n} \langle f, g\rangle,$$

and thus

$$\lim_{n\to\infty}\int_a^b E[f_n(t)g_n(t)]dt = \int_a^b E[f(t)g(t)]dt. \tag{3.1}$$

Moreover, by point 2 of Lemma 3.13:

$$\int_a^b E[f_n(t)g_n(t)]dt = E\left[\int_a^b f_n(t)dW_t \int_a^b g_n(t)dW_t\right]. \tag{3.2}$$

From the fact that $f_n \xrightarrow{n} f$ in $L^2([a,b]\times\Omega)$, it also follows that $\Phi(f_n) \xrightarrow{n} \Phi(f)$ in $L^2(\Omega)$ (by the continuity of Φ) and, analogously, since $g_n \xrightarrow{n} g$ in $L^2([a,b]\times\Omega)$, it follows that $\Phi(g_n) \xrightarrow{n} \Phi(g)$ in $L^2(\Omega)$. Then, by the continuity of the scalar product in $L^2(\Omega)$, we get

$$\langle \Phi(f_n), \Phi(g_n)\rangle \xrightarrow{n} \langle \Phi(f), \Phi(g)\rangle,$$

and hence

$$\lim_{n\to\infty} E\left[\int_a^b f_n(t)dW_t \int_a^b g_n(t)dW_t\right] = E\left[\int_a^b f(t)dW_t \int_a^b g(t)dW_t\right]. \tag{3.3}$$

The assertion finally follows from (3.1), (3.2), and (3.3).

Point 3 is a direct consequence of 2. □

Remark 3.20. If $f \in \mathcal{C}$ and $(f_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ such that $f_n \xrightarrow{n} f$ in $L^2([a,b]\times\Omega)$, then

1. $\Phi(f_n) \xrightarrow{n} \Phi(f)$ in $L^2(\Omega)$ (by the continuity of Φ),
2. $\Phi(f_n) \xrightarrow{n} \Phi(f)$ in probability.

In fact, as already mentioned, with P being a finite measure, convergence in $L^2(\Omega)$ implies convergence in $L^1(\Omega)$ and, furthermore, convergence in $L^1(\Omega)$ implies convergence in probability, by Theorem 1.157.

An alternative approach to the concept of a stochastic integral is the following.

Definition 3.21. Let $\mathcal{C}_1$ be the set of functions $f : [a,b]\times\Omega \to \mathbb{R}$ such that the conditions 1 and 2 of the characterization of the class $\mathcal{C}$ are satisfied, but, instead of condition 3, we have

$$P\left(\int_a^b |f(t)|^2 dt < \infty\right) = 1. \tag{3.4}$$

Remark 3.22. It is obvious that $\mathcal{C} \subset \mathcal{C}_1$ and thus $\mathcal{S} \subset \mathcal{C}_1$. We will show that it is also possible to define a stochastic integral in $\mathcal{C}_1$, which, in $\mathcal{C}$, is identical to the (stochastic) Itô integral as defined above.

Lemma 3.23. *If $f \in \mathcal{S} \subset \mathcal{C}_1$, then for all $c > 0$ and for all $N > 0$:*

$$P\left(\left|\int_a^b f(t)dW_t\right| > c\right) \leq P\left(\int_a^b |f(t)|^2 dt > N\right) + \frac{N}{c^2}. \tag{3.5}$$

Proof: See, e.g., Friedman (1975). □

Lemma 3.24. *If $f \in \mathcal{C}_1$, then there exists $(f_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ such that*

$$\lim_{n\to\infty} \int_a^b |f(t) - f_n(t)|^2 dt = 0 \qquad \textit{almost surely.}$$

Proof: See, e.g., Friedman (1975). □

Remark 3.25. Resorting to the same notation as in the preceding lemma, we also have that $P-\lim_{n\to\infty} \int_a^b |f_n(t)-f(t)|^2 dt = 0$, because almost sure convergence implies convergence in probability. Let $f \in \mathcal{C}_1$. Then, by the preceding lemma, there exists $(f_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ such that $\lim_{n\to\infty} \int_a^b |f(t) - f_n(t)|^2 dt = 0$ almost surely. Let $(n, m) \in \mathbb{N} \times \mathbb{N}$. Then, because $(a + b)^2 \leq 2(a^2 + b^2)$, we obtain

$$\int_a^b |f_n(t) - f_m(t)|^2 dt \leq 2\left(\int_a^b |f_n(t) - f(t)|^2 dt + \int_a^b |f_m(t) - f(t)|^2 dt\right)$$

and hence $\lim_{m,n\to\infty} \int_a^b |f_n(t) - f_m(t)|^2 dt = 0$ almost surely. Consequently

$$P - \lim_{n\to\infty} \int_a^b |f(t) - f_n(t)|^2 dt = 0.$$

But $(f_n - f_m) \in \mathcal{S} \cap \mathcal{C}_1$ (for all $n, m \in \mathbb{N}$) and by Lemma 3.23, for all $\rho > 0$ and all $\epsilon > 0$:

$$P\left(\left|\int_a^b (f_n(t) - f_m(t))dW_t\right| > \epsilon\right) \leq P\left(\int_a^b |f_n - f_m|^2 dt > \rho\epsilon^2\right) + \rho.$$

Finally, by the arbitrary nature of ρ, we have that

$$\lim_{m,n\to\infty} P\left(\left|\int_a^b (f_n(t) - f_m(t))dW_t\right| > \epsilon\right) = 0.$$

Hence the sequence of random variables $(\int_a^b f_n(t)dW_t)_{n\in\mathbb{N}}$ is Cauchy in probability and therefore admits a limit in probability (see, e.g., Baldi (1984) for details). This limit will be denoted by $\int_a^b f(t)dW_t$.

Definition 3.26. If $f \in \mathcal{C}_1$ and $(f_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ such that $\lim_{n\to\infty} \int_a^b |f(t) - f_n(t)|^2 dt = 0$ almost surely, then the limit in probability to which the sequence of random varibles $(\int_a^b f_n(t) dW_t)_{n\in\mathbb{N}}$ converges is the *(stochastic) Itô integral of* f.

Remark 3.27. The preceding definition is well posed, because it can be shown that $\int_a^b f(t) dW_t$ is independent of the particular approximating sequence $(f_n)_{n\in\mathbb{N}}$. (See, e.g., Baldi (1984) for details.)

Theorem 3.28. *If* $f \in \mathcal{C}_1$*, then (3.5) applies again.*

Proof: See, e.g., Friedman (1975). □

Theorem 3.29. *Let* $f \in \mathcal{C}_1$ *and* $(f_n)_{n\in\mathbb{N}} \in \mathcal{C}_1^{\mathbb{N}}$. *If*

$$P - \lim_{n\to\infty} \int_a^b |f_n(t) - f(t)|^2 dt = 0,$$

then

$$P - \lim_{n\to\infty} \int_a^b f_n(t) dW_t = \int_a^b f(t) dW_t.$$

Proof: Fixing $c > 0, \rho > 0$, by Theorem 3.28, we obtain

$$P\left(\left| \int_a^b (f_n(t) - f(t)) dW_t \right| > c \right) \leq P\left(\int_a^b |f_n(t) - f(t)|^2 dt > c^2 \rho \right) + \rho.$$

Now, the proof follows for $n \to \infty$. □

Now we are able to show that the stochastic integral in $\mathcal{C}_1$ of Definition 3.26 is identical to the one of Theorem 3.18 in $\mathcal{C}$. In fact, for $f \in \mathcal{C}$, because $\mathcal{S}$ is dense in $\mathcal{C}$, there exists $(f_n)_{n\in\mathbb{N}} \in \mathcal{S}^{\mathbb{N}}$ such that

$$\lim_{n\to\infty} E\left[\int_a^b |f_n(t) - f(t)|^2 dt \right] = 0. \tag{3.6}$$

Putting $X_n = \int_a^b |f_n(t) - f(t)|^2 dt$ for all $n \in \mathbb{N}$, by the Markov inequality, we obtain

$$\forall \lambda > 0\colon \qquad P(X_n \geq \lambda E[X_n]) \leq \frac{1}{\lambda} \qquad (n \in \mathbb{N}),$$

and thus $P(X_n \geq \epsilon) \leq \frac{E[X_n]}{\epsilon}$ for $\epsilon = \lambda E[X_n]$. But by (3.6), $\lim_{n\to\infty} E[X_n] = 0$, and therefore also $\lim_{n\to\infty} P(X_n \geq \epsilon) = 0$ and

$$P - \lim_{n\to\infty} \int_a^b |f_n(t) - f(t)|^2 dt = 0. \tag{3.7}$$

From (3.7) and by Theorem 3.29, it follows that

$$P-\lim_{n\to\infty}\int_a^b f_n(t)dW_t = \int_a^b f(t)dW_t, \tag{3.8}$$

where the limit $\int_a^b f(t)dW_t$ is the stochastic integral of f in $\mathcal{C}_1$. But, on the other hand, (3.6) implies, by point 2 of Remark 3.20, that $\Phi(f_n) \xrightarrow{n} \Phi(f)$ in probability (Φ is the linear continuous extension in $\mathcal{C}$) and thus again

$$P-\lim_{n\to\infty}\int_a^b f_n(t)dW_t = \int_a^b f(t)dW_t. \tag{3.9}$$

Now by (3.8) and (3.9) as well as the uniqueness of the limit, the proof is complete. □

Remark 3.30. If $f \in \mathcal{C}_1$ and $P(\int_a^b |f(t)|^2 dt = 0) = 1$, then

$$\forall N > 0: \qquad P\left(\int_a^b |f(t)|^2 dt > N\right) = 0$$

and, by Theorem 3.28,

$$P\left(\left|\int_a^b f(t)dW_t\right| > c\right) = 0 \qquad \forall c > 0,$$

so that

$$P\left(\left|\int_a^b f(t)dW_t\right| = 0\right) = 1.$$

Theorem 3.31. *If $f \in \mathcal{C}_1$ and continuous for almost every ω, then, for every sequence $(\pi_n)_{n\in\mathbb{N}}$ of the partitions $\pi_n : a = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = b$ of the interval $[a, b]$ such that*

$$|\pi_n| = \sup_{k\in\{0,\ldots,n\}} \left|t_{k+1}^{(n)} - t_k^{(n)}\right| \xrightarrow{n} 0,$$

we have

$$P-\lim_{n\to\infty}\sum_{k=0}^{n-1} f\left(t_k^{(n)}\right)\left(W_{t_{k+1}^{(n)}} - W_{t_k^{(n)}}\right) = \int_a^b f(t)dW_t.$$

Proof: By definition of the piecewise function

$$f(t,\omega) = \sum_{k=0}^{n-1} f_k(\omega) I_{[t_k, t_{k+1}[}(t),$$

we have that

$$\sum_{k=0}^{n-1} f\left(t_k^{(n)}\right)\left(W_{t_{k+1}^{(n)}} - W_{t_k^{(n)}}\right) = \int_a^b f_n(t)dW_t.$$

Now by Theorem 3.29 all that needs to be shown is that

$$P - \lim_{n\to\infty}\int_a^b |f_n(t) - f(t)|^2 dt = 0,$$

which follows by the continuity of f for almost every ω. □

Proposition 3.32. *Let $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of the partitions $\pi_n : a = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = b$ of the interval $[a,b]$ such that $|\pi_n| \xrightarrow{n} 0$ and, for all $n \in \mathbb{N}$, let $S_n = \sum_{j=0}^{n-1}(W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}})^2$, i.e., the quadratic variation of $(W_t)_{t\in[a,b]}$ with respect to the partition π_n. Then we have that*

1. *$E[S_n] = b - a$ for all $n \in \mathbb{N}$;*
2. *$Var[S_n] = E[(S_n - (b-a))^2] \xrightarrow{n} 0$.*

Proof: 1.

$$\begin{aligned} E[S_n] &= \sum_{j=0}^{n-1} E\left[\left(W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}}\right)^2\right] = \sum_{j=0}^{n-1} Var\left[W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}}\right] \\ &= \sum_{j=0}^{n-1}\left(t_{j+1}^{(n)} - t_j^{(n)}\right) = b - a. \end{aligned}$$

2. Because Brownian motion, by definition, has independent increments, we have that

$$Var[S_n] = \sum_{j=0}^{n-1} Var\left[\left(W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}}\right)^2\right].$$

Writing $\delta_j = W_{t_{j+1}^{(n)}} - W_{t_j^{(n)}}$, then, by (1.4),

$$\sum_{j=0}^{n-1} Var\left[(\delta_j)^2\right] = \sum_{j=0}^{n-1}\left(E\left[(\delta_j)^4\right] - \left(E\left[(\delta_j)^2\right]\right)^2\right) \leq \sum_{j=0}^{n-1} E\left[(\delta_j)^4\right].$$

Now, by the definition of Brownian motion, the increments δ_j are Gaussian, i.e., $N(0, t_{j+1} - t_j)$, and direct calculation results in

$$E\left[(\delta_j)^4\right] = \int_{-\infty}^{+\infty} (\delta_j)^4 \frac{\exp\left\{-\frac{(\delta_j)^2}{2(t_{j+1}-t_j)}\right\}}{\sqrt{2\pi(t_{j+1}-t_j)}} d\delta_j = 3(t_{j+1} - t_j)^2 \xrightarrow{n} 0.$$

□

Remark 3.33. Given the hypotheses of the preceding proposition, by the Chebychev inequality,

$$P(|S_n - (b-a)| > \epsilon) \leq \frac{Var[S_n]}{\epsilon^2} \stackrel{n}{\rightarrow} 0 \qquad (\epsilon > 0).$$

It follows that $P - \lim_{n\to\infty} S_n = b - a$. On the other hand, if we compare it to the classical Lebesgue integral, we obtain

$$\lim_{n\to\infty} \sum_{j=0}^{n-1} \left(t_{j+1}^{(n)} - t_j^{(n)}\right)^2 \leq \lim_{n\to\infty} |\pi_n| \sum_{j=0}^{n-1} \left(t_{j+1}^{(n)} - t_j^{(n)}\right) = \lim_{n\to\infty} |\pi_n|(b-a) = 0.$$

Remark 3.34. Because the Brownian motion $(W_t)_{t\geq 0}$ is continuous for almost every ω, we can apply Theorem 3.31 with $f(t) = W_t$, obtaining the result of Proposition 3.35.

Proposition 3.35. $\int_a^b W_t dW_t = \frac{1}{2}(W_b^2 - W_a^2) - \frac{b-a}{2}$.

Proof: Let $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of the partitions $\pi_n : a = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = b$ of the interval $[a, b]$ such that $|\pi_n| \stackrel{n}{\rightarrow} 0$. Then, by Theorem 3.31, we have

$$\int_a^b W_t dW_t = P - \lim_{n\to\infty} \sum_{k=0}^{n-1} W_{t_k^{(n)}} \left(W_{t_{k+1}^{(n)}} - W_{t_k^{(n)}}\right). \tag{3.10}$$

Because, in general, $a(b-a) = \frac{1}{2}(b^2 - a^2 - (b-a)^2)$, therefore

$$W_{t_k^{(n)}} \left(W_{t_{k+1}^{(n)}} - W_{t_k^{(n)}}\right) = \frac{1}{2}\left(W_{t_{k+1}^{(n)}}^2 - W_{t_k^{(n)}}^2 - \left(W_{t_{k+1}^{(n)}} - W_{t_k^{(n)}}\right)^2\right).$$

Substitution into (3.10) results in

$$\begin{aligned}\int_a^b W_t dW_t &= P - \lim_{n\to\infty} \frac{1}{2} \sum_{k=0}^{n-1} \left(W_{t_{k+1}^{(n)}}^2 - W_{t_k^{(n)}}^2 - \left(W_{t_{k+1}^{(n)}} - W_{t_k^{(n)}}\right)^2\right) \\ &= \frac{1}{2}(W_b^2 - W_a^2) - P - \lim_{n\to\infty} \frac{1}{2} \sum_{k=0}^{n-1} \left(W_{t_{k+1}^{(n)}} - W_{t_k^{(n)}}\right)^2 \\ &= \frac{1}{2}(W_b^2 - W_a^2) - P - \lim_{n\to\infty} \frac{1}{2} S_n = \frac{1}{2}(W_b^2 - W_a^2) - \frac{b-a}{2},\end{aligned}$$

by Remark 3.33. □

Remark 3.36. The classical Lebesgue integral results in $\int_a^b t dt = \frac{b^2 - a^2}{2}$. However, in the (stochastic) Itô integral we obtain an additional term $(-\frac{b-a}{2})$. Generally, in certain practical applications involving stochastic models, the Stratonovich integral is employed. In the latter, $t_k^{(n)}$ is replaced by $r_k^{(n)} =$

$\frac{t_k^{(n)}+t_{k+1}^{(n)}}{2}$, thus eliminating the additional term $(-\frac{b-a}{2})$. Therefore, in general, one obtains a new family of integrals by varying the chosen point of the partition. In particular, the Stratonovich integral has the advantage that its rules of calculus are identical with the ones of the classical integral. But, nonetheless, the Itô integral is often a more appropriate model for many applications.

3.2 Stochastic Integrals as Martingales

Theorem 3.37. *If $f \in \mathcal{C}$ and, for all $t \in [a,b]$: $X(t) = \int_a^t f(s)dW_s$, then $(X_t)_{t\in[a,b]}$ is a martingale with respect to $\mathcal{F}_t = \sigma(W_s, 0 \leq s \leq t)$.*

Proof: Initially, let $f \in \mathcal{C} \cap \mathcal{S}$. Then there exists a π, a partition of $[a,b]$, $\pi : a = t_0 < t_1 < \cdots < t_n = b$, such that

$$f(t,\omega) = \sum_{i=0}^{n-1} f(t_i,\omega) I_{[t_i,t_{i+1}[}(t), \qquad t \in [a,b], \omega \in \Omega,$$

and for all $t \in [a,b]$:

$$X(t) = \int_a^t f(s)dW_s = \sum_{i=0}^{k-1} f(t_i)(W_{t_{i+1}} - W_{t_i}) + f(t_k)(W_t - W_{t_k})$$

for k, such that $t_k \leq t < t_{k+1}$. Because for all $i \in \{0,\ldots,k\}$, $f(t_i)$ is $\mathcal{F}_{t_i}$-measurable (by Remark 3.11), $X(t)$ is obviously $\mathcal{F}$-measurable, for all $t \in [a,b]$. Now, let $(s,t) \in [a,b] \times [a,b]$ and $s < t$. Then it needs to be shown that

$$E[X(t)|\mathcal{F}_s] = X(s) \text{ a.s.}$$

and thus

$$E[X(t) - X(s)|\mathcal{F}_s] = 0 \text{ a.s.}$$

We observe that

$$\begin{aligned}
&X(t) - X(s) \\
&= \int_s^t f(u)dW_u = \sum_{i=0}^{k-1} f(t_i)(W_{t_{i+1}} - W_{t_i}) \\
&\quad + f(t_k)(W_t - W_{t_k}) - \sum_{j=0}^{h-1} f(t_j)(W_{t_{j+1}} - W_{t_j}) - f(t_h)(W_s - W_{t_h})
\end{aligned}$$

if $t_h \leq s < t_{h+1}$ and $t_k \leq t < t_{k+1}$, where $h \leq k$. Therefore,

$$
\begin{aligned}
&X(t) - X(s) \\
&= \sum_{i=h}^{k-1} f(t_i)(W_{t_{i+1}} - W_{t_i}) + f(t_k)(W_t - W_{t_k}) - f(t_h)(W_s - W_{t_h}) \\
&= \sum_{i=h+1}^{k-1} f(t_i)(W_{t_{i+1}} - W_{t_i}) + f(t_k)(W_t - W_{t_k}) - f(t_h)(W_{t_{h+1}} - W_s).
\end{aligned}
$$

Because $s < t_j$, for $j = h+1, \ldots, k$, thus $\mathcal{F}_s \subset \mathcal{F}_{t_j}$, and by the properties of conditional expectations we obtain

$$
\begin{aligned}
&E[X(t) - X(s)|\mathcal{F}_s] \\
&= \sum_{i=h+1}^{k-1} E[f(t_i)(W_{t_{i+1}} - W_{t_i})|\mathcal{F}_s] \\
&\quad + E[f(t_k)(W_t - W_{t_k})|\mathcal{F}_s] + E[f(t_h)(W_{h+1} - W_{t_s})|\mathcal{F}_s] \\
&= \sum_{i=h+1}^{k-1} E[E[f(t_i)(W_{t_{i+1}} - W_{t_i})|\mathcal{F}_{t_i}]|\mathcal{F}_s] \\
&\quad + E[E[f(t_k)(W_t - W_{t_k})|\mathcal{F}_{t_k}]|\mathcal{F}_s] + E[f(t_h)(W_{t_{h+1}} - W_{t_s})|\mathcal{F}_s] \\
&= \sum_{i=h+1}^{k-1} E[f(t_i)E[W_{t_{i+1}} - W_{t_i}|\mathcal{F}_{t_i}]|\mathcal{F}_s] \\
&\quad + E[f(t_k)E[W_t - W_{t_k}|\mathcal{F}_{t_k}]|\mathcal{F}_s] + f(t_h)E[W_{t_{h+1}} - W_{t_s}|\mathcal{F}_s] \\
&= \sum_{i=h+1}^{k-1} E[f(t_i)E[W_{t_{i+1}} - W_{t_i}]|\mathcal{F}_s] \\
&\quad + E[f(t_k)E[W_t - W_{t_k}]|\mathcal{F}_s] + f(t_h)E[W_{t_{h+1}} - W_{t_s}] \\
&= 0,
\end{aligned}
$$

since $E[W_t] = 0$ for all t and $(W_t)_{t \geq 0}$ has independent increments. This completes the proof for the case $f \in \mathcal{C} \cap \mathcal{S}$.

Now, let $f \in \mathcal{C}$, then $\exists (f_n)_{n \in \mathbb{N}} \in (\mathcal{C} \cap \mathcal{S})^{\mathbb{N}}$ such that $\lim_{n\to\infty} \int_a^b |f(t) - f_n(t)|^2 dt = 0$, by Lemma 3.24. We put

$$
X_n(t) = \int_a^t f_n(s) dW_s \qquad \forall n \in \mathbb{N}, \forall t \in [a,b],
$$

for which we have just shown that $((X_n(t))_{t\in[a,b]})_{n\in\mathbb{N}}$ is a sequence of martingales. Now, let $(s,t) \in [a,b] \times [a,b]$ and $s < t$. Then it will be shown that

$$
E[X(t) - X(s)|\mathcal{F}_s] = 0 \text{ a.s.} \tag{3.11}
$$

We obtain for all $n \in \mathbb{N}$:

$$
\begin{aligned}
&E[X(t) - X(s)|\mathcal{F}_s] \\
&= E[X(t) - X_n(t)|\mathcal{F}_s] + E[X_n(t) - X_n(s)|\mathcal{F}_s] + E[X_n(s) - X(s)|\mathcal{F}_s].
\end{aligned}
$$

Because $(X_n(t))_{t\in[a,b]}$ is a martingale, $E[X_n(t) - X_n(s)|\mathcal{F}_s] = 0$. We also observe that

$$
\begin{aligned}
&E\left[(E[X(t) - X_n(t)|\mathcal{F}_s])^2\right] \\
&\leq E\left[E\left[|X(t) - X_n(t)|^2 \middle| \mathcal{F}_s\right]\right] = E\left[|X(t) - X_n(t)|^2\right] \\
&= E\left[\left|\int_a^t (f(u) - f_n(u))dW_u\right|^2\right] = \int_a^t E\left[|f(u) - f_n(u)|^2\right] du \overset{n}{\to} 0,
\end{aligned}
$$

following the properties of conditional expectations and by point 3 of Proposition 3.19. Hence $E[X(t) - X_n(t)|\mathcal{F}_s]$ converges to zero in $L^2(\Omega)$, analogously, and so does $E[X(s) - X_n(s)|\mathcal{F}_s]$, proving equation (3.11). Finally we need to show that $X(t)$ is $\mathcal{F}_t$-measurable for $t \in [a, b]$. This follows from the fact that $X_n(t)$ is $\mathcal{F}_t$-measurable for $n \in \mathbb{N}$ and moreover

$$
E\left[|X(t) - X_n(t)|^2\right] = \int_a^t E\left[|f(u) - f_n(u)|^2\right] du \overset{n}{\to} 0,
$$

following the above derivation. Hence $X_n(t) \to X(t)$ in $L^2(\Omega)$. □

Proposition 3.38. *Resorting to the notation of the preceding theorem, the martingale $(X_t)_{t\in[a,b]}$ is continuous (in $L^2(\Omega)$).*

Proof: If $t, s \in [a, b]$, then

$$
\begin{aligned}
\lim_{t\to s} E\left[|X(t) - X(s)|^2\right] &= \lim_{t\to s} E\left[\left|\int_s^t f(u)dW_u\right|^2\right] \\
&= \lim_{t\to s} \int_s^t E\left[(f(u))^2\right] du = 0,
\end{aligned}
$$

by point 3 of Proposition 3.19 and following the continuity of the Lebesgue integral. □

Theorem 3.39. *If $f \in \mathcal{C}_1$, then $(X_t)_{t\in[a,b]}$ admits a continuous version and thus admits a modified form with almost every trajectory being continuous.*

Proof: See, e.g., Baldi (1984) or Friedman (1975). □

Following Theorems 2.27 and 3.39, from now on we can always consider continuous and separable versions of $(X_t)_{t\in[a,b]}$. If $f \in \mathcal{C}$ and $X(t) = \int_a^t f(u)dW_u$, $t \in [a, b]$, then because (by Theorem 3.37) $(X_t)_{t\in[a,b]}$ is a martingale, it satisfies Doob's inequality (Proposition 2.69) and the following proposition holds.

Proposition 3.40. *If $f \in \mathcal{C}$, then*

1. $E[\max_{a \leq s \leq b} |\int_a^s f(u)dW_u|^2] \leq 4E[|\int_a^b f(u)dW_u|^2] = 4E[\int_a^b |f(u)|^2 du]$;
2. $P(\max_{a \leq s \leq b} |\int_a^s f(u)dW_u| > \lambda) \leq \frac{1}{\lambda^2} E[\int_a^b |f(u)|^2 du]$, $\lambda > 0$.

Proof: Point 1 follows directly from 2 of Proposition 2.65 with $p = 2$.

Point 2 follows by continuity

$$\left(\max_{a \leq s \leq b} \left| \int_a^s f(u)dW_u \right| \right)^2 = \max_{a \leq s \leq b} \left| \int_a^s f(u)dW_u \right|^2;$$

therefore

$$P\left(\max_{a \leq s \leq b} \left| \int_a^s f(u)dW_u \right| > \lambda \right) = P\left(\left(\max_{a \leq s \leq b} \left| \int_a^s f(u)dW_u \right| \right)^2 > \lambda^2 \right)$$
$$= P\left(\max_{a \leq s \leq b} \left| \int_a^s f(u)dW_u \right|^2 > \lambda^2 \right)$$

and the proof follows from 1 of Proposition 2.69. □

Remark 3.41. Generally, $\max_{a \leq s \leq b} X_s$, almost everywhere with respect to P, is defined due to the continuity of Brownian motion.

Stochastic Integrals with Stopping Times

Let $f \in \mathcal{C}_1([0,T])$, $(W_t)_{t \in \mathbb{R}_+}$ a Wiener process and τ_1, τ_2 two random variables representing stopping times, such that $0 \leq \tau_1 \leq \tau_2 \leq T$. Then

$$\int_{\tau_1}^{\tau_2} f(t)dW_t = \int_0^{\tau_2} f(t)dW_t - \int_0^{\tau_1} f(t)dW_t.$$

Lemma 3.42. *Defining the characteristic function as*

$$\chi_i(t) = \begin{cases} 1 & \text{if } t < \tau_i, \\ 0 & \text{if } t \geq \tau_i, \end{cases} \quad i = 1, 2,$$

we have that

1. $\chi_i(t)$ *is* $\mathcal{F}_t = \sigma(W_s, 0 \leq s \leq t)$*-measurable* $(i = 1, 2)$;
2. $\int_{\tau_1}^{\tau_2} f(t)dW_t = \int_0^T \chi_2(t) f(t)dW_t - \int_0^T \chi_1(t) f(t)dW_t$.

Proof: See, e.g., Friedman (1975). □

Theorem 3.43. *Let $f \in \mathcal{C}_1([0,T])$ and let τ_1, τ_2 be two stopping times, such that $0 \leq \tau_1 \leq \tau_2 \leq T$. Then*

1. $E[\int_{\tau_1}^{\tau_2} f(t)dW_t] = 0$;
2. $E[(\int_{\tau_1}^{\tau_2} f(t)dW_t)^2] = E[\int_{\tau_1}^{\tau_2} |f(t)|^2 dt]$.

Proof: By Lemma 3.42, we get

$$\int_{\tau_1}^{\tau_2} f(t)dW_t = \int_0^T (\chi_2(t) - \chi_1(t))f(t)dW_t,$$

and after applying Proposition 3.19 the proof is completed. This theorem is, in fact, just a generalization of Proposition 3.19. □

3.3 Itô Integrals of Multidimensional Wiener Processes

We denote by $\mathbb{R}^{mn}$ all real-valued $m \times n$ matrices and by

$$\mathbf{W}(t) = (W_1(t), \ldots, W_n(t))', \qquad t \geq 0,$$

an n-dimensional Wiener process. Let $[a,b] \subset [0, +\infty[$ and we put

$$\begin{aligned}
&\mathcal{C}_{\mathbf{W}}([a,b]) \\
&= \{f : [a,b] \times \Omega \to \mathbb{R}^{mn} | \forall 1 \leq i \leq m, \forall 1 \leq j \leq n : f_{ij} \in \mathcal{C}_{W_j}([a,b])\}, \\
&\mathcal{C}_{1\mathbf{W}}([a,b]) \\
&= \{f : [a,b] \times \Omega \to \mathbb{R}^{mn} | \forall 1 \leq i \leq m, \forall 1 \leq j \leq n : f_{ij} \in \mathcal{C}_{1W_j}([a,b])\},
\end{aligned}$$

where $\mathcal{C}_{W_j}([a,b])$ and $\mathcal{C}_{1W_j}([a,b])$ correspond to the classes $\mathcal{C}([a,b])$ and $\mathcal{C}_1([a,b])$ respectively, as defined in part 3.1.

Definition 3.44. If $f : [a,b] \times \Omega \to \mathbb{R}^{mn}$ belongs to $\mathcal{C}_{1\mathbf{W}}([a,b])$, then the *stochastic integral with respect to* $\mathbf{W}$ is the m-dimensional vector defined by

$$\int_a^b f(t)d\mathbf{W}(t) = \left(\sum_{j=1}^n \int_a^b f_{ij}(t)dW_j(t) \right)'_{1 \leq i \leq m}, \tag{3.12}$$

where each of the integrals on the right-hand side is defined in the sense of Itô.

Proposition 3.45. *If* $(i,j) \in \{1, \ldots, n\}^2$ *and*

$$\begin{aligned}
&f_i : [a,b] \times \Omega \to \mathbb{R} \text{ belongs to } \mathcal{C}_{W_i}([a,b]); \\
&f_j : [a,b] \times \Omega \to \mathbb{R} \text{ belongs to } \mathcal{C}_{W_j}([a,b]),
\end{aligned}$$

then

$$E\left[\int_a^b f_i(t)dW_i(t) \int_a^b f_j(t)dW_j(t) \right] = \delta_{ij} E\left[\int_a^b f_i(t)f_j(t)dt \right], \tag{3.13}$$

where $\delta_{ij} = 1$*, if* $i = j$ *or* $\delta_{ij} = 0$*, if* $i \neq j$*, is the Kronecker delta.*

Proof: Suppose $i \neq j$. Then the processes $(W_i(t))_{t\geq 0}$ and $(W_j(t))_{t\geq 0}$ are independent. Hence so are the σ-algebras $\mathcal{F}^{(i)} = \sigma(W_i(s), s \geq 0)$ and $\mathcal{F}^{(j)} = \sigma(W_j(s), s \geq 0)$. Moreover, for all $t \in [a,b]: f_i(t)$ is $\mathcal{F}^{(i)}$-measurable, $f_j(t)$ is $\mathcal{F}^{(j)}$-measurable, and $\mathcal{F}_t^{(i)} = \sigma(W_i(s), 0 \leq s \leq t) \subset \mathcal{F}^{(i)}$ as well as $\mathcal{F}_t^{(j)} = \sigma(W_j(s), 0 \leq s \leq t) \subset \mathcal{F}^{(j)}$. Therefore, $f_i = (f_i(t))_{t\in[a,b]}$ and $f_j = (f_j(t))_{t\in[a,b]}$ are independent. So are $\int_a^b f_i(t)dW_i(t)$ and $\int_a^b f_j(t)dW_j(t)$, and therefore

$$E\left[\int_a^b f_i(t)dW_i(t) \int_a^b f_j(t)dW_j(t)\right]$$
$$= E\left[\int_a^b f_i(t)dW_i(t)\right] E\left[\int_a^b f_j(t)dW_j(t)\right] = 0,$$

by Proposition 3.19. If instead $i = j$, then the proof immediately follows by Proposition 3.19. □

Proposition 3.46. *Let $f : [a,b] \times \Omega \to \mathbb{R}^{mn}$ and $g : [a,b] \times \Omega \to \mathbb{R}^{mn}$. Then*

1. *if $f \in \mathcal{C}_{\mathbf{W}}([a,b])$, then*

$$E\left[\int_a^b f(t)d\mathbf{W}(t)\right] = \mathbf{0} \in \mathbb{R}^m;$$

2. *if $f, g \in \mathcal{C}_{\mathbf{W}}([a,b])$, then*

$$E\left[\left(\int_a^b f(t)d\mathbf{W}(t)\right)\left(\int_a^b g(t)d\mathbf{W}(t)\right)'\right] = E\left[\int_a^b (f(t))(g(t))'dt\right];$$

3. *if $f \in \mathcal{C}_{\mathbf{W}}([a,b])$, then*

$$E\left[\left|\int_a^b f(t)d\mathbf{W}(t)\right|^2\right] = E\left[\int_a^b |f(t)|^2 dt\right],$$

where

$$|f|^2 = \sum_{i=1}^m \sum_{j=1}^n (f_{ij})^2$$

and

$$\left|\int_a^b f(t)d\mathbf{W}(t)\right|^2 = \sum_{i=1}^m \left(\sum_{j=1}^n \int_a^b f_{ij}(t)dW_j(t)\right)^2.$$

Proof: 1. Let $f \in \mathcal{C}_{\mathbf{W}}([a,b]) (\subset \mathcal{C}_{1\mathbf{W}}([a,b]))$. Then

$$E\left[\int_a^b f(t)d\mathbf{W}(t)\right] = \left(E\left[\sum_{j=1}^n \int_a^b f_{ij}(t)dW_j(t)\right]\right)'_{1\leq i\leq m}$$
$$= \left(\sum_{j=1}^n E\left[\int_a^b f_{ij}(t)dW_j(t)\right]\right)'_{1\leq i\leq m} = \mathbf{0} \in \mathbb{R}^m,$$

by Proposition 3.19.

2. Let $f, g \in \mathcal{C}_{\mathbf{W}}([a,b])$ and $(1,k) \in \{1,\dots,m\}^2$. Then

$$E\left[\left(\int_a^b f(t)d\mathbf{W}(t)\right)\left(\int_a^b g(t)d\mathbf{W}(t)\right)'\right]_{lk}$$
$$= E\left[\left(\sum_{j=1}^n \int_a^b f_{lj}(t)dW_j(t)\right)\left(\sum_{j'=1}^n \int_a^b g_{j'k}(t)dW_{j'}(t)\right)\right]$$
$$= \sum_{j=1}^n \sum_{j'=1}^n E\left[\int_a^b f_{lj}(t)dW_j(t)\int_a^b g_{j'k}dW_{j'}(t)\right] = \sum_{j=1}^n E\left[\int_a^b f_{lj}(t)g_{jk}(t)dt\right]$$
$$= E\left[\sum_{j=1}^n \int_a^b f_{lj}(t)g_{jk}(t)dt\right] = E\left[\int_a^b \sum_{j=1}^n (f_{lj}(t)g_{jk}(t))dt\right]$$
$$= E\left[\int_a^b ((f(t))(g(t))')_{lk}dt\right],$$

by Proposition 3.45. Having verified each of the components, the proof of 2 is complete.

3. Let $f \in \mathcal{C}_{\mathbf{W}}([a,b])$. Then by 2 we have

$$E\left[\left(\int_a^b f(t)d\mathbf{W}(t)\right)\left(\int_a^b f(t)d\mathbf{W}(t)\right)'\right] = E\left[\int_a^b (f(t))(f(t))'dt\right]. \quad (3.14)$$

Furthermore, it is easily verified that if a generic $b \in \mathbb{R}^{mn}$, then

$$|b|^2 = \sum_{i=1}^m \sum_{j=1}^n (b_{ij})^2 = trace(bb'),$$

and if a generic $\mathbf{a} \in \mathbb{R}^m$, then

$$|\mathbf{a}|^2 = \sum_{i=1}^m (a_i)^2 = trace(\mathbf{a}\mathbf{a}').$$

Therefore, if in equation (3.14) we consider the trace of both the former and the latter term, we obtain 3. □

3.4 The Stochastic Differential

Definition 3.47. Let $(u(t))_{0\leq t\leq T}$ be a process such that for every $(t_1,t_2)\in [0,T]\times[0,T], t_1<t_2$:

$$u(t_2)-u(t_1)=\int_{t_1}^{t_2}a(t)dt+\int_{t_1}^{t_2}b(t)dW_t, \tag{3.15}$$

where $a\in\mathcal{C}_1([0,T])$ and $b\in\mathcal{C}_1([0,T])$. Then $u(t)$ is said to have the *stochastic differential*

$$du(t)=a(t)dt+b(t)dW_t \tag{3.16}$$

on $[0,T]$.

Remark 3.48. If $u(t)$ has the stochastic differential in the form of (3.16), then for all $t>0$, we have

$$u(t)=u(0)+\int_0^t a(s)ds+\int_0^t b(s)dW_s.$$

Hence

1. the trajectories of $(u(t))_{0\leq t\leq T}$ are continuous almost everywhere (see Theorem 3.39);
2. for $t\in[0,T]$, $u(t)$ is $\mathcal{F}_t=\sigma(W_s, 0\leq s\leq t)$-measurable, thus $u(t)\in\mathcal{C}_1([0,T])$.

Example 3.49. The stochastic differential of $(W_t^2)_{t\geq 0}$ is given by

$$dW_t^2=dt+2W_tdW_t. \tag{3.17}$$

In fact, if $0\leq t_1<t_2$, then, by Proposition 3.35, it follows that

$$\int_{t_1}^{t_2}W_tdW_t=\frac{1}{2}(W_{t_2}^2-W_{t_1}^2)-\frac{t_2-t_1}{2}.$$

Therefore $W_{t_2}^2-W_{t_1}^2=t_2-t_1+2\int_{t_1}^{t_2}W_tdW_t$, which is of the form (3.15) with $a(t)=1$ and $b(t)=2W_t$, $t\geq 0$.

Example 3.50. The stochastic differential of the process $(tW_t)_{t\geq 0}$ is given by

$$d(tW_t)=W_tdt+tdW_t. \tag{3.18}$$

Let $0\leq t_1<t_2$ and $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of partitions of $[t_1,t_2]$, where $\pi_n: t_1=r_1^{(n)}<\cdots<r_n^{(n)}=t_2$, such that $|\pi_n|\xrightarrow{n}0$. Then, by Theorem 3.31,

$$\int_{t_1}^{t_2}tdW_t=P-\lim_{n\to\infty}\sum_{k=1}^{n-1}r_k^{(n)}\left(W_{r_{k+1}^{(n)}}-W_{r_k^{(n)}}\right). \tag{3.19}$$

Moreover, because $(W_t)_{t\geq 0}$ is continuous almost surely, we can consider $\int_{t_1}^{t_2} W_t dt$, obtaining

$$\int_{t_1}^{t_2} W_t dt = \lim_{n\to\infty} \sum_{k=1}^{n-1} W_{r_{k+1}^{(n)}} \left(r_{k+1}^{(n)} - r_k^{(n)} \right) \text{ almost surely.}$$

But since almost sure convergence implies convergence in probability, we have

$$\int_{t_1}^{t_2} W_t dt = P - \lim_{n\to\infty} \sum_{k=1}^{n-1} W_{r_{k+1}^{(n)}} \left(r_{k+1}^{(n)} - r_k^{(n)} \right). \tag{3.20}$$

Combining the relevant terms of (3.19) and (3.20), we obtain

$$\begin{aligned}\int_{t_1}^{t_2} t dW_t + \int_{t_1}^{t_2} W_t dt &= P - \lim_{n\to\infty} \sum_{k=1}^{n-1} \left(r_{k+1}^{(n)} W_{r_{k+1}^{(n)}} - r_k^{(n)} W_{r_k^{(n)}} \right) \\ &= t_2 W_{t_2} - t_1 W_{t_1},\end{aligned}$$

which is of form (3.15) with $a(t) = W_t$ and $b(t) = t$, proving equation (3.18).

Proposition 3.51. *If the stochastic differential of* $(u_i(t))_{t\in[0,T]}$ *is given by*

$$du_i(t) = a_i(t)dt + b_i(t)dW_t, \qquad i = 1, 2,$$

then $(u_1(t)u_2(t))_{t\in[0,T]}$ *has the stochastic differential*

$$d(u_1(t)u_2(t)) = u_1(t)du_2(t) + u_2(t)du_1(t) + b_1(t)b_2(t)dt, \tag{3.21}$$

and thus, for all $0 \leq t_1 < t_2 < T$

$$\begin{aligned}&u_1(t_2)u_2(t_2) - u_1(t_1)u_2(t_1) \\ &= \int_{t_1}^{t_2} u_1(t)a_2(t)dt + \int_{t_1}^{t_2} u_1(t)b_2(t)dW_t \\ &\quad + \int_{t_1}^{t_2} u_2(t)a_1(t)dt + \int_{t_1}^{t_2} u_2(t)b_1(t)dW_t + \int_{t_1}^{t_2} b_1(t)b_2(t)dt.\end{aligned} \tag{3.22}$$

Proof (see, e.g., Baldi (1984)): Case 1: a_i, b_i constant on $[t_1, t_2]$, i.e., $a_i(t) = a_i$, $b_i(t) = b_i$, for all $t \in [t_1, t_2], i = 1, 2,$ a_i, b_i in $\mathcal{C}_1([0, T])$. Then

$$u_1(t_2) = u_1(t_1) + a_1(t_2 - t_1) + b_1(W_{t_2} - W_{t_1}), \tag{3.23}$$

$$u_2(t_2) = u_2(t_1) + a_2(t_2 - t_1) + b_2(W_{t_2} - W_{t_1}). \tag{3.24}$$

The proof of formula (3.22) is complete by employing equations (3.17), (3.18), (3.23), (3.24), and the definitions of both the Lebesgue and stochastic integrals.

Case 2: It can be shown that (3.22) holds for a_i, b_i, $i = 1, 2$, being piecewise functions.

Case 3: Eventually it can be shown that (3.22) holds for any a_i, b_i ($a_i, b_i \in \mathcal{C}$, $i = 1, 2$). □

Remark 3.52. Generally, if $u(t), b(t) \in \mathcal{C}_1([0,T])$, then, by the Cauchy–Schwarz inequality (see Proposition 1.147), $u(t)b(t) \in \mathcal{C}_1([0,T])$ as well, and so $\int_0^T u(t)b(t)dW_t$ is well defined.

Remark 3.53. If $f : \mathbb{R} \to \mathbb{R}$ is a continuous function, then $f(W_t) \in \mathcal{C}_1([0,T])$; in fact, the trajectories of $(f(W_t))_{t\in[0,T]}$ are continuous almost everywhere and thus condition (3.4) is certainly verified. In particular, we have

$$(W_t^n)_{t\in[0,T]} \in \mathcal{C}_1([0,T]) \qquad \forall n \in \mathbb{N}^*.$$

Corollary 3.54. For every integer $n \geq 2$ we get

$$d(W_t^n) = nW_t^{n-1}dW_t + \frac{1}{2}(n-1)nW_t^{n-2}dt. \tag{3.25}$$

Proof: The proof follows from Proposition 3.51 by induction. □

Corollary 3.55. For every polynomial $P(x)$:

$$dP(W_t) = P'(W_t)dW_t + \frac{1}{2}P''(W_t)dt. \tag{3.26}$$

Remark 3.56. The second derivative of $P(W_t)$ is required for its differential.

Proposition 3.57. *If* $f \in C^2(\mathbb{R})$, *then*

$$df(W_t) = f'(W_t)dW_t + \frac{1}{2}f''(W_t)dt. \tag{3.27}$$

Proof: Given the integration-by-parts formula

$$f(x) = f(0) + f'(0)x + \int_0^x (x-y)f''(y)dy, \tag{3.28}$$

and because $f \in C^2(\mathbb{R})$, it follows that $f'' \in C^0(\mathbb{R})$. Then, by the Weierstrass theorem, we can approximate it with polynomials. Hence

$$\exists (q_n(x))_{n\in\mathbb{N}}, \tag{3.29}$$

a sequence of polynomials uniformly converging to f'' on compacts. If we now write

$$Q_n(x) = f(0) + f'(0)x + \int_0^x (x-y)q_n(y)dy, \qquad n \in \mathbb{N},$$

it is evident that $Q_n(x)$ is a polynomial with $Q_n'' = q_n(x)$; thus $(Q_n''(x))_{n\in\mathbb{N}}$ uniformly converges to f'' on compacts. Moreover, $Q_n(x) \xrightarrow{n} f(x)$, $Q_n'(x) \xrightarrow{n} f'(x)$ uniformly on its compacts. In fact, by (3.29), it is possible to replace the limit with the integral in (3.28). Applying (3.26) to the polynomials Q_n, we obtain that for $t_1 < t_2$

$$Q_n(W_{t_2}) - Q_n(W_{t_1}) = \int_{t_1}^{t_2} Q_n'(W_t)dW_t + \frac{1}{2}\int_{t_1}^{t_2} Q_n''(W_t)dt. \tag{3.30}$$

Therefore we observe that

$$Q_n(W_{t_2}) \xrightarrow{n} f(W_{t_2}) \text{ a.s. (and thus in probability),}$$
$$Q_n(W_{t_1}) \xrightarrow{n} f(W_{t_1}) \text{ a.s. (and thus in probability),}$$
$$\frac{1}{2}\int_{t_1}^{t_2} Q_n''(W_t)dt \xrightarrow{n} \frac{1}{2}\int_{t_1}^{t_2} f''(W_t)dt \text{ a.s. (and thus in probability),}$$

and also

$$\lim_{n\to\infty}\int_{t_1}^{t_2}[Q_n'(W_t) - f'(W_t)]^2 dt = 0 \text{ a.s. (and thus in probability).}$$

Hence by Theorem 3.29 we have that

$$P - \lim_{n\to\infty}\int_{t_1}^{t_2} Q_n'(W_t)dW_t = \int_{t_1}^{t_2} f'(W_t)dW_t.$$

Finally, by equation (3.30)

$$f(W_{t_2}) - f(W_{t_1}) = \int_{t_1}^{t_2} f'(W_t)dW_t + \frac{1}{2}\int_{t_1}^{t_2} f''(W_t)dt.$$

□

3.5 Itô's Formula

As one of the most important topics on Brownian motion, Itô's formula represents the stochastic equivalent of Taylor's theorem about the expansion of functions. It is the key concept that connects classical and stochastic theory.

Proposition 3.58. *If $u(t,x) : [0,T]\times\mathbb{R}\to\mathbb{R}$ is continuous with the derivatives u_x, u_{xx}, and u_t, then*

$$du(t,W_t) = \left(u_t(t,W_t) + \frac{1}{2}u_{xx}(t,W_t)\right)dt + u_x(t,W_t)dW_t. \tag{3.31}$$

Proof: Case 1: We suppose $u(t,x) = g(t)\psi(x)$, with $g \in C^1([0,T])$ and $\psi \in C^2(\mathbb{R})$. Then by Proposition 3.57,

$$d\psi(W_t) = \psi'(W_t)dW_t + \frac{1}{2}\psi''(W_t)dt$$

and, by formula (3.21), we obtain an expression for (3.31), namely

$$d(g(t)\psi(W_t)) = g(t)\psi'(W_t)dW_t + \frac{1}{2}g(t)\psi''(W_t)dt + \psi(W_t)g'(t)dt.$$

Case 2: If

$$u(t,x) = \sum_{i=1}^{n} g_i(t)\psi_i(x), \qquad g \in C^1([0,T]), \psi \in C^2(\mathbb{R}), i = 1,\ldots,n, \quad (3.32)$$

then (3.31) is an immediate consequence of the first case.

Case 3: If u is a generic function, satisfying the hypotheses of the proposition, it can be shown that there exists $(u_n)_{n\in\mathbb{N}}$, a sequence of functions of type (3.32), such that for all $K > 0$:

$$\lim_{n\to\infty} \sup_{|x|\le K} \sup_{t\in[0,T]} \{|u_n - u| + |(u_n)_t - u_t| + |(u_n)_x - u_x| + |(u_n)_{xx} - u_{xx}|\} = 0.$$

Therefore, we can approximate u uniformly through the sequence u_n and the proof follows from the second case. □

Remark 3.59. We note that, contrary to what is obtained for an ordinary differential, (3.31) contains the additional term $\frac{1}{2}u_{xx}(t, W_t)dt$. This is due to the presence of Brownian motion.

Remark 3.60. If $u(t,z,\omega) : [0,T] \times \mathbb{R} \times \Omega \to \mathbb{R}$ is continuous with the derivatives u_z, u_{zz}, and u_t such that, for all (t,z), $u(t,z,\cdot)$ is $\mathcal{F}_t = \sigma(W_s, 0 \le s \le t)$-measurable, then formula (3.31) holds for every $\omega \in \Omega$.

Theorem 3.61. (Itô's formula). *If $du(t) = a(t)dt + b(t)dW_t$ and if $f(t,x)$: $[0,T] \times \mathbb{R} \to \mathbb{R}$ is continuous with the derivatives f_x, f_{xx}, and f_t, then the stochastic differential of the process $f(t,u(t))$ is given by*

$$df(t,u(t)) = \left(f_t(t,u(t)) + \frac{1}{2}f_{xx}(t,u(t))b^2(t) + f_x(t,u(t))a(t)\right)dt + f_x(t,u(t))b(t)dW_t. \quad (3.33)$$

Proof: See, e.g., Karatzas and Shreve (1991). □

3.6 Martingale Representation Theorem

Theorem 3.37 stated that, given a process $(f_t)_{t\in[0,T]} \in \mathcal{C}([0,T])$, the Itô integral $\int_0^t f_s dW_s$ is a zero mean $\mathcal{L}^2$-martingale. The martingale representation theorem establishes the relationship between a martingale and the existence of a process vice versa.

Theorem 3.62. (martingale representation theorem I). *Let $(M_t)_{t\in[0,T]}$ be an $\mathcal{L}^2$-martingale with respect to the Wiener process $(W_t)_{t\in[0,T]}$ and $(\mathcal{F}_t)_{t\in[0,T]}$ its natural filtration. Then there exists a unique process $(f_t)_{t\in[0,T]} \in \mathcal{C}([0,T])$, so that*

$$\forall t \in [0,T]: \qquad M(t) = M(0) + \int_0^t f(s)dW_s \qquad a.s. \quad (3.34)$$

Theorem 3.63. (Martingale representation theorem II). *Let $(M_t)_{t\in[0,T]}$ be a martingale with respect to the Wiener process $(W_t)_{t\in[0,T]}$ and $(\mathcal{F}_t)_{t\in[0,T]}$ its natural filtration. Then there exists a unique process $(f_t)_{t\in[0,T]} \in \mathcal{C}_1([0,T])$ so that (3.34) holds.*

The martingale representation theorems are a direct consequence of the following theorem (see Øksendal (1998)).

Theorem 3.64. (Itô representation theorem). *Let $(X_t)_{t\in[0,T]} \in L^2(\Omega, \mathcal{F}_T, P)$ be a stochastic process. Then there exists a unique process $(f_t)_{t\in[0,T]} \in \mathcal{C}([0,T])$, so that*

$$\forall t \in [0,T]: \qquad X(t) = E[X(0)] + \int_0^t f(s)dW_s.$$

For the proof of the Itô representation theorem we require the following lemma.

Lemma 3.65. *The linear span of random variables of the Doléans exponential type*

$$\exp\left\{\int_0^T h(t)dW_t - \frac{1}{2}\int_0^T (h(t))^2 dt\right\}$$

for a deterministic process $(h_t)_{t\in[0,T]} \in L^2([0,T])$ is dense in $L^2(\Omega, \mathcal{F}_T, P)$.

Proof (of the Itô representation theorem): Initially suppose that $(X_t)_{t\in[0,T]}$ has the Doléans exponential form

$$X_t = \exp\left\{\int_0^t h(s)dW_s - \frac{1}{2}\int_0^t (h(s))^2 ds\right\} \qquad \forall t \in [0,T],$$

for a deterministic process $(h_t)_{t\in[0,T]} \in L^2([0,T])$. Also define

$$Y(t) = \exp\left\{\int_0^t h(s)dW_s - \frac{1}{2}\int_0^t (h(s))^2 ds\right\} \qquad \forall t \in [0,T].$$

Then, invoking Itô's formula we obtain

$$\begin{aligned} &dY(t) \\ &= Y(t)\left(h(t)dW_t - \frac{1}{2}(h(t))^2 dt\right) + \frac{1}{2}Y(t)(h(t))^2 dt = Y(t)h(t)dW_t \end{aligned} \tag{3.35}$$

Therefore

$$Y(t) = 1 + \int_0^t Y(s)h(s)dW_s, \qquad t \in [0,T],$$

and in particular

$$X(T) = Y(T) = 1 + \int_0^T Y(s)h(s)dW_s,$$

so that, after taking expectations, we obtain $E[X(T)] = 1$. Now by Lemma 3.65 we may extend the proof to any arbitrary $(X_t)_{t\in[0,T]} \in L^2(\Omega, \mathcal{F}_T, P)$. To prove that the process $(h_t)_{t\in[0,T]}$ is unique, suppose that two processes $h_t^1, h_t^2 \in \mathcal{C}([0,T])$ exist with

$$X(T) = E[X(0)] + \int_0^T h^1(t)dW_t = E[X(0)] + \int_0^T h^2(t)dW_t.$$

Subtracting the two integrals and taking expectation of the squared difference, we obtain

$$E\left[\left(\int_0^T \left(h^1(t) - h^2(t)\right) dW_t\right)^2\right] = 0,$$

and using the Itô isometry we obtain

$$\int_0^T E\left[h^1(t) - h^2(t)\right]^2 dt = 0,$$

implying that $h_t^1 = h_t^2$ almost surely for all $t \in [0,T]$. □

3.7 Multidimensional Stochastic Differentials

Definition 3.66. Let $(\mathbf{u}_t)_{0\leq t\leq T}$ be an m-dimensional process and

$$\mathbf{a} : [0,T] \times \Omega \to \mathbb{R}^m, \mathbf{a} \in \mathcal{C}_{1\mathbf{W}}([0,T]),$$
$$b : [0,T] \times \Omega \to \mathbb{R}^{mn}, b \in \mathcal{C}_{1\mathbf{W}}([0,T]).$$

The *stochastic differential* $d\mathbf{u}(t)$ of $\mathbf{u}(t)$ is given by

$$d\mathbf{u}(t) = \mathbf{a}(t)dt + b(t)d\mathbf{W}(t) \tag{3.36}$$

if, for all $0 \leq t_1 < t_2 \leq T$

$$\mathbf{u}(t_2) - \mathbf{u}(t_1) = \int_{t_1}^{t_2} \mathbf{a}(t)dt + \int_{t_1}^{t_2} b(t)d\mathbf{W}(t).$$

Remark 3.67. Under the assumptions of the preceding definition, we obtain for $1 \leq i \leq m$

$$du_i(t) = a_i(t)dt + \sum_{j=1}^n (b_{ij}(t)dW_j(t)).$$

Example 3.68. Suppose that the coefficients a_{11} and a_{12} of the system

$$\begin{cases} du_1(t) = a_{11}(t)u_1(t)dt + a_{12}(t)u_2(t)dt, \\ du_2(t) = a_{21}(t)u_1(t)dt + a_{22}(t)u_2(t)dt \end{cases} \tag{3.37}$$

are subject to the noise

$$a_{11}(t)dt = a_{11}^0(t)dt + \tilde{a}_{11}(t)dW_1(t),$$
$$a_{12}(t)dt = a_{12}^0(t)dt + \tilde{a}_{12}(t)dW_2(t).$$

The first equation of (3.37) becomes

$$\begin{aligned} &du_1(t) \\ &= (a_{11}^0(t)u_1(t) + a_{12}^0(t)u_2(t))dt + \tilde{a}_{11}(t)u_1(t)dW_1(t) + \tilde{a}_{12}(t)u_2(t)dW_2(t) \\ &= \bar{a}_1(t)dt + b_{11}(t)dW_1(t) + b_{12}(t)dW_2(t), \end{aligned}$$

where the meaning of the new parameters $\bar{a}_1, b_{11}$, and b_{12} is obvious. Now, if both a_{21} and a_{22} are affected by the noise

$$a_{21}(t)dt = a_{21}^0(t)dt + \tilde{a}_{21}(t)dW_3(t),$$
$$a_{22}(t)dt = a_{22}^0(t)dt + \tilde{a}_{22}(t)dW_4(t),$$

then the second equation of (3.37) becomes

$$du_2(t) = \bar{a}_2(t)dt + b_{23}(t)dW_3(t) + b_{24}(t)dW_4(t).$$

In this case the matrix

$$b = \begin{pmatrix} b_{11} & b_{12} & 0 & 0 \\ 0 & 0 & b_{23} & b_{24} \end{pmatrix}$$

is of order 2×4, but, in general, it is possible that $m > n$.

Theorem 3.69. (multidimensional Itô formula). *Let $f(t, \mathbf{x}) : \mathbb{R}_+ \times \mathbb{R}^m \to \mathbb{R}$ be continuous with the derivatives $f_{x_i}, f_{x_i x_j}$, and f_t. Let $\mathbf{u}(t)$ be an m-dimensional process, endowed with the stochastic differential*

$$d\mathbf{u}(t) = \mathbf{a}(t)dt + b(t)d\mathbf{W}(t),$$

where $\mathbf{a} = (a_1, \ldots, a_m)' \in \mathcal{C}_{\mathbf{W}}([0, T])$ and $b = (b_{ij})_{1 \le i \le m, 1 \le j \le n} \in \mathcal{C}_{\mathbf{W}}([0, T])$. Then $f(t, \mathbf{u}(t))$ has the stochastic differential

$$\begin{aligned} df(t, \mathbf{u}(t)) = &\left(f_t(t, \mathbf{u}(t)) + \sum_{i=1}^m f_{x_i}(t, \mathbf{u}(t)) a_i(t) \right. \\ &\left. \quad + \frac{1}{2} \sum_{l=1}^n \sum_{i,j=1}^m f_{x_i x_j}(t, \mathbf{u}(t)) b_{il}(t) b_{jl}(t) \right) dt \\ &+ \sum_{l=1}^n \sum_{i=1}^m f_{x_i}(t, \mathbf{u}(t)) b_{il}(t) dW_l(t). \end{aligned} \tag{3.38}$$

If we put $a_{ij} = (bb')_{ij}$*,* $i, j = 1, \ldots, m$*:*

$$L = \frac{1}{2}\sum_{i,j=1}^{m} a_{ij}\frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{m} a_i \frac{\partial}{\partial x_i} + \frac{\partial}{\partial t}$$

and introduce the gradient operator

$$\nabla_{\mathbf{x}} = \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_m}\right)',$$

then, in vector notation, equation (3.38) can be written as

$$df(t, \mathbf{u}(t)) = Lf(t, \mathbf{u}(t))dt + \nabla_{\mathbf{x}} f(t, \mathbf{u}(t)) \cdot b(t)d\mathbf{W}(t), \tag{3.39}$$

where $\nabla_{\mathbf{x}} f(t, \mathbf{u}(t)) \cdot b(t)d\mathbf{W}(t)$ *is the scalar product of two* m*-dimensional vectors.*

Proof: Employing the following two lemmas the proof is similar to the one-dimensional case. (See, e.g., Baldi (1984).) □

Lemma 3.70. *If* $(W_1(t))_{t \geq 0}$ *and* $(W_2(t))_{t \geq 0}$ *are two independent Wiener processes, then*

$$d(W_1(t)W_2(t)) = W_1(t)dW_2(t) + W_2(t)dW_1(t). \tag{3.40}$$

Proof: Since $W_1(t)$ and $W_2(t)$ are independent, it is easily shown that $W_t = \frac{1}{\sqrt{2}}(W_1(t) + W_2(t))$ is also a Wiener process. Moreover, for a Wiener process $W(t)$ we have

$$dW^2(t) = dt + 2W(t)dW(t). \tag{3.41}$$

Hence from

$$W_1(t)W_2(t) = W^2(t) - \frac{1}{2}W_1^2(t) - \frac{1}{2}W_2^2(t),$$

it follows that $W_1(t)W_2(t)$ is endowed with the differential

$$\begin{aligned}
&d(W_1(t)W_2(t)) \\
&= dW^2(t) - \frac{1}{2}dW_1^2(t) - \frac{1}{2}dW_2^2(t) \\
&= dt + 2W(t)dW(t) - \frac{1}{2}dt - W_1(t)dW_1(t) - \frac{1}{2}dt - W_2(t)dW_2(t) \\
&= 2\left(\frac{1}{2}W_1(t)dW_1(t) + \frac{1}{2}W_1(t)dW_2(t) + \frac{1}{2}W_2(t)dW_1(t) + \frac{1}{2}W_2(t)dW_2(t)\right) \\
&\quad -W_1(t)dW_1(t) - W_2(t)dW_2(t),
\end{aligned}$$

completing the proof. □

Lemma 3.71. *If $W_1, \ldots, W_n$ are independent Wiener processes and*

$$du_i(t) = a_i(t)dt + \sum_{j=1}^{n}(b_{ij}(t)dW_j(t)), \qquad i = 1, 2,$$

then

$$d(u_1u_2)(t) = u_1(t)du_2(t) + u_2(t)du_1(t) + \sum_{j=1}^{n} b_{1j}b_{2j}dt. \tag{3.42}$$

Proof: It is analogous to the proof of Proposition 3.51 (see, e.g., Baldi (1984)). Use equations (3.40), (3.41), (3.18), and approximate the resulting polynomials. □

Remark 3.72. Note that equation (3.41) is not a particular case of (3.40) (in the latter, independence is not given), whereas equation (3.42) generalizes both.

Remark 3.73. The multidimensional Itô formula (3.39) asserts that the processes

$$f(t, \mathbf{u}(t)) - f(0, \mathbf{u}(0))$$

and

$$\int_0^t Lf(s, \mathbf{u}(s))ds + \int_0^t \nabla_{\mathbf{x}} f(s, \mathbf{u}(s)) \cdot b(s)d\mathbf{W}(s)$$

are stochastically equivalent. They are both continuous and so their trajectories coincide almost surely. Taking expectations on both sides, we therefore get

$$E[f(t, \mathbf{u}(t))] - E[f(0, \mathbf{u}(0))] = E\left[\int_0^t Lf(s, \mathbf{u}(s))ds\right].$$

3.8 Exercises and Additions

3.1. Let $(X_t)_{t\in\mathbb{R}_+}$ be a Brownian motion in $\mathbb{R}$, $X_0 = 0$. Prove directly from the definition of Itô integrals that

$$\int_0^t X_s^2 dX_s = \frac{1}{3}X_t^3 - \int_0^t X_s ds.$$

3.2. Prove Corollary 3.54.

3.3. Prove Lemma 3.71.

3.4. Prove the multidimensional Itô formula (3.39).

3.5. Let $(W_t)_{t\in\mathbb{R}_+}$ denote an n-dimensional Brownian motion and let $f:\mathbb{R}^n\to\mathbb{R}$ be C^2. Use Itô's formula to prove that

$$df(W_t)=\nabla f(W_t)dW_t+\frac{1}{2}\triangle f(W_t)dt,$$

where ∇ denotes the gradient and $\triangle=\sum_{i=1}^n\frac{\partial^2}{\partial x_i^2}$ is the Laplace operator.

3.6. Let $(W_t)_{t\in\mathbb{R}_+}$ be a one-dimensional Brownian motion with $W_0=0$. Using Itô's formula, show that

$$E[W_t^k]=\frac{1}{2}k(k-1)\int_0^t E[W_s^{k-2}]ds,\qquad k\geq 2,\ t\geq 0.$$

3.7. Use Itô's formula to write the following stochastic process u_t in the standard form

$$du_t=\mathbf{a}(t)dt+b(t)dW_t$$

for a suitable choice of $\mathbf{a}\in\mathbb{R}^n$, $b\in\mathbb{R}^{nm}$, and dimensions n,m:

1. $u_1(t,W_1(t))=3+2t+e^{2W_1(t)}$ ($W_1(t)$ is one-dimensional);
2. $u_2(t,\mathbf{W}_t)=W_2^2(t)+W_3^2(t)$ ($\mathbf{W}_t=(W_2(t),W_3(t))$ is two-dimensional);
3. $u_3(t,\mathbf{W}_t)=\ln(u_1(t)u_2(t))$;
4. $u_4(t,\mathbf{W}_t)=\exp\left\{\frac{u_1(t)}{u_2(t)}\right\}$;
5. $u_5(t,W_t)=(5+t,t+4W_t)$ (W_t one-dimensional);
6. $u_6(t,\mathbf{W}_t)=(W_1(t)+W_2(t)-W_3(t),W_2^2(t)-W_1(t)W_2(t)+W_3(t))$ ($\mathbf{W}_t=(W_1(t),W_2(t),W_3(t))$ is three-dimensional).

3.8. Let $(\mathbf{W}_t)_{t\in\mathbb{R}_+}$ be an n-dimensional Brownian motion starting at $x\neq 0$. Are the processes

$$u_t=\ln\left(|\mathbf{W}_t|^2\right),$$

$$v_t=\frac{1}{|\mathbf{W}_t|}$$

martingales? If not, find two processes $(\overline{u}_t)_{t\in\mathbb{R}_+}$, $(\overline{v}_t)_{t\in\mathbb{R}_+}$ such that

$$u_t-\overline{u}_t,$$

$$v_t-\overline{v}_t$$

are martingales.

3.9. Let $(X_t)_{t\in\mathbb{R}_+}$ be an Itô integral

$$dX_t=\mathbf{v}_t d\mathbf{W}_t,$$

where $\mathbf{v}_t\in\mathbb{R}^n$ and $(\mathbf{W}_t)_{t\in\mathbb{R}_+}$ is an n-dimensional Brownian motion.

1. Give an example to show that $\mathbf{X}_t^2$, in general, is not a martingale.

2. Prove that
$$M_t = \mathbf{X}_t^2 - \int_0^t |\mathbf{v}_s|^2 ds$$
is a martingale. The process $\langle \mathbf{X}, \mathbf{X} \rangle_t = \int_0^t |\mathbf{v}_s|^2 ds$ is called the *quadratic variation process* of the martingale $\mathbf{X}_t$. (See the next chapter for a more comprehensive definition.)

3.10. (*exponential martingales*). Let $dZ_t = \alpha dt + \beta dW_t$, $Z_0 = 0$ where α, β are constants and $(W_t)_{t\in\mathbb{R}_+}$ is a one-dimensional Brownian motion. Define
$$M_t = \exp\left\{ Z_t - \left(\alpha + \frac{1}{2}\beta^2 \right) t \right\} = \exp\left\{ -\frac{1}{2}\beta^2 t + \beta W_t \right\}.$$
Use Itô's formula to prove that
$$dM_t = \beta M_t dW_t.$$
In particular, $M = (M_t)_{t\in\mathbb{R}_+}$ is a martingale.

3.11. Let $(W_t)_{t\in\mathbb{R}_+}$ be a one-dimensional Brownian motion, and let $\phi \in L^2_{loc}[0,T]$ for any $T \in \mathbb{R}_+$. Show that for any $\theta \in \mathbb{R}$,
$$X_t := \exp\left\{ i\theta \int_0^t \phi(s) dW_s + \frac{1}{2}\theta^2 \int_0^t \phi^2(s) ds \right\}$$
is a local martingale.

3.12. With reference to the preceding problem 3.11, assume now that
$$P\left(\int_0^{+\infty} \phi^2(s) ds = +\infty \right) = 1,$$
and let
$$\tau_t := \min\left\{ u\mathbb{R}_+ | \int_0^u \phi^2(s) ds \geq t \right\}, \quad t \in \mathbb{R}_+.$$
Show that $(X_{\tau_t})_{t\in\mathbb{R}_+}$ is an $\mathcal{F}_{\tau_t}$-martingale.

3.13. With reference to problem 3.12, let
$$Z_t := \int_0^t \phi(s) dW_s, \quad t \in \mathbb{R}_+.$$
Show that $(Z_{\tau_t})_{t\in\mathbb{R}_+}$ has independent increments and $Z_{\tau_t} - Z_{\tau_s} \sim N(0, t-s)$ for any $0 < s < t < +\infty$. (*Hint:* Show that if $\mathcal{F}' \subset \mathcal{F}'' \subset \mathcal{F}$ are σ-fields on the probability space $(\Omega, \mathcal{F}, P)$ and Z is an $\mathcal{F}''$-measurable random variable, such that
$$E\left[e^{i\theta Z} \middle| \mathcal{F}' \right] = e^{-\theta^2\sigma^2/2},$$
then Z is independent of $\mathcal{F}'$ and $Z \sim N(0, \sigma^2)$.)

3.14. With reference to problem 3.13, show that the process $(Z_{\tau_t})_{t\in\mathbb{R}_+}$ is a standard Brownian motion.

3.15. Let $(W_t)_{t\in\mathbb{R}_+}$ be a one-dimensional Brownian motion. Formulate suitable conditions on u, v such that the following holds:

Let $dZ_t = u_t dt + v_t dW_t$, $Z_0 = 0$ be a stochastic integral with values in $\mathbb{R}$. Define

$$M_t = \exp\left\{ Z_t - \int_0^t \left[u_s + \frac{1}{2} v_s v_s' \right] ds \right\}.$$

Then $M = (M_t)_{t\in\mathbb{R}_+}$ is a martingale.

3.16. Let $(W_t)_{t\in\mathbb{R}_+}$ be a one-dimensional Brownian motion. Show that for any real function, which is continuous up to its second derivative, the process

$$\left(f(W_t) - \frac{1}{2}\int_0^t f''(W_s)ds \right)_{t\in\mathbb{R}_+}$$

is a local martingale.

3.17. Let $\mathbf{X}$ be a time-homogeneous Markov process with transition probability measure $P_t(\mathbf{x}, A)$, $\mathbf{x} \in \mathbb{R}^d$, $A \in \mathcal{B}_{\mathbb{R}^d}$, with $d \geq 1$. Given a test function φ, let

$$u(t, \mathbf{x}) := E_{\mathbf{x}}[\varphi(\mathbf{X}(t))] = \int_{\mathbb{R}^d} \varphi(\mathbf{y}) P_t(\mathbf{x}, d\mathbf{y}),\ t \in \mathbb{R}_+,\ \mathbf{x} \in \mathbb{R}^d.$$

Show that, under rather general assumptions, the function u satisfies the so-called Kolmogorov equation

$$\begin{aligned}
&\frac{\partial}{\partial t} u(t, \mathbf{x}) \\
&= \frac{1}{2} \sum_{i,j=1}^d q_{ij}(\mathbf{x}) \frac{\partial^2}{\partial x_i \partial x_j} u(t, \mathbf{x}) \\
&+ \sum_{j=1}^d f_j(\mathbf{x}) \frac{\partial}{\partial x_j} u(t, \mathbf{x}) \\
&+ \int_{\mathbb{R}^d} \left(u(t, \mathbf{x}+\mathbf{y}) - u(t, \mathbf{x}) - \frac{1}{1+|\mathbf{y}|^2} \sum_{j=1}^d y_j \frac{\partial}{\partial x_j} u(t, \mathbf{x}) \right) \nu(\mathbf{x}, d\mathbf{y})
\end{aligned}$$

for $t > 0$, $\mathbf{x} \in \mathbb{R}^d$, subject to the initial condition

$$u(0, \mathbf{x}) = \phi(\mathbf{x}),\ \mathbf{x} \in \mathbb{R}^d.$$

Here f and Q are functions with values being, respectively, vectors in $\mathbb{R}^d$ and symmetric, nonnegative $d \times d$ matrices, $f : \mathbb{R}^d \to \mathbb{R}^d$ and $Q := (q_{ij}) : \mathbb{R}^d \to$

$L_+(\mathbb{R}^d, \mathbb{R}^d)$, and ν is a Lévy measure, i.e., $\nu : \mathbb{R}^d \to \mathcal{M}(\mathbb{R}^d \setminus \{0\})$, being $\mathcal{M}(\mathbb{R}^d \setminus \{0\})$ the set of nonnegative measures on $\mathbb{R}^d \setminus \{0\}$ such that

$$\int_{\mathbb{R}^d} (|\mathbf{y}^2| \wedge 1)\nu(\mathbf{x}, d\mathbf{y}) < +\infty.$$

The functions f, Q, and ν are known as the *drift vector*, *diffusion matrix*, and *jump measure*, respectively. Show that the process $\mathbf{X}$ has continuous trajectories whenever $\nu \equiv 0$.

4

Stochastic Differential Equations

4.1 Existence and Uniqueness of Solutions

Definition 4.1. Let $(W_t)_{t\in\mathbb{R}_+}$ be a Wiener process on the probability space $(\Omega, \mathcal{F}, P)$, equipped with the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, $\mathcal{F}_t = \sigma(W_s, 0 \leq s \leq t)$. Furthermore, let $a(t,x)$, $b(t,x)$ be measurable functions in $[0,T] \times \mathbb{R}$ and $(u(t))_{t\in[0,T]}$ a stochastic process. Now $u(t)$ is said to be the *solution of the stochastic differential equation*

$$du(t) = a(t, u(t))dt + b(t, u(t))dW_t, \tag{4.1}$$

with the initial condition

$$u(0) = u^0 \text{ a.s. } (u^0 \text{ a random variable}), \tag{4.2}$$

if

1. $u(0)$ is $\mathcal{F}_0$-measurable;
2. $|a(t, u(t))|^{\frac{1}{2}}, b(t, u(t)) \in \mathcal{C}_1([0,T])$;
3. $u(t)$ is differentiable and $du(t) = a(t, u(t))dt + b(t, u(t))dW_t$, thus $u(t) = u(0) + \int_0^t a(s, u(s))ds + \int_0^t b(s, u(s))dW_s$, $t \in]0,T]$.

Remark 4.2. If $u(t)$ is the solution of (4.1), (4.2), then it is nonanticipatory (by point 3 of the preceding definition and as already observed in Remark 3.48).

Lemma 4.3. (Gronwall). *If $\phi(t)$ is an integrable, nonnegative function, defined on $t \in [0,T]$, with*

$$\phi(t) \leq \alpha(t) + L\int_0^t \phi(s)ds, \tag{4.3}$$

where L is a positive constant and $\alpha(t)$ is an integrable function, then

$$\phi(t) \leq \alpha(t) + L\int_0^t e^{L(t-s)}\alpha(s)ds.$$

Proof: Putting $\psi(t) = L\int_0^t \phi(s)ds$ as well as $z(t) = \psi(t)e^{-Lt}$, then $z(0) = \psi(0) = 0$, and moreover

$$\begin{aligned} z'(t) &= \psi'(t)e^{-Lt} - L\psi(t)e^{-Lt} = L\phi(t)e^{-Lt} - L\psi(t)e^{-Lt} \\ &\le L\alpha(t)e^{-Lt} + L\psi(t)e^{-Lt} - L\psi(t)e^{-Lt}. \end{aligned}$$

Therefore $z'(t) \le L\alpha(t)e^{-Lt}$ and after integration, $z(t) \le L\int_0^t \alpha(s)e^{-Ls}ds$. Hence

$$\psi(t)e^{-Lt} \le L\int_0^t \alpha(s)e^{-Ls}ds \Rightarrow \psi(t) \le L\int_0^t e^{L(t-s)}\alpha(s)ds,$$

but, by (4.3), $\psi(t) = L\int_0^t \phi(s)ds \ge \phi(t) - \alpha(t)$, completing the proof. □

Theorem 4.4. (existence and uniqueness). *Resorting to the notation of the preceding definition, if the following conditions are satisfied:*

1. *for all* $t \in [0,T]$ *and all* $(x,y) \in \mathbb{R}\times\mathbb{R}$: $|a(t,x)-a(t,y)|+|b(t,x)-b(t,y)| \le K^*|x-y|$;
2. *for all* $t \in [0,T]$ *and all* $x \in \mathbb{R}$: $|a(t,x)| \le K(1+|x|), |b(t,x)| \le K(1+|x|)$ *(*K^*, K *constants);*
3. $E[|u^0|^2] < \infty$;
4. u^0 *is independent of* $\mathcal{F}_T$ *(which is equivalent to requiring* u^0 *to be* $\mathcal{F}_0$*-measurable),*

then there exists a unique $(u(t))_{t\in[0,T]}$*, solution of (4.1), (4.2), such that*

- $(u(t))_{t\in[0,T]}$ *is continuous almost surely (thus almost every trajectory is continuous);*
- $(u(t))_{t\in[0,T]} \in \mathcal{C}([0,T])$.

Remark 4.5. If $(u_1(t))_{t\in[0,T]}$ and $(u_2(t))_{t\in[0,T]}$ are two solutions of (4.1), (4.2), belonging to $\mathcal{C}([0,T])$, then the uniqueness of a solution is understood in the sense that

$$P\left(\sup_{0\le t\le T} |u_1(t) - u_2(t)| = 0\right) = 1.$$

Proof (of Theorem 4.4): *Uniqueness.* Let $u_1(t)$ and $u_2(t)$ be solutions of (4.1), (4.2) belonging to $\mathcal{C}([0,T])$. Then, by point 3 of Definition 4.1,

$$\begin{aligned} &u_1(t) - u_2(t) \\ &= \int_0^t [a(s,u_1(s)) - a(s,u_2(s))]ds + \int_0^t [b(s,u_1(s)) - b(s,u_2(s))]dW_s \\ &= \int_0^t \tilde{a}(s)ds + \int_0^t \tilde{b}(s)dW_s, \qquad t \in]0,T], \end{aligned}$$

where $\tilde{a}(s) = a(s,u_1(s)) - a(s,u_2(s))$ and $\tilde{b}(s) = b(s,u_1(s)) - b(s,u_2(s))$. Because, in general, $(a+b)^2 \le 2(a^2+b^2)$, we obtain

$$|u_1(t) - u_2(t)|^2 \leq 2 \left(\int_0^t \tilde{a}(s) ds \right)^2 + 2 \left(\int_0^t \tilde{b}(s) dW_s \right)^2,$$

and by the Cauchy–Schwarz inequality

$$\left(\int_0^t \tilde{a}(s) ds \right)^2 \leq t \left(\int_0^t |\tilde{a}(s)|^2 ds \right),$$

therefore

$$E\left[\left(\int_0^t \tilde{a}(s) ds \right)^2 \right] \leq tE\left[\int_0^t |\tilde{a}(s)|^2 ds \right].$$

Moreover, by point 2 of Theorem 4.4

$$\begin{aligned} E\left[\int_0^T (b(s, u_i(s)))^2 ds \right] &\leq E\left[\int_0^T (K(1 + |u_i(s)|))^2 ds \right] \\ &\leq 2K^2 E\left[\int_0^T (1 + |u_i(s)|^2) ds \right] < +\infty \end{aligned}$$

for $i = 1, 2$ and because $u_i(s) \in \mathcal{C}$. Now, this shows $b(s, u_i(s)) \in \mathcal{C}$ for $i = 1, 2$ and thus $\tilde{b}(s) \in \mathcal{C}$. Then, by Proposition 3.19

$$E\left[\left(\int_0^t \tilde{b}(s) dW_s \right)^2 \right] = E\left[\int_0^t (\tilde{b}(s))^2 ds \right],$$

from which it follows that

$$E[(u_1(t) - u_2(t))^2] \leq 2tE\left[\int_0^t (\tilde{a}(s))^2 ds \right] + 2E\left[\int_0^t (\tilde{b}(s))^2 ds \right].$$

By point 1 of Theorem 4.4, we have that

$$\begin{aligned} |\tilde{a}(s)|^2 &\leq (K^*)^2 |u_1(s) - u_2(s)|^2, \\ |\tilde{b}(s)|^2 &\leq (K^*)^2 |u_1(s) - u_2(s)|^2, \end{aligned}$$

and therefore

$$\begin{aligned} &E[|u_1(t) - u_2(t)|^2] \\ &\leq 2t(K^*)^2 \int_0^t E[|u_1(t) - u_2(t)|^2] ds + 2(K^*)^2 \int_0^t E[|u_1(t) - u_2(t)|^2] ds \\ &\leq 2T(K^*)^2 \int_0^t E[|u_1(t) - u_2(t)|^2] ds + 2(K^*)^2 \int_0^t E[|u_1(t) - u_2(t)|^2] ds \\ &= 2(K^*)^2 (T + 1) \int_0^t E[|u_1(t) - u_2(t)|^2] ds. \end{aligned}$$

Since, by Gronwall's Lemma 4.3,

$$E[|u_1(t) - u_2(t)|^2] = 0 \qquad \forall t \in [0,T],$$

we get

$$u_1(t) - u_2(t) = 0,\ P\text{-a.s. } \forall t \in [0,T]$$

or, equivalently, for all $t \in [0,T]$:

$$\exists N_t \subset \Omega, P(N_t) = 0 \text{ such that } \forall \omega \notin N_t\colon\ u_1(t)(\omega) - u_2(t)(\omega) = 0.$$

Because the type of processes that we consider are separable, there exists an $M \subset [0,T]$, a separating set of $(u_1(t) - u_2(t))_{t\in[0,T]}$, countable and dense in $[0,T]$, such that, for all $t \in [0,T]$

$$\exists (t_n)_{n\in\mathbb{N}} \in M^{\mathbb{N}} \text{ such that } \lim_n t_n = t,\ P\text{-a.s.}$$

and

$$\lim_n (u_1(t_n) - u_2(t_n)) = u_1(t) - u_2(t),\ P\text{-a.s.}$$

(and the empty set A, where this does not hold, does not depend on t). Putting $N = \bigcup_{t\in M} N_t$, we obtain $P(N) = 0$ and

$$\forall \omega \notin N\colon\ u_1(t) - u_2(t) = 0, t \in M,$$

hence

$$\forall t \in [0,T], \forall \omega \notin N \cup A\colon\ u_1(t) - u_2(t) = 0$$

and thus

$$P\left(\sup_{0\le t\le T} |u_1(t) - u_2(t)| = 0\right) = 1.$$

Existence. We will prove the existence of a solution $u(t)$ by the method of sequential approximations. We define

$$\begin{cases} u_0(t) = u^0, \\ u_n(t) = u^0 + \int_0^t a(s, u_{n-1}(s))ds + \int_0^t b(s, u_{n-1}(s))dW_s, \forall t \in [0,T], n \in \mathbb{N}^*. \end{cases}$$

Assuming u^0 being $\mathcal{F}_0$-measurable and by point 3 of Theorem 4.4, it is obvious that $u^0 \in \mathcal{C}([0,T])$. By induction, we will now show both that

$$\forall n \in \mathbb{N} : E[|u_{n+1}(t) - u_n(t)|^2] \le \frac{(ct)^{n+1}}{(n+1)!}, \tag{4.4}$$

where $c = \max\{4K^2(T+1)(1+E[|u^0|^2]), 2(K^*)^2(T+1)\}$, and

$$\forall n \in \mathbb{N} : u_{n+1} \in \mathcal{C}([0,T]). \tag{4.5}$$

By conditions 1 and 2 of Theorem 4.4, we obtain

$$E[|b(s,u^0)|^2] \le E[K^2(1+|u^0|)^2] \le 2K^2(1+E[|u^0|^2]) < +\infty,$$

where we make use of the generic inequality

$$(|x|+|y|)^2 \le 2|x|^2 + 2|y|^2, \tag{4.6}$$

and thus

$$b(s,u^0) \in \mathcal{C}([0,T]).$$

Analogously $a(s,u^0) \in \mathcal{C}([0,T])$, resulting in u_1 being nonanticipatory and well posed. As a further result of (4.6), we have

$$\begin{aligned}|u_1(t)-u^0|^2 &= \left|\int_0^t a(s,u^0)ds + \int_0^t b(s,u^0)dW_s\right|^2 \\ &\le 2\left|\int_0^t a(s,u^0)ds\right|^2 + 2\left|\int_0^t b(s,u^0)dW_s\right|^2\end{aligned}$$

and by the Schwarz inequality

$$\left|\int_0^t a(s,u^0)ds\right|^2 \le t\int_0^t |a(s,u^0)|^2 ds \le T\int_0^t |a(s,u^0)|^2 ds.$$

Moreover, by Proposition 3.19, we have

$$E\left[\left|\int_0^t b(s,u^0)dW_s\right|^2\right] = E\left[\int_0^t |b(s,u^0)|^2 ds\right].$$

Therefore, as a conclusion and by point 2 of Theorem 4.4

$$\begin{aligned}E[|u_1(t)-u^0|^2] &\le 2TE\left[\int_0^t |a(s,u^0)|^2 ds\right] + 2E\left[\int_0^t |b(s,u^0)|^2 ds\right] \\ &\le 2TE\left[\int_0^t K^2(1+|u^0|)^2 ds\right] + 2E\left[\int_0^t K^2(1+|u^0|)^2 ds\right] \\ &= (2TK^2+2K^2)E\left[\int_0^t (1+|u^0|)^2 ds\right] \\ &= 2K^2(T+1)tE[(1+|u^0|)^2] \\ &\le 4K^2(T+1)t(1+E[|u^0|^2]) = ct,\end{aligned}$$

where the last inequality is a direct result of (4.6). Hence (4.4) holds for $n=1$, from which it follows that $u_1 \in \mathcal{C}([0,T])$. Supposing now that (4.4) and (4.5) hold for n, we will show that this implies that they also hold for $n+1$. By the induction hypotheses $u_n \in \mathcal{C}([0,T])$. Then, by point 2 of Theorem 4.4 and proceeding as before, we obtain that

$$a(s,u_n(s)) \in \mathcal{C}([0,T]) \text{ and } b(s,u_n(s)) \in \mathcal{C}([0,T]).$$

Therefore u_{n+1} is well posed and nonanticipatory. We thus get

$$
\begin{aligned}
&| u_{n+1}(t) - u_n(t)|^2 \\
&\leq \left(\int_0^t |a(s, u_n(s)) - a(s, u_{n-1}(s))| ds \right. \\
&\quad \left. + \left| \int_0^t (b(s, u_n(s)) - b(s, u_{n-1}(s))) dW_s \right| \right)^2 \\
&\leq 2 \left(\int_0^t |a(s, u_n(s)) - a(s, u_{n-1}(s))| ds \right)^2 \\
&\quad +2 \left(\int_0^t (b(s, u_n(s)) - b(s, u_{n-1}(s))) dW_s \right)^2, \qquad (4.7)
\end{aligned}
$$

and by the Schwarz inequality

$$
\begin{aligned}
\left(\int_0^t |a(s, u_n(s)) - a(s, u_{n-1}(s))| ds \right)^2 &\leq t \int_0^t |a(s, u_n(s)) - a(s, u_{n-1}(s))|^2 ds \\
&\leq T(K^*)^2 \int_0^t |u_n(s) - u_{n-1}(s)|^2 ds,
\end{aligned}
$$

where the last inequality is due to point 1 of Theorem 4.4. Moreover, by Proposition 3.19,

$$
\begin{aligned}
&E\left[\left(\int_0^t (b(s, u_n(s)) - b(s, u_{n-1}(s))) dW_s \right)^2 \right] \\
&= E\left[\int_0^t |b(s, u_n(s)) - b(s, u_{n-1}(s))|^2 ds \right] \\
&\leq (K^*)^2 E\left[\int_0^t |u_n(s) - u_{n-1}(s)|^2 ds \right],
\end{aligned}
$$

again by point 1. Now we obtain

$$
\begin{aligned}
E[|u_{n+1}(t) - u_n(t)|^2] &\leq 2T(K^*)^2 E\left[\int_0^t |u_n(s) - u_{n-1}(s)|^2 ds \right] \\
&\quad +2(K^*)^2 E\left[\int_0^t |u_n(s) - u_{n-1}(s)|^2 ds \right] \\
&\leq cE\left[\int_0^t |u_n(s) - u_{n-1}(s)|^2 ds \right] \\
&\leq c \int_0^t \frac{(cs)^n}{n!} ds = \frac{(ct)^{n+1}}{(n+1)!},
\end{aligned}
$$

where the last inequality is due to the induction hypotheses. Hence the proof of (4.4) is complete and so $u_{n+1} \in \mathcal{C}([0, T])$. Moreover, from (4.6) it follows that

$$\sup_{0\le t\le T} |u_{n+1}(t) - u_n(t)|^2 \le 2\left(\int_0^t |a(s,u_n(s)) - a(s,u_{n-1}(s))|ds\right)^2$$
$$+2\sup_{0\le t\le T}\left|\int_0^t (b(s,u_n(s)) - b(s,u_{n-1}(s)))dW_s\right|^2,$$

where, after taking expectations and recalling point 1 of Proposition 3.40

$$\begin{aligned}
E\left[\sup_{0\le t\le T} |u_{n+1}(t) - u_n(t)|^2\right] &\le 2E\left[\left(\int_0^T |a(s,u_n(s)) - a(s,u_{n-1}(s))|ds\right)^2\right] \\
&\quad +8E\left[\int_0^T |b(s,u_n(s)) - b(s,u_{n-1}(s))|^2 ds\right] \\
&\le 2T(K^*)^2 E\left[\int_0^T |u_n(s) - u_{n-1}(s)|^2 ds\right] \\
&\quad +8(K^*)^2 E\left[\int_0^T |u_n(s) - u_{n-1}(s)|^2 ds\right] \\
&= 2T(K^*)^2 \int_0^T E[|u_n(s) - u_{n-1}(s)|^2]ds \\
&\quad +8(K^*)^2 \int_0^T E[|u_n(s) - u_{n-1}(s)|^2]ds \\
&\le \frac{(cT)^n}{n!}(2(K^*)T^2 + 8(K^*)^2 T),
\end{aligned}$$

where the last equality is due to 1 of Theorem 4.4 as well as the Schwarz inequality, and the last inequality, is due to (4.4). Hence

$$E\left[\sup_{0\le t\le T} |u_{n+1}(t) - u_n(t)|^2\right] \le c^* \frac{(cT)^n}{n!}, \tag{4.8}$$

with $c^* = 2(K^*)^2T^2 + 8(K^*)^2T$. Because the terms are positive

$$\sup_{0\le t\le T} |u_{n+1}(t) - u_n(t)|^2 = \left(\sup_{0\le t\le T} |u_{n+1}(t) - u_n(t)|\right)^2,$$

and therefore

$$\begin{aligned}
P\left(\sup_{0\le t\le T} |u_{n+1}(t) - u_n(t)| > \frac{1}{2^n}\right) &= P\left(\sup_{0\le t\le T} |u_{n+1}(t) - u_n(t)|^2 > \frac{1}{2^{2n}}\right) \\
&\le E\left[\sup_{0\le t\le T} |u_{n+1}(t) - u_n(t)|^2\right] 2^{2n} \\
&\le c^* \frac{(cT)^n}{n!} 2^{2n},
\end{aligned}$$

where the last two inequalities are due to the Markov inequality and (4.8), respectively. Because the the progression $\sum_{n=1}^{\infty} \frac{(cT)^n}{n!} 2^{2n}$ converges, so does

$$\sum_{n=1}^{\infty} P\left(\sup_{0 \le t \le T} |u_{n+1}(t) - u_n(t)| > \frac{1}{2^n} \right),$$

and, by the Borel–Cantelli Lemma 1.98, we have that

$$P\left(\limsup_n \left\{ \sup_{0 \le t \le T} |u_{n+1}(t) - u_n(t)| > \frac{1}{2^n} \right\} \right) = 0.$$

Therefore, putting $A = \limsup_n \{\sup_{0 \le t \le T} |u_{n+1}(t) - u_n(t)| > \frac{1}{2^n}\}$, for all $\omega \in (\Omega - A)$:

$$\exists N = N(\omega) \text{ such that } \forall n \in N, n \ge N(\omega) \Rightarrow \sup_{0 \le t \le T} |u_{n+1}(t) - u_n(t)| \le \frac{1}{2^n},$$

and $u^0 + \sum_{n=0}^{\infty} (u_{n+1}(t) - u_n(t))$ converges uniformly on $t \in [0, T]$ with probability 1. Thus, given the sum $u(t)$ and observing that $u^0 + \sum_{k=0}^{n-1} (u_{k+1}(t) - u_k(t)) = u_n(t)$, it follows that the sequence $(u_n(t))_n$ of the nth partial sum of $u^0 + \sum_{n=0}^{\infty} (u_{n+1}(t) - u_n(t))$ has the limit

$$\lim_{n \to \infty} u_n(t) = u(t), \; P\text{-a.s., uniformly on } t \in [0, T]. \tag{4.9}$$

Analogous to the property of the processes u_n, it follows that the trajectories of $u(t)$ are continuous almost surely and nonanticipatory. We will now demonstrate that $u(t)$ is the solution of (4.1), (4.2). By point 1 of the same theorem, we have

$$\left| \int_0^t a(s, u_{n-1}(s)) ds - \int_0^t a(s, u(s)) ds \right| \le K^* \int_0^t |u_{n-1}(s) - u(s)| ds,$$

and since we can take the limit of (4.9) inside the integral sign,

$$\int_0^t a(s, u_{n-1}(s)) ds \xrightarrow{n} \int_0^t a(s, u(s)) ds, \; P\text{-a.s., uniformly on } t \in [0, T],$$

and therefore also in probability. Moreover,

$$|b(s, u_{n-1}(s)) - b(s, u(s))|^2 \le (K^*)^2 |u_{n-1}(s) - u(s)|^2,$$

and thus

$$\int_0^t |b(s, u_{n-1}(s)) - b(s, u(s))|^2 ds \xrightarrow{n} 0, \; P\text{-a.s., uniformly on } t \in [0, T],$$

and therefore also in probability. Hence, by Theorem 3.29, we also have

$$P - \lim_{n\to\infty} \int_0^t b(s, u_{n-1}(s))dW_s = \int_0^t b(s, u(s))dW_s.$$

Then if we take the limit $n \to \infty$ of

$$u_n(t) = u^0 + \int_0^t a(s, u_{n-1}(s))ds + \int_0^t b(s, u_{n-1}(s))dW_s, \tag{4.10}$$

by the uniqueness of the limit in probability, we obtain

$$u(t) = u^0 + \int_0^t a(s, u(s))ds + \int_0^t b(s, u(s))dW_s,$$

with $u(t)$ as the solution of (4.1), (4.2). It remains to show that

$$E[u^2(t)] < \infty, \text{ for all } t \in [0, T].$$

Because, in general, $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$, by (4.10), it follows that

$$\begin{aligned}
E[u_n^2(t)] &\leq 3\left(E[(u^0)^2] + E\left[\left|\int_0^t a(s, u_{n-1}(s))ds\right|^2\right]\right. \\
&\qquad \left. +E\left[\left|\int_0^t b(s, u_{n-1}(s))dW_s\right|^2\right]\right) \\
&\leq 3\left(E[(u^0)^2] + TE\left[\int_0^t |a(s, u_{n-1}(s))|^2 ds\right]\right. \\
&\qquad \left. +E\left[\int_0^t |b(s, u_{n-1}(s))|^2 ds\right]\right),
\end{aligned}$$

where the last relation holds due to the Schwarz inequality as well as point 3 of Proposition 3.19. From 2 of Theorem 4.4 and inequality (4.6), it further follows that

$$\begin{aligned}
|a(s, u_{n-1}(s))|^2 &\leq K^2(1 + |u_{n-1}(s)|)^2 \leq 2K^2(1 + |u_{n-1}(s)|^2), \\
|b(s, u_{n-1}(s))|^2 &\leq K^2(1 + |u_{n-1}(s)|)^2 \leq 2K^2(1 + |u_{n-1}(s)|^2).
\end{aligned}$$

Therefore

$$\begin{aligned}
E[u_n^2(t)] &\leq 3\left(E[(u^0)^2] + 2K^2(T+1)\int_0^t (1 + E[|u_{n-1}(s)|^2])ds\right) \\
&\leq 3\left(E[(u^0)^2] + 2K^2T(T+1) + 2K^2(T+1)\int_0^t E[|u_{n-1}(s)|^2]ds\right) \\
&\leq c(1 + E[(u^0)^2]) + c\int_0^t E[|u_{n-1}(s)|^2]ds,
\end{aligned}$$

where c is a constant that only depends on K and T. Continuing with the induction, we have

$$E[u_n^2(t)] \leq \left(c + c^2 t + c^3 \frac{t^2}{2} + \cdots + c^{n+1} \frac{t^n}{n!}\right)(1 + E[(u^0)^2]),$$

and taking the limit $n \to \infty$,

$$\lim_{n\to\infty} E\left[u_n^2(t)\right] \leq ce^{ct}\left(1 + E\left[(u^0)^2\right]\right) \leq ce^{cT}\left(1 + E\left[(u^0)^2\right]\right).$$

Therefore, by Fatou's Lemma A.26 and by 3 of Theorem 4.4, we obtain

$$E\left[u^2(t)\right] \leq ce^{cT}\left(1 + E\left[(u^0)^2\right]\right) < +\infty, \tag{4.11}$$

and hence $(u(t))_{t\in[0,T]} \in \mathcal{C}([0,T])$, completing the proof. □

Remark 4.6. By (4.11), it also follows that

$$\sup_{0\leq t\leq T} E\left[u^2(t)\right] \leq ce^{cT}\left(1 + E\left[(u^0)^2\right]\right) < +\infty.$$

Remark 4.7. Theorem 4.4 continues to hold if its hypothesis 1 is substituted by the following condition.

1′. For all $n > 0$, there exists a $K_n > 0$ such that, for all $(x_1, x_2) \in \mathbb{R}^2$, $|x_i| \leq n$ $i = 1, 2$:

$$|a(t, x_1) - a(t, x_2)| \leq K_n|x_1 - x_2|,$$
$$|b(t, x_1) - b(t, x_2)| \leq K_n|x_1 - x_2|.$$

Proof: See, e.g., Friedman (1975). □

Example 4.8. We suppose that in (4.1) $a(t, u(t)) = 0$ and $b(t, u(t)) = g(t)u(t)$. Then the stochastic differential equation

$$\begin{cases} u_0(t) = u^0, \\ du(t) = g(t)u(t)dW_t \end{cases}$$

has the solution

$$u(t) = u^0 \exp\left\{\int_0^t g(s)dW_s - \frac{1}{2}\int_0^t g^2(s)ds\right\}.$$

Putting

$$X(t) = \int_0^t g(s)dW_s - \frac{1}{2}\int_0^t g^2(s)ds$$

and $Y(t) = \exp\{X(t)\} = f(X(t))$, then $u(t) = u^0 Y(t)$ and thus $du(t) = u^0 dY(t)$. We will further show that $u^0 dY(t) = g(t)u(t)dW_t$. Because

$$dX(t) = -\frac{1}{2}g^2(t)dt + g(t)dW_t,$$

with the help of Itô's formula, we obtain

$$\begin{aligned} dY(t) &= \left(-\frac{1}{2}g^2(t)f_x(X(t)) + \frac{1}{2}g^2(t)f_{xx}(X(t))\right)dt + g(t)f_x(X(t))dW_t \\ &= \left(-\frac{1}{2}g^2(t)\exp\{X(t)\} + \frac{1}{2}g^2(t)\exp\{X(t)\}\right)dt + g(t)\exp\{X(t)\}dW_t \\ &= Y(t)g(t)dW_t, \end{aligned}$$

resulting in $du(t) = u^0 Y(t)g(t)dW_t = u(t)g(t)dW_t$.

Example 4.9. Three important stochastic differential equations that have wide applicability, for instance in financial modeling, are

1. arithmetic Brownian motion

$$\begin{cases} u_0(t) = u^0, \\ du(t) = adt + bdW_t; \end{cases}$$

2. geometric Brownian motion

$$\begin{cases} u_0(t) = u^0, \\ du(t) = au(t)dt + bu(t)dW_t; \end{cases}$$

3. (mean-reverting) Ornstein–Uhlenbeck process

$$\begin{cases} u_0(t) = u^0, \\ du(t) = (a - bu(t))dt + cdW_t. \end{cases}$$

The derivations of the solutions of 1–3 resort to a number of standard solution techniques for stochastic differential equations. Throughout, for the sake of generality, we will denote the initial time by t_0, but still assume that $W_{t_0} = 0$.

1. Direct integration gives

$$u(t) = u^0 + a(t - t_0) + bW_t,$$

so that we can take the expectation and variance directly to obtain

$$E[u(t)] = u^0 + a(t - t_0), \qquad Var[u(t)] = b^2(t - t_0).$$

2. We calculate the stochastic differential $d \ln u(t)$ with the help of Itô's formula (3.33) and obtain

$$d \ln u(t) = \left(a - \frac{1}{2}b^2\right)dt + bdW_t.$$

We can then integrate both sides directly, which results in

$$\begin{aligned} \ln u(t) &= \ln u^0 + \left(a - \frac{1}{2}b^2\right)(t - t_0) + bW_t, \\ \Leftrightarrow u(t) &= u^0 \exp\left\{\left(a - \frac{1}{2}b^2\right)(t - t_0) + bW_t\right\}. \end{aligned} \tag{4.12}$$

To calculate its expectation we will require the expected value of $\tilde{u}(t) = \exp\{bW_t\}$. We apply Itô's formula to calculate the latter's differential as

$$d\exp\{bW_t\} = d\tilde{u}(t) = b\tilde{u}(t)dW_t + \frac{1}{2}b^2\tilde{u}(t)dt,$$

which after direct integration, rearrangement and the taking of expectations results in

$$E[\tilde{u}(t)] = \tilde{u}(0) + \int_{t_0}^{t} \frac{1}{2}b^2 E[\tilde{u}(s)]ds.$$

Differentiating both sides with respect to t gives

$$\frac{dE[\tilde{u}(t)]}{dt} = \frac{1}{2}b^2 E[\tilde{u}(t)],$$

which, after rearrangement and integration, results in

$$E[\tilde{u}(t)] = e^{\frac{1}{2}b^2(t-t_0)}.$$

Therefore, the expectation of (4.12) is

$$E[u(t)] = u^0 e^{\left(a-\frac{1}{2}b^2\right)(t-t_0)} E\left[e^{bW_t}\right] = u^0 e^{a(t-t_0)}. \tag{4.13}$$

For the variance we employ the standard general result (1.4), so that we only need to calculate $E[(u(t))^2]$. For this, we proceed as above in deriving the stochastic differential of $(u(t))^2$, differentiating twice with respect to t, integrating and taking expectations, to get

$$E[(u(t))^2] = (u^0)^2\exp\{2a(t-t_0)\} + \frac{b^2}{2a}(\exp\{2a(t-t_0)\} - 1).$$

Therefore the variance of (4.12) is

$$Var[u(t)] = E[(u(t))^2] - (E[u(t)])^2 = \frac{b^2}{2a}(\exp\{2a(t-t_0)\} - 1).$$

3. To find the solution of the Ornstein–Uhlenbeck process, we require an integrating factor $\phi = \exp\{bt\}$, so that

$$d(\phi u(t)) = \phi(bu(t) + du(t)) = \phi(adt + cdW_t).$$

Because the drift term, which depended on $u(t)$, has dropped out, we can integrate directly and, after rearrangement, obtain

$$u(t) = \frac{a}{b}\exp\{bt_0\} + u^0\exp\{-b(t-t_0)\} + c\int_{t_0}^{t}\exp\{-b(t-s)\}dW_s. \tag{4.14}$$

Therefore, the expectation of (4.14) is

$$E[u(t)] = \frac{a}{b}\exp\{bt_0\} + u^0\exp\{-b(t-t_0)\}$$

and for the variance we again resort to (1.4), so that we require $E[(u(t))^2]$. Squaring (4.14) and taking expectations yields

$$\begin{aligned} E[(u(t))^2] &= \left(\frac{a}{b}e^{bt_0} + u^0 e^{-b(t-t_0)}\right)^2 + \left(c\int_{t_0}^t e^{-b(t-s)}dW_s\right)^2 \\ &= (E[u(t)])^2 + c^2\int_{t_0}^t e^{-2b(t-s)}ds, \end{aligned}$$

where the last step is due to the Itô isometry (point 3 of Proposition 3.19). Hence the variance of (1.4) is

$$\begin{aligned} Var[u(t)] &= (E[u(t)])^2 + c^2\int_{t_0}^t \exp\{-2b(t-s)\}ds - (E[u(t)])^2 \\ &= \frac{c^2}{2b}(1 - \exp\{-2b(t-t_0)\}). \end{aligned}$$

Remark 4.10. Let $(X_t)_t$ be a process that is continuous in probability, stationary, Gaussian, and Markovian. Then it is of the form $Y_t + c$, where Y_t is an Ornstein–Uhlenbeck process and c a constant.

Proof: See Breiman (1968). □

We have seen in the proof of Theorem 4.4 that if $E[(u^0)^2] < +\infty$, then $E[(u(t))^2] < +\infty$. This result can be generalized as follows.

Theorem 4.11. *Given the hypotheses of Theorem 4.4, if* $E[(u^0)^{2n}] < +\infty$ *for* $n \in \mathbb{N}$, *then*

1. $E[(u(t))^{2n}] \leq (1 + E[(u^0)^{2n}])e^{ct}$,
2. $E[\sup_{0\leq s\leq t}|u(s) - u^0|^{2n}] \leq \bar{c}(1 + E[(u^0)^{2n}])t^n e^{ct}$,

where c *and* $\bar{c}$ *are constants that only depend on* K, T, *and* n.

Proof: For all $N \in \mathbb{N}$ we put

$$u_N^0(\omega) = \begin{cases} u_0(\omega) & \text{for } |u^0(\omega)| \leq N, \\ N\mathrm{sgn}\{u^0(\omega)\} & \text{for } |u^0(\omega)| > N; \end{cases}$$

$$a_N(t,x) = \begin{cases} a(t,x) & \text{for } |x| \leq N, \\ a(t, N\mathrm{sgn}\{x\}) & \text{for } |x| > N; \end{cases}$$

$$b_N(t,x) = \begin{cases} b(t,x) & \text{for } |x| \leq N, \\ b(t, N\mathrm{sgn}\{x\}) & \text{for } |x| > N \end{cases}$$

and denote by $u_N(t)$ the solution of

$$\begin{cases} u_N(0) = u_N^0, \\ du_N(t) = a_N(t, u_N(t))dt + b_N(t, u_N(t))dW_t \end{cases}$$

(the solution will exist due to Theorem 4.4). Then, applying Itô's formula to $f(u_N(t)) = (u_N(t))^{2n}$, we obtain

$$\begin{aligned} & d(u_N(t))^{2n} \\ & = (n(2n-1)(u_N(t))^{2n-2} b_N^2(t, u_N(t)) \\ & \quad + 2n(u_N(t))^{2n-1} a_N(t, u_N(t)))dt + 2n(u_N(t))^{2n-1} b_N(t, u_N(t))dW_t. \end{aligned}$$

Hence

$$\begin{aligned} & (u_N(t))^{2n} \\ & = (u_N^0)^{2n} + n(2n-1)\int_0^t (u_N(s))^{2n-2} b_N^2(s, u_N(s))ds \\ & \quad + 2n\int_0^t (u_N(s))^{2n-1} a_N(s, u_N(s))ds + 2n\int_0^t (u_N(s))^{2n-1} b_N(s, u_N(s))dW_s. \end{aligned}$$

Since $u_N(t) = u_N^0 + \int_0^t a_N(s, u_N(s))ds + \int_0^t b_N(s, u_N(s))dW_s$, $E[(u_N^0)^{2n}] < +\infty$ and both $a_N(t, x)$ and $b_N(t, x)$ are bounded, we have

$$E[(u_N(t))^{2n}] < +\infty,$$

meaning[8] $(u_N(t))^n \in \mathcal{C}([0, T])$. By 2 of Theorem 4.4 and by $(a+b)^2 \leq 2(a^2+b^2)$ it follows that

$$\begin{aligned} |a_N(s, u_n(s))| &\leq K(1 + |u_N(s)|), \\ |b_N(s, u_n(s))|^2 &\leq 2K^2(1 + |u_N(s)|^2). \end{aligned}$$

Moreover, because $(u_N(t))^n \in \mathcal{C}([0, T])$ we have

$$E\left[2n\int_0^t |u_N(s)|^{2n-1} |b_N(s, u_N(s))| dW_s\right] = 0,$$

and therefore

$$\begin{aligned} E[u(t)^{2n}] = E[(u_N^0)^{2n}] + \int_0^t & E[(2n u_N(s) a_N(s, u_N(s)) \\ & + n(2n-1) b_N^2(s, u_N(s))) u_N(s)^{2n-2}] ds \end{aligned}$$

---

[8] It suffices to make use of the following theorem for $E[\int_0^t b_N(s, u_N(s))dW_s]^{2n}$:
Theorem. *If $f^n \in \mathcal{C}([0, T])$ for $n \in \mathbb{N}^*$, then*

$$E\left[\int_0^T f(t)dW_t\right]^{2n} \leq [n(2n-1)]^n T^{n-1} E\left[\int_0^T f^{2n}(t)dt\right].$$

Proof: See, e.g., Friedman (1975). □

$$\leq E[(u_N^0)^{2n}] + n(2n+1)\int_0^t E[(u_N(s)a_N(s,u_N(s)) \\ + b_N^2(s,u_N(s)))u_N(s)^{2n-2}]ds$$
$$\leq E[(u_N^0)^{2n}] + n(2n+1)K^2\int_0^t E[(1+u_N^2(s))u_N(s)^{2n-2}]ds,$$

where the first inequality follows when condition 2 of Theorem 4.4 is substituted by $xa(t,x) + b^2(t,x) \leq K^2(1+x^2)$ for all $t \in [0,T]$, and all $x \in \mathbb{R}$. Now since, in general, $x^{2n-2} \leq 1 + x^{2n}$, we have

$$u_N(s)^{2n-2}(1+u_N^2(s)) \leq 1 + 2u_N(s)^{2n}.$$

Therefore,

$$E[u_N(t)^{2n}] \leq E[(u_N^0)^{2n}] + n(2n+1)K^2\int_0^t E[1+2u_N(s)^{2n}]ds$$

and, by putting $\phi(t) = E[u_N(t)^{2n}]$, we can write

$$\phi(t) \leq \phi(0) + n(2n+1)K^2\int_0^t (1+2\phi(s))ds$$
$$= \phi(0) + n(2n+1)K^2 t + 2n(2n+1)K^2\int_0^t \phi(s) = \alpha(t) + L\int_0^t \phi(s)ds,$$

where $\alpha(t) = \phi(0) + n(2n+1)K^2 t$ and $L = 2n(2n+1)K^2$. By Gronwall's Lemma 4.3, we have that

$$\phi(t) \leq \alpha(t) + L\int_0^t e^{L(t-s)}\alpha(s)ds,$$

and thus

$$E[u_N(t)^{2n}]$$
$$\leq E[(u_N^0)^{2n}] + \frac{L}{2}t + L\int_0^t e^{L(t-s)}\left(E[(u_N^0)^{2n}] + \frac{L}{2}s\right)ds$$
$$= E[(u_N^0)^{2n}] + \frac{L}{2}t - E[(u_N^0)^{2n}] + E[(u_N^0)^{2n}]e^{Lt} + Le^{Lt}\int_0^t e^{-Ls}\frac{L}{2}sds$$
$$= \frac{L}{2}t + E[(u_N^0)^{2n}]e^{Lt} - \frac{L}{2}t - \frac{1}{2}e^{Lt}(e^{-Lt}-1) \leq e^{Lt}(1+E[(u_N^0)^{2n}]).$$

Therefore, point 1 holds for $u_N(t)$ ($N \in \mathbb{N}^*$) and, taking the limit $N \to \infty$, it also holds for $u(t)$. For the proof of 2, see, e.g., Gihman and Skorohod (1972). □

4.2 The Markov Property of Solutions

In the preceding section we have shown that if $a(t,x)$ and $b(t,x)$ are measurable functions on $(t,x) \in [0,T] \times \mathbb{R}$ that satisfy conditions 1 and 2 of Theorem 4.4, then there exists a unique solution in $\mathcal{C}([0,T])$ of

$$\begin{cases} u(0) = u^0 \text{ a.s.}, \\ du(t) = a(t,u(t))dt + b(t,u(t))dW_t, \end{cases} \tag{4.15}$$

provided that the random variable u^0 is independent of $\mathcal{F}_T = \sigma(W_s, 0 \leq s \leq T)$ and $E[(u^0)^2] < +\infty$. Analogously, for all $s \in]0,T]$, there exists a unique solution in $\mathcal{C}([s,T])$ of

$$\begin{cases} u(s) = u_s \text{ a.s.}, \\ du(t) = a(t,u(t))dt + b(t,u(t))dW_t, \end{cases} \tag{4.16}$$

provided that the random variable u_s is independent of $\mathcal{F}_{s,T} = \sigma(W_t - W_s, t \in [s,T])$ and $E[(u_s)^2] < +\infty$. (The proof is left to the reader as a useful exercise.) Now, let $t_0 \geq 0$ and c be a random variable with $u(t_0) = c$ almost surely and, moreover, c be independent of $\mathcal{F}_{t_0,T} = \sigma(W_t - W_{t_0}, t \geq t_0)$ as well as $E[c^2] < +\infty$. Under conditions 1 and 2 of Theorem 4.4 there exists a unique solution $\{u(t), t \in [t_0,T]\}$ of the stochastic differential equation (4.15) with the initial condition $u(t_0) = c$ almost surely, and the following holds.

Lemma 4.12. *If $h(x,\omega)$ is a real-valued function defined, for all $(x,\omega) \in \mathbb{R} \times \Omega$ such that*

1. *h is $\mathcal{B}_{\mathbb{R}} \otimes \mathcal{F}$-measurable,*
2. *h is bounded,*
3. *for all $x \in \mathbb{R}$: $h(x,\cdot)$ is independent of $\mathcal{F}_s$ for all $s \in [t_0,T]$,*

then

$$\forall s \in [t_0,T]: \qquad E[h(u(s),\cdot)|\mathcal{F}_s] = E[h(u(s),\cdot)|u(s)] \;\; a.s. \tag{4.17}$$

Proof: We limit ourselves to the case of h being decomposable of the form

$$h(x,\omega) = \sum_{i=1}^{n} Y_i(x) Z_i(\omega), \tag{4.18}$$

with the Z_i independent of $\mathcal{F}_s$. In that case

$$E[h(u(s),\cdot)|\mathcal{F}_s] = \sum_{i=1}^{n} E[Y_i(u(s))Z_i(\cdot)|\mathcal{F}_s] = \sum_{i=1}^{n} Y_i(u(s)) E[Z_i(\cdot)|\mathcal{F}_s],$$

because $Y_i(u(s))$ is $\mathcal{F}_s$-measurable. Therefore

$$E[h(u(s),\cdot)|\mathcal{F}_s] = \sum_{i=1}^{n} Y_i(u(s)) E[Z_i(\cdot)],$$

and recapitulating, because $\sigma(u(s)) \subset \mathcal{F}_s$, we have

$$E[h(u(s),\cdot)|\mathcal{F}_s] = \sum_{i=1}^{n} Y_i(u(s))E[Z_i(\cdot)|u(s)]$$
$$= \sum_{i=1}^{n} E[Y_i(u(s))Z_i(\cdot)|u(s)] = E[h(u(s),\cdot)|u(s)].$$

It can be shown that every h that satisfies conditions 1, 2, and 3 can be approximated by functions that are decomposable as in (4.18). □

Theorem 4.13. *If $(u(t))_{t\in[t_0,T]}$ is a Markov process with respect to the filtration $\mathcal{U}_t = \sigma(u(s), t_0 \le s \le t)$, then it satisfies the condition*

$$\forall B \in \mathcal{B}_{\mathbb{R}}, \forall s \in [t_0, t[: P(u(t) \in B|\mathcal{U}_s) = P(u(t) \in B|u(s)) \text{ a.s.} \tag{4.19}$$

Proof: Putting $\mathcal{F}_t = \sigma(c, W_s, t_0 \le s \le t)$, then $u(t)$ is $\mathcal{F}_t$-measurable, as can be deduced from Theorem 4.4. Therefore, $\sigma(u(t)) \subset \mathcal{F}_t$ and thus $\mathcal{U}_t \subset \mathcal{F}_t$. In order to prove (4.19), it is now sufficient to show that

$$\forall B \in \mathcal{B}_{\mathbb{R}}, \forall s \in [t_0, t[: P(u(t) \in B|\mathcal{F}_s) = P(u(t) \in B|u(s)) \text{ a.s.} \tag{4.20}$$

Fixing $B \in \mathcal{B}_{\mathbb{R}}$ and $s < t$, we denote by $u(t, s, x)$ the solution of (4.15) with the initial condition $u(s) = x$ a.s. ($x \in \mathbb{R}$), and we define the mapping $h : \mathbb{R} \times \Omega \to \mathbb{R}$ as

$$h(x, \omega) = I_B(u(t, s, x; \omega)) \text{ for } (x, \omega) \in \mathbb{R} \times \Omega.$$

h is bounded, and moreover, for all $x \in \mathbb{R}$, $h(x, \cdot)$ is independent of $\mathcal{F}_s$, because so is $u(t, s, x; \omega)$ (given that $u(s) = x \in \mathbb{R}$ is a certain event). Furthermore, observing that if $t_0 < s$, $s \in [0, T]$, we obtain

$$u(t, t_0, c) = u(t, s, u(s, t_0, c)) \text{ for } t \ge s, \tag{4.21}$$

where c is the chosen random value. Equation (4.21) states the fact that the solution of (4.15) with the initial condition $u(t_0) = c$ is identical to the solution of the same equations with the initial condition $u(s) = u(s, t_0, c)$ for $t \ge s$ (see, e.g., Baldi (1984)). Equation (4.21) is called the *semigroup property* or *dynamic system*. (The proof of the property is left to the reader as an exercise.) Now, because $h(x, \omega) = I_B(u(t, s, x; \omega))$ satisfies conditions 1, 2, and 3 of Lemma 4.12 and by (4.21) we have $h(u(s), \omega) = I_B(u(t; \omega))$. Then, by (4.17), we obtain

$$P(u(t) \in B|\mathcal{F}_s) = P(u(t) \in B|u(s)) \text{ a.s.},$$

completing the proof. □

Remark 4.14. By (4.21) and (4.20) we also have

$$P(u(t) \in B|u(s)) = P(u(t,s,u(s)) \in B|u(s))$$

and, in particular,

$$P(u(t) \in B|u(s) = x) = P(u(t,s,u(s)) \in B|u(s) = x), \qquad x \in \mathbb{R}.$$

Hence

$$P(u(t) \in B|u(s) = x) = P(u(t,s,x) \in B), \qquad x \in \mathbb{R}. \tag{4.22}$$

Theorem 4.15. *If* $(u(t))_{t\in[t_0,T]}$ *is the solution of*

$$\begin{cases} u(t_0) = c \ a.s., \\ du(t) = a(t,u(t))dt + b(t,u(t))dW_t, \end{cases}$$

defining, for all $B \in \mathcal{B}_{\mathbb{R}}$ *and all* $t_0 \leq s < t \leq T$ *and all* $x \in \mathbb{R}$*:*

$$p(s,x,t,B) = P(u(t) \in B|u(s) = x) = P(u(t,s,x) \in B),$$

then p *is a transition probability (of the Markov process* $u(t)$*).*

Proof: We have to show that the conditions 1, 2 and 3 of Definition 2.97 are satisfied.

Point 1. Fixing s and t such that $t_0 \leq s < t \leq T$ and $B \in \mathcal{B}_{\mathbb{R}}$,

$$p(s,x,t,B) = P(u(t) \in B|u(s) = x) = E[I_B(u(t))|u(s) = x], \qquad x \in \mathbb{R}.$$

Then, as a property of conditional probabilities, $p(s,\cdot,t,B)$ is $\mathcal{B}_{\mathbb{R}}$-measurable.

Point 2 is true by the definition of $p(s,x,t,B)$.

Point 3. Fixing s and t such that $t_0 \leq s < t \leq T$ and $x \in \mathbb{R}$, $p(s,x,t,B) = P(u(t,s,x) \in B)$, for all $B \in \mathcal{B}_{\mathbb{R}}$. This is the induced probability P of $u(t,s,x)$. Therefore, if $\psi : \mathbb{R} \to \mathbb{R}$ is a bounded $\mathcal{B}_{\mathbb{R}}$-measurable function, then

$$\int_{\mathbb{R}} \psi(y)p(s,x,t,dy) = \int_{\Omega} \psi(u(t,s,x,\omega))dP(\omega).$$

Now, let $\psi(y) = p(r,y,t,B)$ with $B \in \mathcal{B}_{\mathbb{R}}$, $y \in \mathbb{R}$, $t_0 \leq r < t \leq T$. Then, for $s < r$, we have

$$\begin{aligned} &\int_{\mathbb{R}} p(r,y,t,B)p(s,x,r,dy) \\ &= \int_{\Omega} p(r,u(r,s,x,\omega),t,B)dP(\omega) \\ &= E[p(r,u(r,s,x),t,B)] = E[P(u(t) \in B|u(r) = u(r,s,x))] \\ &= E[E[I_B(u(t))|u(r) = u(r,s,x)]] = E[I_B(u(t))] = P(u(t) \in B) \\ &= p(s,x,t,B). \end{aligned}$$

In fact, $u(t)$ satisfies (4.15) with the initial condition $u(s) = x$. □

Remark 4.16. By Theorem 2.99, the knowledge of the solution $u(t)$ of

$$\begin{cases} u(t_0) = c \text{ a.s.,} \\ du(t) = a(t,u(t))dt + b(t,u(t))dW_t \end{cases}$$

is equivalent to assigning the transition probability p to the process $u(t)$ and the distribution P_0 of c.

Remark 4.17. Every stochastic differential equation generates Markov processes in the sense that every solution is a Markov process.

Now, if we assume that in the assumptions of Definition 4.1 the coefficients a and b do not explicitly depend upon time, i.e., (4.1) becomes

$$du(t) = a(u(t))dt + b(u(t))dW_t, \tag{4.23}$$

then the stochastic differential equation is called *autonomous* and the existence and uniqueness Theorem 4.4 can be restated in the following way.

Theorem 4.18. *Let $a(x)$, $b(x)$ be measurable functions in $\mathbb{R}$ with the property that for some constant $K > 0$:*

$$|a(x) - a(y)| + |b(x) - b(y)| \leq K|x - y|, \qquad x, y \in \mathbb{R}.$$

Then for any $u^0 \in L^2(\Omega, \mathcal{F}_0, P)$, independent of $\mathcal{F}_T$, there exists a unique $(u(t))_{t\in[0,T]}$, solution of the system (4.23) with initial condition (4.2), such that

- *$(u(t))_{t\in[0,T]}$ is continuous almost surely;*
- *$(u(t))_{t\in[0,T]} \in \mathcal{C}([0,T])$.*

Theorem 4.19. *Implicit in the underlying hypotheses of Theorem 4.4 is that if the stochastic differential equation is autonomous of form (4.23), then the Markov process $\{u(t,t_0,c), t \in [t_0,T]\}$ is homogeneous.*

Remark 4.20. The transition measure of the homogeneous process $(u_t)_{t\in[t_0,T]}$ is time-homogeneous, i.e.,

$$P(u(t+s) \in A | u(t) = x) = P(u(s) \in A | u(0) = x) \text{ almost surely,}$$

for any $s, t \in \mathbb{R}_+$, $x \in \mathbb{R}$ and $A \in \mathcal{B}_{\mathbb{R}}$.

Theorem 4.21. *If for*

$$\begin{cases} u(t_0) = c \ a.s., \\ du(t) = a(t,u(t))dt + b(t,u(t))dW_t, \end{cases}$$

the hypotheses of Theorem 4.4 are satisfied, with $a(t,x)$ and $b(t,x)$ being continuous in $(t,x) \in [0,\infty] \times \mathbb{R}$, then the solution $u(t)$ is a diffusion process with drift coefficient $a(t,x)$ and diffusion coefficient $b^2(t,x)$.

Proof: We prove point 1 of Lemma 2.123. Let $u(t,s,x)$ be a solution of the problem with initial value

$$u(s) = x \text{ a.s., } x \in \mathbb{R} \text{ (fixed), } t \geq s \text{ (s fixed).}$$

By (4.22):

$$p(t,x,t+h,A) = P(u(t+h,t,u(t)) \in A | u(t) = x) = P(u(t+h,t,x) \in A).$$

Hence $p(t,x,t+h,A)$ is the probability distribution of the random variable $u(t+h,t,x)$ and thus

$$E[f(u(t+h,t,x)-x)] = \int_{\mathbb{R}} f(y-x)p(t,x,t+h,dy),$$

for every function $f(z)$ such that[9] $|f(z)| \leq K(1+|z|^{2n})$, with $\alpha \geq 1$, $K > 0$, and $f(z)$ continuous. It is now sufficient to prove that

$$\lim_{h \downarrow 0} \frac{1}{h} E[|u(t+h,t,x)-x|^4] = 0.$$

Given that z^4 is of the preceding form $f(z)$, the above limit follows from

$$\frac{1}{h} E[|u(t+h,t,x)-x|^4] \leq \frac{1}{h} K h^2 (1+|x|^4)$$

by 2 of Theorem 4.11.

Now we prove 2 of Lemma 2.123. This is equivalent to showing that

$$\lim_{h \downarrow 0} \frac{1}{h} E[u(t+h,t,x)-x] = a(t,x).$$

Because $u(t,t,x) = x$ almost surely, due to the definition of the stochastic differential we obtain

$$E[u(t+h,t,x)-x] = E\left[\int_t^{t+h} a(s,u(s,t,x))ds + \int_t^{t+h} b(s,u(s,t,x))dW_s\right].$$

But since $E[\int_t^{t+h} b(s,u(s,t,x))dW_s] = 0$, we get

$$\begin{aligned} E[u(t+h,t,x)-x] &= E\left[\int_t^{t+h} a(s,u(s,t,x))ds\right] \\ &= E\left[\int_t^{t+h} (a(s,u(s,t,x)) - a(s,x))ds\right] + \int_t^{t+h} a(s,x)ds \\ &= \int_t^{t+h} E[a(s,u(s,t,x)) - a(s,x)]ds + \int_t^{t+h} a(s,x)ds, \end{aligned}$$

[9] The assumption $|f(z)| \leq K(1+|z|^{2n})$ implies that $E[|f(z)|] \leq K(1+E[|z|^{2n}])$ and, by Theorem 4.11, $E[|u(t+h,t,x)|^{2n}] < +\infty$. Therefore, $f(u(t+h,t,x)-x)$ is integrable.

after adding and subtracting the term $a(s,x)$. Moreover, $|\cdot|$ being a convex function, by the Schwarz inequality:

$$\begin{aligned}
&\left|\int_t^{t+h} E[a(s,u(s,t,x)) - a(s,x)]ds\right| \\
&\leq \int_t^{t+h} E[|a(s,u(s,t,x)) - a(s,x)|]ds \\
&\leq h^{\frac{1}{2}}\left(\int_t^{t+h} (E[|a(s,u(s,t,x)) - a(s,x)|])^2 ds\right)^{\frac{1}{2}} \\
&\leq h^{\frac{1}{2}}\left(\int_t^{t+h} E[|a(s,u(s,t,x)) - a(s,x)|^2]ds\right)^{\frac{1}{2}}.
\end{aligned}$$

Then, by hypothesis 1 of Theorem 4.4,

$$|a(s,u(s,t,x)) - a(s,x)|^2 \leq (K^*)^2|u(s,t,x) - x|^2,$$

and, by 2 of 4.11:

$$E[|u(s,t,x) - x|^2] \leq Kh(1+|x|^2), \; K \text{ constant, positive,}$$

and thus for $h \downarrow 0$

$$\frac{1}{h}\left|\int_t^{t+h} E[a(s,u(s,t,x)) - a(s,x)]ds\right| \leq \frac{1}{h}h^{\frac{1}{2}}K^*(hKh(1+|x|^2))^{\frac{1}{2}} \to 0.$$

Hence, as a conclusion, by the mean value theorem for $t \leq r \leq t+h$:

$$\lim_{h\downarrow 0}\frac{1}{h}E[u(t+h,t,x) - x] = \lim_{h\downarrow 0}\frac{1}{h}\int_t^{t+h} a(s,x)ds = \lim_{h\downarrow 0}\frac{1}{h}a(r,x) = a(t,x).$$

Lastly, we have to show that the assumptions of Lemma 2.123 are satisfied (see, e.g., Friedman (1975)). □

The Strong Markov Property of Solutions of Stochastic Differential Equations

Lemma 4.22. *By hypotheses 1. and 2. of Theorem 4.4, we have that*

$$\forall R > 0, \forall T > 0: \qquad E\left[\sup_{r\leq t\leq T}|u(t,s,x) - u(t,r,y)|^2\right] \leq C(|x-y|^2 + |s-r|),$$

for $|x| \leq R$, $|y| \leq R$, $0 \leq s \leq r \leq T$, where C is a constant that depends on R and T.

Proof: See, e.g., Friedman (1975). □

Theorem 4.23. *By hypotheses 1 and 2 of Theorem 4.4,* $(u(t,s,x))_{t\in[s,T]}$*, the solution of*

$$du(t) = a(t,u(t))dt + b(t,u(t))dW_t$$

satisfies the Feller property and hence the strong Markov property.

Proof: Let $f \in BC(\mathbb{R})$. By the Lebesgue theorem, we have

$$E[f(u(t+r,s,x))] \to E[f(u(t+s,s,x))] \text{ for } r \to s. \tag{4.24}$$

Moreover, by Lemma 4.22, and again by the Lebesgue theorem:

$$E[f(u(t+r,r,y))] - E[f(u(t+r,s,x))] \to 0 \text{ for } y \to x, r \to s; \tag{4.25}$$

therefore,

$$E[f(u(t+r,r,y))] - E[f(u(t+s,s,x))] \to 0 \text{ for } y \to x, r \to s.$$

Hence $(s,x) \to \int_{\mathbb{R}} p(s,x,s+t,dy)f(y)$ is continuous and so $(u(t,s,x))_{t\in[s,T]}$ satisfies the Feller property and, by Theorem 2.99 (because it is continuous) has the strong Markov property. □

4.3 Girsanov Theorem

Theorem 4.24. (Lévy characterization of Brownian motion). *Let* $(X_t)_{t\in\mathbb{R}_+}$ *be a real-valued continuous random process on a probability space* $(\Omega, \mathcal{F}, P)$*. Then the following two statements are equivalent:*

1. $(X_t)_{t\in\mathbb{R}_+}$ *is a* P*-Brownian motion;*
2. $(X_t)_{t\in\mathbb{R}_+}$ *and* $X_t^2 - t$ *are* P*-martingales (and with respect to their respective natural filtrations).*

Proof: See, e.g., Ikeda and Watanabe (1989). □

Example 4.25. The Wiener process $(W_t)_{t\in\mathbb{R}_+}$ is a continuous square-integrable martingale, with $W_t - W_s \sim N(0, t-s)$, for all $0 \le s < t$. To show that $W_t^2 - t$ is also a martingale we need to show that either

$$E[W_t^2 - t|\mathcal{F}_s] = W_s^2 - s \qquad \forall 0 \le s < t,$$

or, equivalently, that

$$E[W_t^2 - W_s^2|\mathcal{F}_s] = t - s \qquad \forall 0 \le s < t.$$

In fact,

$$E[W_t^2 - W_s^2|\mathcal{F}_s] = E\left[(W_t - W_s)^2 \middle| \mathcal{F}_s\right] = Var[W_t - W_s] = t - s.$$

Because of uniqueness, we can say that $\langle W_t \rangle = t$, for all $t \ge 0$ by indistinguishability.

Example 4.26. Consider the Itô integral

$$X_t = \int_0^t h_s dW_s, \qquad t \in [0,T],$$

for bounded $h_s \in C([0,T])$. Then

$$M_t = X_t^2 - \int_0^t |f_s|^2 ds, \qquad t \in [0,T]$$

is a martingale and the compensator $\langle X_t \rangle = \int_0^t |f_s|^2 ds$ is the quadratic variation process of the martingale X_t.

Lemma 4.27. *Let Z be a strictly positive random variable on $(\Omega, \mathcal{F}, P)$ with $E[Z] \equiv E_P[Z] = 1$. Furthermore, define the random measure $dQ = ZdP$. If $\mathcal{G}$ is a σ-algebra with $\mathcal{G} \subseteq \mathcal{F}$, then for any adapted random variable $X \in \mathcal{L}^1(Q)$ we have that*

$$E_Q[X|\mathcal{G}] = \frac{E[XZ|\mathcal{G}]}{E[Z|\mathcal{G}]}.$$

Lemma 4.28. *Let $(\mathcal{F}_t)_{t\in[0,T]}$, for $T > 0$, be a filtration on the probability space $(\Omega, \mathcal{F}, P)$, and let $(Z_t)_{t\in[0,T]}$ be a strictly positive $\mathcal{F}_t$-martingale with respect to the probability measure P such that $E_P[Z_t] = 1$ for any $t \in [0,T]$. A sufficient condition for an adapted stochastic process $(Y_t)_{t\in[0,T]}$ to be an $\mathcal{F}_t$-martingale with respect to the measure $dQ = Z_T dP$ is that the process $(Z_t Y_t)_{t\in[0,T]}$ is an $\mathcal{F}_t$-martingale with respect to P.*

Proof: Because $(Z_t Y_t)_{t\in[0,T]}$ is an $\mathcal{F}_t$-martingale with respect to P, for $s \leq t \leq T$, by the tower law of probability we have that

$$\begin{aligned} E[Z_T Y_t|\mathcal{F}_s] &= E[E[Z_T Y_t|\mathcal{F}_t]|\mathcal{F}_s] = E[Y_t E[Z_T|\mathcal{F}_t]|\mathcal{F}_s] = E[Y_t Z_t|\mathcal{F}_s] \\ &= Y_s Z_s. \end{aligned}$$

As a consequence we have that

$$E_Q[Y_t|\mathcal{F}_s] = \frac{E[Z_T Y_t|\mathcal{F}_s]}{E[Z_T|\mathcal{F}_s]} = \frac{Z_s Y_s}{Z_s} = Y_s.$$

□

Proposition 4.29. *1. Let $h_t \in L^2([0,T])$ be a Q-deterministic function, $W_t(\omega)$ a Brownian motion, and define*

$$Y_t(\omega) = \exp\left\{\int_0^t h_s dW_s(\omega) - \frac{1}{2}\int_0^t h_s^2 ds\right\}, \qquad t \in [0,T].$$

Then, by Itô's formula (see (3.35)),

$$dY_t = Y_t h_t dW_t.$$

2. Let $\vartheta_s(\omega) \in C([0,T])$ with $T \leq \infty$ and define

$$Z_t(\omega) = \exp\left\{\int_0^t \vartheta_s(\omega)dW_s(\omega) - \frac{1}{2}\int_0^t \vartheta_s^2(\omega)ds\right\}, \qquad t \in [0,T].$$

Then, by Itô's formula,

$$dZ_t = Z_t\vartheta_t dW_t.$$

Lemma 4.30. (Novikov condition). *Under the assumptions of point 2 of Proposition 4.29, if*

$$E\left[\exp\left\{\frac{1}{2}\int_0^T |\vartheta(s)|^2 ds\right\}\right] < +\infty,$$

then $(Z_t)_{t\in[0,T]}$ is a martingale and $E[Z_t] = E[Z_0] = 1$.

Theorem 4.31. (Girsanov). *Let $(Z_t)_{t\in[0,T]}$ be a P-martingale and let ϑ_s satisfy the Novikov condition. Then the process*

$$Y_t = W_t - \int_0^t \vartheta_s ds$$

is a Brownian motion with respect to the measure $dQ = Z_T dP$.

Proof: We resort to the Lévy characterization of Brownian motion, Theorem 4.24 and prove point 2. Let $M_t = Z_t Y_t$. Then, by Lemma 4.28, to prove that $(Y_t)_{t\in[0,T]}$ is a Q-martingale it is sufficient to show that $(M_t)_{t\in[0,T]}$ is a P-martingale. Assuming that $(\vartheta_t)_{t\in[0,T]}$ satisfies the Novikov condition and that $(Z_t)_{t\in[0,T]}$ is a martingale with $E[Z_t] = 1$, by Itô's formula we obtain

$$\begin{aligned} dM_t &= Z_t dY_t + Y_t dZ_t + Z_t\vartheta_t dt = Z_t(dW_t - \vartheta_t dt) + Y_t Z_t \vartheta_t dW_t + Z_t\vartheta_t dt \\ &= Z_t(dW_t + Y_t\vartheta_t dW_t) = Z_t(1 + \vartheta_t Y_t)dW_t. \end{aligned}$$

Hence $(M_t)_{t\in[0,T]}$ is a martingale. To further show that $Y_t^2 - t$ is a martingale is left as an exercise. □

Remark 4.32. The Girsanov theorem implies that for all $F_1, \ldots, F_n \in \mathcal{B}$, where $\mathcal{B}$ is the state space of the processes and for all $t_1, \ldots, t_n \in [0,T]$:

$$Q(Y_{t_1} \in F_1, \ldots, Y_{t_k} \in F_k) = P(W_{t_1} \in F_1, \ldots, W_{t_k} \in F_k)$$

and $Q \ll P$ as well as with Radon–Nikodym derivative

$$\frac{dQ}{dP} = M_T, \qquad \text{on } \mathcal{F}_T.$$

Furthermore, because by the Radon–Nikodym Theorem A.53

$$Q(F) = \int_F M_T(\omega)P(d\omega)$$

and $M_T > 0$, we have that

$$Q(F) > 0 \Rightarrow P(F) > 0$$

and vice versa, so that

$$Q(F) = 0 \Rightarrow P(F) = 0,$$

and thus $P \ll Q$. Therefore, the two measures are equivalent.

4.4 Kolmogorov Equations

We will consider the stochastic differential equation

$$du(t) = a(t, u(t))dt + b(t, u(t))dW_t \tag{4.26}$$

and suppose that the coefficients a and b satisfy the assumptions of the existence and uniqueness Theorem 4.4. We will denote by $u(t, x)$, for $s \leq t \leq T$, the solution of (4.26) subject to the initial condition

$$u(s, s, x) = x \text{ a.s. } (x \in \mathbb{R}).$$

Remark 4.33. Under the assumptions 1 and 2 of Theorem 4.4 on the coefficients a and b, if $f(t, x)$ is continuous in both variables as well as $|f(t, x)| \leq K(1 + |x|^m)$ with k, m positive constants, it can be shown that

$$\lim_{h \downarrow 0} \frac{1}{h} \int_h^{t+h} E[f(s, u(s, t, x))]ds = f(t, x), \tag{4.27}$$

$$\lim_{h \downarrow 0} \frac{1}{h} \int_{t-h}^{t} E[f(s, u(s, t, x))]ds = f(t, x). \tag{4.28}$$

The proof employs similar arguments as the proofs of Theorems 4.21 and 4.11.

Lemma 4.34. *If $f : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function and if there exist $C > 0$ and $m > 0$ such that*

$$|f(x)| + |f'(x)| + |f''(x)| \leq C(1 + |x|^m), \qquad x \in \mathbb{R},$$

and if the coefficients $a(t, x)$ and $b(t, x)$ satisfy assumptions 1 and 2 of Theorem 4.4, then

$$\lim_{h \downarrow 0} \frac{1}{h} \left(E[f(u(t, t-h, x))] - f(x) \right) = a(t, x)f'(x) + \frac{1}{2}b^2(t, x)f''(x). \tag{4.29}$$

Proof: By Itô's formula, we get

$$f(u(t,t-h,x)) - f(x) = \int_{t-h}^{t} a(s,u(s,t-h,x))f'(u(s,t-h,x))ds + \int_{t-h}^{t} \frac{1}{2}b^2(s,u(s,t-h,x))f''(u(s,t-h,x))ds + \int_{t-h}^{t} b(s,u(s,t-h,x))f'(u(s,t-h,x))dW_s,$$

and after taking expectations

$$E[f(u(t,t-h,x))] - f(x) = E\left[\int_{t-h}^{t} a(s,u(s,t-h,x))f'(u(s,t-h,x))ds + \int_{t-h}^{t} \frac{1}{2}b^2(s,u(s,t-h,x))f''(u(s,t-h,x))ds\right],$$

hence

$$\frac{1}{h}(E[f(u(t,t-h,x))] - f(x)) = \frac{1}{h}\int_{t-h}^{t} E[a(s,u(s,t-h,x))f'(u(s,t-h,x))]ds + \int_{t-h}^{t} E\left[\frac{1}{2}b^2(s,u(s,t-h,x))f''(u(s,t-h,x))\right] ds.$$

Then Proposition 4.29 follows from Lemma 4.27 because $u(t,t,x) = x$. □

Remark 4.35. Resorting to the notation of Definitions 2.103 and 2.104, equation (4.29) can also be written as

$$\mathcal{A}_t f = \lim_{h\downarrow 0} \frac{T_{t-h,t}f - f}{h} = f'a(t,\cdot) + \frac{1}{2}f''b^2(t,\cdot). \qquad (4.30)$$

Moreover, by Theorem 4.21 and Proposition 2.124, we also have

$$\mathcal{A}_s f = \lim_{h\downarrow 0} \frac{T_{s,s+h}f - f}{h} = f'a(s,\cdot) + \frac{1}{2}f''b^2(s,\cdot), \qquad (4.31)$$

if $f \in BC(\mathbb{R}) \cap C^2(\mathbb{R})$. On the other hand, in the time-homogeneous case we have

$$\mathcal{A}f = \lim_{h\downarrow 0} \frac{T_h f - f}{h} = f'a(\cdot) + \frac{1}{2}f''b^2(\cdot).$$

Theorem 4.36. *If $u(t)$ is the Markovian solution of the homogeneous stochastic differential equation (4.23) and $\mathcal{A}$ the associated infinitesimal generator, then for $f \in BC(\mathbb{R}) \cap C^2(\mathbb{R})$ the process*

$$M_t = f(u(t)) - \int_0^t [\mathcal{A}f](u(s))ds \qquad (4.32)$$

is a martingale.

Proof: By Itô's formula, we have that

$$f(u(t)) = f(u^0) + \int_0^t [\mathcal{A}f](u(s))ds + \int_0^t b(u(s))f'(u(s))dW_s,$$

which, substituted into (4.32), results in

$$M_t = f(u^0) + \int_0^t b(u(s))f'(u(s))dW_s.$$

Since an Itô integral is a martingale with respect to filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ generated by the Wiener process $(W_t)_{t\in\mathbb{R}_+}$, therefore

$$E[M_t|\mathcal{F}_s] = M_s.$$

If we now consider the filtration $(\mathcal{M}_t)_{t\in[0,T]}$, generated by $(M_t)_{t\in[0,T]}$, then

$$E[M_t|\mathcal{M}_s] = E[E[M_t|\mathcal{F}_s]|\mathcal{M}_s] = E[M_t|\mathcal{F}_s] = M_s,$$

because $\mathcal{F}_s \subset \mathcal{M}_s$. □

Furthermore, we note that it is valid to reverse the argumentation of Theorem 4.21, as the following theorem states.

Theorem 4.37. *If $(u(t))_{t\in[0,T]}$ is a diffusion process with drift $a(t,x)$ and diffusion coefficient $c(t,x)$, where*

1. *$a(t,x)$ is continuous in both variables as well as $|a(t,x)| \leq K(1+|x|)$, K a positive constant;*
2. *$c(t,x)$ is continuous in both variables and has continuous as well as bounded derivatives $\frac{\partial}{\partial t}c(t,x)$ and $\frac{\partial}{\partial x}c(t,x)$, and moreover $\frac{1}{c(t,x)}$ is bounded;*
3. *there exists a function $\psi(x)$ that is independent of t and where*

$$\psi(x) > 1 + |x|, \qquad \sup_{0\leq t\leq T} E[\psi(u(t))] < +\infty,$$

as well as

$$\left|\int_\Omega (y-x)p(t,x,t+h,dy)\right| + \left|\int_\Omega (y-x)^2 p(t,x,t+h,dy)\right| \leq \psi(x)h,$$
$$\int_\Omega (|y|+y^2)p(t,x,t+h,dy) \leq \psi(x),$$

then there exists a Wiener process W_t, so that $u(t)$ satisfies the stochastic differential equation

$$du(t) = a(t,u(t))dt + \sqrt{c(t,u(t))}dW_t.$$

Proof: See, e.g., Gihman and Skorohod (1972). □

Remark 4.38. Equation (4.31) can also be shown by Itô's formula, as in the proof of Lemma 4.34.

Proposition 4.39. Let $f(x)$ be r times differentiable and suppose that there exists an $m > 0$ so that $|f^{(k)}(x)| \leq L(1+|x|^m)$. If $a(t,x)$ and $b(t,x)$ both satisfy the assumptions of Theorem 4.4 and there exist $\frac{\partial^k}{\partial x^k}a(t,x), \frac{\partial^k}{\partial x^k}b(t,x)$, $k = 1, \ldots, r$, that are continuous as well as

$$\left|\frac{\partial^k}{\partial x^k}a(t,x)\right| + \left|\frac{\partial^k}{\partial x^k}b(t,x)\right| \leq C_k(1+|x|^{m_k}), \qquad k = 1, \ldots, r$$

(with C_k and m_k being positive constants), then the function $\phi_s(z) = E[f(u(t,s,z))]$ is r times differentiable with respect to z (i.e., with respect to the initial condition).

Proof: See, e.g., Gihman and Skorohod (1972). □

Theorem 4.40. *If the coefficients* $a(t,x)$ *and* $b(t,x)$ *are continuous and have continuous partial derivatives* $a'_x(t,x)$, $a''_x(t,x)$, $b'_x(t,x)$, *and* $b''_{xx}(t,x)$, *and moreover, if there exist a* $k > 0$ *and an* $m > 0$ *such that*

$$\begin{aligned} |a(t,x)| + |b(t,x)| &\leq k(1+|x|), \\ |a'_x(t,x)| + |a''_{xx}(t,x)| + |b'_x(t,x)| + |b''_{xx}(t,x)| &\leq k(1+|x|^m), \end{aligned}$$

and furthermore if the function $f(x)$ *is twice continuously differentiable with*

$$|f(x)| + |f'(x)| + |f''(x)| \leq k(1+|x|^m),$$

then the function

$$q(t,x) \equiv E[f(u(s,t,x))], \qquad 0 < t < s, \qquad x \in \mathbb{R}, s \in]0,T[, \tag{4.33}$$

satisfies the equation

$$\frac{\partial}{\partial t}q(t,x) + a(t,x)\frac{\partial}{\partial x}q(t,x) + \frac{1}{2}b^2(t,x)\frac{\partial^2}{\partial x^2}q(t,x) = 0, \tag{4.34}$$

with the boundary condition

$$\lim_{t \uparrow s} q(t,x) = f(x). \tag{4.35}$$

Equation (4.34) is called Kolmogorov's backward differential equation.

Proof: Since, by the semigroup property, $u(s,t-h,x) = u(s,t,u(t,t-h,x))$ and in general $E[f(Y(\cdot,X))|X=x] = E[f(Y(\cdot,x))]$, we have

$$\begin{aligned} q(t-h,x) &= E[f(u(s,t-h,x))] \\ &= E[E[f(u(s,t-h,x))|u(t,t-h,x)]] \\ &= E[E[f(u(s,t,u(t,t-h,x)))|u(t,t-h,x)]] \\ &= E[E[f(u(s,t,u(t,t-h,x)))]] = E[q(t,u(t,t-h,x))]. \end{aligned} \tag{4.36}$$

By Proposition 4.39, $q(t,x)$ is twice differentiable with respect to x and, by Lemma 4.34, we get

$$\lim_{h\downarrow 0} \frac{E[q(t,u(t,t-h,x))]-q(t,x)}{h} = a(t,x)\frac{\partial}{\partial x}q(t,x) + \frac{1}{2}b^2(t,x)\frac{\partial^2}{\partial x^2}q(t,x).$$

Therefore, by equation (4.36) the limit

$$\lim_{h\downarrow 0} \frac{q(t,x)-q(t-h,x)}{h} = \lim_{h\downarrow 0} \frac{q(t,x)-E[q(t,u(t,t-h,x))]}{h},$$

and thus

$$\frac{\partial}{\partial t}q(t,x) = \lim_{h\downarrow 0} \frac{q(t,x)-q(t-h,x)}{h} = -a(t,x)\frac{\partial}{\partial x}q(t,x) - \frac{1}{2}b^2(t,x)\frac{\partial^2}{\partial x^2}q(t,x).$$

It can further be shown that $\frac{\partial}{\partial t}q(t,x)$ is continuous in t and so are $\frac{\partial q}{\partial x}$ as well as $\frac{\partial^2 q}{\partial x^2}$. We observe that

$$|E[f(u(s,t,x))-f(x)]| \leq E[|f(u(s,t,x))-f(x)|],$$

and, by Lagrange's theorem (also known as the *mean value theorem*),

$$|f(u(s,t,x))-f(x)| = |u(s,t,x)-x||f'(\xi)|,$$

with ξ related to $u(s,t,x)$ and x through the assumptions $|f'(\xi)| \leq k(1+|\xi|^m)$ and

$$(1+|\xi|^m) \leq \begin{cases} 1+|x|^m & \text{if } u(s,t,x) \leq \xi \leq x, \\ 1+|u(s,t,x)|^m & \text{if } x \leq \xi \leq u(s,t,x). \end{cases}$$

Therefore, by both the Schwarz inequality and the fact that

$$E[(u(s,t,x)-x)^2] \leq \widetilde{L}(1+|x|^2)(s-t)^2,$$

we obtain

$$\begin{aligned} &|E[f(u(s,t,x))-f(x)]| \\ &\leq LE[|u(s,t,x)-x|(1+|x|^m+|u(s,t,x)|^m)] \\ &\leq L(E[(u(s,t,x)-x)^2])^{\frac{1}{2}}(E[(1+|x|^m+|u(s,t,x)|^m)^2])^{\frac{1}{2}}, \end{aligned}$$

where L is a positive constant. Since $\widetilde{L}(1+|x|^2)(s-t)^2 \to 0$ for $t \uparrow s$, it follows that

$$\lim_{t\uparrow s} E[f(u(s,t,x))] = f(x).$$

□

Remark 4.41. If we put $\tilde{t} = s-t$ for $0<t<s$, then $\frac{\partial}{\partial \tilde{t}} = -\frac{\partial}{\partial t}$ and the limit $\lim_{t\uparrow s}$ is equivalent to $\lim_{\tilde{t}\downarrow 0}$. Hence (4.34) takes us back to a classic parabolic differential equation with initial condition (4.35) given by $\lim_{\tilde{t}\downarrow 0} q(\tilde{t},x) = f(x)$.

Theorem 4.42. (Feynman–Kac formula). *Under the assumptions of Theorem 4.40, let c be a real-valued, nonnegative continuous function in* $]0, T[\times\mathbb{R}$. *Then the function, for* $x \in \mathbb{R}$,

$$q(t,x) = E\left[f(u(s,t,x))e^{-\int_t^s c(u(\tau,t,x),\tau)d\tau}\right], \qquad 0 < t < s < T, \tag{4.37}$$

satisfies the equation

$$\frac{\partial}{\partial t}q(t,x) + a(t,x)\frac{\partial}{\partial x}q(t,x) + \frac{1}{2}b^2(t,x)\frac{\partial^2}{\partial x^2}q(t,x) + c(t,x)q(t,x) = 0,$$

subject to the boundary condition $\lim_{t\uparrow s} q(t,x) = f(x)$. *Equation (4.37) is called the Feynman–Kac formula.*

Proof: The proof is a direct consequence of Theorem 4.40 and Itô's formula, considering that the process

$$Z(t) = e^{-\int_t^s c(u(\tau,t,x),\tau)d\tau}, \qquad 0 < t < s < T, x \in \mathbb{R},$$

satisfies the stochastic differential equation

$$dZ(t) = c(u(t,t_0,x)t)Z(t)dt$$

with initial condition $Z(t_0) = 1$. □

Remark 4.43. We can interpret the exponential term in the Feynman–Kac formula as a "killing" process. Let $(W_t)_{t\in\mathbb{R}_+}$ be a Wiener process that may disappear into a "coffin" state at a random killing time T. Let the killing probability over an interval $]t, t+dt]$ be equal to $c(W_t)dt + o(dt)$. Then the survival probability until T is given by

$$(1 - c(W_{t_1}))dt(1 - c(W_{t_2}))dt \cdots (1 - c(W_T))dt + o(1), \tag{4.38}$$

where $0 = t_0 < t_1 < \cdots < t_n = T$, $dt = t_{i+1} - t_i$. As $dt \to 0$, (4.38) tends to

$$e^{-\int_0^T c(W_t)dt}.$$

Hence for any function $f \in BC(\mathbb{R})$:

$$\begin{aligned} q(t,x) &= E[f(W_t), T > t] = E[f(W_t)P(T > t)] \\ &= E\left[f(W_t)e^{-\int_t^s c(u(\tau,t,x),\tau)d\tau}\right]. \end{aligned}$$

Proposition 4.44. *Consider the Cauchy problem:*

$$\begin{cases} L_0[q] + \frac{\partial q}{\partial t} = 0 & \text{in } [0,T[\times\mathbb{R}, \\ \lim_{t\uparrow T} q(t,x) = \phi(x) & \text{in } \mathbb{R}, \end{cases} \tag{4.39}$$

where $L_0[\cdot] = \frac{1}{2}b^2(t,x)\frac{\partial^2}{\partial x^2} + a(t,x)\frac{\partial}{\partial x}$, *and suppose that*

(B_1) $\phi(x)$ is continuous in $\mathbb{R}$ and $\exists A > 0, a > 0$ such that $|\phi(x)| \leq A(1 + |x|^a)$;

(B_2) a and b are bounded in $[0,T] \times \mathbb{R}$ and uniformly Lipschitz in (t,x) on compact subsets of $[0,T] \times \mathbb{R}$;

(B_3) b is Hölder continuous in x and uniform with respect to (t,x) on $[0,T] \times \mathbb{R}$;

under the conditions $(B_1), (B_2), (B_3)$, and (A_1) (see Appendix C), the Cauchy problem (4.36) admits a unique solution $q(t,x)$ in $[0,T] \times \mathbb{R}$ such that

$$|q(t,x)| \leq C(1 + |x|^a), \tag{4.40}$$

where C is a constant.

Proof: The uniqueness is shown through Corollary C.17 and existence follows from Theorem C.20. Then (4.39) follows, by (B_1) and by Theorem C.19, with $m = 0$. □

Theorem 4.45. *Under the conditions $(B_1), (B_2), (B_3)$, and (A_1), the solution of the Cauchy problem (4.39) is given by*

$$q(t,x) = E[\phi(u(T,t,x))] \equiv E_{t,x}[\phi(u(T))]. \tag{4.41}$$

Proof: The proof follows directly by Theorem 4.40, recalling the uniqueness of the solution of (4.39). □

We denote by $\Gamma_0^*(x,s;y,t)$ the fundamental solution of $L_0 + \frac{\partial}{\partial s}$ $(s < t)$. By Theorem C.20, where we replace t by $T - t$, $q(t,x)$ can be expressed by means of the fundamental solution Γ_0^* as

$$q(t,x) = \int_{\mathbb{R}} \Gamma_0^*(x,s;y,T)\phi(y)dy. \tag{4.42}$$

From equations (4.41) and (4.42), it then follows that

$$E[\phi(u(T,t,x))] = \int_{\mathbb{R}} \Gamma_0^*(x,s;y,T)\phi(y)dy. \tag{4.43}$$

Analogously, for all $0 \leq s < t \leq T$:

$$E[\phi(u(t,s,x))] = \int_{\mathbb{R}} \Gamma_0^*(x,s;y,t)\phi(y)dy \tag{4.44}$$

or, equivalently,

$$\int_{\mathbb{R}} \phi(y)p(s,x,t,dy) = \int_{\mathbb{R}} \Gamma_0^*(x,s;y,T)\phi(y)dy, \tag{4.45}$$

and because equation (4.45) holds for every ϕ that satisfies (B_1), it will certainly hold for every $\phi \equiv I_{]-\infty,z]}$, $z \in \mathbb{R}$. We obtain the following theorem.

Theorem 4.46. *Under the conditions* (A_1) *and* (B_1)*, the transition probability* $p(s,x,t,A) = P(u(t,s,x) \in A)$ *of the Markov process* $u(t,s,x)$ *(the solution of the differential equation (4.26)) is endowed with density. The latter is given by* $\Gamma_0^*(x,s;y,t)$ *and thus*

$$p(s,x,t,A) = \int_A \Gamma_0^*(x,s;y,t)dy \qquad (s < t), \text{ for all } A \in \mathcal{B}_{\mathbb{R}}. \tag{4.46}$$

Definition 4.47. The density $\Gamma_0^*(x,s;y,t)$ of $p(s,x,t,A)$ is the *transition density* of the solution $u(t)$ of (4.26).

Remark 4.48. Following the explanations in Appendix C, we can assert that $\Gamma_0^*(x,s;y,t)$ is the solution of Kolmogorov's backward equation

$$\begin{cases} L_0[\Gamma_0^*] + \frac{\partial}{\partial t}\Gamma_0^* = 0, \\ \lim_{t\to T} \Gamma_0^*(x,s;y,T) = \delta(x-y). \end{cases} \tag{4.47}$$

Example 4.49. The Brownian motion $(W_t)_{t\geq 0}$ is the solution of

$$\begin{cases} du(t) = dW_t, \\ u(0) = 0 \text{ a.s.} \end{cases}$$

We define the operator L_0 by $\frac{1}{2}\Delta$, where Δ is the Laplacian $\frac{\partial^2}{\partial x^2}$. The fundamental solution $\Gamma_0^*(x,s;y,t)$ of the operator $\frac{1}{2}\Delta + \frac{\partial}{\partial t}$, $s < t$, corresponds to the fundamental solution $\Gamma_0(y,t;x,s)$ of the operator $\frac{1}{2}\Delta - \frac{\partial}{\partial t}$, $s < t$, which, apart from the coefficient $\frac{1}{2}$, is the diffusion or heat operator. We therefore find that

$$\Gamma_0^*(x,s;y,t) = \Gamma(y,t;x,s) = \frac{1}{\sqrt{2\pi(t-s)}}e^{-\frac{(x-y)^2}{2(t-s)}},$$

the probability density function of $W_t - W_s$.

Under the assumptions of Theorem 4.46, the transition probability

$$p(s,x,t,A) = P(u(t,s,x) \in A)$$

of the Markov diffusion process $u(t,s,x)$, the latter being the solution of the stochastic differential equation (4.26), subject to the initial condition $u(s,s,x) = x$ a.s. $(x \in \mathbb{R})$, admits a density $f(s,x,t,y)$, which is the solution of system (4.47). Under these conditions the following also holds (see Gihman and Skorohod (1974), pp. 374 onwards):

Theorem 4.50. *In addition to the assumptions of Theorem 4.46, if the transition density is sufficiently regular so that there exist continuous derivatives*

$$\frac{\partial f}{\partial t}(s,x,t,y), \qquad \frac{\partial}{\partial y}(a(t,y)f(s,x,t,y)), \qquad \frac{\partial^2}{\partial y^2}(b(t,y)f(s,x,t,y)),$$

then $f(s,x,t,y)$*, as a function of* t *and* y*, satisfies the equation*

$$\frac{\partial f}{\partial t}(s,x,t,y) + \frac{\partial}{\partial y}(a(t,y)f(s,x,t,y)) - \frac{\partial^2}{\partial y^2}(b(t,y)f(s,x,t,y)) = 0 \quad (4.48)$$

in the region $t \in]s,T]$, $y \in \mathbb{R}$.

Proof: Let $g \in C_0^2(\mathbb{R})$ denote a sufficiently smooth function with compact support. By proceeding as in Lemma 4.34 (see also equation (4.31)),

$$\lim_{h\to 0} \frac{1}{h}\left(\int g(y)f(t,x,t+h,y)dy - g(x)\right) = a(t,x)g'(x) + \frac{1}{2}b(t,y)g"(x)$$

uniformly with respect to x. The Chapman–Kolmogorov equation for the transition densities is

$$f(t_1,x,t_3,y) = \int f(t_1,x,t_2,z)f(t_2,z,t_3,y)dz \qquad \text{for } t_1 < t_2 < t_3.$$

If we take $t_1 = s$, $t_2 = t$, $t_3 = t+h$, we obtain

$$\begin{aligned}
&\frac{\partial}{\partial t}\int f(s,x,t,y)g(y)dy \\
&= \int \frac{\partial}{\partial t} f(s,x,t,y)g(y)dy \\
&= \lim_{h\to 0}\frac{1}{h}\left(\int g(y)f(s,x,t+h,y)dy - \int g(z)f(s,x,t,z)dz\right) \\
&= \lim_{h\to 0}\frac{1}{h}\left(\int f(s,x,t,z)\left(\int g(y)f(s,z,t+h,y)dy - g(z)\right)\right)dz \\
&= \int f(s,x,t,z)\left(a(t,z)g'(z) + \frac{1}{2}b(t,z)g''(z)\right)dz.
\end{aligned}$$

An integration by parts leads to

$$\begin{aligned}
&\int \frac{\partial}{\partial t} f(s,x,t,y)g(y)dy \\
&= \int \left(-\frac{\partial}{\partial y}(a(t,y)f(s,x,t,y)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(b(t,y)f(s,x,t,z))\right)dy,
\end{aligned}$$

which represents a weak formulation of (4.48). □

Equation (4.48) is known as the *forward Kolmogorov equation* or *Fokker–Planck equation.* It is worth pointing out that while the forward equation has a more intuitive interpretation than the backward equation, the regularity conditions on the functions a and b are more stringent than those needed in the backward case. The problem of existence and uniqueness of the solution of the Fokker–Planck equation is not of an elementary nature, especially in the presence of boundary conditions. This suggests that the backward approach is more convenient than the forward approach from the viewpoint of analysis. For a discussion on the subject we refer to Feller (1971), page 326 onwards, Sobczyk (1991), page 34, and Taira (1988), page 9.

4.5 Multidimensional Stochastic Differential Equations

Let $\mathbf{a}(t,\mathbf{x}) = (a_1(t,\mathbf{x}),\ldots,a_m(t,\mathbf{x}))'$ and $b(t,\mathbf{x}) = (b_{ij}(t,\mathbf{x}))_{i=1,\ldots,m,j=1,\ldots,n}$ be measurable functions with respect to $(t,\mathbf{x}) \in [0,T]\times\mathbb{R}^n$. An *m-dimensional stochastic differential equation* is of the form

$$d\mathbf{u}(t) = \mathbf{a}(t,\mathbf{u}(t))dt + b(t,\mathbf{u}(t))d\mathbf{W}(t), \tag{4.49}$$

with the initial condition

$$\mathbf{u}(0) = \mathbf{u}^0 \text{ a.s.}, \tag{4.50}$$

where $\mathbf{u}^0$ is a fixed m-dimensional random vector. The entire theory of the one-dimensional case translates to the multidimensional case, with the norms redefined as

$$|\mathbf{b}|^2 = \sum_{i=1}^m |b_i|^2 \quad \text{if } \mathbf{b}\in\mathbb{R}^m,$$
$$|b|^2 = \sum_{i=1}^m\sum_{j=1}^n |b_{ij}|^2 \quad \text{if } b\in\mathbb{R}^{mn}.$$

Further, we introduce the notation

$$D_{\mathbf{x}}^\alpha = \frac{\partial^\alpha}{\partial\mathbf{x}^\alpha} = \frac{\partial^{\alpha_1+\cdots+\alpha_m}}{\partial x_1^{\alpha_1}\cdots\partial x_m^{\alpha_m}}, \qquad |\alpha| = \alpha_1+\cdots+\alpha_m,$$

which, as an application of Itô's formula, gives the following result.

Theorem 4.51. *If for a system of stochastic differential equations the conditions of the existence and uniqueness theorem (analogous to Theorem 4.4) are satisfied and if*

1. *there exist $D_{\mathbf{x}}^\alpha\mathbf{a}(t,\mathbf{x})$ and $D_{\mathbf{x}}^\alpha b(t,\mathbf{x})$ continuous for $|\alpha|\le 2$, with*
$$|D_{\mathbf{x}}^\alpha\mathbf{a}(t,\mathbf{x})| + |D_{\mathbf{x}}^\alpha\mathbf{b}(t,\mathbf{x})| \le k_0(1+|\mathbf{x}|^\beta), \qquad |\alpha|\le 2,$$
where k_0,β are strictly positive constants;
2. *$f:\mathbb{R}^m\to\mathbb{R}$ is a function endowed with continuous derivatives to second order, with*
$$|D_{\mathbf{x}}^\alpha f(\mathbf{x})| \le c(1+|\mathbf{x}|^{\beta'}), \qquad |\alpha|\le 2,$$
where c,β' are strictly positive constants;

then, putting $q(t,\mathbf{x}) = E[f(\mathbf{u}(s,t,\mathbf{x}))]$ for $\mathbf{x}\in\mathbb{R}^m$ and $t\in]0,s[$, we have that q_t, q_{x_i}, $q_{x_ix_j}$ are continuous in $(t,\mathbf{x})\in]0,s[\times\mathbb{R}^m$ and q satisfies the backward Kolmogorov equation

$$\frac{\partial q}{\partial t} + \sum_{i=1}^m a_i\frac{\partial q}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^m (bb')_{ij}\frac{\partial^2 q}{\partial x_i\partial x_j} = 0 \ \text{in }]0,s[\times\mathbb{R}^m, \tag{4.51}$$

$$\lim_{t\uparrow s} q(t,\mathbf{x}) = f(\mathbf{x}). \tag{4.52}$$

Applications of Itô's Formula: First Hitting Times

Let $\Omega \subset \mathbb{R}^m$ and $\mathbf{u}(t)$ be the solution of (4.49) with the initial condition $\mathbf{u}(s) = \mathbf{x}$ almost surely, $\mathbf{x} \in \Omega$. Putting

$$\tau_{\mathbf{x},s} = \inf\{t \geq s | \mathbf{u}(t) \in \partial\Omega\},$$

then $\tau_{\mathbf{x},s}$ is the *first hitting time of the boundary of* Ω or the *first exit time from* Ω. Because $\partial\Omega$ is a closed set, by Theorem 2.99, $\tau_{\mathbf{x},s}$ is a stopping time. Following Theorem 3.69, if $\phi : (\mathbf{x}, t) \in \mathbb{R}^m \times \mathbb{R}_+ \to \phi(\mathbf{x}, t) \in \mathbb{R}$ is sufficiently regular, then we obtain Itô's formula

$$d\phi(\mathbf{u}(t), t) = L\phi(\mathbf{u}(t), t)dt + \nabla_{\mathbf{x}}\phi(\mathbf{u}(t), t) \cdot b(t)d\mathbf{W}(t),$$

which we can apply on the interval $[s, \tau_{\mathbf{x},s}]$:

$$\phi(\mathbf{u}(\tau_{\mathbf{x},s}), \tau_{\mathbf{x},s}) = \phi(\mathbf{x}, s) + \int_s^{\tau_{\mathbf{x},s}} L\phi(\mathbf{u}(t'), t')dt' + \int_s^{\tau_{\mathbf{x},s}} \nabla_{\mathbf{x}}\phi(\mathbf{u}(t'), t') \cdot b(t')d\mathbf{W}(t')$$

and after taking expectations

$$E[\phi(\mathbf{u}(\tau_{\mathbf{x},s}), \tau_{\mathbf{x},s})] = \phi(\mathbf{x}, s) + E\left[\int_s^{\tau_{\mathbf{x},s}} L\phi(\mathbf{u}(t'), t')dt'\right]. \tag{4.53}$$

The value of the stochastic integral is 0 by Theorem 3.43. If we now suppose that ϕ satisfies the conditions

$$\begin{cases} L\phi(\mathbf{x}, t) = -1 \ \forall t \geq s, \forall \mathbf{x} \in \Omega, \\ \phi(\mathbf{x}, t) = 0 \qquad \forall \mathbf{x} \in \partial\Omega, \end{cases} \tag{4.54}$$

then, by (4.53), we get

$$E[\phi(\mathbf{u}(\tau_{\mathbf{x},s}), \tau_{\mathbf{x},s})] = \phi(\mathbf{x}, s) - E[\tau_{\mathbf{x},s}] + s,$$

and by (4.54),

$$E[\phi(\mathbf{u}(\tau_{\mathbf{x},s}), \tau_{\mathbf{x},s})] = 0.$$

Thus

$$E[\tau_{\mathbf{x},s}] = s + \phi(\mathbf{x}, s). \tag{4.55}$$

Remark 4.52. If equation (4.49) is homogeneous (i.e., $\mathbf{a} = \mathbf{a}(\mathbf{x})$ and $b = b(\mathbf{x})$ do not explicitly depend on time), then the process $\mathbf{u}(t)$, namely, the solution of (4.49), is time-homogeneous. Without loss of generality we can assume that $s = 0$. Then (4.55) becomes

$$E[\tau_{\mathbf{x}}] = \phi(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \tag{4.56}$$

which is *Dynkin's formula*. Notably, in this case, $\phi(\mathbf{x})$ is the solution of the elliptic problem

$$\begin{cases} M[\phi] = -1 \text{ in } \Omega, \\ \phi = 0 \qquad \text{on } \partial\Omega, \end{cases} \tag{4.57}$$

where $M = \sum_{i=1}^m a_i \frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^m (bb')_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$.

Crossing Points

If we now suppose that ϕ satisfies the conditions

$$\begin{cases} L[\phi](\mathbf{x},t) = 0 \ \ \forall t \geq s, \forall \mathbf{x} \in \Omega, \\ \phi(\mathbf{x},t) = f(\mathbf{x}) \ \forall \mathbf{x} \in \partial\Omega, \end{cases} \tag{4.58}$$

then, by (4.53), we obtain

$$E[\phi(\mathbf{u}(\tau_{\mathbf{x},s}), \tau_{\mathbf{x},s})] = \phi(\mathbf{x}, s), \tag{4.59}$$

which is *Kolmogorov's formula.*

Remark 4.53. In the homogeneous case, if $\phi(\mathbf{x})$ is the solution of the elliptic problem

$$\begin{cases} M[\phi] = 0 \text{ in } \Omega \\ \phi = f \qquad \text{on } \partial\Omega \end{cases}, \tag{4.60}$$

then equation (4.49) becomes

$$E[\phi(\mathbf{u}(\tau_{\mathbf{x},s}))] = \phi(\mathbf{x}), \qquad \mathbf{x} \in \Omega. \tag{4.61}$$

4.6 Stability of Stochastic Differential Equations

We consider the autonomous system of stochastic differential equations

$$d\mathbf{u}(t) = \mathbf{a}(\mathbf{u}(t))dt + b(\mathbf{u}(t))d\mathbf{W}(t) \tag{4.62}$$

and suppose that $b(\mathbf{x}) \neq \mathbf{0}$, for all $\mathbf{x} \in \bar{\Omega}$ (compact sets of $\mathbb{R}^m$). In this case the operator

$$M = \sum_{i=1}^{m} a_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{m} (bb')_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$$

is uniformly elliptic (see Appendix C) and the elliptic problem

$$\begin{cases} M[\phi] = -1 \text{ in } \Omega, \\ \phi = 0 \qquad \text{on } \partial\Omega \end{cases}$$

has a bounded solution. Therefore, by Dynkin's formula $E[\tau_{\mathbf{x}}] = \phi(\mathbf{x})$, it follows that $\tau_{\mathbf{x}} < +\infty$ almost surely and thus the system exits from Ω (to which $\mathbf{0}$ belongs) in a finite time with probability 1 (for all Ω bounded). Therefore, for any $\mathbf{b}$ the system is unstable, even if the deterministic system was asymptotically stable.

We now consider the case in which $\mathbf{0}$ is also an equilibrium for b. We let $\mathbf{a}(\mathbf{0}) = \mathbf{0}$, $b(\mathbf{0}) = \mathbf{0}$ and look for a suitable definition of stability in this case. Let

$$\begin{cases} d\mathbf{u}(t) = \mathbf{a}(t, \mathbf{u}(t))dt + b(t, \mathbf{u}(t))d\mathbf{W}(t), \ t > t_0, \\ \mathbf{u}(t_0) = \mathbf{c}, \end{cases} \tag{4.63}$$

be a system of stochastic differential equations, where

1. the conditions of the existence and uniqueness theorem are satisfied globally on $[t_0, +\infty]$;
2. $\mathbf{a}$ and b are continuous with respect to t;
3. $\mathbf{c} \in \mathbb{R}^m$ is a constant.

Then there exists a unique solution $\mathbf{u}(t, t_0, \mathbf{c})$, $t \in [t_0, +\infty[$, which is a Markov diffusion process with drift $\mathbf{a}$ and diffusion matrix bb'. With $\mathbf{c}$ being constant, the moments $E[|\mathbf{u}(t)|^k]$, $k > 0$, exist for every t. If we suppose that $\mathbf{a}(t, \mathbf{0}) = b(t, \mathbf{0}) = \mathbf{0}$ for all $t \geq t_0$ and let $v : [t_0, +\infty[\times\mathbb{R}^m \to \mathbb{R}$ be a positive definite function equipped with the continuous partial derivatives $\frac{\partial}{\partial t}$, $\frac{\partial}{\partial x_i}$, and $\frac{\partial^2}{\partial x_i \partial x_j}$, then we can apply Itô's formula to the process $V(t) = v(t, \mathbf{u}(t, t_0, \mathbf{c}))$, so that

$$dV(t) = L[v](t, \mathbf{u}(t))dt + \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\partial}{\partial x_i} v(t, \mathbf{u}(t)) b_{ij}(t, \mathbf{u}(t)) dW_j(t). \quad (4.64)$$

For the origin to be a stable point we require that for all t: $dV(t) \leq 0$ for every trajectory. But this is not possible due to the presence of the random term. At least we require that

$$E[dV(t)] \leq 0, \quad (4.65)$$

and hence

$$E[L[v](t, \mathbf{u}(t))dt] \leq 0. \quad (4.66)$$

If,

$$\forall t \geq t_0, \forall \mathbf{x} \in \mathbb{R}^m: \; L[v](t, \mathbf{x}) \leq 0, \quad (4.67)$$

then condition (4.66) is certainly satisfied. The functions $v(t, \mathbf{x})$ that satisfy (4.67) are the stochastic equivalents of Lyapunov functions. Integrating equation (4.64), we obtain

$$\begin{aligned} V(t) &= v(t_0, \mathbf{c}) + \int_{t_0}^{t} L[v](r, \mathbf{u}(r))dr \\ &\quad + \int_{t_0}^{t} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\partial}{\partial x_i} v(r, \mathbf{u}(r)) b_{ij}(r, \mathbf{u}(r)) dW_j(r), \\ V(s) &= v(t_0, \mathbf{c}) + \int_{t_0}^{s} L[v](r, \mathbf{u}(r))dr \\ &\quad + \int_{t_0}^{s} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\partial}{\partial x_i} v(r, \mathbf{u}(r)) b_{ij}(r, \mathbf{u}(r)) dW_j(r), \end{aligned}$$

and subtracting one from the other gives

$$V(t) - V(s) = \int_{s}^{t} L[v](r, \mathbf{u}(r))dr + \int_{s}^{t} \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\partial}{\partial x_i} v(r, \mathbf{u}(r)) b_{ij}(r, \mathbf{u}(r)) dW_j(r).$$

Putting

$$H(t) = \int_s^t \sum_{i=1}^m \sum_{j=1}^n \frac{\partial}{\partial x_i} v(r, \mathbf{u}(r)) b_{ij}(r, \mathbf{u}(r)) dW_j(r),$$

$$\mathcal{F}_t = \sigma\left(\mathcal{F}_s^{(i)}, 0 \leq s \leq t, i = 1, \ldots, n\right),$$

we obtain

$$E[V(t) - V(s)|\mathcal{F}_s] = E\left[\int_s^t L[v](r, \mathbf{u}(r))dr|\mathcal{F}_s\right] + E[H(t)|\mathcal{F}_s]. \tag{4.68}$$

It can be shown that $H(t)$ is a martingale with respect to $\{\mathcal{F}_t, t \in \mathbb{R}_+\}$ (the proof is equivalent as in the scalar case of Theorem 3.37). Therefore

$$E[H(t)|\mathcal{F}_s] = H(s) = 0.$$

Then (4.68) can be written as

$$E[V(t) - V(s)|\mathcal{F}_s] = E\left[\int_s^t L[v](r, \mathbf{u}(r))dr \mid \mathcal{F}_s\right],$$

and by (4.67)

$$E[V(t) - V(s)|\mathcal{F}_s] \leq 0.$$

Thus $V(t)$ is a supermartingale with respect to $\{\mathcal{F}_t, t \in \mathbb{R}_+\}$. By the supermartingale inequality

$$\forall [a, b] \subset [t_0, +\infty) : P\left(\sup_{a \leq t \leq b} v(t, \mathbf{u}(t)) \geq \epsilon\right) \leq \frac{1}{\epsilon} E[v(a, \mathbf{u}(a))]$$

and, for $a = t_0$, $\mathbf{u}(a) = \mathbf{c}$ (constant), $b \to +\infty$ we obtain

$$P\left(\sup_{t_0 \leq t \leq +\infty} v(t, \mathbf{u}(t)) \geq \epsilon\right) \leq \frac{1}{\epsilon} v(t_0, \mathbf{c}) \qquad \forall \epsilon > 0, \mathbf{c} \in \mathbb{R}^m.$$

If we suppose that $\lim_{\mathbf{c} \to \mathbf{0}} v(t_0, \mathbf{c}) = 0$, then

$$\lim_{\mathbf{c} \to \mathbf{0}} P\left(\sup_{t_0 \leq t \leq +\infty} v(t, \mathbf{u}(t)) \geq \epsilon\right) \leq \frac{1}{\epsilon} v(t_0, \mathbf{c}) = 0 \qquad \forall \epsilon > 0, \tag{4.69}$$

and hence, for all $\epsilon_1 > 0$, there exists a $\delta(\epsilon_1, t_0)$ such that

$$\forall |\mathbf{c}| < \delta : \ P\left(\sup_{t_0 \leq t \leq +\infty} v(t, \mathbf{u}(t)) \geq \epsilon\right) \leq \epsilon_1.$$

If we suppose that

$$|\mathbf{u}(t)| > \epsilon_2 \Rightarrow v(t, \mathbf{u}(t)) > \epsilon,$$

as, for example, if v is the Euclidean norm, then (4.69) can be written as

$$\lim_{\mathbf{c} \to \mathbf{0}} P\left(\sup_{t_0 \leq t \leq +\infty} u(t, t_0, \mathbf{c}) \geq \epsilon\right) = 0 \qquad \forall \epsilon > 0.$$

Definition 4.54. The point $\mathbf{0}$ is a *stochastically stable* equilibrium of (4.63) if

$$\lim_{\mathbf{c}\to\mathbf{0}} P\left(\sup_{t_0\le t\le+\infty} |u(t,t_0,\mathbf{c})| \ge \epsilon\right) = 0 \qquad \forall\epsilon > 0.$$

The point $\mathbf{0}$ is *asymptotically stochastically stable* if

$$\begin{cases} \mathbf{0} \text{ is stochastically stable,} \\ \lim_{\mathbf{c}\to\mathbf{0}} P(\lim_{t\to+\infty} u(t,t,\mathbf{c}) = \mathbf{0}) = 1. \end{cases}$$

The point $\mathbf{0}$ is *globally asymptotically stochastically stable* if

$$\begin{cases} \mathbf{0} \text{ is stochastically stable,} \\ P(\lim_{t\to+\infty} u(t,t_0,\mathbf{c}) = \mathbf{0}) = 1 \ \forall\mathbf{c}\in\mathbb{R}^m. \end{cases}$$

Theorem 4.55. *The following two statements can be shown to be true (see also Arnold (1974) and Schuss (1980)):*

1. *If* $L[v](t,\mathbf{x}) \le 0$*, for all* $t \ge t_0$*,* $\mathbf{x}\in B_h$ *(see Appendix D), then* $\mathbf{0}$ *is stochastically stable.*
2. *If* $v(t,\mathbf{x}) \le \omega(\mathbf{x})$ *for all* $t\ge t_0$*, with positive definite* $\omega(\mathbf{x})$ *and negative definite* $L[v]$*, then* $\mathbf{0}$ *is asymptotically stochastically stable.*

Example 4.56. Consider, for $a, b\in\mathbb{R}$, the one-dimensional linear equation

$$du(t) = au(t)dt + bu(t)dW(t), \tag{4.70}$$

subject to a given initial condition $u(0) = u_0$. We know that the solution is given by

$$u(t) = u_0 \exp\left\{\left(a - \frac{b^2}{2}\right)t + bW(t)\right\}.$$

By the strong law of large numbers (see Proposition 2.149)

$$\frac{W(t)}{t} \to 0 \text{ a.s.} \qquad \text{for } t\to+\infty,$$

and we have

- $u(t)\to 0$ almost surely, if $a - \frac{b^2}{2} < 0$,
- $u(t)\to+\infty$ almost surely, if $a - \frac{b^2}{2} > 0$.

If $a = \frac{b^2}{2}$, then

$$u(t) = u_0 \exp\{bW(t)\},$$

and therefore

$$P\left(\limsup_{t\to+\infty} u(t) = +\infty\right) = 1.$$

Let us now consider the function $v(x) = |x|^\alpha$ for some $\alpha\in\mathbb{R}_+$. Then

$$L[v](x) = \left(a + \frac{1}{2}b^2(\alpha - 1)\right)\alpha|x|^\alpha.$$

It is easily seen that, if $a - \frac{b^2}{2} < 0$, we can choose α such that $0 < \alpha < 1 - \frac{2a}{b^2}$ and obtain a Lyapunov function v with

$$L[v](x) \leq -kv(x)$$

for $k > 0$. This confirms the global asymptotic stability of 0 for the stochastic differential equation.

The result in the preceding example may be extended to the nonlinear case by local linearization techniques (see Gard (1988), page 139).

Theorem 4.57. *Consider the scalar stochastic differential equation*

$$du(t) = a(t, u(t))dt + b(t, u(t))dW(t), \tag{4.71}$$

where, in addition to the existence and uniqueness conditions, the functions a and b are such that two real constants a_0 and b_0 exist so that

$$a(t, x) = a_0 x + \bar{a}(t, x),$$
$$b(t, x) = b_0 x + \bar{b}(t, x),$$

for any $t \in \mathbb{R}_+$ and any $x \in \mathbb{R}$, with $\bar{a}(t, x) = o(x)$ and $\bar{b}(t, x) = o(x)$, uniformly in t. Then, if $a_0 - \frac{b_0^2}{2} < 0$, the equilibrium solution $u^{eq} \equiv 0$ of equation (4.57) is stochastically asymptotically stable.

Proof: Consider again the function

$$v(x) = |x|^\alpha$$

for some $\alpha > 0$. From Itô's formula we obtain

$$\begin{aligned}
&L[v](x)\\
&= \left(a_0 + \frac{\bar{a}(t,x)}{x} + \frac{1}{2}(\alpha - 1)\left(b_0 + \frac{\bar{b}(t,x)}{x}\right)^2\right)\alpha|x|^\alpha\\
&= \left(a_0 - \frac{b_0^2}{2} + \frac{\bar{a}(t,x)}{x} + \frac{1}{2}\alpha b_0^2 + (\alpha - 1)\left(b_0\frac{\bar{b}(t,x)}{x} + \frac{\bar{b}^2(t,x)}{2x^2}\right)\right)\alpha|x|^\alpha.
\end{aligned}$$

Choose $\alpha > 0$ and $r > 0$ sufficiently small so that for $x \in]-r, 0[\cup]0, r[$ we have

$$\left|\frac{\bar{a}(t,x)}{x}\right| + \frac{1}{2}\alpha b_0^2 + \left|(\alpha - 1)\left(b_0\frac{\bar{b}(t,x)}{x} + \frac{\bar{b}^2(t,x)}{2x^2}\right)\right| < \left|a_0 - \frac{b_0^2}{2}\right|.$$

We may then claim that a constant $k > 0$ exists such that

$$L[v](x) \leq -kv(x),$$

from which the required result follows. □

We now consider the autonomous multidimensional case, i.e., a stochastic differential equation in $\mathbb{R}^n$ of the form

$$d\mathbf{u}(t) = \mathbf{a}(\mathbf{u}(t))dt + b(\mathbf{u}(t))d\mathbf{W}(t). \tag{4.72}$$

The preceding results provide conditions for the asymptotic stability of $\mathbf{0}$ as an equilibrium solution. In particular, we obtain that, for a suitable initial condition $\mathbf{c} \in \mathbb{R}^n$, we have

$$\lim_{t\to+\infty} \mathbf{u}(t,0,\mathbf{c}) = \mathbf{0}, \qquad \text{a.s.}$$

We may notice that almost sure convergence implies convergence in law of $\mathbf{u}(t,0,\mathbf{c})$ to the degenerate random variable $\mathbf{u}^{eq} \equiv \mathbf{0}$, i.e., the convergence of the transition probability to a degenerate invariant distribution with density $\delta_0(\mathbf{x})$, the standard Dirac delta function:

$$P(t,\mathbf{x},\mathbf{B}) \to \int_{\mathbf{B}} \delta_0(\mathbf{x})d\mathbf{x} \text{ for any } \mathbf{B} \in \mathcal{B}_{\mathbb{R}^n}.$$

If (4.72) does not have an equilibrium, we may still investigate the possibility that an asymptotically invariant (but not necessarily degenerate) distribution exists for the solution of the stochastic differential equation; still in terms of a Lyapunov function. The following theorem (see Gard (1988)) provides an answer, which is from an analysis of Has'minskii (1980).

Theorem 4.58. *Consider a stochastic differential equation in* $\mathbb{R}^n$:

$$d\mathbf{u}(t) = \mathbf{a}(t,\mathbf{u}(t))dt + b(t,\mathbf{u}(t))d\mathbf{W}(t), \tag{4.73}$$

where $\mathbf{W}(t)$ *is an* m*-dimensional vector of independent Wiener processes. Let* D *and* $(D_n)_{n\in\mathbb{N}}$ *be open sets in* $\mathbb{R}^n$ *such that*

$$D_n \subset D_{n+1}, \bar{D}_n \subset D, D = \bigcup_n D_n,$$

and suppose $\mathbf{a}$ *and* b *satisfy the conditions of existence and uniqueness for equation (4.72), on each set* $\{(t,x)\} \in \{[t_0,+\infty[\times D_n\}$ *for some* $t_0 \in \mathbb{R}_+$*. Suppose further that a nonnegative function* $v \in C^{1,2}([t_0,+\infty[\times D)$ *exists with*

$$\lim_{n\to\infty} \inf_{\substack{t>t_0 \\ x\in D\setminus D_n}} v(t,x) = +\infty.$$

Then, for any initial condition $\mathbf{c}$ *independent of* $\mathbf{W}$*, such that* $P(\mathbf{c} \in D) = 1$*, there is a unique solution* $\mathbf{u}(t)$ *of equation (4.73), subject to* $\mathbf{u}(t_0) = \mathbf{c}$*, so that* $\mathbf{u}(t) \in D$ *almost surely for all* $t > t_0$*. Thus*

$$P(\tau_D = +\infty) = 1,$$

where τ_D *is the first exit time of* $\mathbf{u}(t,t_0,\mathbf{c})$ *from* D*.*

For autonomous systems

$$d\mathbf{u}(t) = \mathbf{a}(\mathbf{u}(t))dt + b(\mathbf{u}(t))d\mathbf{W}(t),$$

we have the following theorem.

Theorem 4.59. *Given the same assumptions as in Theorem 4.58, suppose further that* $n_0 \in \mathbb{N}$ *and* $M, k \in \mathbb{R}_+ \setminus \{0\}$ *exist, such that*

1. $\sum_{i,j=1}^{n} \left(\sum_{k=1}^{m} b_{ik}(\mathbf{x}) b_{kj}(\mathbf{x})\right) \xi_i \xi_j \geq M|\xi|^2$ *for all* $\mathbf{x} \in \bar{D}_{n_0}, \xi \in \mathbb{R}^n$;
2. $L[v](\mathbf{x}) \leq -k$ *for all* $\mathbf{x} \in D \setminus \bar{D}_{n_0}$.

Then there exists an invariant distribution $\tilde{P}$ *with nowhere-zero density in* D*, such that for any* $\mathbf{B} \in \mathcal{B}_{\mathbb{R}^n}, \mathbf{B} \subset D$:

$$P(t, \mathbf{x}, \mathbf{B}) \to \tilde{P}(B) \text{ as } t \to +\infty,$$

where $P(t, \mathbf{x}, \mathbf{B})$ *is the transition probability* $P(t, \mathbf{x}, \mathbf{B}) = P(\mathbf{u}(t, \mathbf{x}) \in \mathbf{B})$ *for the solution of the given stochastic differential equation.*

Application: A Stochastic Food Chain

As a foretaste of the next part on applications of stochastic processes we take an example from Gard (1988), page 177. Consider the system

$$du_1 = u_1[(a_1 + \sigma_1 dW_1) - b_{11}u_1 - b_{12}u_2]dt, \tag{4.74}$$

$$du_2 = u_2[(-a_2 + \sigma_2 dW_2) + b_{21}u_1 - b_{22}u_2 - b_{23}u_3]dt, \tag{4.75}$$

$$du_3 = u_3[(-a_3 + \sigma_3 dW_3) + b_{32}u_2 - b_{33}u_3]dt \tag{4.76}$$

subject to suitable initial conditions. This system represents a food chain in which the three species' growth rates exhibit independent Wiener noises with scaling parameters $\sigma_i > 0$, $i = 1, 2, 3$, respectively. If we assume that all the parameters a_i and b_{ij} are strictly positive and constant for any $i, j = 1, 2, 3$, it can be shown that, in the absence of noise, the corresponding deterministic system admits, in addition to the trivial one, a unique nontrivial feasible equilibrium $\mathbf{x}^{eq} \in \mathbb{R}^3_+$. This one is globally asymptotically stable in the so-called feasible region $\mathbb{R}^3_+ \setminus \{\mathbf{0}\}$, provided that the parameters satisfy the inequality

$$a_1 - \left(\frac{b_{11}}{b_{21}}\right) a_2 - \left(\frac{b_{11}b_{22} + b_{12}b_{21}}{b_{21}b_{32}}\right) a_3 > 0.$$

This result is obtained through the Lyapunov function

$$v(\mathbf{x}) = \sum_{i=1}^{n} c_i \left(x_i - x_i^{eq} - x_i^{eq} \ln \frac{x_i}{x_i^{eq}}\right),$$

provided that the $c_i > 0$, $i = 1, 2, 3$, are chosen to satisfy

$$c_1 b_{12} - c_2 b_{21} = 0 = c_2 b_{23} - c_3 b_{32}.$$

In fact, if one denotes by B the interaction matrix $(b_{ij})_{1 \leq i,j \leq 3}$ and $C = \mathrm{diag}(c_1, c_2, c_3)$, the matrix

$$CB + B'C = -2 \begin{pmatrix} c_1 b_{11} & 0 & 0 \\ 0 & c_2 b_{22} & 0 \\ 0 & 0 & c_3 b_{33} \end{pmatrix}$$

is negative definite. The derivative of v along a trajectory of the deterministic system is given by

$$\dot{v}(\mathbf{x}) = \frac{1}{2} \left(\mathbf{x} - \mathbf{x}^{eq}\right) \cdot \left[CB + B'C\right] \left(\mathbf{x} - \mathbf{x}^{eq}\right),$$

which is then negative definite, thus implying the global asymptotic stability of $\mathbf{x}^{eq} \in \mathbb{R}^3_+$.

Returning to the stochastic system, consider the same Lyapunov function as for the deterministic case. By means of Itô's formula we obtain

$$L[v](\mathbf{x}) = \frac{1}{2} \left(\left(\mathbf{x} - \mathbf{x}^{eq}\right) \cdot \left[CB + B'C\right] \left(\mathbf{x} - \mathbf{x}^{eq}\right) + \sum_{i=1}^{3} c_i \sigma_i^2 x_i^{eq} \right).$$

It can now be shown that, if the σ_i, $i = 1, 2, 3$, satisfy

$$\sum_{i=1}^{3} c_i \sigma_i^2 x_i < 2 \min_i \left\{ c_i b_{ii} x_i^{eq} \right\},$$

then the ellipsoid

$$\left(\mathbf{x} - \mathbf{x}^{eq}\right) \cdot \left[CB + B'C\right] \left(\mathbf{x} - \mathbf{x}^{eq}\right) + \sum_{i=1}^{3} c_i \sigma_i^2 x_i^{eq} = 0$$

lies entirely in $\mathbb{R}^3_+$. One can then take as D_{n_0} any neighborhood of the ellipsoid such that $\bar{D}_{n_0} \subset \mathbb{R}^3_+$ and the conditions of Theorem 4.58 are met. As a consequence the stochastic system (4.74)–(4.76) admits an invariant distribution with nowhere-zero density in $\mathbb{R}^3_+$. An additional interesting application to stochastic population dynamics can be found in Roozen (1987).

4.7 Exercises and Additions

4.1. Prove Remark 4.7.

4.2. Prove Remark 4.10.

4.3. Prove that if $a(t,x)$ and $b(t,x)$ are measurable functions in $[0,T]\times\mathbb{R}$ that satisfy conditions 1 and 2 of Theorem 4.4, then, for all $s \in]0,T]$, there exists a unique solution in $\mathcal{C}([s,T])$ of

$$\begin{cases} u(s) = u_s \quad \text{a.s.}, \\ du(t) = a(t,u(t))dt + b(t,u(t))dW_t, \end{cases}$$

provided that the random variable u_s is independent of $\mathcal{F}_{s,T} = \sigma(W_t - W_s, t \in [s,T])$ and $E[(u_s)^2] < \infty$.

4.4. Complete the proof of Theorem 4.13 by proving the *semigroup property*: If $t_0 < s$, $s \in [0,T]$, denote by $u(t,s,x)$ the solution of

$$\begin{cases} u(s) = x \text{ a.s.}, \\ du(t) = a(t,u(t))dt + b(t,u(t))dW_t. \end{cases}$$

Then

$$u(t,t_0,c) = u(t,s,u(s,t_0,c)) \text{ for } t \geq s,$$

where x is a fixed real number and c is a random variable.

4.5. Complete the proof of Theorem 4.31 (Girsanov) showing that $(Y_t^2 - t)_{t\in[0,T]}$ is a martingale where

$$Y_t = W_t - \int_0^t \vartheta_s ds,$$

$(W_t)_{t\in[0,T]}$ is a Brownian motion, and $(\vartheta_t)_{t\in[0,T]}$ satisfies the Novikov condition.

4.6. Show that

$$\Gamma_0^*(x,s;y,t) = \int_{\mathbb{R}} \Gamma_0^*(x,s;z,r)\Gamma_0^*(z,r;y,t)dz \qquad (s < r < t). \tag{4.77}$$

Expression (4.77) is in general true for the fundamental solution $\Gamma(x,t;\xi,r)$ $(r < t)$ constructed in Theorem C.19.

4.7. Let $(W_t)_{t\in\mathbb{R}_+}$ be a Brownian motion. Consider the population growth model:

$$\frac{dN_t}{dt} = (r_t + \alpha W_t)N_t, \tag{4.78}$$

where N_t is the size of population at time t ($N_0 > 0$ given) and $(r_t + \alpha \cdot W_t)$ is the relative rate of growth at time t. Suppose the process $r_t = r$ is constant.

1. Solve the stochastic differential equation (4.78).
2. Estimate the limit behavior of N_t when $t \to \infty$.
3. Show that if W_t is independent of N_0, then

$$E[N_t] = E[N_0]e^{rt}.$$

An extension model of (4.78) for exponential growth with several independent white noise sources in the relative growth rate is given as follows: Let $(W_1(t), \ldots, W_n(t))_{t\in\mathbb{R}_+}$ be Brownian motion in $\mathbb{R}^d$, with $\alpha_1, \ldots, \alpha_n$ constants. Then

$$dN_t = \left(rdt + \sum_{k=1}^{n} \alpha_k dW_k(t) \right) N_t, \tag{4.79}$$

where N_t is, again, the size of population at time t with $N_0 > 0$ is given.

4. Solve the stochastic differential equation (4.79).

4.8. Let $(W_t)_{t\in\mathbb{R}_+}$ be a one-dimensional Brownian motion. Show that the process (*Brownian motion on the unit circle*)

$$u_t = (\cos W_t, \sin W_t)$$

is the solution of the stochastic differential equations (in matrix notation)

$$du_t = -\frac{1}{2} u_t dt + K u_t dW_t, \tag{4.80}$$

where $K = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

More generally, show that the process (*Brownian motion on the ellipse*)

$$u_t = (a \cos W_t, b \sin W_t)$$

is a solution of (4.80), where $K = \begin{bmatrix} 0 & -\frac{a}{b} \\ \frac{b}{a} & 0 \end{bmatrix}$.

4.9. (*Brownian bridge*). For fixed $a, b \in \mathbb{R}$ consider the one-dimensional equation:

$$\begin{cases} u(0) = a, \\ du_t = \dfrac{b - u_t}{1 - t} dt - dW_t \qquad (0 \le t < 1). \end{cases}$$

Verify that

$$u_t = a(1-t) + bt + (1-t) \int_0^t \frac{dW_s}{1-s} \qquad (0 \le t < 1)$$

solves the equation and prove that $\lim_{t\to 1} u_t = b$ almost surely. The process $(u_t)_{t\in[0,1[}$ is called the *Brownian bridge* (from a to b).

4.10. Solve the following stochastic differential equations:

1. $\begin{bmatrix} du_1 \\ du_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} dt + \begin{bmatrix} 1 & 0 \\ 0 & u_1 \end{bmatrix} \begin{bmatrix} dW_1 \\ dW_2 \end{bmatrix}$.
2. $du_t = u_t dt + dW_t$. (*Hint:* Multiply both sides with e^{-t} and compare with $d(e^{-t}u_t)$.)

3. $du_t = -u_t dt + e^{-t} dW_t$.

4.11. Consider n-dimensional Brownian motion $\mathbf{W} = (W_1, \ldots, W_n)$ starting at $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ $(n \geq 2)$ and assume $|\mathbf{a}| < R$. What is the expected value of the first exit time τ_K of B from the ball

$$K = K_R = \{\mathbf{x} \in \mathbb{R}^n; |\mathbf{x}| < R\}?$$

(*Hint:* Use Dynkin's formula.)

4.12. Find the generators of the following processes:

1. Brownian motion on an ellipse (see problem 4.8).
2. Arithmetic Brownian motion:

$$\begin{cases} u(0) = u_0, \\ du(t) = a dt + b dW_t. \end{cases}$$

3. Geometric Brownian motion:

$$\begin{cases} u(0) = u_0, \\ du(t) = au(t) dt + bu(t) dW_t. \end{cases}$$

4. (Mean-reverting) Ornstein–Uhlenbeck process:

$$\begin{cases} u(0) = u_0 \\ du(t) = (a - bu(t)) dt + c dW_t. \end{cases}$$

4.13. Find a process $(u_t)_{t \in \mathbb{R}_+}$ whose generator is the following:

1. $\mathcal{A}f(x) = f'(x) + f''(x)$, where $f \in BC(\mathbb{R}) \cap C^2(\mathbb{R})$;
2. $\mathcal{A}f(t,x) = \frac{\partial f}{\partial t} + cx\frac{\partial f}{\partial x} + \frac{1}{2}\alpha^2 x^2 \frac{\partial^2 f}{\partial x^2}$, where $f \in BC(\mathbb{R}^2) \cap C^2(\mathbb{R}^2)$ and c, α are constants.

4.14. Let $\triangle$ denote the Laplace operator on $\mathbb{R}^n$, $\phi \in BC(\mathbb{R}^n)$ and $\alpha > 0$. Find a solution $(u_t)_{t \in \mathbb{R}_+}$ of the equation

$$\left(\alpha - \frac{1}{2}\triangle\right) u = \phi \qquad \text{in } \mathbb{R}^n.$$

Is the solution unique?

4.15. Consider a linear stochastic differential equation

$$du(t) = [a(t) + b(t)u(t)]dt + [c(t) + d(t)u(t)]dW(t), \tag{4.81}$$

where the functions a, b, c, d are bounded and measurable. Prove:

1. If $a \equiv c \equiv 0$, then the solution $u(t) = u_0(t)$ is given by

$$u_0(t) = u_0(0) \exp\left\{\int_0^t \left[b(s) - \frac{1}{2}d^2(s)\right] ds + \int_0^t d(s) dW_s\right\}.$$

2. Setting $u(t) = u_0(t)v(t)$, show that $u(t)$ is a solution of (4.81) if and only if
$$v(t) = v(0) + \int_0^t [u_0(s)a(s) - c(s)d(s)]ds + \int_0^t c(s)u_0(s)ds.$$
Thus the solution of (4.81) is $u_0(t)v(t)$ with $u(0) = u_0(0)v(0)$.

4.16. Consider a diffusion process X associated with a stochastic differential equation with drift $\mu(x,t)$ and diffusion coefficient $\sigma^2(x,t)$. Show that for any $\theta \in \mathbb{R}$ the process

$$Y_\theta(t) = \exp\left\{\theta X(t) - \theta \int_0^t \mu(X(s),s)ds - \frac{1}{2}\int_0^t \sigma^2(X(s),s)ds\right\},\ t \in \mathbb{R}_+,$$

is a martingale.

4.17. Consider a diffusion process X associated with a stochastic differential equation with drift $\mu(x,t) = \alpha t$ and diffusion coefficient $\sigma^2(x,t) = \beta t$, with $\alpha \geq 0$ and $\beta > 0$. Let T_a be the first passage time to the level $a \in \mathbb{R}$; evaluate

$$E\left[e^{-\lambda T_a^2}\middle| X(0) = 0\right] \text{ for } \lambda > 0.$$

(*Hint:* Use the result of problem 4.16)

4.18. Let $u(t),\ t \in \mathbb{R}_+$, be the solution of the stochastic differential equation

$$du(t) = a(u(t))dt + \sigma(u(t))dW(t)$$

subject to the initial condition

$$u(0) = u_0 > 0.$$

Provided that $a(0) = \sigma(0) = 0$, show that, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that
$$P_{u_0}\left(\lim_{t\to+\infty} u(t) = 0\right) \geq 1 - \varepsilon$$
whenever $0 < u_0 < \delta$ if and only if

$$\int_0^\delta \exp\left\{\int_0^y \frac{2a(x)}{\sigma^2(x)}\right\} dy < \infty.$$

Further, if $\sigma(x) = \sigma_0 x + o(x)$, and similarly $a(x) = a_0 x + o(x)$, the stability condition is
$$\frac{a_0}{\sigma_0^2} < \frac{1}{2}.$$

4.19. Let X be a diffusion process associated with a stochastic differential equation with drift $\mu(x,t) = -\alpha x$ and constant diffusion coefficient $\sigma^2(x,t) = \beta$, with $\alpha \in \mathbb{R}_+^*$ and $\beta \in \mathbb{R}$. Show that the moments $q_r(t) = E[X(t)^r]$, $r = 1, 2, \ldots$ of $X(t)$ satisfy the system of ordinary differential equations

$$\frac{d}{dt} q_r(t) = -\alpha r q_r(t) + \frac{\beta^2 r(r-1)}{2} q_{r-2}(t), \quad r = 1, 2, \ldots$$

with the assumption $q_r(t) = 0$ for any integer $r \leq -1$.

4.20. Let X be the diffusion process defined in problem 4.18. Show that the characteristic function of $X(t)$, defined as $\varphi(v;t) = E[\exp\{ivX(t)\}], v \in \mathbb{R}$, satisfies the partial differential equation

$$\frac{\partial}{\partial t}\varphi(v;t) = -\alpha v \frac{\partial}{\partial v}\varphi(v;t) - \frac{1}{2}\beta^2 v^2 \varphi(v;t).$$

4.21. Let $u(t)$ be the solution of the stochastic differential equation

$$du(t) = a(u(t))dt + b(u(t))dW(t)$$

subject to an initial condition

$$u(0) = u_0 \in (\alpha, \beta) \subset \mathbb{R}.$$

Show that the mean $\mu_T(u_0)$ of the first exit time

$$T = \inf\{t \geq 0 \mid u(t) \notin (\alpha, \beta)\}$$

is the solution of the ordinary differential equation

$$-1 = a(u_0)\frac{d\mu_T}{du_0} + \frac{1}{2}b(u_0)^2 \frac{d^2\mu_T}{du_0^2}$$

subject to the boundary conditions

$$\mu_T(\alpha) = \mu_T(\beta) = 0.$$

Part II

The Applications of Stochastic Processes

5

Applications to Finance and Insurance

Mathematical finance is one of the most influential driving forces behind the research into stochastic processes. This is due to the fact that a significant part of the world's financial markets relies on stochastic models as the underlying basis for valuation and risk management. But, perhaps more surprisingly, the financial market was also one of the main drivers that led to their discovery.

As early as 1900, Louis Bachelier, a young doctorate researcher, analyzed financial contracts, also called *financial derivatives*, traded on the Paris bourse and in his thesis (Bachelier (1900)) attempted to lay down a mathematical foundation for their valuation. This was some years before Einstein, in the context of physics, discovered Brownian motion, later formalized by Wiener, which in turn led to the development of Itô theory in the 1950's, representing the interface of classical and stochastic mathematics. All these then came to prominence through Robert Merton's (1973) as well as Black and Scholes' (1973) derivation of their partial differential equation and formula for the pricing of financial option contracts. These represented direct applications of the then already known backward Kolmogorov equation and Kac–Feynman formula. It serves as the most widely used basic model of mathematical finance.

In his work, Bachelier concluded that prices of assets traded on the exchange are random and represent market-clearing equilibria under which there are equal numbers of buyers and sellers, whose riskless profit expectations must be zero. The latter foreshadowed the economic concept of no-arbitrage and, mathematically related, martingales. Both are fundamental building blocks of all financial modeling involving stochastic processes, as was shown by Harrison and Kreps (1979) and Harrison and Pliska (1981).

Many books on mathematical finance commence in describing discrete-time stochastic models before deriving the continuous-time equivalent. However, in line with all the preceding chapters on the theory of stochastic processes, we will only focus on continuous-time (and space) models. Discrete-time models in practice serve, primarily, for numerical solutions of continuous processes, but also for an intuitive introduction to the topic. We refer to

Wilmott, Howison, and Dewynne (1993) for the former and Pliska (1997) as well as Cox, Ross and Rubinstein (1979) for the latter.

5.1 Arbitrage-Free Markets

In economic theory the usual definition of a market is a physical or conceptual place where supply meets demand for goods or services and they are exchanged in certain ratios. These exchange ratios are typically formulated in terms of a base monetary measuring unit, namely a currency, and called prices. This motivates the following definition of a market for the purpose of (continuous-time) stochastic modeling.

Definition 5.1. A filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$ endowed with adapted stochastic processes $(S_t^{(i)})_{t\in[0,T]}$, $i = 0, \ldots, n$, representing *asset prices* in terms of particular currencies, is called a *market*.

The asset prices are usually considered stochastic, because they do change over time and often unpredictably so, depending on whether supply outweighs demand or vice versa.

Remark 5.2. The risky assets $(S_t^{(i)})_{t\in[0,T]}$, $i = 1, \ldots, n$, are RCLL stochastic processes, thus their future values are not predictable.

Nonetheless, it is often convenient to consider the concept of a riskless asset.

Remark 5.3. If we define, say, $S_t^{(0)} := B_t$ as the riskless asset, then $(B_t)_{t\in[0,T]}$ is a deterministic, and thus predictable process.

Furthermore, in a market it is possible to exchange assets. This is represented by defining holding and portfolio processes.

Definition 5.4. A *holding process* $\mathbf{H}_t = (H_t^{(0)}, H_t^{(1)}, \ldots, H_t^{(n)})$, which is adapted and predictable to the filtration $(\mathcal{F}_t)_{t\in[0,T]}$, together with the asset processes generates the *portfolio process*

$$\Pi_t = \mathbf{H}_t \cdot \left(B_t, S_t^{(1)}, \ldots, S_t^{(n)}\right)',$$

where $(\Pi_t)_{t\in[0,T]}$ is also adapted to $(\mathcal{F}_t)_{t\in[0,T]}$.

Note that the drivers of the asset and holding processes are different. The former are exogenously driven by aggregate supply and demand in the market, whereas the latter are controlled by a particular market participant. Usually, but not always, the latter is considered to have no influence on the former. The respective underlying random variables also have different dimensions, S_t are stated in prices per unit, and H_t represent number of units. It is also often important to distinguish the following two cases.

Definition 5.5. If $T < \infty$, then the market has a *finite horizon*. Otherwise, if $T = +\infty$, then the market is said to have an *infinite horizon*.

But the definition of a market and its properties, so far, are insufficient to guarantee that the mathematical model is a realistic one in terms of economics. For this purpose, conditions have to be imposed on the processes constituting the market.

Proposition 5.6. *A realistic mathematical model of a market has to satisfy the following conditions.*

1. *(Conservation of funds and nonexplosive portfolios). For every $T \geq 0$ the holding process $\mathbf{H}_t$ has to satisfy:*
$$\Pi_T = \Pi_0 + \int_0^T H_t^{(0)} dB_t + \sum_{i=1}^n \int_0^T H_t^{(i)} dS_t^{(i)}, \qquad (5.1)$$
along with the nonexplosion condition
$$\int_0^T d\Pi_t < \infty \qquad a.s.$$
The conservation of funds condition is also called the self-financing portfolios *property.*
2. *(Nonarbitrage). A deflated portfolio process $(\Pi_t^*)_{t\in[0,T]}$ with almost surely $\Pi_0^* = 0$ and $\Pi_T^* > 0$ or, equivalently, with almost surely $\Pi_0^* < 0$ and $\Pi_T^* \geq 0$ is inadmissible. Here $\Pi_t^* = \Pi_t / S_t^{(j)}$ for any arbitrary numeraire asset j.*
3. *(Trading and/or credit limits). Either $(\mathbf{H}_t)_{t\in[0,T]}$ is square-integrable and of bounded variance or $\Pi_t \geq c$ for all t, with $-\infty < c \leq 0$ constant and arbitrary.*

Here condition 1 is obvious, as (in theory, at least) no money may suddenly disappear and neither can there be infinite wealth. For condition 3 there is a standard example (see Exercise 5.4) demonstrating that in continuous time there exist arbitrage opportunities, if it is not satisfied. Finally, condition 2 is also obvious, in the sense that if an investor could create riskless wealth above the return of the riskless asset (in economic language: "a free lunch"), which could potentially lead to infinite profits. Hence the model would be ill posed. To avoid this, the first fundamental theorem of asset pricing has to be satisfied.

Theorem 5.7. (First fundamental theorem of asset pricing). *If there exists an equivalent martingale (probability) measure $Q \sim P$ (see Definition A.52) for any arbitrary deflated portfolio process $(\Pi_t^*)_{t\in[0,T]}$ in a particular market, namely*
$$\Pi_0^* = E_P\left[\Pi_t^* \Lambda_t\right] = E_Q\left[\Pi_t^*\right] \qquad \forall t \in [0,T],$$

where Λ_t is the Radon–Nikodym derivative (see Remark 4.32)

$$\frac{dQ}{dP} = \Lambda_t \qquad on\ \mathcal{F}_t.$$

Then the market is free of arbitrage opportunities, if the assumptions of Girsanov's Theorem 4.31 are satisfied.

Proof: As the proof is quite lengthy and involved in the general continuous-time case, we refer to Delbaen and Schachermayer (1994). □

Theorem 5.8. (Second fundamental theorem of asset pricing). *If there exists a unique equivalent martingale (probability) measure $Q \sim P$ for any arbitrary deflated portfolio process $(\Pi_t^*)_{t\in[0,T]}$ in a particular market, then the market is complete.*

Proof: This proof is also very involved for the general continuous-time case, and we refer to Cherny and Shiryaev (2001). □

We attempt to make the significance of the two fundamental theorems more intuitive and thereby demonstrate the duality between the concepts of nonarbitrage and the existence of a martingale measure. Assume a particular portfolio in an arbitrage-free market has value $\tilde{\Pi}_T(\omega)$ for each $\omega \in \mathcal{F}_T$. If another portfolio $\hat{\Pi}_0$ can be created so that a self-financing trading strategy $(\hat{\mathbf{H}}_t)_{t\in[0,T]}$ exists that replicates $\tilde{\Pi}_T(\omega)$, namely

$$\hat{\Pi}_T(\omega) = \hat{\mathbf{H}}_0 \cdot \mathbf{S}_0 + \sum_{i=0}^{n} \int_0^T H_t^{(i)} dS_t^{(i)} \geq \tilde{\Pi}_T(\omega) \qquad \forall \omega \in \mathcal{F}_T,$$

then, necessarily,

$$\hat{\Pi}_t \geq \tilde{\Pi}_t \qquad \forall t \in [0,T], \tag{5.2}$$

and, in particular

$$\hat{\Pi}_0 \geq \tilde{\Pi}_0.$$

Otherwise there exists an arbitrage opportunity by buying the cheaper portfolio and selling the overvalued one. In fact, by this argumentation, the value of $\tilde{\Pi}_0$ has to be the solution of the constrained optimization problem

$$\tilde{\Pi}_0 = \min_{(H_t)_{t\in[0,T]}} \hat{\Pi}_0$$

subject to the value conservation condition (5.1) and the (super)replication condition (5.2). Hence if we can find an equivalent measure Q under which

$$E_Q\left[\sum_{i=0}^{n} \int_0^T H_t^{(i)} dS_t^{(i)}\right] \leq 0$$

then the value of the replicated portfolio has to satisfy

$$\tilde{\Pi}_0 = \max_Q E_Q\left[\tilde{\Pi}_T\right],$$

subject to, again, the value-conservation condition and the (super)martingale condition

$$E_Q\left[\hat{\Pi}_t\right] \geq \hat{\Pi}_0.$$

The latter can be considered as the so-called dual formulation of the replication problem. By the second fundamental theorem of asset pricing, if Q is unique, then all inequalities turn to equalities and

$$\tilde{\Pi}_0 = E_Q\left[\tilde{\Pi}_T\right] = \hat{\Pi}_0. \tag{5.3}$$

This result states that the nonarbitrage value of an arbitrary portfolio in an arbitrage-free and complete market is its expectation under the unique equivalent martingale measure.

Here we have implicitly assumed that the values of the portfolios are stated in terms of a numeraire of value 1. Generally, a *numeraire* asset or *deflator* serves as a meain whose units all other assets are stated. The following theorem states that numeraires can be changed.

Theorem 5.9. (Numeraire invariance theorem). *A self-financing holding strategy* $(\mathbf{H}_t)_{t\in[0,T]}$ *remains self-financing under a change of almost surely positive numeraire asset; i.e., if*

$$\frac{\Pi_T}{S_T^{(i)}} = \frac{\Pi_0}{S_T^{(i)}} + \int_0^T d\left(\frac{\Pi_t}{S_t^{(i)}}\right),$$

then

$$\frac{\Pi_T}{S_T^{(j)}} = \frac{\Pi_0}{S_T^{(j)}} + \int_0^T d\left(\frac{\Pi_t}{S_t^{(j)}}\right),$$

with $i \neq j$*, provided* $\int_0^T d\Pi_t < \infty$.

Proof: We arbitrarily choose $S_t^{(i)} = 1$ for all $t \in [0, T]$, and for notational simplicity write $S_t^{(j)} \equiv S_t$. Now it suffices to show that if

$$\Pi_T = \Pi_0 + \int_0^T d\Pi_t = \Pi_0 + \int_0^T \mathbf{H}_t \cdot d\mathbf{S}_t, \tag{5.4}$$

this implies

$$\frac{\Pi_T}{S_T} = \frac{\Pi_0}{S_0} + \int_0^T \mathbf{H}_t \cdot d\left(\frac{\mathbf{S}_t}{S_t}\right). \tag{5.5}$$

Taking the differential and substituting (5.4):

$$d\left(\frac{\Pi_t}{S_t}\right) = \frac{d\Pi_t}{S_t} + \Pi_t d\left(\frac{1}{S_t}\right) + d\Pi_t d\left(\frac{1}{S_t}\right)$$
$$= \mathbf{H}_t \cdot \left(\left(\frac{d\mathbf{S}_t}{S_t}\right) + \mathbf{S}_t d\left(\frac{1}{S_t}\right) + d\mathbf{S}_t d\left(\frac{1}{S_t}\right)\right),$$

after integration gives (5.5). □

In fact, in an arbitrage-free complete market, for every choice of numeraire there will be a distinct equivalent martingale measure. As we will demonstrate, the change of numeraire may be a convenient valuation technique of portfolios.

5.2 The Standard Black–Scholes Model

The Black–Scholes–Merton market has a particularly nice and simple, yet very intuitive form. It consists of a *riskless account* process $(B_t)_{t\in[0,T]}$, following

$$\frac{dB_t}{B_t} = rdt,$$

with constant *instantaneous riskless interest rate* r, so that

$$B_t = B_0 e^{rt}$$

and typically $B_0 \equiv 1$. Here r describes the instantaneous time value of money, namely how much relative wealth can be earned when saved over an infinitesimal instance $t + dt$, or, conversely, how it is discounted if received in the future.

Furthermore there exists a *risky asset* process $(S_t)_{t\in[0,T]}$, following geometric Brownian motion (see Example 4.9)

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t$$

with a constant drift μ and a constant volatility σ scaling a Wiener process dW_t, resulting in

$$S_T = S_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)T + \sigma W_T\right\}.$$

Both assets are adapted to the filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$. The market has a finite horizon and is free of arbitrage as well as complete. To demonstrate this we take B_t as the numeraire asset and attempt to find an equivalent measure Q, for which the discounted process

$$S_t^* := \frac{S_t}{B_t} \tag{5.6}$$

is a local martingale. Invoking Itô's formula gives

$$dS_t^* = S_t^*((\mu - r)dt + \sigma dW_t), \tag{5.7}$$

which, by Girsanov's Theorem 4.31, shows that

$$W_t^Q = W_t + \frac{\mu - r}{\sigma} t$$

turns (5.6) into a martingale, namely

$$S_0^* = E_Q\left[S_t^*\right],$$

under the equivalent measure Q, given by

$$\frac{dQ}{dP} = \exp\left\{-\frac{\mu - r}{\sigma} W_T - \left(\frac{\mu - r}{\sigma}\right)^2 \frac{T}{2}\right\} \qquad \text{on } \mathcal{F}_T.$$

Now by the numeraire invariance theorem this means that there will be unique martingale measures for all possible deflated portfolios, and hence there is no arbitrage in the Black–Scholes model and it is complete. This now allows us to price arbitrary replicable portfolios according to formula (5.3). But going back to the primal replication problem, we can derive the Black–Scholes partial differential equation from the conservation-of-funds condition (5.1). Explicitly, the replication constraints for a particular portfolio

$$V_t := \Pi_t$$

in the Black–Scholes model are

$$\begin{aligned} V_t &= V_0 + \int_0^t H_s^{(S)} dS_s + \int_0^t H_s^{(B)} dB_s \\ &= H_t^{(S)} S_t + H_t^{(B)} B_t \end{aligned} \tag{5.8}$$

subject to the sufficient nonexplosion condition

$$\int_0^t \left|H_s^{(B)}\right| ds + \int_0^t \left|H_s^{(S)}\right|^2 ds < \infty \qquad \text{a.s.}$$

Now, invoking Itô's formula, we obtain

$$dV_t = \frac{\partial V}{\partial t} + \mu S_t \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S^2} \tag{5.9}$$

on the left side of equation (5.8) and

$$dV_t = H_t^{(S)} dS_t + H_t^{(B)} dB_t \tag{5.10}$$

on the right. Now, by equating (5.9) and (5.10) as well as choosing

$$H_t = \frac{\partial V}{\partial S} \qquad \text{and} \qquad \hat{H}_t = V_t - \frac{\partial V}{\partial S} S_t,$$

the *hedging strategy* is entirely risk free, as the Wiener process has cancelled. Rearranging results in the so-called Black–Scholes equation

$$\mathcal{L}_{BS}V_t := \frac{\partial V}{\partial S} + \frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 V}{\partial S^2} + rS_t \frac{\partial V}{\partial S} - rV_t = 0. \tag{5.11}$$

First, it is notable that the physical drift is not present. This is given by the logic that the hedger will always be riskless and thus the statistical properties of the process are irrelevant as risk cancels out. Second, the partial differential equation is a backward Kolmogorov equation (see equation (4.34)) with killing rate r. As such we know that we require a suitable terminal condition and should look for a solution given by the Feynman–Kac formula (4.37). In fact, the valuation formula (5.3) provides us with exactly that. First, let us define the concept of an important financial instrument.

Definition 5.10. A *financial derivative* or *contingent claim* $(V_t)_{t\in[0,T]}$ on an *underlying* asset S_t is an $\mathbb{R}$-valued function of the process $(S_t)_{t\in[0,T]}$ adapted to the filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$.

Remark 5.11. Common derivatives are *options*. So-called *vanilla options* are *Calls* and *Puts*. They have the time T value, also called the *payoff*

$$V_T = \max\{S_T - K, 0\} \qquad \text{(Call)}$$

and

$$V_T = \max\{K - S_T, 0\} \qquad \text{(Put)},$$

where K is a positive constant call the *strike* price. In fact, options can be regarded as a synthetic portfolio $(\Pi_t)_{t\in[0,T]}$, which provides a certain payoff

$$V_T(\omega) = \Pi_T(\omega) \qquad \forall \omega \in \mathcal{F}_T.$$

Hence, substituting the payoff of a call option into formula (5.3) and employing B_t as numeraire, we obtain

$$\begin{aligned} V_0 &= E_Q\left[\frac{V_T}{B_T}\right] \\ &= E_Q\left[e^{-rT}\max\{S_T - K, 0\}\right] \\ &= e^{-rT}\left(E_Q\left[S_T I_{[S_T>K]}(S_T)\right] - KE_Q\left[I_{[S_T>K]}(S_T)\right]\right). \end{aligned} \tag{5.12}$$

Now by (5.7) it becomes obvious that to change to the martingale measure implies setting the drift of the risky asset to r. Hence,

$$\begin{aligned} E_Q\left[S_T I_{[S_T>K]}(S_T)\right] &= \int_K^\infty S_T f(S_T) dS_T \\ &= \int_{-d_2}^\infty S_0 e^{\left(r-\frac{1}{2}\sigma^2\right)T+\sigma\sqrt{T}x}\varphi(x)dx \\ &= Se^{rT}\Phi(d_1) \end{aligned} \tag{5.13}$$

and similarly

$$E_Q\left[S_T I_{[S_T>K]}(S_T)\right] = \Phi(d_2), \tag{5.14}$$

where $f(x)$ is the log-normal density of S_T, $\varphi(x)$ the standard normal density (1.2), $\Phi(x)$ its cumulative distribution, and

$$d_1 = \frac{\ln\frac{S_0}{K} + \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}, \tag{5.15}$$

as well as $d_2 = d_1 - \sigma\sqrt{T}$. Hence the so-called *Black–Scholes formula* for a Call option is

$$V_{BS}(S_0) := V_0 = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2), \tag{5.16}$$

and similarly, the Black–Scholes Put formula is

$$V_0 = Ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1). \tag{5.17}$$

In fact, both are related through the so-called *Put-Call parity*:

$$\text{Put} + S_0 = \text{Call} + Ke^{-rT}. \tag{5.18}$$

Binary Options and Martingale Probabilities

In fact, the payoff kernel of equation can be arbitrary, and in today's financial markets, a vast and ever increasing variety of option payoffs are available. We will look at some standard structures under the Black–Scholes assumptions.

A *binary* or *digital* Call option has the simple payoff $V_T = I_{[S_T \geq K]}(S_T)$. Hence its value is

$$\begin{aligned} V_0 &= e^{-rT} E_Q[I_{[S_T\geq K]}(S_T)] \\ &= e^{-rT}\Phi(d_2), \end{aligned}$$

as has already been demonstrated in (5.14). In fact, the option has the interpretation

$$V_0 = e^{-rT} Q(S_T > K), \tag{5.19}$$

i.e., the probability under the martingale measure of the risky asset exceeding the strike at T. In fact, if $(C_t)_{t\in[0,T]}$ is a call option, then

$$-\frac{\partial C}{\partial K} \equiv -\frac{\partial V_{BS}}{\partial K} = V_t; \tag{5.20}$$

i.e., the derivative of a Call option with respect to strike is the negative discounted probability of being in the money at expiry under the risk-neutral martingale measure.

Barrier Options and Exit Times

A common example of derivatives that depend on the entire path of an underlying random variable $(S_t)_{t\in[0,T]}$ are so-called barrier options. They are often typical Put or Call options with the additional feature that, if the underlying random variable hits a particular upper or lower barrier (or both) at any time in $[0,T]$, an event is triggered. One particular example is a so-called "down-and-out knockout Call option" $(D_t)_{t\in[0,T]}$ that becomes worthless when a lower level b is hit. Hence the payoff is

$$D_T = \max\{S_T - K, 0\} I_{\left[\min_{t\in[0,T]} S_t > b\right]},$$

thus, slotting the payoff into the standard valuation formula, we need to calculate

$$\begin{aligned} D_0 &= e^{-rT} E_Q \left[\max\{S_T - K, 0\} I_{\left[\min_{t\in[0,T]} S_t > b\right]}\right] \\ &= e^{-rT} \left(E_Q \left[(S_T - K) I_{\left[\min_{t\in[0,T]} S_t > b \cap S_T > K\right]}\right]\right) \\ &= e^{-rT} \left(E_Q \left[S_T I_{\left[\min_{t\in[0,T]} S_t > b \cap S_T > K\right]}\right]\right. \\ &\qquad \left. - KQ \left(\min_{t\in[0,T]} S_t > b \cap S_T > K\right)\right). \end{aligned} \tag{5.21}$$

It is not difficult to see that the latter probability can be transformed as

$$\begin{aligned} &Q\left(\min_{t\in[0,T]} W_t^Q > g(b) \cap W_T^Q > g(K)\right) \\ &= Q\left(W_T^Q < -g(K)\right) - Q\left(\min_{t\in[0,T]} W_t^Q < g(b) \cap W_T^Q > g(K)\right), \end{aligned} \tag{5.22}$$

where

$$g(x) = \frac{\ln \frac{x}{S_0} - \left(r - \frac{1}{2}\sigma^2\right) T}{\sigma}.$$

Now using the reflection principle of Lemma 2.144, we see that the last term of (5.22) can be rewritten as

$$\begin{aligned} &Q\left(\left(\tilde{W}_T^Q < g(b) \cup W_T^Q < g(b)\right) \cap W_T^Q > g(K)\right) \\ &= Q\left(\tilde{W}_T^Q < g(b) \cap W_T^Q > g(K)\right) \\ &= Q(W_T^Q < 2g(b) - g(K)). \end{aligned}$$

Since W_T^Q is a standard Brownian motion under Q, we obviously have the probability law

$$Q\left(W_T^Q < y\right) = \Phi\left(\frac{y}{\sqrt{T}}\right)$$

for any $y \in \mathbb{R}$. Backsubstitution gives the solution of the last term of (5.21). We leave the remaining (rather cumbersome) steps of the derivation as an exercise (5.6). Eventually, the result turns out as

$$D_0 = V_{BS}(S_0) - \left(\frac{S_0}{b}\right)^{1-\frac{2r}{\sigma^2}} V_{BS}\left(\frac{b^2}{S_0}\right)$$

in terms of the Black–Scholes price (5.16).

American Options and Stopping Times

So far we have only considered so-called *European* options where a sole payoff occurs at time T. The type of option where the payoff may occur at any time up to expiry at the holder's discretion is called *American*. It can be shown through replication nonarbitrage arguments (see, e.g., Øksendal (1998) or Musiela and Rutkowski (1998)) that their valuation formula is

$$V_0^* = \sup_{\tau \in [0,T]} E_Q\left[V_\tau^*\right].$$

Here τ clearly is a stopping time. In general we are dealing with an optimal stopping or free boundary problem and there are usually no closed-form solutions. The American option value can be posed in terms of a linear complementary problem (see, e.g., Wilmott, Howison, and Dewynne (1993)). Defining the value of immediate exercise as P_t, we have

$$\mathcal{L}_{BS} V_t \leq 0 \qquad \text{and} \qquad V_t \geq P_t, \tag{5.23}$$

with

$$\mathcal{L}_{BS} V(V - P) = 0 \qquad \text{and} \qquad V_T = P_T.$$

Now, if there exists an early exercise region $\mathcal{R} = \{S_\tau | \tau < T\}$, we necessarily have $V_\tau = P_\tau$, if $\mathcal{L}_{BS} V_\tau = \mathcal{L}_{BS} P_\tau > 0$, then this represents a contradiction to (5.23). Therefore, in this case early exercise can never be optimal, as, for instance, for a Call option with payoff $P_\tau = \max\{S_\tau - K, 0\}$, and thus

$$\mathcal{L}_{BS} \max\{S_\tau - K, 0\} = \frac{1}{2}\sigma^2 K^2 \delta(S_\tau - K) + rK I_{S>K}(S_\tau) \geq 0,$$

where δ represent the Dirac-delta. Conversely, if $V(\mathcal{A}) < P(\mathcal{A})$ for some region $\mathcal{A}$, then $\mathcal{A} \subseteq \mathcal{R}$; i.e., it would certainly be optimal to exercise within this region and probably even within a larger one. As an example, for a Put option with $P_\tau = \max\{K - S_\tau, 0\}$, we have that

$$V(0, 0) = Ke^{-rT} < P(0, T) = K.$$

Typically, American options are valued employing numerical methods.

5.3 Models of Interest Rates

The Black–Scholes model incorporates the concept of the *time value of money* through the instantaneous short rate r. However, it assumes that this rate is deterministic and even constant throughout time or, in other words, the term structure (of interest rates) is flat and with zero volatility. But in reality it is neither. In fact, a significant part of the financial markets is related to debt or, as it is more commonly called, fixed income instruments. The latter, in their simplest form, are future cash flows promised to a beneficiary by an emitter, who may be a government, corporation, sovereign, etc. The buyer of the debt hopes to pay as little up front as possible; vice versa for the counterparty. These securities can be regarded as derivatives on interest rates. The latter are used as a tool of expressing the discount between money received today and money received in the future. In reality, this discount tends to be a function of the time to maturity T of the debt, and moreover it changes continuously and unpredictably. These concepts can be formalized in a simple discount bond market.

Definition 5.12. A filtered probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in[0,T]})$ endowed with adapted stochastic processes $(B_t^{(i)})_{t\in[0,T_i]}$, $i = 0, \ldots, n$, $T_n \leq T$, with

$$B_{T_i}^{(i)} = 1 \qquad \forall i = 0, \ldots, n,$$

representing *discount bond prices* is called a *discount bond market*. The *term structure* of (continuously compounded) *zero rates* $(r(t,T))_{\forall t,T;t\leq T}$ is given by the relationship

$$B_t^{(i)} = e^{-r(t,T_i)(T_i-t)} \qquad \forall i.$$

By the fundamental theorems of asset pricing, the discount bond market is free of arbitrage if there exist equivalent martingale measures for all discount bond ratios $B_t^{(i)}/B_t^{(j)}$, $i, j \in \{0, \ldots, n\}$. But instead of evolving the discount bond prices directly, models for fixed income derivatives focus on the dynamics of the underlying interest rates. We will give brief summaries of the main approaches to interest rate modeling.

Short Rate Models

Motivated by the Black–Scholes model, the first stochastic modeling approaches were performed on the concept of the short rate.

Definition 5.13. The instantaneous *short rate*

$$r_t := r(t,t) \qquad \forall t \in [0,T]$$

is connected to the value of a discount bound through

$$B_t^{(i)} = E_Q\left[e^{-\int_t^{T_i} r_s ds}\right] \qquad \forall i = 0, \ldots, n, \tag{5.24}$$

under the risk-neutral measure Q.

Vasicek (1977) proposed that the short rate follows a Gaussian process

$$dr_t = \mu_r(t, r_t)dt + \sigma_r(t, r_t)dW_t^P$$

under the *physical* or empirical measure P. This then results in a nonarbitrage relationship between the short rate and bond processes of different maturities based on the concept of a market price of risk process $(\lambda_t)_{t\in[0,T]}$.

Proposition 5.14. *Let the short rate r_t follow the diffusion process*

$$dr_t = \mu_r(r_t, t)dt + \sigma_r(r_t, t)dW_t^P.$$

Furthermore, assume that the discount bonds $B_t^{(i)}$ with $t \leq T_i$ for all i have interest rates as their sole risky factor and follow the sufficiently regular stochastic processes

$$dB_t^{(i)} = \mu_i(r, t, T_i)dt + \sigma_i(r, t, T_i)dW_t^P \qquad \forall i.$$

Then the nonarbitrage bond drifts are given by

$$\mu_i = r_t + \sigma_i\lambda(r_t, t),$$

where $\lambda(r_t, t)$ is the market price of the interest rate risk process.

Proof: Let us define the portfolio process $(\Pi_t)_t$ as

$$\Pi_t = H_t^{(1)}B_t^{(1)} + H_t^{(2)}B_t^{(2)} \tag{5.25}$$

and normalize it by putting $H_t^{(1)} \equiv 1$ and $H_t^{(2)} := H_t$ for all t. The dynamics over a time interval dt are then given by

$$d\Pi_t = dB_t^{(1)} + H_t dB_t^{(2)}. \tag{5.26}$$

Invoking Itô's formula we have

$$\mu_i = \frac{\partial B^{(i)}}{\partial t} + \mu_r\frac{\partial B^{(i)}}{\partial r} + \frac{1}{2}\sigma_r\frac{\partial^2 B^{(i)}}{\partial r^2}$$

for the bond drift and

$$\sigma_i = \sigma_r\frac{\partial B^{(i)}}{\partial r} \tag{5.27}$$

for the bond volatility. Substituting both into (5.26) after cancellations, we obtain

$$d\Pi_t = (\mu_1 - H_t\mu_2) + \left(\sigma_r\frac{\partial B^{(1)}}{\partial r} - H_t\sigma_r\frac{\partial B^{(2)}}{\partial r}\right)dW_t^P.$$

It becomes obvious that when choosing the hedge ratio as

$$H_t = \sigma_r \frac{\partial B^{(1)}}{\partial r} \left(\sigma_r \frac{\partial B^{(2)}}{\partial r} \right)^{-1}, \tag{5.28}$$

the Wiener process dW_t^P and hence all risk vanishes so that

$$d\Pi_t = r_t dt, \tag{5.29}$$

meaning that the bond must earn the riskless rate. Now, substituting (5.27), (5.28), and (5.29) into (5.25), after rearrangement we get the relationship

$$\frac{\mu_1 - r_t B_t^{(1)}}{\sigma_1} = \frac{\mu_2 - r_t B_t^{(2)}}{\sigma_2}.$$

By observing that the two sides do not depend on the opposite index and we can write

$$\frac{\mu_i - r_t B_t^{(i)}}{\sigma_i} = \lambda(r_t, t) \qquad \forall i,$$

where $\lambda(r_t, t)$ is an adapted process, independent of T_i. □

Corollary 5.15. By changing to the risk-neutral measure Q given by

$$\frac{dQ}{dP} = \exp\left\{ -\int_0^t \lambda dW_s^P - \int_0^t \frac{\lambda^2}{2} ds \right\} \qquad \text{on } \mathcal{F}_t,$$

the risk-neutralized short rate process is given by

$$dr_t = (\mu_r - \sigma_r \lambda) dt + \sigma_r dW_t^Q,$$

where

$$W_t^Q = W_t^P + \int_0^t \lambda ds.$$

The reason why λ arises is that the short rate, representing the stochastic variable, contrary to the asset price process S_t in the Black–Scholes model, is not directly tradeable, meaning that a portfolio $H_t r_t$ is meaningless. One cannot buy units of it directly for hedging.

In practice, however, λ is rarely calculated explicitly. Instead, in a short rate modeling framework some functional forms of μ_r and σ_r are specified and their parameters *calibrated* to observed market prices. This implies that one is moving from a physical measure P to a risk-neutral measure Q. For that purpose it is useful to choose the short rate processes such that there exists a tractable analytic solution for the bond price. In fact, the Vasicek stochastic differential equation under the measure Q for the short rate is chosen to be the mean-reverting Ornstein–Uhlenbeck process (see Example 4.9)

$$dr_t = (a - br_t) dt + \sigma dW_t^Q,$$

which, by substituting it into (5.24), leads one to conjecture that the solution of a discount bond maturing at time T, namely with terminal condition $B_T = 1$, is of the form

$$B_t = e^{C(t,T)-D(t,T)r_t},$$

thus preserving the Markov property of the process. Some cumbersome, yet straightforward, calculations show that

$$D(t,T) = \frac{1}{b}\left(1 - e^{-b(T-t)}\right) \tag{5.30}$$

and

$$C(t,T) = \frac{\sigma^2}{2}\int_t^T (D(s,T))^2 ds - a\int_t^T D(s,T)ds. \tag{5.31}$$

It becomes apparent that the model only provides three parameters to describe the dynamics of a potentially complex term structure. Therefore, another common model is that of Hull and White (1990), also called the extended Vasicek model, which makes all the parameters time-dependent, namely,

$$dr_t = (a_t - b_t r_t)dt + \sigma_t dW_t^Q,$$

thereby allowing a richer description of the yield curve dynamics.

Heath–Jarrow–Morton Approach

As an evolution in interest rate modeling Heath, Jarrow, and Morton (1992) defined an approach assuming a yield curve to be specified by a continuum of traded bonds and evolved it through *instantaneous forward rates* $f(t,T)$ instead of the short rate. The former are defined through the expression

$$B_t^{(T)} = e^{-\int_t^T f(t,s)ds}, \tag{5.32}$$

and thus

$$f(t,T) = -\frac{\partial \ln B_t^{(T)}}{\partial T} \tag{5.33}$$

and

$$f(t,t) = r_t. \tag{5.34}$$

In fact, the Heath–Jarrow–Morton approach is very generic, and most other models are just specializations of it. It assumes that forward rates, under the risk-neutral measure Q associated with the riskless account numeraire, follow the stochastic differential equation

$$df(t,T) = \mu(t,T)dt + \sigma(t,T)\cdot d\mathbf{W}_t^Q, \tag{5.35}$$

where $\sigma(t,T)$ and $d\mathbf{W}_t^Q$ are n-dimensional. In fact, due to nonarbitrage arguments, the drift function $\mu(t,T)$ can be fully specified. Invoking Itô's formula on (5.32), we obtain the relationship

$$\frac{dB_t^{(T)}}{B_t^{(T)}} = \left(r_t - \int_t^T \mu(t,s)ds + \frac{1}{2} \left| \int_t^T \sigma(t,s)ds \right|^2 \right) dt$$
$$- \int_t^T \sigma(t,s)dsd\mathbf{W}_t^Q$$

because

$$\int_t^T \frac{\partial f(t,s)}{\partial t} dsdt = \int_t^T \mu(t,s)dsdt + \int_t^T \sigma(t,s)dsd\mathbf{W}_t^Q,$$

and by noting (5.34), (5.35) as well as Fubini's Theorem A.41. But now, for the deflated discount bond to be a martingale, the drift has to be r_t. Thus

$$\int_t^T \mu(t,s)ds = \frac{1}{2} \left| \int_t^T \sigma(t,s)ds \right|^2$$

and so

$$\mu(t,T) = \sigma(t,T) \cdot \int_t^T \sigma(t,s)ds \tag{5.36}$$

Substituting (5.36) into (5.35), we obtain arbitrage-free processes of a continuum of forward rates, driven by one or more Wiener processes:

$$df(t,T) = \sigma(t,T) \cdot \int_t^T \sigma(t,s)ds + \sigma(t,T) \cdot d\mathbf{W}_t^Q.$$

It can be noted that, unlike for short rate models, no market price of risk appears. This is due to the fact that forward rates are actually tradeable, as the following section will demonstrate.

Brace–Gatarek–Musiela Approach

As a very intuitive and simple yet powerful approach, Brace, Gatarek, and Musiela (1997) and other authors (Miltersen, Sandmann, and Sondermann (1997), Jamshidian (1997)) in parallel introduced a model of discrete forward rates $(F_t^{(i)})_{t\in[0,T]}$, $i = 1, \ldots, n$, that span a yield curve through the discrete discount bonds

$$B_t^{(k)} = \prod_{i=1}^{k} \left(1 + F_t^{(i)}(T_i - T_{i-1})\right)^{-1}, \qquad 1 \leq k \leq n. \tag{5.37}$$

The forward rates are assumed to follow the system of stochastic differential equations

$$d\mathbf{F}_t = \mu(t, \mathbf{F}_t)dt + \Sigma(t, \mathbf{F}_t)d\mathbf{W}_t,$$

where Σ is a diagonal matrix containing the respective volatilities and $d\mathbf{W}_t$ is a vector of Wiener processes with correlations

$$E\left[dW_t^{(i)} dW_t^{(j)}\right] = \rho_{ij} dt.$$

In particular, all forward rate processes are considered of the lognormal form

$$\frac{dF_t^{(i)}}{F_t^{(i)}} = \mu^{(i)}(t, \mathbf{F}_t) + \sigma_t^{(i)} dW_t^{(i)} \qquad \forall i. \tag{5.38}$$

Again, similar to the Heath–Jarrow–Morton model, a martingale nonarbitrage argument determines the drift $\mu^{(i)}$ for each forward rate $F_t^{(i)}$. To see this we can write (5.37) as a recurrence relation, and after rearrangement we obtain

$$F_t^{(i)} B_t^{(i)} = \frac{B_t^{(i-1)} - B_t^{(i)}}{T_i - T_{i-1}}, \tag{5.39}$$

which states that the left-hand side is equivalent to a portfolio of traded assets and has to be driftless under the martingale measure associated with a numeraire asset. In fact, we have a choice of numeraire asset among all combinations of available bonds (5.37). We arbitrarily choose a bond $B_t^{(N)}$, $1 \le N \le n$, with associated *forward measure* Q_N and thus

$$E^{Q_N}\left[d\left(F_t^{(i)} \frac{B_t^{(i)}}{B_t^{(N)}}\right)\right] = 0. \tag{5.40}$$

The derivation is left as an exercise, and the end-result is

$$\mu_t^{(i)} = \begin{cases} -\sum_{j=i+1}^{N} \frac{(T_{j+1}-T_j)F_t^{(j)}\sigma_t^{(i)}\sigma_t^{(j)}\rho_{ij}}{1+(T_{j+1}-T_j)F_t^{(j)}} & \text{if } i < N, \\ 0 & \text{if } i = N, \\ \sum_{j=N+1}^{n} \frac{(T_{j+1}-T_j)F_t^{(j)}\sigma_t^{(i)}\sigma_t^{(j)}\rho_{ij}}{1+(T_{j+1}-T_j)F_t^{(j)}} & \text{if } i > N. \end{cases} \tag{5.41}$$

This model is particularly appealing as it directly takes real-world observable inputs like forward rates and their volatilities and also discrete compounding/discounting. But the potentially large number of Brownian motions makes the model difficult to handle computationally, as it may require large-scale simulations.

5.4 Contingent Claims under Alternative Stochastic Processes

In practice, Black–Scholes is the most commonly used model, despite its simplicity. Nonetheless, there are significant modeling extensions that try to make it more realistic. As already discussed, the introduction of stochastic interest rate processes is a significant step. But there exist other issues. The most significant of them is the fact that the volatility parameter σ is constant. In practice, Put and Call options of different strikes K and expiries T are traded on exchanges and their prices $\hat{V}(T,K)$ are directly observable. This fact motivates one to determine so-called *implied volatilities* $\sigma_{imp}(T,K)$, because the simple Black–Scholes formulae (5.16) and (5.17) are one-to-one mappings between prices of options with respective T and K to the volatilities σ. The implied volatility is then such that

$$V_{BS}(\sigma_{imp}(T,K)) = \hat{V}(T,K).$$

Stating option prices in terms of their implied volatility makes them directly comparable in the sense that an option with a higher implied volatility is more expensive than one with a lower.[10]

If the Black–Scholes model were an accurate description of the real world, then $\sigma_{imp}(T,K) = \sigma$ constant. But in the real world this is not the case. Usually implied volatilities are both dependent on K and T. Typical shapes of the implied volatility surface across the strike are so-called skews or smiles. The former usually means that $\sigma_{imp}(K_1,T) > \sigma_{imp}(K_2,T)$, with $K_1 < K_2$ for Put options and vice versa for Calls, implying that the market believes there is a greater risk in a downward move in S_t and thus sees a negative correlation between S_t and σ_{imp}. A smile shape is usually due to the fact that out-of-the-money[11] options are relatively more expensive. There are various modeling approaches to overcome this deficiency of the Black–Scholes model.

Local Volatility

Dupire (1994) demonstrated that extending the risky asset process under the martingale measure to

$$\frac{dS_t}{S_t} = rdt + \sigma(t,S_t)dW_t \tag{5.42}$$

results in a probability distribution that recovers all observed option prices $\hat{V}(T,K)$. By (5.12) and (5.13) it is clear that we can write the observed (Call) option price as

$$\hat{V}(T,K) = e^{-rT}\int_K^\infty (S_T - K)f(0,S_0,T,S_T)dS_T.$$

Having differentiated with respect to K we obtained the cumulative distribution as (5.19) and (5.20). Repeatedly differentiating, we obtain the so-called *risk-neutral transition density*

[10] By Put-Call parity (5.18), the implied volatility for Puts and Calls in an arbitrage-free market has to be identical for all pairs T,K.

[11] Approximately, a Put option with $K < S_0$ and a Call option with $K > S_0$.

$$f(0, S_0, T, K) = e^{rT} \frac{\partial^2 \hat{V}}{\partial K^2}. \tag{5.43}$$

Now, by Theorem 4.50 $f(0, S_0, T, K)$ has to satisfy the Kolmogorov forward equation

$$\frac{\partial f}{\partial T} = \frac{1}{2} \frac{\partial^2(\sigma^2(T, K)K^2 f)}{\partial K^2} - \frac{\partial(rKf)}{\partial K} \tag{5.44}$$

with initial condition

$$f(S_0, 0, x, 0) = \delta(x - S_0).$$

Substituting (5.43) into (5.44), integrating twice with respect to K (after applying Fubini's Theorem A.41 when changing the order of integration), and noting the boundary condition

$$\lim_{K \to \infty} \frac{\partial \hat{V}}{\partial T} = 0,$$

we obtain

$$\frac{\partial \hat{V}}{\partial T} = \frac{1}{2}\sigma^2(T, K)K^2 \frac{\partial^2 \hat{V}}{\partial K^2} - rK \frac{\partial \hat{V}}{\partial K}.$$

Thus

$$\sigma(T, K) = \sqrt{\frac{\frac{\partial \hat{V}}{\partial T} + rK \frac{\partial \hat{V}}{\partial K}}{\frac{1}{2}\sigma^2 K^2 \frac{\partial^2 \hat{V}}{\partial K^2}}},$$

fully specifying the process (5.42).

Jump Diffusions

Merton (1976) introduced an extension to the Black–Scholes model, which appended the risky asset process by a Poisson process $(N_t)_{t \in [0,T]}$, with $N_0 = 0$ and constant intensity λ, independent of $(W_t)_{t \in [0,T]}$, to allow asset prices to move discontinuously. The compensated risky asset price process now follows

$$\frac{dS_t}{S_{t-}} = (r - \lambda m)dt + \sigma dW_t + J_t dN_t, \tag{5.45}$$

under the risk-neutral equivalent martingale measure, with $(J_t)_{t \in [0,T]}$ an independent and identically distributed sequence of random variables valued in $]-1, \infty[$ of the form $J_t = J_i I_{[t > \tau_i[}(t)$ with $J_0 = 1$, τ_i an increasing sequence of times and where

$$E[dN_t] = \lambda dt \qquad \text{and} \qquad E[J_t] = m.$$

Then the solution to (5.45) can be written as

$$S_T = S_0 \exp\left\{\left(r - \frac{\sigma^2}{2} - \lambda m\right) T + \sigma W_T\right\} \prod_{i=1}^{N_T} J_i.$$

Defining an option value process by $(V(t, S_t))_{t\in[0,T]}$, we apply Itô's formula along with its extension to Poisson processes and assume that jump risk in the market is diversifiable (see Merton (1976) and references therein) so that we can use the chosen risk-neutral measure Q, we obtain

$$[\mathcal{L}_{BS}V](0, S_0) = \lambda \left(m\frac{\partial V}{\partial S} - E[V(0, J_0 S_0)] + V(0, S_0) \right).$$

The solution to this partial differential equation can still be written in the form (5.12), namely

$$V(0, S_0) = e^{-rT} \left(E_Q \left[S_T I_{[S_T>K]}(S_T) \right] - KQ(S_T > K) \right),$$

but closed form expressions of the expectation and probability terms only exist for special cases. Two such cases were identified by Merton (1976), first when $N_t \in \{0, 1\}$ and $J_1 = -1$, i.e., the case when there exists the possibility of a single jump that puts the risky asset into the absorbing state 0. Then the solution for, say, a Call option is V_{BS} but with a modified risk-free rate $r + \lambda$. The second case is when

$$\ln J_t \sim N(\mu, \gamma^2)$$

so that

$$m = e^{\mu+\frac{1}{2}\gamma^2}.$$

Then

$$V(0, S_0) = \sum_{i=0}^{\infty} \frac{e^{-\lambda mT}(\lambda mT)^i}{i!} V_{BS}(\sigma_i, r_i),$$

where the risk-free rate is given by

$$r_i = r + \frac{i}{T}\left(\mu + \frac{\gamma^2}{2} \right) - \lambda(m - 1),$$

and the volatility by

$$\sigma_n = \sqrt{\sigma^2 + \frac{i}{T}\gamma^2}.$$

Another, semiclosed form expression exists when J_t are exponentially distributed (see Kou (2002)), but usually the solution has to be written in terms of Fourier transforms that need to be solved numerically.

5.5 Insurance Risk

Ruin Probabilities

A typical one-company insurance portfolio is modelled as follows. The initial value of the portfolio is the so-called *initial reserve* $u \in \mathbb{R}^*_+$. At random times

$\sigma_n \in \mathbb{R}_+^*$, a random claim $U_n \in \mathbb{R}^*$ occurs for $n \in \mathbb{N}^*$. During the time interval $]0,t] \subset \mathbb{R}_+^*$ an amount $\Pi_t \in \mathbb{R}_+^*$ of income is collected through *premiums*. The *cumulative claims process* up to time $t > 0$ is then given by

$$X_t = \sum_{k=1}^{\infty} U_k I_{[\sigma_k \leq t]}(t).$$

In this way the value of the portfolio at time t, the so-called *risk reserve*, is given by

$$R_t = u + \Pi_t - X_t.$$

The *claims surplus process* is given by

$$S_t = X_t - \Pi_t.$$

If we assume that premiums are collected at a constant rate $\beta > 0$, then

$$\Pi_t = \beta t, \qquad t > 0.$$

Now, the *time of ruin* $\tau(u)$ of the insurance company is a function of the initial reserve level. It is the first time when the claim surplus process crosses this level, namely

$$\tau(u) := \min\{t > 0 | R_t < 0\} = \min\{t > 0 | S_t > u\}.$$

Hence, an insurance company is interested in the ruin probabilities; first the *finite horizon ruin probability*, which is defined as

$$\psi(u,x) := P(\tau(u) \leq x) \qquad \forall x \geq 0;$$

second, the *probability of ultimate ruin*, defined as

$$\psi(u) := \lim_{x \to +\infty} \psi(u,x) = P(\tau(u) \leq +\infty).$$

It may also be interested in the *survival probability* defined as

$$\bar{\psi}(u) = 1 - \psi(u).$$

It is clear that

$$\psi(u,x) = P\left(\max_{0 \leq t \leq x} S_t > u\right).$$

The above model shows that the marked point process $(\sigma_n, U_n)_{n \in \mathbb{N}^*}$ on $(\mathbb{R}_+^* \times \mathbb{R}_+^*)$ plays an important role. As a particular case, we consider the marked Poisson process with *independent marking*, i.e., the case in which $(\sigma_n)_{n \in \mathbb{N}^*}$ is a Poisson process on $\mathbb{R}_+^*$ and $(U_n)_{n \in \mathbb{N}^*}$ is a family of independent and identically distributed $\mathbb{R}_+^*$-valued random variables, independent of the underlying point process $(\sigma_n)_{n \in \mathbb{N}^*}$. In this case, we have that the interoccurrence times between claims $T_n = \sigma_n - \sigma_{n-1}$ (with $\sigma_0 = 0$) are independent and identically exponentially distributed random variables with a common parameter $\lambda > 0$ (see Rolski et al. (1999)). In this way the number of claims N_t during $]0, t]$, $t > 0$, i.e., the underlying counting process

$$N_t = \sum_{k=1}^{\infty} I_{[\sigma_k \leq t]}(t)$$

is a Poisson process on $\mathbb{R}_+^*$ with intensity λ. Now, let the claim sizes U_n be independent and identically distributed with common cumulative distribution function F_U and let $(U_n)_{n \in \mathbb{N}^*}$ be independent of $(N_t)_{t \in \mathbb{R}_+}$. We may notice that in this case the cumulative claim process

$$X_t = \sum_{k=1}^{N_t} U_k = \sum_{k=1}^{\infty} U_k I_{[\sigma_k \leq t]}(t), \qquad t > 0,$$

is a so-called *compound Poisson process*. Clearly, the latter has stationary independent increments and, in fact, it is a Lévy process, so that we can state the following theorem.

Theorem 5.16. *(*Karlin and Taylor (1981), page 428*). Let $(X_t)_{t \in \mathbb{R}_+^*}$ be a stochastic process having stationary independent increments and let $X_0 = 0$. It is then a compound Poisson process if and only if its characteristic function $\phi_{X_t}(z)$ is of the form*

$$\phi_{X_t}(z) = \exp\{-\lambda t(1 - \phi(z))\}, \qquad z \in \mathbb{R},$$

where $\lambda > 0$ and ϕ is a characteristic function.

With respect to the above model, ϕ is the common characteristic function of the claims $U_n, n \in \mathbb{N}^*$. If μ and σ^2 are the mean and the variance of U_1, respectively, we have

$$E[X_t] = \mu \lambda t,$$
$$Var[X_t] = \left(\sigma^2 + \mu^2\right) \lambda t.$$

We may also obtain the cumulative distribution function of X_t through the following argument:

$$\begin{aligned} P(X_t \leq x) &= P\left(\sum_{k=1}^{N_t} U_k \leq x\right) \\ &= \sum_{n=0}^{\infty} P\left(\sum_{k=1}^{N_t} U_k \leq x \,\middle|\, N_t = n\right) P(N_t = n) \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} F^{(n)}(x) \end{aligned}$$

for $x \geq 0$ (it is zero otherwise), where

$$F^{(n)}(x) = P(U_1 + \cdots + U_n \leq x), \qquad x \geq 0,$$

with

$$F^{(0)}(x) = \begin{cases} 1 & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

In the special case of exponentially distributed claims, with common parameter $\mu > 0$, we have

$$F_U(u) = P(U_1 \leq u) = 1 - e^{-\mu u}, \qquad u \geq 0,$$

so that $U_1 + \cdots + U_n$ follows a gamma distribution with

$$F_U^{(n)}(u) = 1 - \sum_{k=0}^{n-1} \frac{(\mu u)^k e^{-\mu u}}{k!} = \frac{\mu^n}{(n-1)!} \int_0^u e^{-\mu v} v^{n-1} dv$$

for $n \geq 1$, $u \geq 0$. The following theorem holds for exponentially distributed claim sizes.

Theorem 5.17. *Let*

$$F_U(u) = 1 - e^{-\mu u}, \qquad u \geq 0.$$

Then

$$\psi(u, x) = 1 - e^{-\mu u - (1+c)\lambda x} g(\mu u + c\lambda x, \lambda x),$$

where

$$c = \mu \frac{\beta}{\lambda},$$

$$g(z, \theta) = J(\theta z) + \theta J^{(1)}(\theta z) + \int_0^z e^{z-v} J(\theta v) dv - \frac{1}{c} \int_0^{c\theta} e^{c\theta - v} J\left(z c^{-1} v\right) dv,$$

with $\theta > 0$. *Here*

$$J(x) = \sum_{n=0}^{\infty} \frac{x^n}{n! n!}, \qquad x \geq 0,$$

and $J^{(1)}(x)$ *is its first derivative.*

Proof: See, e.g., Rolski et al. (1999), page 196. □

For the general compound Poisson model we may provide information about the finite-horizon ruin probability $P(\tau(u) \leq x)$ by means of martingale methods. We note again that, in terms of the claim surplus process S_t, we have

$$\tau(u) = \min\{t | S_t > u\}, \qquad u \geq 0,$$

and

$$\psi(u,x) = P(\tau(u) \leq x), \qquad x \geq 0, u \geq 0.$$

The claim surplus process is then given by

$$S_t = \sum_{k=1}^{N_t} U_k - \beta t,$$

where $\lambda > 0$ is the arrival rate, β the premium rate, and F_U the claim size distribution. Let

$$Y_t = \sum_{k=1}^{N_{t-}} U_k, \qquad t \geq 0,$$

be the left-continuous version of the cumulative claim size X_t. Based on the notion of reversed martingales (see Rolski et al. (1999), page 434), it can be shown that the process

$$Z_t = X^*_{x-t}, \qquad t \in [0, x[, x > 0,$$

with

$$X^*_t = \frac{Y_t}{u+\beta t} + \int_t^x \frac{Y_v}{v} \frac{u}{(u+\beta v)^2} dv, \qquad 0 < t \leq x,$$

for $u \geq 0$ and $x > 0$, is an $\mathcal{F}_t^X$-martingale. Let

$$\tau^0 = \sup\{v | v \geq x, S_v \geq u\},$$

and $\tau^0 = 0$, if $S(v) < u$ for all $v \in [0, x]$. Then $\tau := x - \tau^0$ is a bounded $\mathcal{F}_t^X$-stopping time. As a consequence,

$$E[Z_\tau] = E[Z_0],$$

i.e.,

$$E\left[\frac{Y_x}{u+\beta x} \middle| Y_x \leq u + \beta x\right] = E\left[\frac{Y_{\tau^0}}{u+\beta\tau^0} + \int_{\tau^0}^x \frac{Y_v}{v}\frac{u}{(u+\beta v)^2} dv \middle| S_{x-} \leq u\right].$$

On the other hand, we have

$$P(\tau(u) > x) = P\left(S_x \leq u \cap \tau^0 = 0\right) = P(S_x \leq u) - P\left(S_x \leq u \cap \tau^0 > 0\right).$$

Now, since

$$Y_{\tau^0} = u + \beta\tau^0 \qquad \text{for } \tau^0 > 0,$$

we have

$$\begin{aligned} E\left[\frac{Y_{\tau^0}}{u+\beta\tau^0} \middle| S_x \leq u\right] &= E\left[\frac{Y_{\tau^0}}{u+\beta\tau^0} \middle| S_x \leq u \cap \tau^0 > 0\right] \\ &= P\left(S_x \leq u \cap \tau^0 > 0\right). \end{aligned}$$

Thus, for $u > 0$, we have the following result.

Theorem 5.18. *(*Rolski et al. (1999), page 434*). For all $u \geq 0$ and $x > 0$,*

$$1 - \psi(u,x) = \max\left\{E\left[1 - \frac{Y_x}{u+\beta x}\right], 0\right\} + E\left[\int_{\tau^0}^{x} \frac{Y_v}{v} \frac{u}{(u+\beta v)^2} dv \,\middle|\, S_x \leq u\right].$$

In particular, for $u = 0$,

$$1 - \psi(0,x) = \max\left\{E\left[1 - \frac{Y_x}{\beta x}\right], 0\right\}.$$

A Stopped Risk Reserve Process

Consider the risk reserve process

$$R_t = u + \beta t - \sum_{k=1}^{N_t} U_k.$$

A useful model for stopping the process is to stop R_t at the time of ruin $\tau(u)$ and let it jump to a *cemetery* state. In other words, consider the process

$$X_t = \begin{cases} (1, R_t) & \text{if } t \leq \tau(u), \\ (0, R_\tau(u)) & \text{if } t > \tau(u). \end{cases}$$

The process $(X_t, t)_{t \in \mathbb{R}_+}$ is a *piecewise deterministic Markov process* as defined in Davis (1984). The infinitesimal generator of $(X_t, t)_{t \in \mathbb{R}_+}$ is given by

$$\begin{aligned} &\mathcal{A}g(y,t) \\ &= \frac{\partial g}{\partial t}(y,t) + I_{[y \geq 0]}(y)\left(\beta \frac{\partial g}{\partial y}(y,t) + \lambda\left(\int_0^y g(y-v,t)dF_U(v) - g(y,t)\right)\right) \end{aligned}$$

for g satisfying sufficient regularity conditions, so that it is in the domain of $\mathcal{A}$ (see Rolski et al. (1999), page 467). If g does not depend explicitly upon time and $g(y) = 0$ for $y < 0$, then the infinitesimal generator reduces to

$$\mathcal{A}g(y) = \beta \frac{dg}{dy}(y) + \lambda\left(\int_0^y g(y-v)dF_U(v) - g(y)\right).$$

The following theorem holds.

Theorem 5.19. *Under the above assumptions,*

1. *the only solution $g(y)$ to $\mathcal{A}g(y) = 0$, such that $g(0) > 0$ and $g(y) = 0$, for $y \in]-\infty, 0[$, is the survival function $\bar{\psi}(y) = P(\tau(u) = +\infty)$;*
2. *let $x > 0$ be fixed and let $g(y,t)$ solve $\mathcal{A}g = 0$ in $(\mathbb{R} \times [0,x])$ with boundary condition $g(y,x) = I_{[y \geq 0]}(y)$. Then $g(y,0) = P(\tau(y) > x)$ for any $y \in \mathbb{R}$, $x \in \mathbb{R}_+^*$.*

5.6 Exercises and Additions

5.1. Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ and $(\mathcal{G}_n)_{n\in\mathbb{N}}$ be two filtrations on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathcal{G}_n \subseteq \mathcal{F}_n \subseteq \mathcal{F}$ for all $n \in \mathbb{N}$; we say that a real-valued discrete time process $(X_n)_{n\in\mathbb{N}}$ is an $(\mathcal{F}_n, \mathcal{G}_n)$-*martingale* if and only if

- $(X_n)_{n\in\mathbb{N}}$ is an $\mathcal{F}_n$-adapted integrable process;
- for any $n \in \mathbb{N}$, $E[X_{n+1} - X_n|\mathcal{G}_n] = 0$.

A process $C = (C_n)_{n\geq}$ is called $\mathcal{G}_n$-*predictable* if C_n is $\mathcal{G}_{n-1}$-measurable. Given $N \in \mathbb{N}$, we say that a $\mathcal{G}_n$-predictable process C is *totally bounded by time* N if

- $C_n = 0$ almost surely for all $n > N$;
- there exists a $K \in \mathbb{R}_+$ such that $C_n < k$ almost surely for all $n \leq N$.

Let C be a $\mathcal{G}_n$-predictable process, totally bounded by time N. We say that it is a risk-free $\{\mathcal{G}_n\}_N$-strategy if, further,

$$\sum_{i=1}^{N} C_i(X_i - X_{i-1}) \geq 0 \qquad \text{a.s.}, \qquad \mathbb{P}\left(\sum_{i=1}^{N} C_i(X_i - X_{i-1}) > 0\right) > 0.$$

Show that there exists a risk-free $\{\mathcal{G}_n\}_N$-strategy for $X = (X_n)_{n\in\mathbb{N}}$ if and only if there does not exist an equivalent measure $\widetilde{\mathbb{P}}$ such that $X_{n\wedge N}$ is a $(\mathcal{F}_n, \mathcal{G}_n)$-martingale under $\widetilde{\mathbb{P}}$. This is an extension of the first fundamental theorem of asset pricing, Theorem 5.7. See also Dalang, Morton, and Willinger (1990).

5.2. Given a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the filtration $\mathcal{F}_n^m := \mathcal{F}_{n-m}$ is called an *m-delayed filtration.* An $\mathcal{F}_n$-adapted, integrable stochastic process $X = (X_n)_{n\in\mathbb{N}}$ is an *m-martingale* if it is an $(\mathcal{F}_n, \mathcal{F}_n^m)$-martingale (see problem 5.1). Find a real-valued (2-martingale) X where no profit is available during any unit of time; i.e.,

$$\forall i, \qquad \mathbb{P}(X_i - X_{i-1} > 0) > 0, \qquad \mathbb{P}(X_i - X_{i-1} < 0) > 0,$$

but admits a risk-free $\{\mathcal{F}_n^1\}_3$-strategy C; i.e.,

$$\sum_{i=1}^{3} C_i(X_i - X_{i-1}) \geq 0 \qquad \text{a.s.}, \qquad \mathbb{P}\left(\sum_{i=1}^{3} C_i(X_i - X_{i-1}) > 0\right) > 0.$$

(See Aletti and Capasso (2003).)

5.3. With reference to problem 5.2, consider a risk-free $\{\mathcal{F}_n\}_N$-strategy. Show that there exists an $n \in \{1, \ldots, N\}$ such that

$$C_n(X_n - X_{n-1}) \geq 0 \qquad \text{a.s.}, \qquad \mathbb{P}\left(C_n(X_n - X_{n-1}) > 0\right) > 0;$$

i.e., if no profit is available during any unit of time, then we cannot have a profit up to time N. (See Aletti and Capasso (2003).)

5.4. Consider a Black–Scholes market with $r = \mu = 0$ and $\sigma = 1$. Then a value-conserving strategy $H_t^{(S)} = 1/\sqrt{T}$ yields a portfolio value of $\Pi_t = \int_0^t dW_s/\sqrt{T-s}$. Show that

$$P(\Pi_\tau \geq c, 0 \leq \tau \leq T) = 1,$$

with c an arbitrary constant and τ a stopping time. Hence any amount can be obtained in finite time. It is easy to see that (unlike conditions 1 and 2) condition 3 of Proposition 5.6 is not automatically satisfied (see, e.g., Duffie (1996)).

5.5. For a drifting Wiener process $X_t = W_t + \mu t$, where W_t is P-Brownian motion and its maximum value attained is

$$M_t = \max_{\tau \in [0,t]} X_\tau,$$

apply the reflection principle and Girsanov's theorem to show that

$$P(X_T \leq a \cap M_T \geq b) = e^{2\mu b} P(X_T \geq 2b - a + 2\mu T)$$

for $a \leq b$ and $b \geq 0$. See also Musiela and Rutkowski (1998) or Borodin and Salminen (1996).

5.6. Referring to the Barrier option problem (5.21), show that

$$Q\left(\min_{t\in[0,T]} W_t^Q > g(b) \cap W_T^Q > g(K)\right)$$
$$= \Phi(d_1) - \left(\frac{b}{S_0}\right)^{\frac{2r}{\sigma^2}-1} \Phi\left(\frac{\ln\frac{b^2}{S_0 K} + \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right), \tag{5.46}$$

where d_1 is given by (5.15). From (5.46) obtain the joint density of S_T and its minimum over $[0, T]$, and thus solve

$$E_Q\left[S_T I_{\left[\min_{t\in[0,T]} S_t > b \cap S_T > K\right]}\right].$$

5.7. Why can it be conjectured that the bond equation in the Vasicek model is of the form

$$B_t = e^{C(t,T) - D(t,T)r_t}? \tag{5.47}$$

Derive the results (5.30) and (5.31). (*Hint:* Derive a partial differential equation for $P(t, T)$ using a similar argumentation as for the Black–Scholes equation. Substitute (5.47) and solve. (See also Hunt and Kennedy (2000).))

5.8. In the Brace–Gatarek–Musiela model, derive the nonarbitrage drifts (5.41) of the lognormal forward rates $F_t^{(i)}$. (*Hint:* In equation (5.40) note that $\frac{B_t^{(i)}}{B_t^{(N)}}$ is a martingale under Q_N. Given this, derive the drift as

$$\mu_i = -\frac{d\left\langle \ln F_t^{(i)}, \ln \frac{B_t^{(i)}}{B_t^{(N)}} \right\rangle}{dt}$$

and solve.)

5.9. A nonarbitrage argument shows that a so-called *swap rate* $S(t, T_s, T_e)$ has to satisfy

$$S(t, T_s, T_e) = \frac{\sum_{i=s+1}^{e} F_t^{(i-1)} B_t^{(i)} (T_i - T_{i-1})}{A_{s,e}}, \tag{5.48}$$

where

$$A_{s,e} = \sum_{i=s+1}^{e} B_t^i (t_i - t_{i-1})$$

is called an *annuity*. If relationship (5.39) holds and the forward rates are driven by (5.38), then show that the swap rate process can approximately be written as

$$dS(t, T_s, T_e) = \sigma_{S(t,T_s,T_e)} S(t, T_s, T_e) dW_t^{A_{s,e}},$$

where $\sigma_{S(t,T_s,T_e)}$ is deterministic and $dW_t^{A_{s,e}}$ is a Brownian motion under the martingale measure induced by taking $A_{s,e}$ as numeraire. (*Hint:* Assume that the coefficients of all the forward rates $F_t^{(i)}$ in (5.48) are approximately deterministic, invoke Itô's formula, and apply Girsanov's theorem.)

5.10. The *constant elasticity of variance* market (see Cox (1996) or Boyle and Tian (1999)) is a Black–Scholes market where the risky asset follows

$$dS_t = \mu S_t dt + \sigma S_t^{\frac{\alpha}{2}} dW_t$$

for $0 \leq \alpha < 2$. Show that this market has no equivalent risk-neutral measure.

6

Applications to Biology and Medicine

6.1 Population Dynamics: Discrete-in-Space–Continuous-in-Time Models

In the chapter on stochastic processes the Poisson process was introduced as an example of an RCLL nonexplosive counting process. Furthermore, we reviewed a general theory of counting processes as point processes on the real line within the framework of martingale theory and dynamics. Indeed, for these processes, under the usual regularity assumptions, we can invoke the Doob–Meyer decomposition theorem (see (2.79) onwards) and claim that any nonexplosive RCLL process $(X_t)_{t\in\mathbb{R}_+}$ satisfies a generalized stochastic differential equation of the form

$$dX_t = dA_t + dM_t \tag{6.1}$$

subject to a suitable initial condition. Here A is the compensator of the process representing the model of "evolution" and M is a martingale representing the "noise."

As was mentioned in the sections on counting and marked point processes, a counting process $(N_t)_{t\in\mathbb{R}_+}$ is a random process that counts the occurrence of certain events over time, namely N_t being the number of such events having occurred during the time interval $]0,t]$. We have noticed that a nonexplosive counting process is RCLL with upward jumps of magnitude 1 and we impose the initial condition $N_0 = 0$, almost surely. Since we deal with those counting processes that satisfy the conditions of Theorem 2.87 (local Doob–Meyer decomposition theorem), a nondecreasing predictable process $(A_t)_{t\in\mathbb{R}_+}$ (the compensator) exists such that $(N_t - A_t)_{t\in\mathbb{R}_+}$ is a right-continuous local martingale. Further, we assume that the compensator is absolutely continuous with respect to the usual Lebesgue measure on $\mathbb{R}_+$. In this case we say that $(N_t)_{t\in\mathbb{R}_+}$ has a (predictable) intensity $(\lambda_t)_{t\in\mathbb{R}_+}$ such that

$$A_t = \int_0^t \lambda_s ds \qquad \text{for any } t \in \mathbb{R}_+,$$

and the stochastic differential equation (6.1) can be rewritten as

$$dX_t = \lambda_t dt + dM_t. \tag{6.2}$$

If the process is integrable and λ is left-continuous with right limits (LCRL), one can easily show that

$$\lambda_t = \lim_{\Delta t \to 0+} \frac{1}{\Delta t} E[N_{t+\delta t} - N_t | \mathcal{F}_{t-}] \qquad \text{a.s.}$$

and, if we further assume the simplicity of the process, we also have

$$\lambda_t = \lim_{\Delta t \to 0+} \frac{1}{\Delta t} P(N_{t+\delta t} - N_t = 1 | \mathcal{F}_{t-}) \qquad \text{a.s.;}$$

i.e., $\lambda_t dt$ is the conditional probability of a new *event* during $[t, t+dt]$ given the history of the process during $[0, t]$. It really represents the model of evolution of the counting process, similar to classical deterministic differential equations.

Example 6.1. Let X be a nonnegative real random variable with absolutely continuous probability law having density f, cumulative distribution function F, survival function $S = 1 - F$, and hazard rate function $\alpha(t) = \frac{f(t)}{S(t)},\ t > 0$. Assume

$$\int_0^t \alpha(s) ds = -\ln(1 - F(t)) < +\infty \qquad \text{for any } t \in \mathbb{R}_+,$$

but

$$\int_0^{+\infty} \alpha(t) dt = +\infty.$$

Define the univariate process N_t by

$$N_t = I_{[X \leq t]}(t)$$

and let $(\mathcal{N}_t)_{t \in \mathbb{R}_+}$ be the filtration the process generates; i.e.,

$$\mathcal{N}_t = \sigma(N_s, s \leq t) = \sigma\left(X \wedge t, I_{[X \leq t]}(t)\right).$$

Define the left-continuous adapted process Y_t by

$$Y_t = I_{[X \geq t]}(t) = 1 - N_{t-}.$$

It can be easily shown (see, e.g., Andersen et al. (1993)) that N_t admits

$$A_t = \int_0^t Y_s \alpha(s) ds$$

as a compensator and hence N_t has stochastic intensity λ_t defined by

$$\lambda_t = Y_t \alpha(t), \qquad t \in \mathbb{R}_+.$$

In other words,

$$N_t - \int_0^{X \wedge t} \alpha(s) ds$$

is a local martingale. Here $\alpha(t)$ is a deterministic function, while Y_t, clearly, is a predictable process. This is a first example of what is known as a multiplicative intensity model.

Example 6.2. Let X be a random time as in the previous example, and let U be another random time, i.e., a nonnegative real random variable. Consider the random variable $T = X \wedge U$ and define the processes

$$N_t = I_{[T \leq t]} I_{[X \leq U]}(t)$$

and

$$N_t^U = I_{[T \leq t]} I_{[U < X]}(t)$$

and the filtration

$$\mathcal{N}_t = \sigma\left(N_s, N_s^U, s \leq t\right).$$

The hazard rate function α of X is known as the *net hazard rate*; it is given by

$$\alpha(t) = \lim_{h \to 0+} \frac{1}{h} P[t \leq X \leq t + h | X > t].$$

On the other hand, the quantity

$$\alpha^+(t) = \lim_{h \to 0+} \frac{1}{h} P[t \leq X \leq t + h | X > t, U > t]$$

is known as the *crude hazard rate*, whenever the limit exists. In this case

$$N_t - \int_0^t I_{[T \geq t]} \alpha(s) ds$$

is a local martingale.

Birth-and-Death Processes

A Markov birth-and-death process provides an example of a bivariate counting process. Let $(X_t)_{t \in \mathbb{R}_+}$ be the size of a population subject to a birth rate λ and a death rate μ. Then the infinitesimal transition probabilities are

$$P\left(X_{t+\delta t} = j | X_t = h\right) = \begin{cases} \lambda h \delta t + o(\delta t) & \text{if } j = h + 1, \\ \mu h \delta t + o(\delta t) & \text{if } j = h - 1, \\ 1 - (\lambda h + \mu h)\delta t + o(\delta t) & \text{if } j = h, \\ o(\delta t) & \text{otherwise.} \end{cases}$$

Let $N_t^{(1)}$ and $N_t^{(2)}$ be the number of births and deaths, respectively, up to time $t \geq 0$, assuming $N_0^{(1)} = 0$ and $N_0^{(2)} = 0$. Then

$$(\mathbf{N}_t)_{t \in \mathbb{R}_+} = \left(N_t^{(1)}, N_t^{(2)}\right)$$

is a bivariate counting process with intensity process $(\lambda X_{t-}, \mu X_{t-})_{t \in \mathbb{R}_+}$ (see Figures 6.1 and 6.2). This is an example of a formulation of a Markov process with countable state space as a counting process. In particular, we may write a stochastic differential equation for X_t as follows:

$$dX_t = \lambda X_{t-} dt - \mu X_{t-} dt + dM_t,$$

where M_t is a suitable martingale noise.

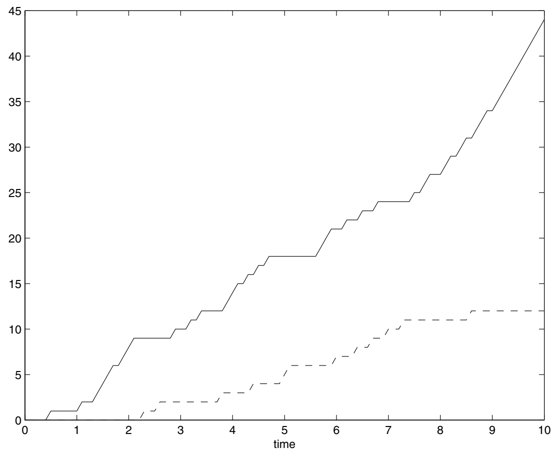

Fig. 6.1. Simulation of a birth-and-death process with birth rate $\lambda = 0.2$, death rate $\mu = 0.05$, initial population $X_0 = 10$, time step $dt = 0.1$, and interval of observation $[0, 10]$. The continuous line represents the number of births $N_t^{(1)}$; the dashed line represents the number of deaths $N_t^{(2)}$.

A Model for Software Reliability

Let N_t denote the number of software failures detected during the time interval $]0, t]$ and suppose that F is the true number of faults existing in the software at time $t = 0$. In the Jelinski–Moranda model (see Jelinski and Moranda (1972)) it is assumed that N_t is a counting process with intensity

$$\lambda_t = \rho(F - N_{t-}),$$

where ρ is the individual failure rate (see Figure 6.3). One may note that this model corresponds to a pure death process in which the total initial population F usually is unknown, as is the rate ρ.

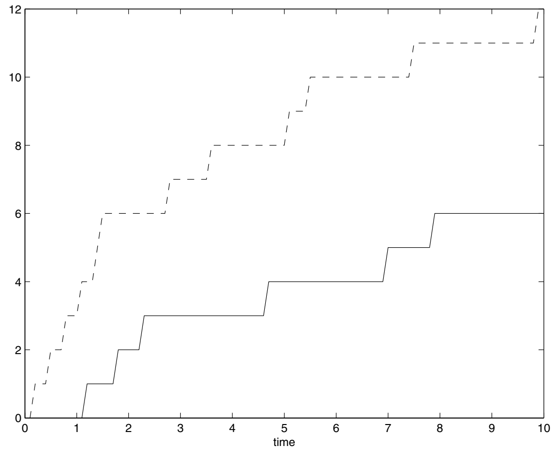

Fig. 6.2. Simulation of a birth-and-death process with birth rate $\lambda = 0.09$, death rate $\mu = 0.2$, initial population $X_0 = 10$, time step $dt = 0.1$, and interval of observation $[0, 10]$. The continuous line represents the number of births $N_t^{(1)}$; the dashed line represents the number of deaths $N_t^{(2)}$.

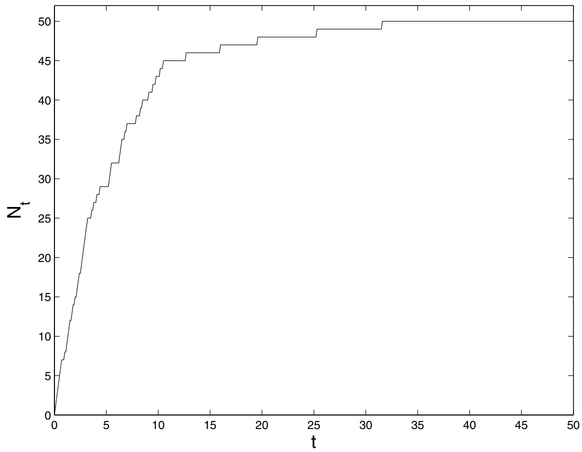

Fig. 6.3. Simulation of a model for software reliability: individual failure rate $\rho = 0.2$, true initial number of faults $F = 50$, time step $dt = 0.1$, and interval of observation $[0, 50]$.

Contagion: The Simple Epidemic Model

Epidemic systems provide models for the transmission of a contagious disease within a population. In the "simple epidemic model" (Bailey (1975) and Becker (1989)) the total population N is divided into two main classes:

(S) the class of susceptibles, including those individuals capable of contracting the disease and becoming infectives themselves;
(I) the class of infectives, including those individuals who, having contracted the disease, are capable of transmitting it to susceptibles.

Let I_t denote the number of individuals who have been infected during the time interval $]0, t]$. Assume that individuals become infectious themselves immediately upon infection and remain so for the entire duration of the epidemic. Suppose that at time $t = 0$ there are S_0 susceptible individuals and I_0 infectives in the community. The classical model based on the *law of mass action* (see, e.g., Bailey (1975) or Capasso (1993)) assumes that the counting process I_t has stochastic intensity

$$\lambda_t = \beta_t(I_0 + I_{t-})(S_0 - I_{t-}),$$

which is appropriate when the community is mixing uniformly. Here β_t is called the *infection rate*(see Figure6.4).

It can be noted that formally this corresponds to writing $N(t)$ with the stochastic differential equation

$$dI_t = \beta_t(I_0 + I_{t-})(S_0 - I_{t-})dt + dM_t,$$

where M_t is a suitable martingale noise. In this case we obtain

$$\langle M \rangle_t = \int_0^t \lambda_s ds$$

for the variation process $\langle M \rangle_t$, so that

$$M_t^2 - \int_0^t \lambda_s ds$$

is a zero mean martingale. As a consequence

$$Var[M_t] = E\left[\int_0^t \lambda_s ds\right] = E[I_t].$$

More general models can be found in Capasso (1990) and references therein.

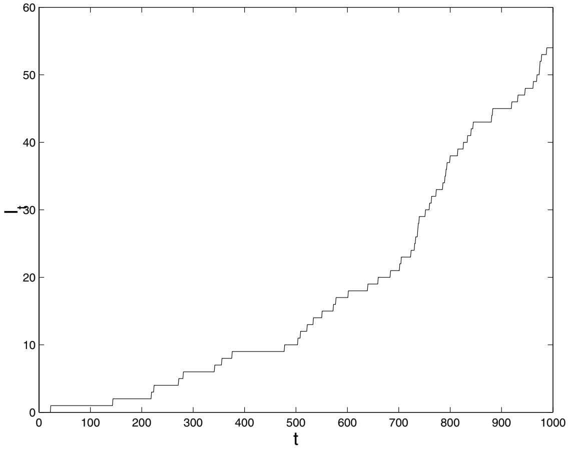

Fig. 6.4. Simulation of a simple epidemic (SI) model: initial number of susceptibles $S_0 = 500$, initial number of infectives $I_0 = 4$, infection rate (constant) $\beta = 5 \cdot 10^{-6}$, time step $dt = 1$, interval of observation $[0, 1000]$.

Contagion: The General Stochastic Epidemic

For a wide class of epidemic models the total population $(N_t)_{t\in\mathbb{R}_+}$ includes three subclasses. In addition to the classes of susceptibles $(S_t)_{t\in\mathbb{R}_+}$ and infectives $(I_t)_{t\in\mathbb{R}_+}$ already introduced in the simple model, a third class is considered, i.e.,

(R), the class of *removals*. This comprises those individuals who, having contracted the disease, and thus being already infectives, are no longer in the position of transmitting the disease to other susceptibles because of death, immunization, or isolation. Let us denote the number of removals as $(R_t)_{t\in\mathbb{R}_+}$.

The process $(S_t, I_t, R_t)_{t\in\mathbb{R}_+}$ is modelled as a multivariate jump Markov process valued in $E' = \mathbb{N}^3$. Actually, if we know the behavior of the total population process N_t, because

$$S_t + I_t + R_t = N_t \qquad \text{for any } t \in \mathbb{R}_+,$$

we need to provide a model only for the bivariate process $(S_t, I_t)_{t\in\mathbb{R}_+}$, which is now valued in $E = \mathbb{N}^2$. The only nontrivial elements of a resulting intensity matrix Q (see chapter on Markov processes) are given by

- $q_{(s,i),(s+1,i)} = \alpha$, birth of a susceptible;
- $q_{(s,i),(s-1,i)} = \gamma s$, death of a susceptible;
- $q_{(s,i),(s,i+1)} = \beta$, birth of an infective;
- $q_{(s,i),(s,i-1)} = \delta i$, removal of an infective;

- $q_{(s,i),(s-1,i+1)} = \kappa si$, infection of a susceptible.

For $\alpha = \beta = \gamma = 0$, we have the so-called *general stochastic epidemic* (see, e.g., Bailey (1975) and Becker(1989)). In this case the total population is constant (assume $R_0 = 0$; see Figure 6.5):

$$N_t \equiv N = S_0 + I_0 \qquad \text{for any } t \in \mathbb{R}_+.$$

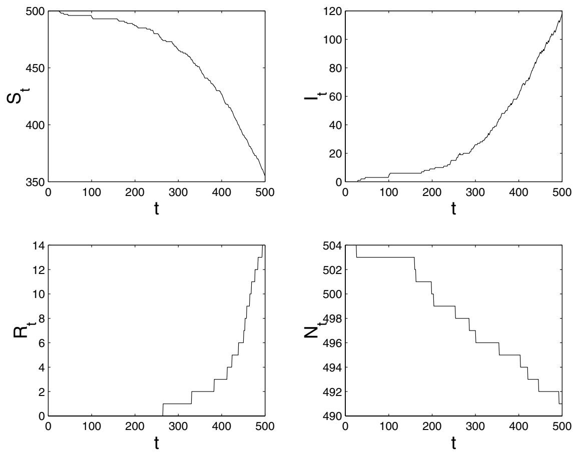

Fig. 6.5. Simulation of an SIR epidemic model with vital dynamics: initial number of susceptibles $S_0 = 500$, initial number of infectives $I_0 = 4$, initial number of removed $R_0 = 0$, birth rate of susceptibles $\alpha = 10^{-4}$, death rate of a susceptible $\gamma = 5 \cdot 10^{-5}$, birth rate of an infective $\beta = 10^{-5}$, rate of removal of an infective $\delta = 8.5 \cdot 10^{-4}$, infection rate of a susceptible $k = 1.9 \cdot 10^{-5}$, time step $dt = 1$, interval of observation $[0, 500]$.

Contagion: Diffusion of Innovations

When a new product is introduced in a market, its diffusion is due to a process of adoption by individuals who are aware of it. Classical models of innovation diffusion are very similar to epidemic systems, even though in this case rates of adoption (infection) depend upon specific marketing and advertising strategies (see Capasso, Di Liddo, and Maddalena (1994) and Mahajan and Wind (1986)). In this case the total population N of possible consumers is divided into the following main classes:

(S) the class of potential adopters, including those individuals capable of adopting the new product, thus themselves becoming adopters;

(A)the class of adopters, those individuals who, having adopted the new product, are capable of transmitting it to potential adopters.

Let A_t denote the number of individuals who, by time $t \geq 0$, have already adopted a new product that has been put on the market at time $t = 0$. Suppose that at time $t = 0$ there are S_0 potential adopters and A_0 adopters in the market. In the basic models it is assumed that all consumers are homogeneous with respect to their inclination to adopt the new product. Moreover, all adopters are homogeneous in their ability to persuade others to try new products, and adopters never lose interest, but continue to inform those consumers who are not aware of the new product. Under these assumptions the classical model for the adoption rate is again based on the law of mass action (see Bartholomew (1976)), apart from an additional parameter $\lambda_0(t)$ that describes adoption induced by external actions, independent of the number of adopters, such as advertising, price reduction policy, etc. Then the stochastic intensity for this process is given by

$$\lambda(t) = (\lambda_0(t) + \beta_t A_{t-})(S_0 - A_{t-}),$$

which is appropriate when the community is mixing uniformly. Here β_t is called the *adoption rate*(see Figure 6.6).

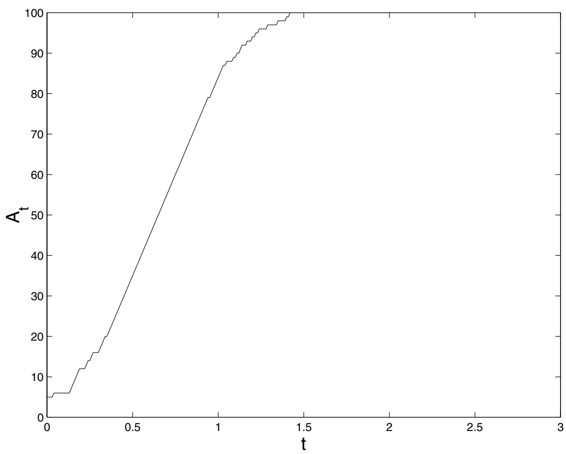

Fig. 6.6. Simulation of the contagion model for diffusion of innovations: external influence $\lambda_0(t) = 5 \cdot 10^{-4} t$, adoption rate (constant) $\beta = 0.05$, initial potential adopters $S_0 = 100$, initial adopters $A_0 = 5$, time step $dt = 0.01$, interval of observation $[0, 3]$.

Inference for Multiplicative Intensity Processes

Let

$$dN_t = \alpha_t Y_t dt + dM_t$$

be a stochastic equation for a counting process N_t, where the noise is a zero mean martingale. Furthermore, let

$$B_s = \frac{J_{s-}}{Y_s} \qquad \text{with} \qquad J_s = I_{[Y_s>0]}(s).$$

B_t is, like Y_t, a predictable process, so that by the integration theorem,

$$M_t^* = \int_0^t B_s dM_s$$

is itself a zero mean martingale. It can be noted that

$$M_t^* = \int_0^t B_s dM_s = \int_0^t B_s dN_s - \int_0^t \alpha_s J_{s-} ds,$$

so that

$$E\left[\int_0^t B_s dN_s\right] = E\left[\int_0^t \alpha_s J_{s-} ds\right];$$

i.e., $\int_0^t B_s dN_s$ is an unbiased estimator of $E[\int_0^t \alpha_s J_{s-} ds]$. If α is constant and we stop the process at a time T such that $Y_t > 0, t \in [0, T]$, then

$$\hat{\alpha} = \frac{1}{T}\int_0^T \frac{dN_s}{Y_s}$$

is an unbiased estimator of α. This method of inference is known as Aalen's method (Aalen (1978)) (the reader may also refer to Andersen et al. (1993) for an extensive application of this method to the statistics of counting processes).

Inference for the Simple Epidemic Model

We may apply the above procedure to the simple epidemic model as discussed in Becker (1989). Let

$$B_s = \frac{I_{[S_s>0]}(s)}{I_{s-}S_{s-}}$$

and suppose β is constant. Let T be such that $S_t > 0, t \in [0, T]$. Then an unbiased estimator for β would be

$$\hat{\beta} = \frac{1}{T}\int_0^T \frac{dI_s}{S_{s-}I_{s-}} = \frac{1}{T}\frac{1}{S_0 I_0} + \frac{1}{(S_0-1)(I_0-1)} + \cdots + \frac{1}{(S_T+1)(I_T+1)}.$$

The standard error (SE) of $\hat{\beta}$ is

$$\frac{1}{T}\left(\int_0^T B_s^2 dI_s\right)^2.$$

By the central limit theorem for martingales (see Rebolledo (1980)) we can also deduce that

$$\frac{\hat{\beta}-\beta}{SE(\hat{\beta})}$$

has an asymptotic $N(0,1)$ distribution, which leads to confidence intervals and hypothesis testing on the model in the usual way (see Becker (1989) and references therein).

Inference for a General Epidemic Model

In Yang (1985) a model was proposed as an extension of the *general epidemic model* presented above. The epidemic process is modelled in terms of a multivariate jump Markov process $(S_t, I_t, R_t)_{t\in\mathbb{R}_+}$, or simply $(S_t, I_t)_{t\in\mathbb{R}_+}$, when the total population is constant, i.e.,

$$N_t := S_t + I_t + R_t = N + 1.$$

In this case, if we further suppose that $S_0 = N$, $I_0 = 1$, $R_0 = 0$, instead of using (S_t, I_t), the epidemic may be described by the number of infected individuals (not including the initial case) $M_1(t)$ and the number of removals $M_2(t) = R_t$ during $]0,t]$, $t \in \mathbb{R}_+^*$. Since we are dealing with a finite total population, the number of infected individuals and the number of removals are bounded, so that

$$E[M_k(t)] \leq N + 1, \qquad k = 1, 2.$$

The processes $M_1(t)$ and $M_2(t)$ are submartingales with respect to the history $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ of the process, i.e., the filtration generated by all relevant processes. We assume that the two processes admit multiplicative stochastic intensities of the form

$$\begin{aligned} \Lambda_1(t) &= \kappa G_1(t-)(N - M_1(t-)),\\ \Lambda_2(t) &= \delta(1 + M_1(t-) - M_2(t-)), \end{aligned}$$

respectively, where $G_1(t)$ is a known function of infectives in circulation at time t. It models the release of pathogen material by infected individuals. Hence

$$Z_k(t) = M_k(t) - \int_0^t \Lambda_k(s)ds, \qquad k = 1, 2,$$

are orthogonal martingales with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$. As a consequence, Aalen's unbiased estimators for the infection rate κ and the removal rate δ are given by

$$\hat{\kappa} = \frac{M_1(t)}{B_1(t-)}, \quad \hat{\delta} = \frac{M_1(t)}{B_1(t-)},$$

where

$$B_1(t) = \int_0^t G_1(s)(N - M_1(s))ds,$$
$$B_2(t) = \int_0^t (1 + M_1(s) - M_2(s))ds.$$

Theorem 1.3 in Jacobsen (1982), page 163, gives conditions for a multivariate martingale sequence to converge to a normal process. If such conditions are met, then as $N \to \infty$,

$$\begin{pmatrix} \sqrt{B_1(t)}(\hat{\kappa} - \kappa) \\ \sqrt{B_2(t)}(\hat{\delta} - \delta) \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Gamma\right),$$

where

$$\Gamma = \begin{pmatrix} \kappa & 0 \\ 0 & \delta \end{pmatrix}.$$

In general, it is not easy to verify the conditions of this theorem. They surely hold for the simple epidemic model presented above, where $\delta = 0$. Related results are given in Ethier and Kurtz (1986) and Wang (1977) for a scaled infection rate $\kappa \to \frac{\kappa}{N}$ (see the following section). See also Capasso (1990) for additional models and related inference problems.

6.2 Population Dynamics: Continuous Approximation of Jump Models

A more realistic model than the general stochastic epidemic of the preceding section, which takes into account a rescaling of the force of infection due to the size of the total population, is the following (see Capasso (1993)):

$$q_{(s,i),(s-1,i+1)} = \frac{\kappa}{N} si = N\kappa \frac{s}{N}\frac{i}{N}.$$

We may also rewrite

$$q_{(s,i),(s,i-1)} = \delta N \frac{i}{N} = \frac{i}{N},$$

so that both transition rates are of the form

$$q_{k,k+l}^{(N)} = N\beta_l\left(\frac{k}{N}\right)$$

for

$$k = (s, i)$$

and

$$k + l = \begin{cases} (s, i-1), \\ (s-1, i+1). \end{cases}$$

This model is a particular case of the following situation:

Let $E = \mathbb{Z}^d \cup \{\Delta\}$, where Δ is the point at infinity of $\mathbb{Z}^d$, $d \geq 1$. Further, let

$$\beta_l : \mathbb{Z}^d \to \mathbb{R}_+, \qquad l \in \mathbb{Z}^d,$$

$$\sum_{l \in \mathbb{Z}^d} \beta_l(k) < +\infty \qquad \text{for each} \qquad k \in \mathbb{Z}^d.$$

For f defined in $\mathbb{Z}^d$, and vanishing outside a finite subset of $\mathbb{Z}^d$, let

$$\mathcal{A}f(x) = \begin{cases} \sum_{l \in \mathbb{Z}^d} \beta_l(x)(f(x+l) - f(x)), & x \in \mathbb{Z}^d, \\ 0, \quad x = \Delta. \end{cases}$$

Let $(Y_l)_{l \in \mathbb{Z}^d}$ be a family of independent standard Poisson processes. Let $X(0) \in \mathbb{Z}^d$ be nonrandom and suppose

$$X(t) = X(0) + \sum_{l \in \mathbb{Z}^d} l Y_l \left(\int_0^t \beta_l(X(s)) ds \right), \qquad t < \tau_\infty, \tag{6.3}$$

$$X(t) = \Delta, \qquad t \geq \tau_\infty, \tag{6.4}$$

where

$$\tau_\infty = \inf\{t | X(t-) = \Delta\}.$$

The following theorem holds (see Ethier and Kurtz (1986), page 327).

Theorem 6.3. *1. Given $X(0)$, the solution of system (6.3)–(6.4) above is unique.*
2. If $\mathcal{A}$ is a bounded operator, then X is a solution of the martingale problem for $\mathcal{A}$.

As a consequence, for our class of models for which

$$q^{(N)}_{k,k+l} = N \beta_l \left(\frac{k}{N} \right), \qquad k \in \mathbb{Z}^d, l \in \mathbb{Z}^d,$$

we have that the corresponding Markov process, which we shall denote by $\hat{X}^{(N)}$, satisfies, for $t < \tau_\infty$:

$$\hat{X}^{(N)}(t) = \hat{X}^{(N)}(0) + \sum_{l \in \mathbb{Z}^d} l Y_l \left(N \int_0^t \beta_l \left(\frac{\hat{X}^{(N)}(s)}{N} \right) ds \right),$$

where the Y_l are independent standard Poisson processes. By setting

$$F(x) = \sum_{l \in \mathbb{Z}^d} l \beta_l(x), \qquad x \in \mathbb{R}^d,$$

and

$$X^{(N)} = \frac{1}{N}\hat{X}^{(N)},$$

we have

$$X^{(N)}(t) = X^{(N)}(0) + \sum_{l\in\mathbb{Z}^d} \frac{l}{N}\tilde{Y}_l\left(N\int_0^t \beta_l\left(X^{(N)}(s)\right)ds\right)$$
$$+\int_0^t F(X^{(N)}(s))ds, \qquad (6.5)$$

where

$$\tilde{Y}_l(u) = Y_l(u) - u$$

is the centered standard Poisson process. The state space for $X^{(N)}$ is

$$E_N = E \cap \left\{\frac{k}{N}, k \in \mathbb{Z}^d\right\}$$

for $E \subset \mathbb{R}^d$. We require that $x \in E_N$ and $\beta_(x) > 0$ imply $x + \frac{l}{N} \in E_N$. The generator for $X^{(N)}$ is

$$\begin{aligned}
&\mathcal{A}^{(N)}f(x)\\
&= \sum_{l\in\mathbb{Z}^d} N\beta_l(x)\left(f\left(x+\frac{l}{N}\right) - f(x)\right)\\
&= \sum_{l\in\mathbb{Z}^d} N\beta_l(x)\left(f\left(x+\frac{l}{N}\right) - f(x) - \frac{l}{N}\nabla f(x)\right) + F(x)\nabla f(x), \qquad x \in E_N.
\end{aligned}$$

By the strong law of large numbers, we know that

$$\lim_{N\to\infty}\sup_{u\leq v}\left|\frac{1}{N}\tilde{Y}_l(Nu)\right| = 0, \qquad \text{a.s.}$$

for any $v \geq 0$. As a consequence, the following theorem holds (Ethier and Kurtz (1986), page 456).

Theorem 6.4. *Suppose that for each compact $K \subset E$,*

$$\sum_{l\in\mathbb{Z}^d} |l| \sup_{x\in K} \beta_l(x) < +\infty,$$

and there exists $M_K > 0$ such that

$$|F(x) - F(y)| \leq M_K|x-y|, \qquad x, y \in K;$$

suppose $X^{(N)}$ satisfies equation (6.5) above, with

$$\lim_{N\to\infty} X^{(N)}(0) = x_0 \in \mathbb{R}^d.$$

Then, for every $t \geq 0$,

$$\lim_{N\to\infty} \sup_{s\leq t} \left| X^{(N)}(s) - x(s) \right| = 0 \qquad a.s.,$$

where $x(t)$, $t \in \mathbb{R}_+$ *is the unique solution of*

$$x(t) = x_0 + \int_0^t F(x(s))ds, \qquad t \geq 0,$$

wherever it exists.

For the application of the above theorem to the general stochastic epidemic introduced at the beginning of this section see problem 6.9. For a graphical illustration of the above see Figures 6.7 and 6.8. Further, and interesting examples may be found in section 6.4 of Tan (2002).

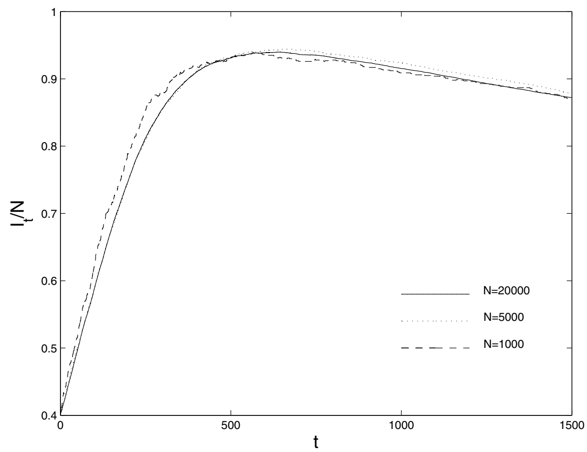

Fig. 6.7. Continuous approximation of a jump model: general stochastic epidemic model with $S_0 = 0.6N$, $I_0 = 0.4N$, $R_0 = 0$, rate of removal of an infective $\delta = 10^{-4}$; infection rate of a susceptible $k = 8 \cdot 10^{-3}N$; time step $dt = 10^{-2}$; interval of observation $[0, 1500]$. The three lines represent the simulated I_t/N as a function of time t for three different values of N.

6.3 Population Dynamics: Individual-Based Models

The scope of this chapter is to introduce the reader to the modeling of a system of a large but still finite population of individuals subject to mutual interaction and random dispersal. These systems may well describe the collective

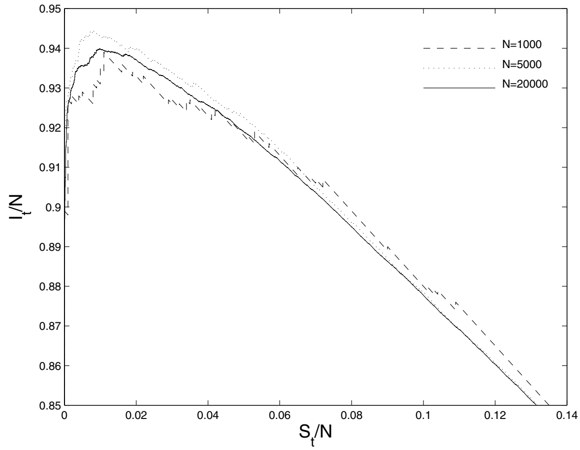

Fig. 6.8. Continuous approximation of a jump model: the same model as in Figure 6.7 of a general stochastic epidemic model with $S_0 = 0.6N$, $I_0 = 0.4N$, $R_0 = 0$, rate of removal of an infective $\delta = 10^{-4}$, infection rate of a susceptible $k = 8 \cdot 10^{-3}N$, time step $dt = 10^{-2}$, interval of observation $[0, 1500]$. The three lines represent the simulated trajectory $(S_t/N, I_t/N)$ for three different values of N.

behavior of individuals in herds, swarms, colonies, armies, etc. (examples can be found in Burger, Capasso, and Morale (2003), Durrett and Levin (1994), Flierl et al. (1999), Gueron, Levin, and Rubenstein (1996), Okubo (1986), Skellam (1951)). It is interesting to observe that under suitable conditions the behavior of such systems in the limit of the number of individuals tending to infinity may be described in terms of nonlinear reaction-diffusion systems. We may then claim that while stochastic differential equations may be utilized for modeling populations at the *microscopic* scale of individuals (Lagrangian approach), partial differential equations provide a *macroscopic* Eulerian description of population densities.

Up to now, Kolmogorov equations like that of Black–Scholes were linear partial differential equations; in this chapter we derive nonlinear partial differential equations for density-dependent diffusions. This field of research, already well established in the general theory of statistical physics (see, e.g., De Masi and Presutti (1991), Donsker and Varadhan (1989), Méléard (1996)), has gained increasing attention, since it also provides the framework for the modelling, analysis, and simulation of agent-based models in economics and finance (see, e.g., Epstein and Axtell (1996)).

The Empirical Distribution

We start from the Lagrangian description of a system of $N \in \mathbb{N} \setminus \{0, 1\}$ particles. Suppose the kth particle ($k \in \{1, \ldots, N\}$) is located at $X_N^k(t)$, at time

$t \geq 0$. Each $(X_N^k(t))_{t\in\mathbb{R}_+}$ is a stochastic process valued in the state space $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, $d \in \mathbb{N} \setminus \{0\}$, on a common probability space $(\Omega, \mathcal{F}, P)$. An equivalent description of the above system may be given in terms of the (random) measures $\epsilon_{X_N^k(t)}$ $(k = 1, 2, \ldots, N)$ on $\mathcal{B}_{\mathbb{R}^d}$ such that, for any real function $f \in C_0(\mathbb{R}^d)$ we have

$$\int_{\mathbb{R}^d} f(y)\epsilon_{X_N^k(t)}(dy) = f\left(X_N^k(t)\right).$$

As a consequence, information about the collective behavior of the N particles is provided by the so-called *empirical measure*, i.e., the random measure on $\mathbb{R}^d$:

$$X_N(t) := \frac{1}{N}\sum_{k=1}^{N} \epsilon_{X_N^k(t)}, \qquad t \in \mathbb{R}_+.$$

This measure may be considered as the empirical spatial distribution of the system. It is such that for any $f \in C_0(\mathbb{R}^d)$:

$$\int_{\mathbb{R}^d} f(y)[X_N(t)](dy) = \frac{1}{N}\sum_{k=1}^{N} f\left(X_N^k(t)\right).$$

In particular, given a region $B \in \mathcal{B}_{\mathbb{R}^d}$, the quantity

$$[X_N(t)](B) := \frac{1}{N}\left(\sharp\left\{X_N^k(t) \in B\right\}\right)$$

denotes the relative frequency of individuals, out of N, that at time t stay in B. This is why the measure-valued process

$$X_N : t \in \mathbb{R}_+ \to X_N(t) = \frac{1}{N}\sum_{k=1}^{N} \epsilon_{X_N^k(t)} \in \mathcal{M}_{\mathbb{R}^d} \tag{6.6}$$

is called the process of *empirical distributions* of the system of N particles.

The Evolution Equations

The Lagrangian description of the dynamics of the system of interacting particles is given via a system of stochastic differential equations. Suppose that for any $k \in \{1, \ldots, N\}$, the process $(X_N^k(t))_{t\in\mathbb{R}_+}$ satisfies the stochastic differential equation

$$dX_N^k(t) = F_N[X_N(t)](X_N^k(t))dt + \sigma_N dW^k(t), \tag{6.7}$$

subject to a suitable initial condition $X_N^k(0)$, which is an $\mathbb{R}^d$-valued random variable. Thus we are assuming that the kth particle is subject to random dispersal, modelled as a Brownian motion W^k. In fact, we suppose that W^k,

$k = 1, \ldots, N$, is a family of independent standard Wiener processes. Furthermore the common variance σ_N^2 may depend on the total number of particles.

The drift term is defined in terms of a given function

$$F_N : \mathcal{M}_{\mathbb{R}^d} \to C(\mathbb{R}^d)$$

and it describes the "interaction" of the kth particle located at $X_N^k(t)$ with the random field $X_N(t)$ generated by the whole system of particles at time t. An evolution equation for the empirical process $(X_N(t))_{t\in\mathbb{R}_+}$ can be obtained thanks to Itô's formula. For each individual particle $k \in \{1, \ldots, N\}$ subject to its stochastic differential equation, given $f \in C_b^2(\mathbb{R}^d \times \mathbb{R}_+)$, we have

$$\begin{aligned} f\left(X_N^k(t), t\right) = f\left(X_N^k(0), 0\right) &+ \int_0^t F_N[X_N(s)]\left(X_N^k(s)\right) \nabla f\left(X_N^k(s), s\right) ds \\ &+ \int_0^t \left[\frac{\partial}{\partial s} f\left(X_N^k(s), s\right) + \frac{\sigma_N^2}{2} \triangle f\left(X_N^k(s), s\right)\right] ds \\ &+ \sigma_N \int_0^t \nabla f\left(X_N^k(s), s\right) dW^k(s). \end{aligned} \tag{6.8}$$

Correspondingly, for the empirical process $(X_N(t))_{t\in\mathbb{R}_+}$, we get the following weak formulation of its evolution equation. For any $f \in C_b^{2,1}(\mathbb{R}^d \times \mathbb{R}_+)$ we have

$$\begin{aligned} \langle X_N(t), f(\cdot, t)\rangle = \langle X_N(0), f(\cdot, 0)\rangle &+ \int_0^t \langle X_N(s), F_N[X_N(s)](\cdot)\nabla f(\cdot, s)\rangle\, ds \\ &+ \int_0^t \left\langle X_N(s), \frac{\sigma_N^2}{2} \triangle f(\cdot, s) + \frac{\partial}{\partial s} f(\cdot, s)\right\rangle ds \\ &+ \frac{\sigma_N}{N} \int_0^t \sum_k \nabla f\left(X_N^k(s), s\right) dW^k(s). \end{aligned} \tag{6.9}$$

In the previous expressions, we have used the notation

$$\langle \mu, f\rangle = \int f(x)\mu(dx), \tag{6.10}$$

for any measure μ on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ and any (sufficiently smooth) function $f : \mathbb{R}^d \to \mathbb{R}$.

The last term of (6.9) is a martingale with respect to the process's $(X_N(t))_{t\in\mathbb{R}_+}$ natural filtration. Hence we may apply Doob's inequality (see Proposition 2.69) such that

$$E\left[\sup_{t\leq T} |M_N(f, t)|^2\right] \leq \frac{4\sigma_N^2 \|\nabla f\|_\infty^2 T}{N}.$$

This shows that, for N sufficiently large, the martingale term, which is the only source of stochasticity of the evolution equation for $(X_N(t))_{t\in\mathbb{R}_+}$, tends to zero, for N tending to infinity, since ∇f is bounded in $[0,T]$, and $\frac{\sigma_N^2}{N} \to 0$ for N tending to infinity. Under these conditions we may conjecture that a limiting measure-valued deterministic process $(X_\infty(t))_{t\in\mathbb{R}_+}$ exists, whose evolution equation (in weak form) is

$$\langle X_\infty(t), f(\cdot,t)\rangle = \langle X_\infty(0), f(\cdot,0)\rangle + \int_0^t \langle X_\infty(s), F[X_\infty(s)](\cdot)\nabla f(\cdot,s)\rangle \, ds$$
$$+ \int_0^t \left\langle X_\infty(s), \frac{\sigma_\infty^2}{2}\triangle f(\cdot,s) + \frac{\partial}{\partial s} f(\cdot,s)\right\rangle ds$$

for $\sigma_\infty^2 \geq 0$. Actually, various nontrivial mathematical problems arise in connection with the existence of a limiting measure-valued process $(X_\infty(t))_{t\in\mathbb{R}_+}$. A typical resolution includes the following:

1. Prove the existence of a deterministic limiting measure-valued process $(X_\infty(t))_{t\in\mathbb{R}_+}$.
2. Prove the absolute continuity of the limiting measure with respect to the usual Lebesgue measure on $\mathbb{R}^d$.
3. Provide an evolution equation for the density $p(x,t)$.

In the following subsections we will show how the above procedure has been carried out in particular cases.

A "Moderate" Repulsion Model

As an example we consider the system (due to Oelschläger (1990))

$$dX_N^k(t) = -\frac{1}{N}\sum_{m=1,m\neq k}^{N} \nabla V_N\left(X_N^k(t) - X_N^m(t)\right) dt + dW^k(t), \tag{6.11}$$

where W^k, $k = 1,\ldots,N$, represent N independent standard Brownian motions valued in $\mathbb{R}^d$ (here all variances are set equal to 1). The kernel V_N is chosen of the form

$$V_N(\mathbf{x}) = \chi_N^d V_1(\chi_N \mathbf{x}), \qquad \mathbf{x} \in \mathbb{R}^d, \tag{6.12}$$

where V_1 is a symmetric probability density with compact support in $\mathbb{R}^d$ and

$$\chi_N = N^{\frac{\beta}{d}}, \qquad \beta \in]0,1[. \tag{6.13}$$

With respect to the general structure introduced in the preceding subsection on evolution equations, we have assumed that the drift term is given by

$$F_N[X_N(t)]\left(X_N^k(t)\right) = [\nabla V_N * X_N(t)]\left(X_N^k(t)\right)$$
$$= -\frac{1}{N}\sum_{m=1,m\neq k}^{N} \nabla V_N\left(X_N^k(t) - X_N^m(t)\right).$$

System (6.11) describes a population of N individuals, subject to random dispersal (Brownian motion) and to repulsion within the range of the kernel V_N. The choice of the scaling (6.12) in terms of the parameter β means that the range of interaction of each individual with the rest of the population is a decreasing function of N (correspondingly, the strength is an increasing function of N). On the other hand, the fact that β is chosen to belong to $]0,1[$ is relevant for the limiting procedure. It is known as *moderate interaction* and allows one to apply suitable convergence results (*laws of large numbers*) (see Oelschläger (1985)).

For the sake of useful regularity conditions, we assume that

$$V_1 = W_1 * W_1,$$

where W_1 is a symmetric probability density with compact support in $\mathbb{R}^d$, satisfying the condition

$$\int_{\mathbb{R}^d} (1+|\lambda|^\alpha)|\widetilde{W_1}(\lambda)|^2 d\lambda < \infty \tag{6.14}$$

for some $\alpha > 0$ (here $\widetilde{W_1}$ denotes the Fourier transform of W_1). Henceforth we also make use of the following notations:

$$W_N(x) = \chi_N^d W_1(\chi_N x), \tag{6.15}$$

$$h_N(x,t) = (X_N(t) * W_N)(x), \tag{6.16}$$

$$V_N(x) = \chi_N^d V_1(\chi_N x) = (W_N * W_N)(x), \tag{6.17}$$

$$g_N(x,t) = (X_N(t) * V_N)(x) = (h_N(\cdot,t) * W_N)(x), \tag{6.18}$$

so that system (6.11) can be rewritten as

$$dX_N^k(t) = -\nabla g_N(X_N^k(t),t)dt + dW^k(t), \qquad k = 1,\dots,N. \tag{6.19}$$

The following theorem holds.

Theorem 6.5. *Let*

$$X_N(t) = \frac{1}{N}\sum_{k=1}^N \epsilon_{X_N^k(t)}$$

be the empirical process associated with system (6.11). Assume that

1. *condition (6.14) holds;*
2. $\beta \in]0, \frac{d}{d+2}[$;
3. $$\sup_{N\in\mathbb{N}} E\left[\langle X_N(0), \varphi_1\rangle\right] < \infty, \qquad \varphi_1(x) = (1+x^2)^{1/2}; \tag{6.20}$$
4. $$\sup_{N\in\mathbb{N}} E\left[||h_N(\cdot,0)||_2^2\right] < \infty; \tag{6.21}$$

5.

$$\lim_{N\to\infty} \mathcal{L}\left(X_N(0)\right) = \epsilon_{\Xi_0} \ \textit{in}\ \mathcal{M}(\mathcal{M}(\mathbb{R}^d)), \tag{6.22}$$

where Ξ_0 *is a probability measure having a density* $p_0 \in C_b^{2+\alpha}(\mathbb{R}^d)$ *with respect to the usual Lebesgue measure on* $\mathbb{R}^d$.

Then the empirical process X_N *converges to* X_∞, *which admits a density satisfying the evolution equation*

$$\begin{aligned} \frac{\partial}{\partial t} p(x,t) &= \frac{1}{2}\left(\Delta p(x,t)\right)^2 + \frac{1}{2}\Delta p(x,t), \\ &= \nabla(p(x,t)\nabla p(x,t)) + \frac{1}{2}\Delta p(x,t), \\ p(\cdot,0) &= p_0. \end{aligned} \tag{6.23}$$

More precisely,

$$\lim_{N\to\infty} \mathcal{L}(X_N) = \epsilon_{\Xi} \ \textit{in}\ \mathcal{M}(C([0,T],\mathcal{M}(\mathbb{R}^d))), \tag{6.24}$$

where

$$\Xi = (\Xi(t))_{0\le t\le T} \in C([0,T],\mathcal{M}(\mathbb{R}^d))$$

admits a density

$$p \in C_b^{2+\alpha,1+\frac{\alpha}{2}}(\mathbb{R}^d \times [0,T]),$$

which satisfies

$$\begin{aligned} \frac{\partial}{\partial t} p(x,t) &= \frac{1}{2}\nabla(1+2p(x,t))\nabla p(x,t), \\ p(x,0) &= p_0(x). \end{aligned} \tag{6.25}$$

It can be observed that equation (6.25) includes nonlinear terms, as in the porous media equation (see Oelschläger (1990)). This is due to the repulsive interaction between particles, which in the limit produces a density-dependent diffusion. A linear diffusion persists because the variance of the Brownian motions in the individual equations was kept constant. We will see in a second example how it may vanish when the individual variances tend to zero for N tending to infinity. We will not provide a detailed proof of Theorem 6.5, even though we are going to provide a significant outline of it, leaving further details to the referred literature.

By proceeding as in the previous subsection, a straightforward application of Doob's inequality for martingales (Proposition 2.69) justifies the vanishing of the noise term in the following evolution equation for the empirical measure $(X_N(t))_{t\in\mathbb{R}_+}$:

$$\langle X_N(t), f(\cdot,t)\rangle = \langle X_N(0), f(\cdot,0)\rangle + \int_0^t \langle X_N(s), \nabla g_N(\cdot,s)\nabla f(\cdot,s)\rangle\, ds$$

$$+\int_0^t \left\langle X_N(s), \frac{\sigma^2}{2}\triangle f(\cdot,s) + \frac{\partial}{\partial s} f(\cdot,s)\right\rangle ds$$
$$+\frac{\sigma}{N}\int_0^t \sum_k \nabla f\left(X_N^k(s), s\right) dW^k(s) \tag{6.26}$$

for a given $T > 0$ and any $f \in C_b^{2,1}(\mathbb{R}^d \times [0,T])$. The major difficulty in a rigorous proof of Theorem 6.5 comes from the nonlinear term

$$\Xi_{N,f}(t) = \int_0^t \langle X_N(s), \nabla g_N(\cdot,s)\nabla f(\cdot,s)\rangle\, ds. \tag{6.27}$$

If we rewrite (6.27) in an explicit form we get

$$\Xi_{N,f}(t) = \int_0^t \frac{1}{N^2}\sum_{k,m=1}^N \nabla V_N\left(X_N^k(s) - X_N^m(s)\right)\nabla f\left(X_N^k(s), s\right) ds. \tag{6.28}$$

Since for $\beta > 0$ the kernel $V_N \to \delta_0$, namely the Dirac delta function, this shows that, in the limit, even small changes of the relative position of neighbouring particles may have a considerable effect on $\Xi_{N,f}(t)$. But in any case, the regularity assumptions made on the kernel V_N let us state the following lemma which provides sufficient estimates about g_N and h_N as defined above.

Lemma 6.6. *Under the assumptions 2 and 4 of Theorem 6.5, the following holds:*

$$E\left[\sup_{t\le T}\|h_N(\cdot,t)\|_2^2 + \int_0^t \left\langle X_N(s), |\nabla g_N(\cdot,s)|^2\right\rangle ds + \int_0^t \|\nabla h_N(\cdot,t)\|_2^2 ds\right] < \infty. \tag{6.29}$$

As a consequence the sequence $\{h_N(\cdot,t) : N \in \mathbb{N}\}$ *is relatively compact in* $L^2(\mathbb{R}^d)$.

A significant consequence of the above lemma is the following one.

Lemma 6.7. *With* X_N *as above, the sequence* $\mathcal{L}(X_N)$ *is relatively compact in the space* $\mathcal{M}(C([0,T],\mathcal{M}(\mathbb{R}^d)))$.

By Lemma 6.7 we may claim that a subsequence of $(\mathcal{L}(X_N))_{N\in\mathbb{N}}$ exists, which converges to a probability measure on the space $\mathcal{M}(C([0,T],\mathcal{M}(\mathbb{R}^d)))$ (see the appendix on convergence of probability measures). The Skorohod representation Theorem 1.158 then assures that a process X_∞^k exists in $C([0,T],\mathcal{M}(\mathbb{R}^d))$ such that

$$\lim_{l\to\infty} X_{N_l} = X_\infty^k, \qquad \text{almost surely with respect to } P.$$

If we can assure the uniqueness of the limit, then all X_∞^k will coincide with some X_∞.

Remark 6.8. We need to notice that a priori the limiting process X_∞ may still be a random process in $C([0,T],\mathcal{M}(\mathbb{R}^d))$.

The proof of Theorem 6.5 is now based on the proof of the two following lemmas. Uniqueness of X_∞ is a consequence of Lemma 6.10.

Lemma 6.9. *Under the assumptions of Theorem 6.5 the random variable $X_\infty(t)$ admits almost surely with respect to P a density $h_\infty(\cdot,t)$ with respect to the usual Lebesgue measure on $\mathbb{R}^d$ for any $t \in [0,T]$. Moreover,*

$$\langle X_\infty(t), f\rangle = \langle X_\infty(0), f\rangle - \frac{1}{2}\int_0^t \langle \nabla h_\infty(\cdot,s), (1+2h_\infty(\cdot,s))\nabla f\rangle\, ds,$$

with $0 \le t \le T$, $f \in C_b^1(\mathbb{R}^d)$, almost surely with respect to P.

This shows that if we assume that $X_\infty(0)$ admits a deterministic density p_0 at time $t=0$, then $(X_\infty(t))_{t\in[0,T]}$ satisfies a deterministic evolution equation and is thus itself a deterministic process on $C([0,T],\mathcal{M}(\mathbb{R}^d))$. From the general theory we know that equation (6.23) admits a unique solution $p \in C_b^{2+\alpha,1+\alpha/2}(\mathbb{R}^d \times [0,T])$. We can now state the following lemma.

Lemma 6.10.

$$\|h_\infty(\cdot,t) - p(\cdot,t)\|_2^2 \le C\int_0^t \|h_\infty(\cdot,s) - p(\cdot,s)\|_2^2 ds.$$

Due to Gronwall's Lemma 4.3 we may then state that

$$\sup_{t\le T}\|h_\infty(\cdot,t) - p(\cdot,t)\|_2^2 = 0,$$

which concludes the proof of Theorem 6.5.

Ant Colonies

As another example, we consider a model for ant colonies. The latter provide an interesting concept of *aggregation* of individuals. According to a model proposed in Morale, Capasso, and Oelschläger (2004), (1998) (see also Burger, Capasso, and Morale (2003)) (based on an earlier model by Grünbaum and Okubo (1994)), in a colony or in an army (in which case the model may be applied to any cross section) ants are assumed to be subject to two conflicting *social forces*: long-range attraction and short-range repulsion. Hence we consider the following basic assumptions:

(i) Particles tend to aggregate subject to their interaction within a range of size $R_a > 0$ (finite or not). This corresponds to the assumption that each particle is capable of perceiving the others only within a suitable sensory range; in other words, each particle has a limited knowledge of the spatial distribution of its neighbors.

(ii) Particles are subject to repulsion when they come "too close" to each other.

We may express assumptions (i) and (ii) by introducing in the drift term F_N in (6.7) two additive components (see Warburton and Lazarus (1991)): F_1, responsible for aggregation, and F_2, for repulsion, such that

$$F_N = F_1 + F_2.$$

The Aggregation Term F_1

We introduce a convolution kernel $G_a : \mathbb{R}^d \to \mathbb{R}_+$, having a support confined to the ball centered at $0 \in \mathbb{R}^d$ and radius $R_a \in \bar{\mathbb{R}}_+$ as the range of sensitivity for aggregation, independent of N. A *generalized gradient* operator is obtained as follows. Given a measure μ on $\mathbb{R}^d$, we define the function

$$[\nabla G_a * \mu](x) = \int_{\mathbb{R}^d} \nabla G_a(x-y)\mu(dy), \qquad x \in \mathbb{R}^d,$$

as the classical convolution of the gradient of the kernel G_a with the measure μ. Furthermore, G_a is such that

$$G_a(x) = \widehat{G}_a(|x|), \tag{6.30}$$

with $\widehat{G}_a$ a decreasing function in $\mathbb{R}_+$. We assume that the aggregation term F_1 depends on such a generalized gradient of $X_N(t)$ at $X_N^k(t)$:

$$F_1[X_N(t)]\left(X_N^k(t)\right) = [\nabla G_a * X_N(t)]\left(X_N^k(t)\right). \tag{6.31}$$

This means that each individual feels this generalized gradient of the measure $X_N(t)$ with respect to the kernel G_a. The positive sign for F_1 and (6.30) expresses a force of attraction of the particle in the direction of increasing concentration of individuals.

We emphasize the great generality provided by this definition of a generalized gradient of a measure μ on $\mathbb{R}^d$. By using particular shapes of G_a, one may include angular ranges of sensitivity, asymmetries, etc. at a finite distance (see Gueron et al (1996)).

The Repulsion Term F_2

As far as repulsion is concerned we proceed in a similar way by introducing a convolution kernel $V_N : \mathbb{R}^d \to \mathbb{R}_+$, which determines the range and the strength of influence of neighbouring particles. We assume (by anticipating a limiting procedure) that V_N depends on the total number N of interacting particles. Let V_1 be a continuous probability density on $\mathbb{R}^d$ and consider the scaled kernel $V_N(x)$ as defined in (6.12), again with $\beta \in]0,1[$. It is clear that

$$\lim_{N\to+\infty} V_N = \delta_0, \tag{6.32}$$

where δ_0 is Dirac's delta function. We define

$$\begin{aligned} F_2[X_N(t)]\left(X_N^k(t)\right) &= -\left(\nabla V_N * X_N(t)\right)\left(X_N^k(t)\right) \\ &= -\frac{1}{N}\sum_{m=1}^{N} \nabla V_N\left(X_N^k(t) - X_N^m(t)\right). \end{aligned} \tag{6.33}$$

This means that each individual feels the gradient of the population in a small neighborhood. The negative sign for F_2 expresses a drift towards decreasing concentration of individuals. In this case the range of the repulsion kernel decreases to zero as the size N of the population increases to infinity.

The Diffusion Term

In this model randomness may be due to both external sources and "social" reasons. The external sources could, for instance, be unpredictable irregularities of the environment (like obstacles, changeable soils, varying visibility). On the other hand, the innate need of interaction with peers is a social reason. As a consequence, randomness can be modelled by a multidimensional Brownian motion $\mathbf{W}_t$.

The coefficient of $d\mathbf{W}_t$ is a matrix function depending upon the distribution of particles or some environmental parameters. Here, we take into account only the intrinsic stochasticity due to the need of each particle to interact with others. In fact, experiments carried out on ants have shown this need. Hence, simplifying the model, we consider only one Brownian motion dW_t with the variance of each particle σ_N depending on the total number of particles, not on their distribution. We could interpret this as an approximation of the model by considering all the stochasticities (also the ones due to the environment) modeled by $\sigma_N dW_t$.

Since σ_N expresses the intrinsic randomness of each individual due to its need for social interaction, it should be decreasing as N increases. Indeed, if the number of particles is large, the mean free path of each particle may reduce down to a limiting value that may eventually be zero:

$$\lim_{N\to\infty} \sigma_N = \sigma_\infty \geq 0. \tag{6.34}$$

Scaling Limits

Let us discuss the two choices for the interaction kernel in the aggregation and repulsion terms, respectively. They anticipate the limiting procedure for N tending to infinity. Here we are focusing on two types of scaling limits, the *McKean–Vlasov limit*, which applies to the long-range aggregation, and the

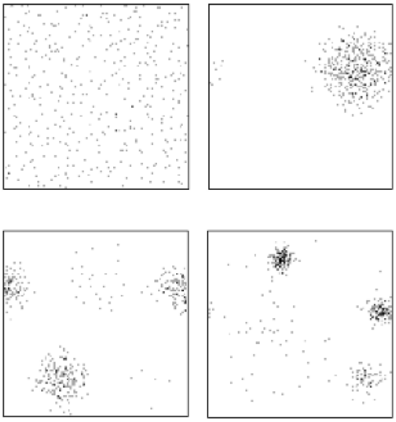

Fig. 6.9. A simulation of the long-range aggregation (6.31) and short-range repulsion (6.33) model for the ant colony with diffusion.

moderate limit, which applies to the short-range repulsion. In the previous subsection we have already considered the moderate limit case.

Mathematically the two cases correspond to the choice made on the interaction kernel. In the moderate limit case (see, e.g., Oelschläger (1985)) the kernel is scaled with respect to the total size of the population N via a parameter $\beta \in]0,1[$. In this case the range of interaction among particles is reduced to zero for N tending to infinity. Thus any particle interacts with many (of order $\frac{N}{\alpha(N)}$) other particles in a small volume (of order $\frac{1}{\alpha(N)}$), where both $\alpha(N)$ and $\frac{N}{\alpha(N)}$ tend to infinity. In the McKean–Vlasov case (see, e.g., Méléard (1996)) $\beta = 0$, so that the range of interaction is independent of N, and as a consequence any particle interacts with order N other particles.

This is why in the moderate limit we may speak of *mesoscale*, which lies between the *microscale* for the typical volume occupied by each individual and the *macroscale* applicable to the typical volume occupied by the total population. Obviously, it would be possible also to consider interacting particle systems rescaled by $\beta = 1$. This case is known as the hydrodynamic case, for which we refer to the literature (De Masi and Presutti (1991), Donsker and Varadhan (1989)).

The case $\beta > 1$ is less significant in population dynamics. It would mean that the range of interaction decreases much faster than the typical distance between neighboring particles. So most of the time particles do not approach sufficiently close to feel the interaction.

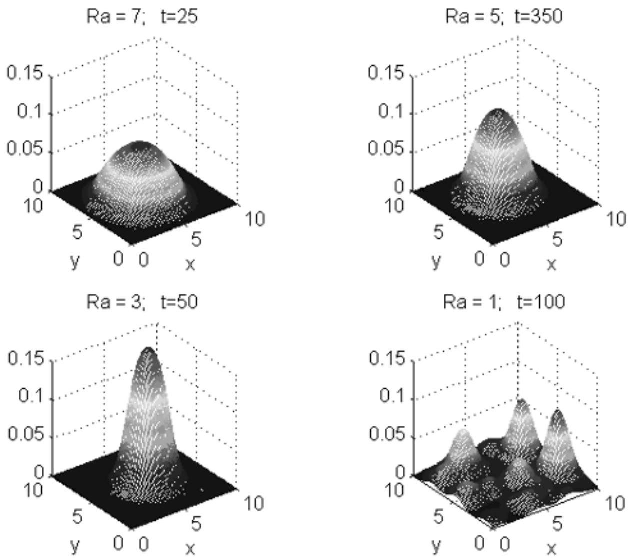

Fig. 6.10. A simulation of the long-range aggregation (6.31) and short-range repulsion (6.33) model for the ant colony with diffusion (smoothed empirical distribution).

Evolution Equations

Again, the fundamental tool for deriving an evolution equation for the empirical measure process is Itô's formula. As in the previous case, the time evolution of any function $f\left(X_N^k(t), t\right)$, $f \in C_b^2(\mathbb{R}^d \times \mathbb{R}_+)$, of the trajectory $\left(X_N^k(t)\right)_{t\in\mathbb{R}_+}$ of the individual particle, subject to the stochastic differential equation (6.7), is given by (6.8). By taking into account expressions (6.31) and (6.33) for F_1 and F_2 and (6.10), then from (6.8), we get the following weak formulation of the time evolution of $X_N(t)$ for any $f \in C_b^{2,1}(\mathbb{R}^d \times [0,\infty[)$:

$$\begin{aligned}\langle X_N(t), f(\cdot,t)\rangle = \langle X_N(0), f(\cdot,0)\rangle &+ \int_0^t \langle X_N(t), (X_N(s) * \nabla G_a) \cdot \nabla f(\cdot,s)\rangle\, ds \\ &- \int_0^t \langle X_N(t), \nabla g_N(\cdot,s) \cdot \nabla f(\cdot,s)\rangle\, ds \\ &+ \int_0^t \left\langle X_N(t), \frac{\sigma_N^2}{2}\triangle f(\cdot,s) + \frac{\partial}{\partial s} f(\cdot,s)\right\rangle ds \\ &+ \frac{\sigma_N}{N}\int_0^t \sum_k \nabla f\left(X_N^k(s), s\right) dW^k(s), \qquad (6.35)\end{aligned}$$

$$g_N(x,t) = (X_N(t) * V_N)(x). \qquad (6.36)$$

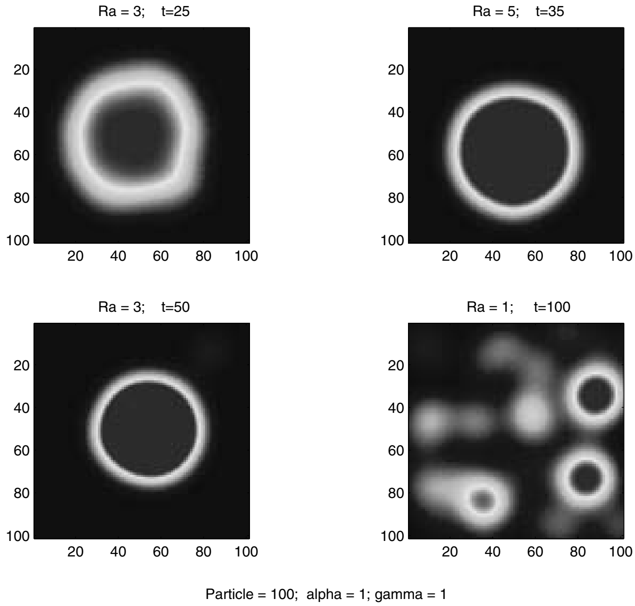

Fig. 6.11. A simulation of the long-range aggregation (6.31) and short-range repulsion (6.33) model for the ant colony with diffusion (two-dimensional projection of the smoothed empirical distribution).

Also for this case we may proceed as in the previous subsection on evolution equations with the analysis of the last term in (6.35). The process

$$M_N(f,t) = \frac{\sigma_N}{N} \int_0^t \sum_k \nabla f \left(X_N^k(s), s \right) dW^k(s), \qquad t \in [0,T],$$

is a martingale with respect to the process's $(X_N(t))_{t\in\mathbb{R}_+}$ natural filtration. By applying Doob's inequality (Proposition 2.69), we obtain

$$E \left[\sup_{t\leq T} |M_N(f,t)| \right]^2 \leq \frac{4\sigma_N^2 \|\nabla f\|_\infty^2 T}{N}.$$

Hence, by assuming that σ_N remains bounded as in (6.34), $M_N(f,\cdot)$ vanishes in the limit $N \to \infty$. This is again the essential reason of the deterministic limiting behavior of the process, since then its evolution equation will no longer be perturbed by Brownian noise.

We will not go into more details at this point. The procedure is the same as for the previous model. But here we confine ourselves to a formal convergence procedure. Indeed, let us suppose that the empirical process $(X_N(t))_{t\in\mathbb{R}_+}$ tends, as $N \to \infty$, to a deterministic process $(X(t))_{t\in\mathbb{R}_+}$, which for any t is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$, with density $\rho(x,t)$:

$$\begin{aligned}\lim_{N\to\infty}\langle X_N(t), f(\cdot,t)\rangle &= \langle X(t), f(\cdot,t)\rangle \\ &= \int f(x,t)\rho(x,t)dx, \qquad t\geq 0.\end{aligned}$$

As a formal consequence we get

$$\begin{aligned}\lim_{N\to\infty} g_N(x,t) &= \lim_{N\to\infty}(X_N(t)*V_N)(x) = \rho(x,t),\\ \lim_{N\to\infty}\nabla g_N(x,t) &= \nabla\rho(x,t),\\ \lim_{N\to\infty}(X_N(t)*\nabla G_a)(x) &= (X(t)*\nabla G_a(x))\\ &= \int \nabla G_a(x-y)\rho(y,t)dy.\end{aligned}$$

Hence, by applying the above limits, from (6.35) we obtain

$$\begin{aligned}&\int_{\mathbb{R}^d} f(x,t)\rho(x,t)dx\\ &= \int_{\mathbb{R}^d} f(x,0)\rho(x,0)dx\\ &\quad+\int_0^t ds\int_{\mathbb{R}^d} dx\,[(\nabla G_a*\rho(\cdot,s))(x)-\nabla\rho(x,s)]\cdot\nabla f(x,s)\rho(x,s)\\ &\quad+\int_0^t ds\int_{\mathbb{R}^d} dx\left[\frac{\partial}{\partial s}f(x,s)\rho(x,s)+\frac{\sigma_\infty^2}{2}\triangle f(x,s)\rho(x,s)\right], \qquad (6.37)\end{aligned}$$

where σ_∞ is defined as in (6.34).

It can be observed that (6.37) is a weak version of the following equation for the spatial density $\rho(x,t)$:

$$\begin{aligned}\frac{\partial}{\partial t}\rho(x,t) &= \frac{\sigma_\infty^2}{2}\triangle\rho(x,t)+\nabla\cdot(\rho(x,t)\nabla\rho(x,t))\\ &\quad-\nabla\cdot[\rho(x,t)(\nabla G_a*\rho(\cdot,t))(x)], \qquad x\in\mathbb{R}^d, t\geq 0, \quad (6.38)\\ \rho(x,0) &= \rho_0(x). \qquad (6.39)\end{aligned}$$

In the degenerate case, i.e., if (6.34) holds with equality, equation (6.38) becomes

$$\frac{\partial}{\partial t}\rho(x,t) = \nabla\cdot(\rho(x,t)\nabla\rho(x,t)) - \nabla\cdot[\rho(x,t)(\nabla G_a*\rho(\cdot,t))(x)]. \qquad (6.40)$$

As in the preceding subsection on moderate repulsion, we need to prove existence and uniqueness of a sufficiently regular solution to equation (6.40). We refer to Burger, Capasso, and Morale (2003) or Nagai and Mimura (1983) and also to Carrillo (1999) for a general discussion of this topic.

A Law of Large Numbers in Path Space

In this section we supplement our results on the asymptotics of the empirical processes by a law of large numbers in path space. This means that we study the *empirical measures in path space*

$$X_N = \frac{1}{N}\sum_{k=1}^{N} \epsilon_{X_N^k(\cdot)},$$

where $X_N^k(\cdot) = (X_N^k(t))_{0\le t\le T}$ denotes the entire path of the kth particle in the time interval $[0,T]$. The particles move continuously in $\mathbb{R}^d$. Moreover, X_N is a measure on the space $\mathcal{C}([0,T],\mathbb{R}^d)$ of continuous functions from $[0,T]$ to $\mathbb{R}^d$. As in the case of empirical processes, one can prove the convergence of X_N to some limit Y. The proof can be achieved with a few additional arguments from the limit theorem for the empirical processes.

By heuristic considerations in Morale, Capasso, and Oelschläger (2004) we get a convergence result for the empirical distribution of the drift $\nabla g_N(\cdot,t)$ of the individual particles

$$\lim_{N\to\infty}\int_0^T \langle X_N(t), |\nabla g_N(\cdot,t) - \nabla\rho(\cdot,t)|\rangle\, dt = 0, \tag{6.41}$$

$$\lim_{N\to\infty}\int_0^T \langle X_N(t), |X_N(t) * \nabla G_a - \nabla G_a * \rho(\cdot,t)|\rangle\, dt = 0.$$

So equation (6.41) allows us to replace the drift

$$\nabla g_N(\cdot,t) - X_N(t) * \nabla G_a$$

with the function

$$\nabla\rho(\cdot,t) - \nabla G_a * \rho(\cdot,t)$$

for large N. Hence, for most k, we have $X_k(t) \sim Y(t)$, uniformly in $t\in[0,T]$, where $Y = Y(t)$, $0\le t\le T$, is the solution of

$$dY(t) = [\nabla G_a * \rho(\cdot,t)(Y(t)) - \nabla\rho(Y(t))]\, dt + \sigma_\infty dW^k(t), \tag{6.42}$$

with the initial condition, for each $k = 1,\ldots,N$,

$$Y(0) = X_N^k(0). \tag{6.43}$$

So, not only does the density follow the deterministic equation (6.38), which presents the memory of the fluctuations by means of the term $\frac{\sigma_\infty}{2}\triangle\rho$, but also the stochasticity of the movement of each particle is preserved.

For the degenerate case $\sigma_\infty = 0$, the Brownian motion vanishes as $N\to\infty$. From (6.42) the dynamics of a single particle depend on the density of the whole system. This density is the solution of (6.40), which does not contain any diffusion term. So, not only do the dynamics of a single particle become deterministic, but neither is there any memory of the fluctuations present, when the number of particles N is finite. The following result confirms these heuristic considerations (see Morale, Capasso, and Oelschläger (2004)).

Theorem 6.11. *For the stochastic system (6.7)–(6.33) make the same assumptions as in Theorem 6.5. Then we obtain*

$$\lim_{N\to\infty} E\left[\frac{1}{N}\sum_{k=1}^{N}\sup_{t\leq T}\left|X_N^k(t)-Y(t)\right|\right]=0, \tag{6.44}$$

where Y is the solution of (6.42) with the initial solution (6.43) for each $k=1,\ldots,N$ and ρ is the density of the limit of the empirical processes; i.e., it is the solution of (6.40).

Price Herding

As an example of herding in economics we present a model for price herding that has been applied to simulate the prices of cars; see Capasso, Morale, and Sioli (2003). The model is based on the assumption that prices of products of a similar nature and within the same market segment tend to aggregate within a given interaction kernel, which characterizes the segment itself. On the other hand, unpredictable behavior of individual prices may be modelled as a family of mutually independent Brownian motions. Hence we suppose that in a segment of N prices, for any $k\in\{1,\ldots,N\}$ the price $X_N^k(t)$, $t\in\mathbb{R}_+$, satisfies

$$\frac{dX_N^k(t)}{X_N^k(t)} = F_k[\mathbf{X}(t)]\left(X_N^k(t)\right)dt+\sigma_k(\mathbf{X}(t))dW^k(t).$$

As usual, for a population of prices it is more convenient to consider the evolution of rates. For the force of interaction F_k, which depends upon the vector of all individual prices

$$\mathbf{X}(t):=\left(X_N^1(t),\ldots,X_N^N(t)\right),$$

we assume the following model, similar to the ant colony of the previous subsection:

$$F_k[\mathbf{X}(t)]\left(X_N^k(t)\right)=\frac{1}{N}\sum_{j=1}^{N}\frac{1}{A_{jk}}\left(\frac{I_j(t)}{I_k(t)}\right)^{\beta_{jk}}\nabla K_a\left(X_N^k(t)-X_N^j(t)\right), \tag{6.45}$$

which includes the following ingredients:

(a) The aggregation kernel

$$K_a(x)=\frac{1}{\sqrt{2\pi a^2}}e^{-\frac{x^2}{2a^2}},$$

$$\nabla K_a(x)=-\frac{x}{a^2}\frac{1}{\sqrt{2\pi a^2}}e^{-\frac{x^2}{2a^2}}.$$

(b) The sensitivity coefficient for aggregation

$$\frac{1}{A_{jk}}\left(\frac{I_j(t)}{I_k(t)}\right)^{\beta_{jk}}$$

depending (via the parameters A_{jk} and β_{jk}) on the relative market share $I_j(t)$ of the product j with respect to the market share $I_k(t)$ of product k. Clearly, a stronger product will be less sensitive to the prices of competing weaker products.

(c) The coefficient $\frac{1}{N}$ takes into account possible crowding effects, which are also modulated by the coefficients A_{jk}.

As an additional feature a model for inflation may be included in F_k. Given a general rate of inflation $(\alpha_t)_{t\in\mathbb{R}_+}$, F_k may include a term $s_k\alpha_t$ to model via s_k the specific sensitivity of price k. We leave the analysis of the model to the reader, who may refer to Capasso, Morale, and Sioli (2003) for details.

Data are shown in Figure 6.12; parameter estimates are given in Tables 6.1, 6.2, and 6.3; Figure 6.13 shows the simulated car prices based on such estimates.

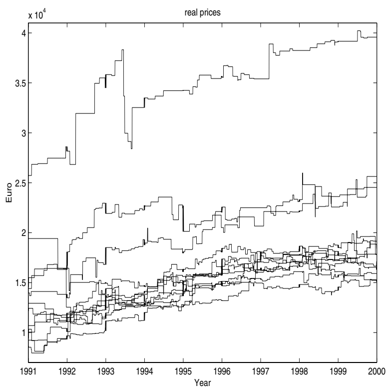

Fig. 6.12. Time series of prices of a segment of cars in Italy during the years 1991–2000 (source: *Quattroruote Magazine, Editoriale Domus, Milan, Italy).*

6.4 Neurosciences

Stein's Model of Neural Activity

The main component of Stein's model (Stein (1965), (1967)) is the depolarization V_t for $t \in \mathbb{R}_+$. A nerve cell is said to be *excited* (or *depolarized*), if $V_t > 0$, and *inhibited*, if $V_t < 0$. In the absence of other events V_t decays according to

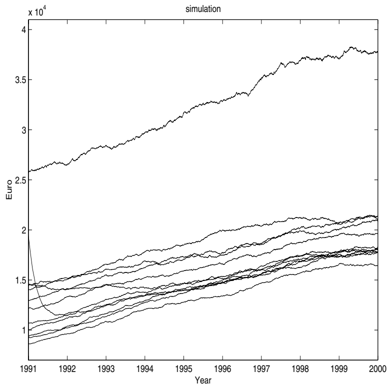

Fig. 6.13. Simulated car prices.

Parameter	Method of Estimation	Estimate	St. Dev.
$X_1(0)$	ML	1.6209E+00	5.8581E-02
$X_2(0)$	ML	8.4813E-01	6.0740E-03
$X_3(0)$	ML	7.4548E-01	2.3420E-02
$X_4(0)$	ML	1.0189E+00	1.2273E-01
$X_5(0)$	ML	1.4164E+00	1.4417E-01
$X_6(0)$	ML	2.4872E+00	6.2947E-02
$X_7(0)$	ML	1.2084E+00	4.7545E-02
$X_8(0)$	ML	1.0918E+00	4.7569E-02
a	ML	5.0767E+03	6.5267E+02

Table 6.1. Estimates for the price herding model (6.45) for the initial conditions $X_k(0)$ and the range of the kernel a.

$$\frac{dV}{dt} = -\alpha V,$$

where $\alpha = 1/\tau$ is the reciprocal of the nerve membrane time constant $\tau > 0$.

In the resting state (initial condition) $V_0 = 0$. Afterwards jumps may occur at random times according to independent Poisson processes $(N_t^E)_{t\in\mathbb{R}_+}$ and $(N_t^I)_{t\in\mathbb{R}_+}$ with intensities λ_E and λ_I, respectively, assumed to be strictly positive real constants. If an excitation (a jump) occurs for N^E, at some time $t_0 > 0$, then

$$V_{t_0} - V_{t_0-} = a_E,$$

whereas if an inhibition (again a jump) occurs for N^I, then

$$V_{t_0} - V_{t_0-} = -a_I,$$

where a_E and a_I are nonnegative real numbers. When V_t attains a given value $\theta > 0$ (the *threshold*), the cell fires. Upon firing V_t is reset to zero along with

Parameter	Method of Estimation	Estimate	St. Dev.
A_{12}	ML	1.0649E-03	3.0865E-02
A_{13}	ML	1.1489E-04	4.1737E-04
A_{14}	ML	1.5779E-03	5.4687E-02
A_{15}	ML	7.6460E-04	1.8381E-02
A_{16}	ML	1.2908E-03	4.0634E-02
A_{17}	ML	1.8114E-03	6.5617E-02
A_{18}	ML	1.5956E-03	5.5572E-02
A_{23}	ML	1.0473E-04	7.2687E-05
A_{24}	ML	1.7397E-04	6.0809E-04
A_{25}	ML	1.7550E-04	5.1100E-04
A_{26}	ML	1.2080E-03	3.7392E-02
A_{27}	ML	9.4809E-04	2.6037E-02
A_{28}	ML	2.7277E-04	2.0135E-03
A_{34}	ML	4.0404E-04	5.5468E-03
A_{35}	ML	1.8136E-04	8.6471E-04
A_{36}	ML	9.5558E-03	4.9764E-01
A_{37}	ML	1.0341E-04	4.4136E-05
A_{38}	ML	7.0953E-04	1.6428E-02
A_{45}	ML	1.0066E-03	2.8485E-02
A_{46}	ML	1.3354E-04	1.3632E-03
A_{47}	ML	2.5239E-04	1.6979E-03
A_{48}	ML	1.1232E-03	3.3652E-02
A_{56}	ML	2.3460E-03	9.2592E-02
A_{57}	ML	1.0143E-03	2.8898E-02
A_{58}	ML	1.1026E-03	3.2724E-02
A_{67}	ML	1.8560E-03	6.8275E-02
A_{68}	ML	2.2820E-03	8.9278E-02
A_{78}	ML	6.4630E-04	1.4003E-02

Table 6.2. Estimates for the price herding model (6.45) for the parameters A_{ij}.

N^E and N^I and the process restarts along the previous model. By collecting all of the above assumptions, the subthreshold evolution equation for V_t may be written in the following form:

$$dV_t = -\alpha V_t dt + a_E dN_t^E - a_I dN_t^I,$$

subject to the initial condition $V_0 = 0$. The model is a particular case of a more general (stochastic) evolution equation of the form

$$dX_t = \alpha(X_t)dt + \int_{\mathbb{R}} \gamma(X_t, u) N(dt, du), \tag{6.46}$$

where N is a marked Poisson process on $\mathbb{R}_+ \times \mathbb{R}$ (in (6.46) the integration is over u). In Stein's model $\alpha(x) = -\alpha x$, with $\alpha > 0$ (or simply $\alpha(x) = -x$, if we assume $\alpha = 1$); $\gamma(x, u) = u$, and the marked Poisson process N has intensity measure

Parameter	Method of Estimation	Estimate	St. Dev.
β_{12}	ML	6.8920E-01	5.8447E+00
β_{13}	ML	2.3463E+00	2.7375E+00
β_{14}	ML	7.2454E-01	6.6182E+00
β_{15}	ML	8.4049E-01	6.2349E+00
β_{16}	ML	7.7929E-01	5.6565E+00
β_{17}	ML	6.6793E-01	5.4208E+00
β_{18}	ML	7.6508E-01	5.8422E+00
β_{23}	ML	2.4531E+00	4.5883E-01
β_{24}	ML	1.6924E+00	6.8734E+00
β_{25}	ML	1.6262E+00	5.7128E+00
β_{26}	ML	1.2122E+00	2.1666E+00
β_{27}	ML	7.5140E-01	7.4760E+00
β_{28}	ML	1.3537E+00	6.0109E+00
β_{34}	ML	1.2444E+00	8.1509E+00
β_{35}	ML	1.7544E+00	8.4976E+00
β_{36}	ML	1.0572E+00	8.0208E+00
β_{37}	ML	2.4730E+00	1.9801E-01
β_{38}	ML	1.0674E+00	8.4626E+00
β_{45}	ML	7.5781E-01	6.7267E+00
β_{46}	ML	2.2121E+00	6.9754E+00
β_{47}	ML	1.7360E+00	6.4971E+00
β_{48}	ML	8.1043E-01	6.1451E+00
β_{56}	ML	7.1269E-01	4.5857E+00
β_{57}	ML	7.7251E-01	6.3947E+00
β_{58}	ML	7.0792E-01	6.5014E+00
β_{67}	ML	8.4060E-01	6.8871E+00
β_{68}	ML	8.1190E-01	6.0759E+00
β_{78}	ML	1.0794E+00	8.4994E+00

Table 6.3. Estimates for the price herding model (6.45) for the parameters β_{ij}.

$$\Lambda((s,t)\times B) = (t-s)\int_B \phi(u)du \qquad \text{for any } s,t \in \mathbb{R}_+, s<t, B \subset \mathcal{B}_{\mathbb{R}}.$$

Here

$$\phi(u) = \lambda_E \delta_0(u-a_E) + \lambda_I \delta_0(u+a_I),$$

with δ_0 the standard Dirac delta distribution. The infinitesimal generator $\mathcal{A}$ of the Markov process $(X_t)_{t\in\mathbb{R}_+}$ given by (6.46) is given by

$$\mathcal{A}f(x) = \alpha(x)\frac{\partial f}{\partial x}(x) + \int_{\mathbb{R}} (f(x+\gamma(x,u)) - f(x))\phi(u)du$$

for any test function f in the domain of $\mathcal{A}$.

The firing problem may be seen as a first passage time through the threshold $\theta > 0$. Let $A =]-\infty, \theta[$. Then the random variable of interest is

$$T_A(x) = \inf\{t \in \mathbb{R}_+ | X_t \in A, X_0 = x \in A\},$$

Parameter	Method of Estimation	Estimate	St. Dev.
s_1	ML	2.0267E-03	2.1858E-04
s_2	ML	5.1134E-03	1.6853E-03
s_3	ML	3.6238E-03	2.5305E-03
s_4	ML	3.6777E-03	2.3698E-03
s_5	ML	1.0644E-04	1.1132E-04
s_6	ML	5.4133E-03	1.2452E-03
s_7	ML	1.0769E-04	1.4414E-04
s_8	ML	2.1597E-03	2.8686E-03
σ_1	MAP	7.0000E-03	2.9073E-06
σ_2	MAP	7.0000E-03	2.9766E-06
σ_3	MAP	7.0000E-03	3.0128E-06
σ_4	MAP	7.0000E-03	2.9799E-06
σ_5	MAP	7.0000E-03	3.0025E-06
σ_6	MAP	7.0000E-03	2.9897E-06
σ_7	MAP	7.0000E-03	2.8795E-06
σ_8	MAP	7.0000E-03	2.9656E-06

Table 6.4. Estimates for the price herding model (6.45) of s_k and σ_k.

which is the first exit time from A. If the indicated set is empty, then we set $T_A(x) = +\infty$. The following result holds

Theorem 6.12. (Tuckwell (1976), Darling and Siegert (1953)). *Let* $(X_t)_{t\in\mathbb{R}_+}$ *be a Markov process satisfying (6.46) and assume that the existence and uniqueness conditions are fulfilled. Then the distribution function*

$$F_A(x,t) = P(T_A(x) \leq t)$$

satisfies

$$\frac{\partial F_A}{\partial t}(x,t) = \mathcal{A}F_A(\cdot,t)(x), \qquad x \in A, t > 0,$$

subject to the initial condition

$$F_A(x,0) = \begin{cases} 0 & \textit{for } x \in A, \\ 1 & \textit{for } x \notin A, \end{cases}$$

and boundary condition

$$F_A(x,t) = 1, \qquad x \notin A, x \geq 0.$$

Corollary 6.13. If the moments

$$\mu_n(x) = E\left[(T_A(x))^n\right], \qquad n \in \mathbb{N}^*,$$

exist, they satisfy the recursive system of equations

$$\mathcal{A}\mu_n(x) = -n\mu_{n-1}(x), \qquad x \in A, \tag{6.47}$$

subject to the boundary conditions

$$\mu_n(x) = 0, \qquad x \notin A.$$

The quantity $\mu_0(x)$, $x \in A$, is the probability of X_t exiting from A in a finite time. It satisfies the equation

$$\mathcal{A}\mu_0(x) = 0, \qquad x \in A, \tag{6.48}$$

subject to

$$\mu_0(x) = 1, \qquad x \notin A.$$

The following lemma is due to Gihman and Skorohod (1972).

Lemma 6.14. *If there exists a bounded function g on $\mathbb{R}$ such that*

$$\mathcal{A}g(x) \leq -1, \qquad x \in A, \tag{6.49}$$

then $\mu_1 < \infty$ and $P(T_A(x) < +\infty) = 1$.

As a consequence of Lemma 6.14 a neuron in Stein's model fires in a finite time with probability 1 and with finite mean interspike interval. This is due to the fact that the solution of (6.48) is $\mu_0(x) = 1$, $x \in \mathbb{R}$, and this satisfies (6.49). The mean first passage time through θ for an initial value x satisfies, by (6.47):

$$-x\frac{d\mu_1}{dx}(x) + \lambda_E\mu_1(x + a_E) + \lambda_I\mu_1(x - a_I) - (\lambda_E + \lambda_I)\mu_1(x) = -1, \tag{6.50}$$

with $x < \theta$ and boundary condition

$$\mu_1(x) = 0, \qquad \text{for } x \geq \theta.$$

The solution of (6.50) is discussed in Tuckwell (1989), where a diffusion approximation of the original Stein's model of neuron firing is also analyzed.

6.5 Exercises and Additions

6.1. Consider a birth-and-death process $(X(t))_{t\in\mathbb{R}_+}$ valued in $\mathbb{N}$, as in section 6.1. In integral form the evolution equation for X will be

$$X(t) = X(0) + \alpha \int X(s-)ds + M(t),$$

where $\alpha = \lambda - \mu$ is the survival rate and $M(t)$ is a martingale. Show that

1.
$$\langle M\rangle(t) = \langle M, M\rangle(t) = (\lambda + \mu)\int_0^t X(s-)ds.$$

2.
$$E[X(t)] = X(0)e^{\alpha t}.$$

3. $X(t)e^{-\alpha t}$ is a square-integrable martingale.
4. $Var[X(t)e^{-\alpha t}] = X(0)\dfrac{\lambda+\mu}{\lambda-\mu}(1-e^{-\alpha t}).$

6.2. (*Age-dependent birth-and-death process*). An age-dependent population can be divided into two subpopulations, described by two marked counting processes. Given $t > 0$, $U^{(1)}(A_0, t)$ describes those individuals who already existed at time $t = 0$ with ages in $A_0 \in \mathcal{B}_{\mathbb{R}_+}$ and are still alive at time t; and $U^{(2)}(T_0, t)$ describes those individuals who are born during $T_0 \in \mathcal{B}_{\mathbb{R}+}$, $T_0 \subset [0, t]$ and are still alive at time t. Assume that the age-specific death rate is $\mu(a),\ a \in \mathbb{R}_+$, and that the birth process $B(T_0), T_0 \in \mathcal{B}_{\mathbb{R}+}$ admits stochastic intensity

$$\alpha(t_0) = \int_0^{+\infty} \beta(a_0 + t_0)U^{(1)}(da_0, t_0-) + \int_0^{t_0-} \beta(t_0 - \tau)U^{(2)}(d\tau, t_0-),$$

where $\beta(a),\ a \in \mathbb{R}_+$ is the age-specific fertility rate. Assume now that suitable densities u_0 and b exist on $\mathbb{R}_+$ such that

$$E[U^{(1)}(A_0, 0)] = \int_{A_0} u_0(a)da$$

and
$$E[B(T_0)] = \int_{T_0} b(\tau)d\tau.$$

Show that the following *renewal equation* holds for any $s \in \mathbb{R}_+$:

$$b(s) = \int_0^{+\infty} da\, u_0(a)\, n(s+a)\, \beta(a+s) + \int_0^s d\tau\, \beta(s-\tau)\, n(s-\tau)\, b(\tau),$$

where $n(t) = \exp\{-\int_0^t \mu(\tau)d\tau\},\ t \in \mathbb{R}_+$. The reader may refer to Capasso (1988).

6.3. Let $\bar{E}$ be the closure of an open set $E \subset \mathbb{R}^d$ for $d \geq 1$. Consider a spatially structured birth-and-death process associated with the marked point process defined by the random measure on $\mathbb{R}^d$:

$$\nu(t) = \sum_{i=1}^{I(t)} \varepsilon_{X^i(t)},$$

where $I(t),\ t \in \mathbb{R}_+$, denotes the number of individuals in the total population at time t; and $X^i(t)$ denotes the random location of the ith individual in $\bar{E}$. Consider the process defined by the following parameters:

1. $\mu : \bar{E} \to \mathbb{R}_+$ is the spatially structured death rate;
2. $\gamma : \bar{E} \to \mathbb{R}_+$ is the spatially structured birth rate;
3. for any $x \in \bar{E}$, $D(x, \cdot) : \mathcal{B}_{\mathbb{R}^d} \to [0, 1]$ is a probability measure such that $\int_{\bar{E}\setminus\{x\}} D(x, dz) = 1$; $D(x, A)$ for $x \in \bar{E}$ and $A \in \mathcal{B}_{\mathbb{R}^d}$ represents the probability that an individual born in x will be dispersed in A.

Show that the infinitesimal generator of the process is the operator L defined as follows: for any sufficiently regular test function ϕ

$$L\phi(\nu) = \int_{\bar{E}} \nu(dx) \int_{\mathbb{R}^d} \gamma(x) D(x, dz)[-\phi(\nu) + \phi(\nu + \varepsilon_{x+z})] \\ +\mu(x)[-\phi(\nu) + \phi(\nu - \varepsilon_x)].$$

(The reader may refer to Fournier and Méléard (2003) for further analysis.)

6.4. Let X be an integer-valued random variable, with probability distribution $p_k = P(X = k),\ k \in \mathbb{N}$. The probability generating function of X is defined as

$$g_X(s) = E[s^X] = \sum_{k=0}^{\infty} s^k p_k,\ \ |s| \leq 1.$$

Consider a homogeneous birth-and-death process $X(t), t \in \mathbb{R}_+$, with birth rate λ and death rate μ, and initial value $X(0) = k_0 > 0$. Show that the probability generating function $G_X(s; t)$ of $X(t)$ satisfies the partial differential equation

$$\frac{\partial}{\partial t} G_X(s; t) + (1 - s)(\lambda s - \mu) \frac{\partial}{\partial s} G_X(s; t) = 0,$$

subject to the initial condition

$$G_X(s; 0) = s^{k_0}.$$

6.5. Consider now a nonhomogeneous birth-and-death process $X(t), t \in \mathbb{R}_+$, with time-dependent birth rate $\lambda(t)$ and death rate $\mu(t)$, and initial value $X(0) = k_0 > 0$. Show that the probability generating function $G_X(s; t)$ of $X(t)$ satisfies the partial differential equation

$$\frac{\partial}{\partial t} G_X(s; t) + (1 - s)(\lambda(t)s - \mu(t)) \frac{\partial}{\partial s} G_X(s; t) = 0,$$

subject to the initial condition

$$G_X(s; 0) = s^{k_0}.$$

Evaluate the probability of extinction of the population. (The reader may refer to Chiang (1968).)

6.6. Consider the general epidemic process as defined in section 6.1 with infection rate $\kappa = 1$ and removal rate δ. Let $G_Z(x, y; t)$ denote the probability generating function of the random vector $\mathbf{Z}(t) = (S(t), I(t))$, where $S(t)$ denotes the number of susceptibles at time $t \geq 0$ and $I(t)$ denotes the number of infectives at time $t \geq 0$. Assume that $S(0) = s_0$ and $I(0) = i_0$, and let $p(m, n; t) = P(S(t) = m, I(t) = n)$. The joint probability generating function G will be defined as

$$G_{\mathbf{Z}}(x, y; t) = E[x^{S(t)} y^{I(t)}] = \sum_{m=0}^{s_0} \sum_{n=0}^{s_0+i_0-m} p(m, n; t)\, x^m\, y^n.$$

Show that it satisfies the partial differential equation

$$\frac{\partial}{\partial t} G_{\mathbf{Z}}(x, y; t) = y(y - x) \frac{\partial^2}{\partial x \partial y} G_{\mathbf{Z}}(x, y; t) + \delta(1 - y) \frac{\partial}{\partial y} G_{\mathbf{Z}}(x, y; t),$$

subject to the initial condition

$$G_{\mathbf{Z}}(x, y; 0) = x^{s_0} y^{i_0}.$$

6.7. Consider a discrete birth-and-death chain $(Y_n^{(\Delta)})_{n\in\mathbb{N}}$ valued in $S = \{0, \pm\Delta, \pm 2\Delta, \ldots\}$, with step size $\Delta > 0$, and denote by $p_{i,j}$ the one-step transition probabilities

$$p_{ij} = P\left(Y_{n+1}^{(\Delta)} = j\Delta \,\middle|\, Y_n^{(\Delta)} = i\Delta\right) \text{ for } i, j \in \mathbb{Z}.$$

Assume that the only nontrivial transition probabilities are

1. $p_{i,i-1} = \gamma_i := \frac{1}{2}\sigma^2 - \frac{1}{2}\mu\Delta$,
2. $p_{i,i+1} = \beta_i := \frac{1}{2}\sigma^2 + \frac{1}{2}\mu\Delta$,
3. $p_{i,i} = 1 - \beta_i - \gamma_i = 1 - \sigma^2$;

where σ^2 and μ are strictly positive real numbers. Note that for Δ sufficiently small, all rates are nonnegative. Consider now the rescaled (in time) process $(Y_{n/\varepsilon}^{(\Delta)})_{n\in\mathbb{N}}$, with $\varepsilon = \Delta^2$; show (formally and possibly rigorously) that the rescaled process weakly converges to a diffusion on $\mathbb{R}$ with drift μ and diffusion coefficient σ^2.

6.8. With reference to the previous problem, show that the same result may be obtained (with suitable modifications) also in the case in which the drift and the diffusion coefficient depend upon the state of the process. For this case show that the probability $\psi(x)$ that the diffusion process reaches c before d, when starting from a point $x \in (c, d) \subset \mathbb{R}$, is given by

$$\psi(x) = \frac{\int_x^d \exp\left\{-\int_c^z \left(2\frac{\mu(y)}{\sigma^2(y)}\right) dy\right\} dz}{\int_c^d \exp\left\{-\int_c^z \left(2\frac{\mu(y)}{\sigma^2(y)}\right) dy\right\} dz}.$$

The reader may refer, e.g., to Bhattacharya and Waymire (1990).

6.9. Consider the general stochastic epidemic with the rescaling proposed at the beginning of section 6.2. Derive the asymptotic ordinary differential system corresponding to Theorem 6.4.

Part III

Appendices

A

Measure and Integration

A.1 Rings and σ-Algebras

Definition A.1. A collection $\mathcal{F}$ of the elements of a set Ω is called a *ring* on Ω if it satisfies the following conditions:

1. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$,
2. $A, B \in \mathcal{F} \Rightarrow A \setminus B \in \mathcal{F}$.

Furthermore, $\mathcal{F}$ is called an *algebra* if $\mathcal{F}$ is both a ring and $\Omega \in \mathcal{F}$.

Definition A.2. A ring $\mathcal{F}$ on Ω is called a σ-*ring* if it satisfies the following additional condition:

3. For every countable family $(A_n)_{n\in\mathbb{N}}$ of the subsets of $\mathcal{F}$: $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

A σ-ring $\mathcal{F}$ on Ω is called a σ-*algebra* if $\Omega \in \mathcal{F}$.

Definition A.3. Every collection $\mathcal{F}$ of the elements of a set Ω, is called a *semiring* on Ω if it satisfies the following conditions:

1. $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$,
2. $A, B \in \mathcal{F} \Rightarrow A \subset B \Rightarrow \exists (A_j)_{i\leq j\leq m} \in \mathcal{F}^{\{1,\ldots,m\}}$ of disjoint sets such that $B \setminus A = \bigcup_{j=1}^{m} A_j$.

If $\mathcal{F}$ is both a semiring and $\Omega \in \mathcal{F}$, then it is called a *semialgebra*.

Proposition A.4. *A set Ω has the following properties:*

1. *If $\mathcal{F}$ is a σ-algebra of the subsets of Ω, then it is an algebra.*
2. *If $\mathcal{F}$ is a σ-algebra of the subsets of Ω, then*
   - $E_1, \ldots, E_n, \ldots \in \mathcal{F} \Rightarrow \bigcap_{n=1}^{\infty} E_n \in \mathcal{F}$,
   - $E_1, \ldots, E_n \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{n} E_i \in \mathcal{F}$,
   - $B \in \mathcal{F} \Rightarrow \Omega \setminus B \in \mathcal{F}$.
3. *If $\mathcal{F}$ is a ring on Ω, then it is also a semiring.*

Definition A.5. Every pair $(\Omega, \mathcal{F})$ consisting of a set Ω and a σ-ring $\mathcal{F}$ of the subsets of Ω is a *measurable space*. Furthermore, if $\mathcal{F}$ is a σ-algebra, then $(\Omega, \mathcal{F})$ is a *measurable space on which a probability measure can be built.* If $(\Omega, \mathcal{F})$ is a measurable space, then the elements of $\mathcal{F}$ are called $\mathcal{F}$*-measurable* or just *measurable sets.* We will henceforth assume that if a space is measurable, then we can build a probability measure on it.

Example A.6.

1. If $\mathcal{B}$ is a σ-algebra on the set E and $X : \Omega \to E$ a generic mapping, then the set
$$X^{-1}(\mathcal{B}) = \{A \subset \Omega | \exists B \in \mathcal{B} \text{ such that } A = X^{-1}(B)\}$$
is a σ-algebra on Ω.
2. *Generated σ-algebra*. If $\mathcal{A}$ is a set of the elements of a set Ω, then there exists a smallest σ-algebra of subsets of Ω that contains $\mathcal{A}$. This is the σ-algebra *generated* by $\mathcal{A}$, denoted $\sigma(\mathcal{A})$. If, now, $\mathcal{G}$ is the set of all σ-algebras of the subsets of Ω containing $\mathcal{A}$, then it is not empty because it has $\sigma(\Omega)$ among its elements, so that $\sigma(\mathcal{A}) = \bigcap_{\mathcal{C} \in \mathcal{G}} \mathcal{C}$.
3. *Borel σ-algebra.* Let Ω be a topological space. Then the *Borel σ-algebra* on Ω, denoted by $\mathcal{B}_\Omega$, is the σ-algebra generated by the set of all open subsets of Ω. Its elements are called Borelian or Borel-measurable.
4. The set of all bounded and unbounded intervals of $\mathbb{R}$ is a semialgebra.
5. If $\mathcal{B}_1$ and $\mathcal{B}_2$ are algebras on Ω_1 and Ω_2, respectively, then the set of rectangles $B_1 \times B_2$, with $B_1 \in \mathcal{B}_1$ and $B_2 \in \mathcal{B}_2$, is a semialgebra.
6. *Product σ-algebra.* Let $(\Omega_i, \mathcal{F}_i)_{1 \le i \le n}$ be a family of measurable spaces and let $\Omega = \prod_{i=1}^n \Omega_i$. Defining
$$\mathcal{R} = \left\{ E \subset \Omega | \forall, i = 1, \ldots, n \ \exists E_i \in \mathcal{F}_i \text{ such that } E = \prod_{i=1}^n E_i \right\},$$
then $\mathcal{R}$ is a semialgebra of the elements of Ω. The σ-algebra generated by $\mathcal{R}$ is called the *product σ-algebra* of the σ-algebras $(\mathcal{F}_i)_{1 \le i \le n}$.

Proposition A.7. *Let $(\Omega_i)_{1 \le i \le n}$ be a family of topological spaces with a countable base and let $\Omega = \prod_{i=1}^n \Omega_i$. Then the Borel σ-algebra $\mathcal{B}_\Omega$ is identical to the product σ-algebra of the family of Borel σ-algebras $(\mathcal{B}_{\Omega_i})_{1 \le i \le n}$.*

A.2 Measurable Functions and Measure

Definition A.8. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces. A function $f : \Omega_1 \to \Omega_2$ is *measurable* if
$$\forall E \in \mathcal{F}_2 \colon f^{-1}(E) \in \mathcal{F}_1.$$

Remark A.9. If $(\Omega, \mathcal{F})$ is not a measurable space, i.e., $\Omega \notin \mathcal{F}$, then there does not exist a measurable mapping from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, because $\mathbb{R} \in \mathcal{B}_{\mathbb{R}}$ and $f^{-1}(\mathbb{R}) = \Omega \notin \mathcal{F}$.

Definition A.10. Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \mathbb{R}^n$ a mapping. If f is measurable with respect to the σ-algebras $\mathcal{F}$ and $\mathcal{B}_{\mathbb{R}^n}$, the latter being the Borel σ-algebra on $\mathbb{R}^n$, then f is *Borel-measurable.*

Proposition A.11. *Let $(E_1, \mathcal{B}_1)$ and $(E_2, \mathcal{B}_2)$ be two measurable spaces, $\mathcal{U}$ a set of the elements of E_2, which generates $\mathcal{B}_2$ and $f : E_1 \to E_2$. The necessary and sufficient condition for f to be measurable is $f^{-1}(\mathcal{U}) \subset \mathcal{B}_1$.*

Remark A.12. If a function $f : \mathbb{R}^k \to \mathbb{R}^n$ is continuous, then it is Borel-measurable.

Definition A.13. Let $(\Omega, \mathcal{F})$ be a measurable space. Every Borel-measurable mapping $h : \Omega \to \bar{\mathbb{R}}$ that can only have a finite number of distinct values is called an *elementary function.* Equivalently, a function $h : \Omega \to \bar{\mathbb{R}}$ is elementary if and only if it can be written as the finite sum

$$\sum_{i=1}^{r} x_i I_{E_i},$$

where, for every $i = 1, \ldots, r$, the E_i are disjoint sets of $\mathcal{F}$ and I_{E_i} is the indicator function on E_i.

Theorem A.14. (Approximation of measurable functions through elementary functions.) *Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \bar{\mathbb{R}}$ a nonnegative measurable function. There exists a sequence of measurable elementary functions $(s_n)_{n\in\mathbb{N}}$ such that*

1. $0 \leq s_1 \leq \cdots \leq s_n \leq \cdots \leq f$,
2. $\lim_{n\to\infty} s_n = f$.

Proposition A.15. *If $f_1, f_2 : \Omega \to \bar{\mathbb{R}}$ are Borel-measurable functions, then so are the functions $f_1 + f_2$, $f_1 - f_2$, $f_1 f_2$, and f_1/f_2, as long as the operations are well defined.*

Lemma A.16. *If $f : (\Omega_1, \mathcal{F}_1) \to (\Omega_2, \mathcal{F}_2)$ and $g : (\Omega_2, \mathcal{F}_2) \to (\Omega_3, \mathcal{F}_3)$ are measurable functions, then so is $g \circ f : (\Omega_1, \mathcal{F}_1) \to (\Omega_3, \mathcal{F}_3)$.*

Proposition A.17. *Let $(\Omega_i, \mathcal{F}_i)_{1\leq i\leq n}$ be a family of measurable spaces, $\Omega = \prod_{i=1}^n \Omega_i$ and $\pi_i : \Omega \to \Omega_i$ for $1 \leq i \leq n$ the ith projection. Then the product σ-algebra $\bigotimes_{i=1}^n \mathcal{F}_i$ of the family of σ-algebras $(\mathcal{F}_i)_{1\leq i\leq n}$ is the smallest σ-algebra on Ω for which every projection π_i is measurable.*

Proposition A.18. *If $h : (E, \mathcal{B}) \to (\Omega = \prod_{i=1}^n \Omega_i, \mathcal{F} = \bigotimes_{i=1}^n \mathcal{F}_i)$ is a mapping, then the following statements are equivalent:*

1. *h is measurable;*
2. *for all $i = 1, \dots n$, $h_i = \pi_i \circ h$ is measurable.*

Proof: $1 \Rightarrow 2$ follows from Proposition A.17 and Lemma A.16. To prove that $2 \Rightarrow 1$, it is sufficient to see that given $\mathcal{R}$, the set of rectangles on Ω, it follows that, for all $B \in \mathcal{R} : h^{-1}(B) \in \mathcal{B}$. Let $B \in \mathcal{R}$. Then for all $i = 1, \dots, n$, there exists a $B_i \in \mathcal{F}_i$ such that $B = \prod_{i=1}^n B_i$. Therefore, by recalling that due to 2 every h_i is measurable, we have that

$$h^{-1}(B) = h^{-1}\left(\prod_{i=1}^n B_i\right) = \bigcap_{i=1}^n h_i^{-1}(B_i) \in \mathcal{B}.$$

□

Corollary A.19. Let $(\Omega, \mathcal{F})$ be a measurable space and $h : \Omega \to \mathbb{R}^n$ a function. Defining $h_i = \pi_i \circ h : \Omega \to \mathbb{R}$ for $1 \leq i \leq n$, the following two propositions are equivalent:

1. h is Borel-measurable;
2. for all $i = 1, \dots, n$, h_i is Borel-measurable.

Definition A.20. Let $(\Omega, \mathcal{F})$ be a measurable space. Every function $\mu : \Omega \to \bar{\mathbb{R}}$ that

1. for all $E \in \mathcal{F}$: $\mu(E) \geq 0$,
2. for all $E_i, \dots, E_n, \dots \in \mathcal{F}$ such that $E_i \cap E_j = \emptyset$, for $i \neq j$, we have that

$$\mu\left(\bigcup_{i=1}^\infty E_i\right) = \sum_{i=1}^\infty \mu(E_i)$$

is a *measure* on $\mathcal{F}$. Moreover, if $(\Omega, \mathcal{F})$ is a measurable space and if

$$\mu(\Omega) = 1, \tag{A.1}$$

then μ is a *probability measure* or *probability*. Furthermore, a measure μ is *finite* if

$$\forall A \in \mathcal{F}: \ \mu(A) < +\infty$$

and σ-*finite*, if

1. there exists an $(A_n)_{n \in \mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ such that $\Omega = \bigcup_{n \in \mathbb{N}} A_n$;
2. for all $n \in \mathbb{N}$: $\mu(A_n) < +\infty$.

Definition A.21. The ordered triple $(\Omega, \mathcal{F}, \mu)$, where Ω denotes a set, $\mathcal{F}$ a σ-ring on Ω, and $\mu : \mathcal{F} \to \bar{\mathbb{R}}$ a measure on $\mathcal{F}$, is a *measure space*. If μ is a probability measure, then $(\Omega, \mathcal{F}, \mu)$ is a *probability space*.[12]

[12] Henceforth we will call every measurable space that has a probability measure assigned to it a probability space.

Definition A.22. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\lambda : \mathcal{F} \to \bar{\mathbb{R}}$ a measure on Ω. Then λ is said to be *absolutely continuous* with respect to μ, denoted $\lambda \ll \mu$, if

$$\forall A \in \mathcal{F}: \ \mu(A) = 0 \Rightarrow \lambda(A) = 0.$$

Proposition A.23. (Characterization of measure). *Let μ be additive on an algebra $\mathcal{F}$ and valued in $\mathbb{R}$ (and not everywhere equal to $+\infty$). The following two statements are equivalent:*

1. *μ is a measure on $\mathcal{F}$.*
2. *For increasing $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$, where $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$, we have that*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n) = \sup_{n\in\mathbb{N}} \mu(A_n).$$

If μ is finite, then 1 and 2 are equivalent to the following.

3. *For decreasing $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$, where $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{F}$, we have*

$$\mu\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mu(A_n) = \inf_{n\in\mathbb{N}} \mu(A_n).$$

4. *For decreasing $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$, where $\bigcap_{n\in\mathbb{N}} A_n = \emptyset$, we have*

$$\lim_{n\to\infty} \mu(A_n) = \inf_{n\in\mathbb{N}} \mu(A_n) = 0.$$

Proposition A.24. (Generalization of a measure). *Let $\mathcal{G}$ be a semiring on E and $\mu : \mathcal{G} \to \mathbb{R}_+$ a function that satisfies the following properties:*

1. *μ is (finitely) additive on $\mathcal{G}$,*
2. *μ is countably additive on $\mathcal{G}$,*
3. *there exists an $(S_n)_{n\in\mathbb{N}} \in \mathcal{G}^{\mathbb{N}}$ such that $E \subset \bigcup_{n\in\mathbb{N}} S_n$.*

Under these assumptions

$$\exists | \bar{\mu} : \mathcal{B} \to \bar{\mathbb{R}}_+ \text{ such that } \bar{\mu}|_{\mathcal{G}} = \mu,$$

where $\mathcal{B}$ is the σ-ring generated by $\mathcal{G}$.[13] *Moreover, if μ is a probability measure, then so is $\bar{\mu}$.*

Proposition A.25. *Let $\mathcal{U}$ be a ring on E and $\mu : \mathcal{U} \to \bar{\mathbb{R}}_+$ (not everywhere equal to $+\infty$) a measure on $\mathcal{U}$. Then, if $\mathcal{B}$ is the σ-ring generated by $\mathcal{U}$,*

$$\exists | \bar{\mu} : \mathcal{B} \to \bar{\mathbb{R}}_+ \text{ such that } \bar{\mu}|_{\mathcal{U}} = \mu.$$

Moreover, if μ is a probability measure, then so is $\bar{\mu}$.

[13] $\mathcal{B}$ is identical to the σ-ring generated by the ring generated by $\mathcal{G}$.

Lemma A.26. (Fatou). *Let $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ be a sequence of random variables and $(\Omega, \mathcal{F}, P)$ a probability space. Then*

$$P(\liminf_n A_n) \leq \liminf_n P(A_n) \leq \limsup_n P(A_n) \leq P(\limsup_n A_n).$$

If $\liminf_n A_n = \limsup_n A_n = A$, then $A_n \to A$.

Corollary A.27. Under the assumptions of Fatou's Lemma A.26, if $A_n \to A$, then $P(A_n) \to P(A)$.

A.3 Lebesgue Integration

Let $(\Omega, \mathcal{F})$ be a measurable space. We will denote by $\mathcal{M}(\mathcal{F}, \bar{\mathbb{R}})$ (or, respectively, by $\mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$) the set of measurable functions on $(\Omega, \mathcal{F})$ and valued in $\bar{\mathbb{R}}$ (or $\bar{\mathbb{R}}_+$).

Proposition A.28. *Let $(\Omega, \mathcal{F})$ be a measurable space and μ a positive measure on $\mathcal{F}$. Then there exists a unique mapping Φ from $\mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$ to $\bar{\mathbb{R}}_+$, such that:*

1. *For every $\alpha \in \mathbb{R}_+$, $f, g \in \mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$,*
   $\Phi(\alpha f) = \alpha\Phi(f)$,
   $\Phi(f + g) = \Phi(f) + \Phi(g)$,
   $f \leq g \Rightarrow \Phi(f) \leq \Phi(g)$.
2. *For every increasing sequence $(f_n)_{n\in\mathbb{N}}$ of elements of $\mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$ we have that $\sup_n \Phi(f_n) = \Phi(\sup_n f_n)$ (Beppo–Levi property).*
3. *For every $B \in \mathcal{F}$, $\Phi(I_B) = \mu(B)$.*

Definition A.29. If Φ is the unique functional associated with μ, the measure on the measurable space $(\Omega, \mathcal{F})$, then for every $f \in \mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$:

$$\Phi(f) = \int^* f(x)d\mu(x) \text{ or } \int^* f(x)\mu(dx) \text{ or } \int^* f(x)d\mu$$

the *upper integral* of μ.

Remark A.30. Let $(\Omega, \mathcal{F})$ be a measurable space and let Φ be the functional canonically associated with μ measure on $\mathcal{F}$.

1. If $s : \Omega \to \bar{\mathbb{R}}_+$ is an elementary function, thus $s = \sum_{i=1}^n x_i I_{E_i}$, then

$$\Phi(s) = \int^* s d\mu = \sum_{i=1}^n x_i \mu(E_i).$$

2. If $f \in \mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$ and defining $\Omega_f = \{s : \Omega \to \bar{\mathbb{R}}_+ | s \text{ elementary }, s \leq f\}$, then Ω_f is nonempty and

$$\Phi(f) = \int^* f d\mu = \sup_{s\in\Omega_f} \int^* s d\mu = \sup_{s\in\Omega_f} \left(\sum_{i=1}^n x_i \mu(E_i)\right).$$

3. If $f \in \mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$ and $B \in \mathcal{F}$, then by definition

$$\int_B^* f d\mu = \int^* I_B \cdot f d\mu.$$

Definition A.31. Let $(\Omega, \mathcal{F})$ be a measurable space and μ a positive measure on $\mathcal{F}$. An $\mathcal{F}$-measurable function f is *μ-integrable* if

$$\int^* f^+ d\mu < +\infty \text{ and } \int^* f^- d\mu < +\infty,$$

where f^+ and f^- denote the positive and negative parts of f, respectively. The real number

$$\int^* f^+ d\mu - \int^* f^- d\mu$$

is therefore the *Lebesgue integral* of f with respect to μ, denoted by

$$\int f d\mu \text{ or } \int f(x) d\mu(x) \text{ or } \int f(x) \mu(dx).$$

Proposition A.32. *Let $(\Omega, \mathcal{F})$ be a measurable space, endowed with measure μ and $f \in \mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$. Then*

1. $\int^* f d\mu = 0 \Leftrightarrow f = 0$ *almost surely with respect to* μ,
2. *for every* $A \in \mathcal{F}, \mu(A) = 0$ *we have*

$$\int_A^* f d\mu = 0;$$

3. *for every* $g \in \mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$ *such that* $f = g$, *almost surely with respect to* μ, *we have*

$$\int^* f d\mu = \int^* g d\mu.$$

Theorem A.33. (Monotone convergence). *Let $(\Omega, \mathcal{F})$ be a measurable space endowed with measure μ, $(f_n)_{n \in \mathbb{N}}$ an increasing sequence of elements of $\mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$, and $f : \Omega \to \bar{\mathbb{R}}_+$ such that*

$$\forall \omega \in \Omega \colon\ f(\omega) = \lim_{n \to \infty} f_n(\omega) = \sup_{n \in \mathbb{N}} f_n(\omega).$$

Then $f \in \mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$ and

$$\int^* f d\mu = \lim_{n \to \infty} \int^* f_n d\mu.$$

Theorem A.34. (Lebesgue's dominated convergence). *Let $(\Omega, \mathcal{F})$ be a measurable space endowed with measure μ, $(f_n)_{n \in \mathbb{N}}$ a sequence of μ-integrable functions defined on Ω, and $g : \Omega \to \bar{\mathbb{R}}_+$ a μ-integrable function, such that*

$|f_n| \leq g$, *for all* $n \in \mathbb{N}$. *If we suppose that* $\lim_{n\to\infty} f_n = f$ *exists almost surely in* Ω, *then* f *is* μ*-integrable and we have*

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

Lemma A.35. (Fatou). *Let* $f_n \in \mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$. *Then*

$$\liminf_n \int^* f_n d\mu \geq \int^* \liminf_n f_n d\mu.$$

Theorem A.36. (Fatou–Lebesgue).

1. *Let* $|f_n| \leq g \in \mathcal{L}^1$. *Then*

$$\limsup_n \int f_n d\mu \leq \int \limsup_n f_n d\mu.$$

2. *Let* $|f_n| \leq g \in \mathcal{L}^1$. *Then*

$$\liminf_n \int f_n d\mu \geq \int \liminf_n f_n d\mu.$$

3. *Let* $|f_n| \leq g$ *and* $f = \lim_n f_n$, *almost surely with respect to* μ. *Then*

$$\lim_n \int f_n d\mu = \int f d\mu.$$

Definition A.37. Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{B})$ be a measurable space, endowed with measure μ, and let $h : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ be a measurable function. The mapping $\mu_h : \mathcal{B} \to \bar{\mathbb{R}}_+$, such that $\mu_h(B) = \mu(h^{-1}(B))$ for all $B \in \mathcal{B}$ is a measure on E, called *induced measure* h on μ, denoted $h(\mu)$.

Proposition A.38. *Given the assumptions of Definition A.37 the function* $g : (E, \mathcal{B}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *is integrable with respect to* μ_h *if and only if* $g \circ h$ *is integrable with respect to* μ *and*

$$\int g \circ g d\mu = \int g d\mu_h.$$

Theorem A.39. (Product measure). *Let* $(\Omega_1, \mathcal{F}_1)$ *and* $(\Omega_2, \mathcal{F}_2)$ *be measurable spaces and the former be endowed with* σ*-finite measure* μ_1 *on* $\mathcal{F}_1$. *Further suppose that for all* $\omega_1 \in \Omega_1$ *a measure* $\mu(\omega_1, \cdot)$ *is assigned on* $\mathcal{F}_1$ *and that for all* $B \in \mathcal{F}_2$, $\mu(\cdot, B) : \Omega_1 \to \mathbb{R}$ *is a Borel-measurable function. If* $\mu(\omega_1, \cdot)$ *is uniformly* σ*-finite, then there exists a* $(B_n)_{n\in\mathbb{N}} \in \mathcal{F}_2^{\mathbb{N}}$ *such that* $\Omega_2 = \bigcup_{n=1}^{\infty} B_n$ *and, for all* $n \in \mathbb{N}$ *there exists a* $K_n \in \mathbb{R}$ *such that* $\mu(\omega_1, B_n) \leq K_n$ *for all* $\omega_1 \in \Omega_1$. *Then there exists a unique measure* μ *on the product* σ*-algebra* $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$ *such that*

$$\forall A \in \mathcal{F}_1, B \in \mathcal{F}_2: \qquad \mu(A \times B) = \int_A \mu(\omega_1, B)\mu_1(d\omega_1),$$

and

$$\forall F \in \mathcal{F}: \qquad \mu(F) = \int_{\Omega_1} \mu(\omega_1, F(\omega_1))\mu_1(d\omega_1).$$

Definition A.40. Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces, endowed with σ-finite measures μ_1, μ_2 on $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively. Defining $\Omega = \Omega_1 \times \Omega_2$ and $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, the function $\mu : \mathcal{F} \to \bar{\mathbb{R}}$ with

$$\forall F \in \mathcal{F}: \qquad \mu(F) = \int_{\Omega_1} \mu_2(F(\omega_1))d\mu_1(\omega_1) = \int_{\Omega_2} \mu_1(F(\omega_2))d\mu_2(\omega_2),$$

is the unique measure on $\mathcal{F}$ with

$$\forall A \in \mathcal{F}_1, B \in \mathcal{F}_2: \qquad \mu(A \times B) = \mu_1(A) \times \mu_2(B).$$

Moreover, μ is σ-finite on $\mathcal{F}$ as well as a probability measure, if so are μ_1 and μ_2. The measure μ is the *product measure* of μ_1 and μ_2, denoted by $\mu_1 \otimes \mu_2$.

Theorem A.41. (Fubini). *Given the assumptions of Definition A.40, let* $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *be a Borel-measurable function, such that* $\int_\Omega f d\mu$ *exists. Then*

$$\int_\Omega f d\mu = \int_{\Omega_1} \int_{\Omega_2} f d\mu_2 d\mu_1 = \int_{\Omega_2} \int_{\Omega_1} f d\mu_1 d\mu_2.$$

Proposition A.42. *Let* $(\Omega_i, \mathcal{F}_i)_{1 \le i \le n}$ *be a family of measurable spaces. Further, let* $\mu_1 : \mathcal{F}_1 \to \bar{\mathbb{R}}$ *be a* σ*-finite measure and let*

$$\forall(\omega_1, \ldots, \omega_j) \in \Omega_1 \times \cdots \times \Omega_j: \qquad \mu(\omega_1, \ldots, \omega_j, \cdot) : \mathcal{F}_{j+1} \to \bar{\mathbb{R}}$$

be a measure on $\mathcal{F}_{j+1}$, $1 \le j \le n-1$. *If* $\mu(\omega_1, \ldots, \omega_j, \cdot)$ *is uniformly* σ*-finite and for every* $c \in \mathcal{F}_{j+1}$

$$\mu(\ldots, c) : (\Omega_1 \times \cdots \times \Omega_j, \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_j) \to (\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}}),$$

such that

$$\forall(\omega_1, \ldots, \omega_j) \in \Omega_1 \times \cdots \times \Omega_j: \qquad \mu(\ldots, c)(\omega_1, \ldots, \omega_j) = \mu(\omega_1, \ldots, \omega_j, c)$$

is measurable, then, defining $\Omega = \Omega_1 \times \cdots \times \Omega_n$ *and* $\mathcal{F} = \mathcal{F}_1 \otimes \cdots \otimes \mathcal{F}_n$*:*

1. *There exists a unique measure* $\mu : \mathcal{F} \to \bar{\mathbb{R}}$ *such that for every measurable rectangle* $A_1 \times \cdots \times A_n \in \mathcal{F}$*:*

$$\begin{aligned} &\mu(A_1 \times \cdots \times A_n) \\ &= \int_{A_1} \mu_1(d\omega_1) \int_{A_2} \mu(\omega_1, d\omega_2) \cdots \int_{A_n} \mu(\omega_1, \ldots, \omega_{n-1}, d\omega_n). \end{aligned}$$

   μ *is* σ*-finite on* $\mathcal{F}$ *and a probability whenever* μ_1 *and all* $\mu(\omega_1, \ldots, \omega_j, \cdot)$ *are probability measures.*

2. *If* $f : (\Omega, \mathcal{F}) \to (\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ *is measurable and nonnegative, then*

$$\int_\Omega f d\mu$$
$$= \int_{\Omega_1} \mu_1(d\omega_1) \int_{\Omega_2} \mu(\omega_1, d\omega_2) \cdots \int_{\Omega_n} f(\omega_1, \ldots, \omega_n)\mu(\omega_1, \ldots, \omega_{n-1}, d\omega_n).$$

Proposition A.43. *1. Given the assumptions and the notation of Proposition A.42, if we assume that* $f = I_F$, *then for every* $F \in \mathcal{F}$*:*

$$\mu(F)$$
$$= \int_{\Omega_1} \mu_1(d\omega_1) \int_{\Omega_2} \mu(\omega_1, d\omega_2) \cdots \int_{\Omega_n} I_F(\omega_1, \ldots, \omega_n)\mu(\omega_1, \ldots, \omega_{n-1}, d\omega_n).$$

2. *For all* $j = 1, \ldots, n-1$, *let* $\mu_{j+1} = \mu(\omega_1, \ldots, \omega_j, \cdot)$. *Then there exists a unique measure* μ *on* $\mathcal{F}$ *such that for every rectangle* $A_1 \times \cdots \times A_n \in \mathcal{F}$ *we have*

$$\mu(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n).$$

If $f : (\Omega, \mathcal{F}) \to (\bar{\mathbb{R}}, \mathcal{B}_{\bar{\mathbb{R}}})$ *is measurable and positive, or else if* $\int_\Omega f d\mu$ *exists, then*

$$\int_\Omega f d\mu = \int_{\Omega_1} d\mu_1 \cdots \int_{\Omega_n} f d\mu_n,$$

and the order of integration is arbitrary. The measure μ *is the product measure of* $\mu_1, \ldots, \mu_n$ *and is denoted by* $\mu_1 \otimes \cdots \otimes \mu_n$.

Definition A.44. Let $(v_i)_{1 \le i \le n}$ be a family of measures defined on $\mathcal{B}_{\mathbb{R}}$ and

$$v^{(n)} = v_1 \otimes \cdots \otimes v_n : \mathbb{R}^n \to \mathbb{R}$$

the product measure. The *convolution product* of $v_1, \ldots, v_n$, denoted by $v_1 * \cdots * v_n$, is the induced measure that, for generic functions $f : \mathbb{R}^n \to \mathbb{R}$, associates $(x_i, \ldots, x_n)$ with $\sum_{i=1}^n x_i$ of $v^{(n)}$.

Proposition A.45. *Let* v_1 *and* v_2 *be measures on* $\mathcal{B}_{\mathbb{R}}$. *Then for every* $B \in \mathcal{B}_{\mathbb{R}}$ *we have*

$$v_1 * v_2(B) = \int_B d(v_1 * v_2) = \int_{\mathbb{R}} I_B(z) d(v_1 * v_2) = \int \int I_B(x_1 + x_2) d(v_1 \otimes v_2).$$

A.4 Lebesgue–Stieltjes Measure and Distributions

Definition A.46. Let $\mu : \mathcal{B}_{\mathbb{R}} \to \bar{\mathbb{R}}$ be a measure. It then represents a *Lebesgue–Stieltjes* measure if for every interval I we have that $\mu(I) < +\infty$.

Definition A.47. Every function $F : \mathbb{R} \to \mathbb{R}$ that is right-continuous and increasing is a *distribution function* on $\mathbb{R}$.

It is in fact possible to establish a one-to-one relationship between the set of Lebesgue–Stieltjes measures and the set of distribution functions in the sense that every Lebesgue–Stieltjes measure can be assigned a distribution function and vice versa.

Proposition A.48. *Let μ be a Lebesgue–Stieltjes measure on $\mathcal{B}_{\mathbb{R}}$ and the function $F : \mathbb{R} \to \mathbb{R}$ defined, apart from a constant, as*

$$F(b) - F(a) = \mu(]a,b]) \qquad \forall a, b \in \mathbb{R}, a < b.$$

Then F is a distribution function, in particular the one assigned to μ.

Proposition A.49. *Let F be a distribution function and*

$$F(b) - F(a) = \mu(]a,b]) \qquad \forall a, b \in \mathbb{R}, a < b.$$

There exists a unique extension of μ, which is a Lebesgue–Stieltjes measure on $\mathcal{B}_{\mathbb{R}}$. This measure is the Lebesgue–Stieltjes measure canonically associated with F.

Definition A.50. Every measure $\mu : \mathcal{B}_{\mathbb{R}^n} \to \bar{\mathbb{R}}$ that for every bounded interval I of $\mathbb{R}^n$ has $\mu(I) < +\infty$ is a Lebesgue–Stieltjes measure on $\mathbb{R}^n$

Definition A.51. Let $f : \mathbb{R} \to \mathbb{R}$ be of constant value 1 and we consider the function $F : \mathbb{R} \to \mathbb{R}$ with

$$F(x) - F(0) = \int_0^x f(t)dt \qquad \forall x > 0,$$
$$F(0) - F(x) = \int_x^0 f(t)dt \qquad \forall x < 0,$$

where $F(0)$ is fixed and arbitrary. This function F is a distribution function and its associated Lebesgue–Stieltjes measure is called *Lebesgue measure* on $\mathbb{R}$.

Definition A.52. Let $(\Omega, \mathcal{F}, \mu)$ be a space with σ-finite measure μ and consider another measure $\lambda : \mathcal{F} \to \bar{\mathbb{R}}_+$. λ is said to be defined through its *density* with respect to μ if there exists a Borel-measurable function $g : \Omega \to \bar{\mathbb{R}}_+$ with

$$\lambda(A) = \int_A g d\mu \qquad \forall A \in \mathcal{F}.$$

This function g is the density of λ with respect to μ. In this case λ is absolutely continuous with respect to μ ($\lambda \ll \mu$). If μ is a Lebesgue measure on $\mathbb{R}$, then g is the density of μ. A measure ν is called *μ-singular* if there exists $N \in \mathcal{F}$ such that $\mu(N) = 0$ and $\nu(N \setminus \mathcal{F}) = 0$. Conversely, if also $\mu(N) = 0$ whenever $\nu(N) = 0$, then the two measures are *equivalent* (denoted $\lambda \sim \mu$).

Theorem A.53. (Radon–Nikodym). *Let $(\Omega, \mathcal{F})$ be a measurable space, μ a σ-finite measure on $\mathcal{F}$, and λ an absolutely continuous measure with respect to μ. Then λ is endowed with density with respect to μ. Hence there exists a Borel-measurable function $g : \Omega \to \bar{\mathbb{R}}_+$ such that*

$$\lambda(A) = \int_A g d\mu, \qquad A \in \mathcal{B}.$$

A necessary and sufficient condition for g to be μ-integrable is that λ is bounded. Moreover, if $h : \Omega \to \bar{\mathbb{R}}_+$ is another density of λ, then $g = h$, almost surely with respect to μ.

Theorem A.54. (Lebesgue–Nikodym). *Let ν and μ be a measure and a σ-finite measure on $(E, \mathcal{B})$, respectively. There exist a $\mathcal{B}$-measurable function $f : E \to \bar{\mathbb{R}}_+$ and a μ-singular measure ν' on $(E, \mathcal{B})$ so that*

$$\nu(B) = \int_B f d\mu + \nu'(B) \qquad \forall B \in \mathcal{B}.$$

Furthermore,

1. *ν' is unique.*
2. *If $h : E \to \bar{\mathbb{R}}_+$ is a $\mathcal{B}$-measurable function with*

$$\nu(B) = \int_B h d\mu + \nu'(B) \qquad \forall B \in \mathcal{B},$$

then $f = h$ almost surely with respect to μ.

Definition A.55. A function $F : \mathbb{R} \to \mathbb{R}$ is *absolutely continuous* if, for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all $]a_i, b_i[\subset \mathbb{R}$ for $1 \leq i \leq n$ with $]a_i, b_i[\cap]a_j, b_j[= \emptyset$, $i \neq j$,

$$b_i - a_i < \delta \Rightarrow \sum_{i=1}^{n} |F(b_i) - F(a_i)| < \epsilon.$$

Proposition A.56. *Let F be a distribution function. Then the following two propositions are equivalent:*

1. *F is absolutely continuous.*
2. *The Lebesgue measure canonically associated with F is absolutely continuous.*

Proposition A.57. *Let $f : [a, b] \to \mathbb{R}$ be a mapping. The following two statements are equivalent:*

1. *f is absolutely continuous.*

2. *There exists a Borel-measurable function* $g : [a, b] \to \mathbb{R}$ *that is integrable with respect to Lebesgue measure and*

$$f(x) - f(a) = \int_a^x g(t)dt \qquad \forall x \in [a, b].$$

This function g *is the density of* f.

Proposition A.58. *If* $f : [a, b] \to \mathbb{R}$ *is absolutely continuous, then*

1. f *is differentiable almost everywhere in* $[a, b]$,
2. f', *the first derivative of* f, *is integrable in* $[a, b]$ *and we have that*

$$f(x) - f(a) = \int_a^x f'(t)dt.$$

Theorem A.59. (Fundamental theorem of calculus). *If* $f : [a, b] \to \mathbb{R}$ *is integrable in* $[a, b]$ *and*

$$F(x) = \int_a^x f(t)dt \qquad \forall x \in [a, b],$$

then

1. F *is absolutely continuous in* $[a, b]$,
2. $F' = f$ *almost everywhere in* $[a, b]$.

Vice versa, if we consider a function $F : [a, b] \to \mathbb{R}$ *that satisfies 1 and 2, then*

$$\int_a^b f(x)dx = F(b) - F(a).$$

Proposition A.60. *If* $f : [a, b] \to \mathbb{R}$ *is differentiable in* $[a, b]$ *and has integrable derivatives, then*

1. f *is absolutely continuous in* $[a, b]$,
2. $f(x) = \int_a^x f'(t)dt$.

Definition A.61. Let $(\Omega, \mathcal{F}, \mu)$ be a space endowed with measure and $p > 0$. The set of Borel-measurable functions defined on Ω, such that $\int_\Omega |f|^p d\mu < +\infty$ is a vector space on $\mathbb{R}$ and is denoted with the symbols $\mathcal{L}^p(\mu)$ or $\mathcal{L}^p(\Omega, \mathcal{F}, \mu)$. Its elements are called functions integrable to the exponent p. In particular, elements of $\mathcal{L}^2(\mu)$ are said to be square-integrable functions. Finally, $\mathcal{L}^1(\mu)$ coincides with the space of functions integrable with respect to μ.

A.5 Stochastic Stieltjes Integration

Suppose $(\Omega, \mathcal{F}, P)$ is a given probability space with $(X_t)_{t \in \mathbb{R}_+}$ a measurable stochastic process whose sample paths $(X_t(\omega))_{t \in \mathbb{R}_+}$ are of locally bounded variation for any $\omega \in \Omega$. Now let $(H_s)_{s \in \mathbb{R}_+}$ be a measurable process, whose sample paths are locally bounded for any $\omega \in \Omega$. Then the process $H \bullet X$ defined by

$$(H \bullet X)_t(\omega) = \int_0^t H(s, \omega) dX_s(\omega), \qquad \omega \in \Omega, t \in \mathbb{R}_+$$

is called the *stochastic Stieltjes integral* of H with respect to X. Clearly, $((H * X)_t)_{t \in \mathbb{R}_+}$ is itself a stochastic process.

If we assume further that X is progressively measurable and H is $\mathcal{F}_t$-predictable with respect to the σ-algebra generated by X, then $H * X$ is progressively measurable. In particular, if $N = \sum_{n \in \mathbb{N}^*} \epsilon_{\tau_n}$ is a point process on $\mathbb{R}_+$, then for any nonnegative process H on $\mathbb{R}_+$, the stochastic integral $H * N$ exists and is given by

$$(H \bullet N)_t = \sum_{n \in \mathbb{N}^*} I_{[\tau_n \leq t]}(t) H(\tau_n).$$

Theorem A.62. *Let M be a martingale of locally integrable variation, i.e., such that*

$$E\left[\int_0^t d|M_s|\right] < \infty \qquad \textit{for any } t > 0,$$

and let C be a predictable process satisfying

$$E\left[\int_0^t |C_s| d|M_s|\right] < \infty \qquad \textit{for any } t > 0.$$

Then the stochastic integral $C * M$ is a martingale.

B

Convergence of Probability Measures on Metric Spaces

B.1 Metric Spaces

For more details on the following and further results refer to Loève (1963), Dieudonné (1960), and Aubin (1977).

Definition B.1. Consider a set R. A *distance (metric)* on R is a mapping $\rho : R \times R \to \mathbb{R}_+$, which satisfies the following properties.

D1.For any $x, y \in R$, $\rho(x, y) = 0 \Leftrightarrow x = y$.
D2.For any $x, y \in R$, $\rho(x, y) = \rho(y, x)$.
D3.For any $x, y, z \in R$, $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$ (triangle inequality).

Definition B.2. A *metric space* is a set R endowed with a metric ρ; we shall write (R, ρ). Elements of a metric space will be called *points*.

Definition B.3. Given a metric space (R, ρ), a point $a \in R$, and a real number $r > 0$, the *open ball* (or the *closed ball*) of center a and radius r is the set $B(a, r) := \{x \in R | \rho(a, x) < r\}$ (or $B'(a, r) := \{x \in R | \rho(a, x) \leq r\}$).

Definition B.4. In a metric space (R, ρ), an *open set* is any subset A of R such that for any $x \in A$ there exists an $r > 0$ such that $B(a, r) \subset A$.

The empty set is open, and so is the entire space R.

Proposition B.5. *The union of any family of open sets is an open set. The intersection of a finite family of open sets is an open set.*

Definition B.6. The family $\mathcal{T}$ of all open sets in a metric space is called its *topology*. In this respect the couple $(R, \mathcal{T})$ is a *topological space*.

Definition B.7. The *interior* of a set A is the largest open subset of A.

Definition B.8. In a metric space (R, ρ), a *closed set* is any subset of R which is a complement of an open set.

The empty set is closed, and so is the entire space R.

Proposition B.9. *The intersection of any family of closed sets is a closed set. The union of a finite family of closed sets is a closed set.*

Definition B.10. In a metric space (R, ρ), the *closure* of a set A is the smallest subset of R containing A. It is denoted by $\bar{A}$. Any element of the closure of A is called a *point of closure* of A.

Proposition B.11. *A closed set is the intersection of a decreasing sequence of open sets. An open set is the union of an increasing sequence of closed sets.*

Definition B.12. *A topological space is called a* Hausdorff topological space *if it satisfies the following property:*

(HT) *For any two distinct points x and y there exist two disjoint open sets A and B such that $x \in A$ and $y \in B$.*

Proposition B.13. *A metric space is a Hausdorff topological space.*

Definition B.14. In a metric space (R, ρ), the *boundary* of a set A is the set $\partial A = \bar{A} \cap (R \setminus A)$. Here $R \setminus A$ is the complement of A.

Definition B.15. Given two metric spaces (R, ρ) and (R', ρ'), a function $f : R \to R'$ is *continuous*, if for any open set A' in (R', ρ'), the set $f^{-1}(A')$ is an open set in (R, ρ).

Definition B.16. Two metric spaces (R, ρ) and (R', ρ') are said to be *homeomorphic* if a function $f : R \to R'$ exists satisfying the following two properties:

1. f is a bijection (an invertible function);
2. f is bicontinuous; i.e., both f and its inverse f^{-1} are continuous.

The function f above is called a *homeomorphism*.

Definition B.17. Given two distances ρ and ρ' on the same set R, we say that they are *equivalent distances* if the identity $i_R : x \in R \mapsto x \in R$ is a homeomorphism between the metric spaces (R, ρ) and (R', ρ').

Remark B.18. We may remark here that the notions of open set, closed set, closure, boundary, and continuous function are *topological notions*. They depend only on the topology induced by the metric. The topological properties of a metric space are invariant with respect to a homeomorphism.

Definition B.19. Given a subset A of a metric space (R, ρ) its *diameter* is given by $\delta(A) = \sup_{x \in A, y \in A} d(x, y)$. A is *bounded* if its diameter is finite.

Definition B.20. Given two metric spaces (R, ρ) and (R', ρ'), a function $f : R \to E'$ is *uniformly continuous* if for any $\epsilon > 0$, a $\delta > 0$ exists such that $x, y \in R$, $\rho(x, y) < \delta$ implies $\rho'(f(x), f(y)) < \epsilon$.

Proposition B.21. *A uniformly continuous function is continuous. (The converse is not true in general.)*

Remark B.22. The notions of diameter of a set and of uniform continuity of a function are *metric notions.*

Definition B.23. Let A, B be two subsets of a metric space R. A is said to be *dense* in B if $B \subseteq \bar{A}$. A is said to be *everywhere dense* in R if $\bar{A} = R$.

Definition B.24. A metric space R is said to be *separable* if it contains an everywhere dense countable subset.

Here are some examples of separable spaces with the corresponding everywhere countable subset.

- The space $\mathbb{R}$ of real numbers with distance function $\rho(x, y) = |x - y|$, with the set $\mathbb{Q}$.
- The space $\mathbb{R}^n$ of ordered n-tuples of real numbers $x = (x_1, x_2, \ldots, x_n)$ with distance function $\rho(x, y) = \{\sum_{k=1}^{n}(y_k - x_k)^2\}^{\frac{1}{2}}$, with the set of all vectors with rational coordinates.
- The space $\mathbb{R}_0^n$ of ordered n-tuples of real numbers $x = (x_1, x_2, \ldots, x_n)$ with distance function $\rho_0(x, y) = \max\{|y_k - x_k|; 1 \leq k \leq n\}$ with the set of all vectors with rational coordinates.
- $C^2([a, b])$, the totality of all continuous functions on the segment $[a, b]$ with distance function $\rho(x, y) = \int_a^b [x(t) - y(t)]^2 dt$ with the set of all polynomials with rational coefficients.

Definition B.25. A family $\{G_\alpha\}$ of open sets in the metric space R is called a *basis* of R if every open set in R can be represented as the union of a (finite or infinite) number of sets belonging to this family.

Definition B.26. R is said to be a space with countable basis if there is at least one basis in R consisting of a countable number of elements.

Theorem B.27. *A necessary and sufficient condition for R to be a space with countable basis is that there exists in R an everywhere dense countable set.*

Corollary B.28. A metric space R is separable if and only if it has a countable basis.

Definition B.29. A *covering* of a set is a family of sets, whose union contains the set. If the number of elements of the family is countable, then we have a *countable covering.* If the sets of the family are open, we have an open *covering.*

Theorem B.30. *If R is a separable space, then we can select a countable covering from each of its open coverings.*

Theorem B.31. *Every separable metric space R is homeomorphic to a subset of* $\mathbb{R}^\infty$.

Definition B.32. In a metric space (R, ρ), a sequence $(x_n)_{n\in\mathbb{N}}$ is any function from $\mathbb{N}$ to R.

Definition B.33. We say that a sequence $(x_n)_{n\in\mathbb{N}}$ admits a *limit* $b \in R$ (is convergent to b), if b is such that for any open set V, with $x \in V$, there exists an $n_V \in \mathbb{N}$ such that for any $n > n_V$ we have $x_n \in V$. We write $\lim_{n\to\infty} x_n = b$.

Definition B.34. A subsequence of a sequence $(x_n)_{n\in\mathbb{N}}$ is any sequence $k \in \mathbb{N} \mapsto x_{n_k} \in R$ such that $(n_k)_{k\in\mathbb{N}}$ is strictly increasing.

Proposition B.35. *If* $\lim_{n\to\infty} x_n = b$*, then* $\lim_{k\to\infty} x_{n_k} = b$ *for any subsequence of* $(x_n)_{n\in\mathbb{N}}$.

Definition B.36. b is called a *cluster point* of a sequence $(x_n)_{n\in\mathbb{N}}$ if a subsequence exists having b as a limit.

Proposition B.37. *Given a subset A of a metric space (R, ρ), for any $a \in \bar{A}$ there exists a sequence of elements of A converging to a.*

Proposition B.38. *If x is the limit of a sequence $(x_n)_{n\in\mathbb{N}}$, then x is the unique cluster point of $(x_n)_{n\in\mathbb{N}}$. Conversely, $(x_n)_{n\in\mathbb{N}}$ may have a unique cluster point x and still this does not imply that x is the limit of $(x_n)_{n\in\mathbb{N}}$ (see Aubin (1977), page 67, for a counterexample).*

Definition B.39. In a metric space (R, ρ), a *Cauchy sequence* is a sequence $(x_n)_{n\in\mathbb{N}}$ such that for any $\epsilon > 0$ an integer $n_0 \in \mathbb{N}$ exists such that $m, n \in \mathbb{N}$, $m, n > n_0$ implies $\rho(x_m, x_n) < \epsilon$.

Proposition B.40. *In a metric space, any convergent sequence is a Cauchy sequence. The converse is not true in general.*

Proposition B.41. *In a metric space, if a Cauchy sequence $(x_n)_{n\in\mathbb{N}}$ has a cluster point x, then x is the limit of $(x_n)_{n\in\mathbb{N}}$.*

Definition B.42. A metric space R is called *complete* if any Cauchy sequence in R is convergent to a point of R.

Definition B.43. A subspace of a metric space (R, ρ) is any nonempty subset F of R endowed with the restriction of ρ to $F \times F$.

Proposition B.44. *If a subspace of a metric space R is complete, then it is closed in R. In a complete metric space any closed subspace is complete.*

Definition B.45. A metric space R is said to be *compact* if any arbitrary open covering $\{O_\alpha\}$ of the space R contains a finite subcovering.

Definition B.46. A metric space R is called *precompact* if, for all $\epsilon > 0$, there is a finite covering of R by sets of diameter $< \epsilon$.

Remark B.47. The notion of compactness is a topological one, while the notion of precompactness is a metric one.

Theorem B.48. *For a metric space R, the following three conditions are equivalent:*

1. *R is compact.*
2. *Any infinite sequence in R has at least a limit point.*
3. *R is precompact and complete.*

Proposition B.49. *Every precompact metric space is separable.*

Proposition B.50. *In a compact metric space any sequence which has only one cluster value, a converges to a.*

Proposition B.51. *Any continuous mapping of a compact metric space into another metric space is uniformly continuous.*

Definition B.52. A *compact set* (or *precompact set*) in a metric space R is any subset of R that is compact (or precompact) as a subspace of R.

Proposition B.53. *Any precompact set is bounded.*

Proposition B.54. *Any compact set in a metric space is closed. In a compact metric space, any closed subset is compact.*

Proposition B.55. *Any compact set in a metric space is complete.*

Definition B.56. A set M in the metric space R is said to be *relatively compact* if $M = \bar{M}$.

Theorem B.57. *A relatively compact set is precompact. In a complete metric space a precompact set is relatively compact.*

Proposition B.58. *A necessary and sufficient condition that a subset M of a metric space R be relatively compact is that every sequence of points of M has a cluster point in R.*

Definition B.59. A metric space R is said to be *locally compact*, if for every point $x \in R$ there exists a compact neighborhood of x in R.

Theorem B.60. *Let R be a locally compact metric space. The following properties are equivalent:*

1. *there exists an increasing sequence (U_n) of open relatively compact sets in R, such that $\bar{U}_n \subset U_{n+1}$ for every n, and $R = \cup_n U_n$;*
2. *R is the countable union of compact subsets;*
3. *R is separable.*

Convergence of Probability Measures

Let (S, ρ) be a metric space and let $\mathcal{S}$ be the σ-algebra of Borel subsets generated by the topology induced by ρ. Let $P, P_1, P_2, \ldots$ be probability measures on $(S, \mathcal{S})$.

Definition B.61. A sequence of probability measures $\{P_n\}_{n\in\mathbb{N}}$ converges weakly to the probability measure P (notation $P_n \xrightarrow{\mathcal{W}} P$) if

$$\int_E f dP_n \to \int_E f dP$$

for every function $f \in C_b(S)$, the class of continuous bounded functions on S.

Definition B.62. A set A in $\mathcal{S}$ such that $P(\partial A) = 0$ is called a P-*continuity set.*

Theorem B.63. *Let P_n and P be probability measures on $(S, \mathcal{S})$. These five conditions are equivalent:*

1. $P_n \xrightarrow{\mathcal{W}} P$,
2. $\lim_n \int f dP_n = \int f dP$ *for all bounded, uniformly continuous real functions* f,
3. $\limsup_n P_n(F) \leq P(F)$ *for all closed* F,
4. $\liminf_n P_n(G) \geq P(G)$ *for all open* G,
5. $\lim_n P_n(A) = P(A)$ *for all* P-continuity sets A.

On the set of probability measures on $(S, \mathcal{S})$, we may refer to the topology of weak convergence.

Definition B.64. Let Π be a family of probability measures on $(S, \mathcal{S})$. Π is said to be *relatively compact* if every sequence of elements of Π contains a weakly convergent subsequence; i.e., for every sequence $\{P_n\}$ in Π there exists a subsequence $\{P_{n_k}\}$ and a probability measure P (defined on $(S, \mathcal{S})$, but not necessarily an element of Π) such that $P_{n_k} \xrightarrow{\mathcal{W}} P$.

Definition B.65. A family Π of probability measures on the general metric space $(S, \mathcal{S})$ is said to be *tight* if, for all $\epsilon > 0$, there exists a compact set K such that

$$P(K) > 1 - \epsilon \qquad \forall P \in \Pi.$$

Consider sequences of random variables (X_n) and (Y_n) valued in a metric separable space (S, ρ) having common domain; it makes sense to speak of the distance $\rho(X_n, Y_n)$, the function with value $\rho(X_n(\omega), Y_n(\omega))$ at ω. Since S is separable, $\rho(X_n, Y_n)$ is a random variable (see Billingsley (1968), page 225), and we have the following theorem.

Theorem B.66. *If $X_n \xrightarrow{\mathcal{D}} X$ and $\rho(X_n, Y_n) \xrightarrow{P} 0$, then $Y_n \xrightarrow{\mathcal{D}} X$.*

Let h be a measurable mapping of the metric space S into another metric space S'. Denote by $h(P)$ the probability measure induced by h on $(S'\mathcal{S}')$, defined by $h(P)(A) = P(h^{-1}(A))$ for any $A \in \mathcal{S}'$. Let D_h be the set of discontinuities of h.

Theorem B.67. *If $P_n \xrightarrow{\mathcal{W}} P$ and $P(D_h) = 0$, then $h(P_n) \xrightarrow{\mathcal{W}} h(P)$.*

For a random element X of S, $h(X)$ is a random element of S' (since h is measurable), and we have the following corollary.

Corollary B.68. If $X_n \xrightarrow{\mathcal{D}} X$ and $P(X \in D_h) = 0$, then $h(X_n) \xrightarrow{\mathcal{D}} h(X)$.

We recall now one of the most frequently used results in analysis.

Theorem B.69. (Helly). *For every sequence (F_n) of distribution functions there exists a subsequence (F_{n_k}) and a nondecreasing, right-continuous function F (a generalized distribution function) such that $0 \leq F \leq 1$ and $\lim_k F_{n_k}(x) = F(x)$ at continuity points x of F.*

Consider a probability measure P on $(\mathbb{R}^\infty, \mathcal{B}_{\mathbb{R}^\infty})$ and let π_k be the projection from $\mathbb{R}^\infty$ to $\mathbb{R}^k$, defined by $\pi_{i_1,\ldots,i_k}(x) = (x_{i_1}, \ldots, x_{i_k})$. The functions $\pi_k(P) : \mathbb{R}^k \to [0,1]$ are called *finite-dimensional distributions* corresponding to P. It is possible to show that probability measures on $(\mathbb{R}^\infty, \mathcal{B}_{\mathbb{R}^\infty})$ converge weakly if and only if all the corresponding finite-dimensional distributions converge weakly.

Let $C := C([0,1])$ be the space of continuous functions on $[0,1]$ with the uniform topology, i.e., the topology obtained by defining the distance between two points $x, y \in C$ as $\rho(x,y) = \sup_t |x(t) - y(t)|$. We shall denote with $(C, \mathcal{C})$ the space C with the topology induced by this metric ρ.

For $t_1, \ldots, t_k$ in $[0,1]$, let $\pi_{t_1 \ldots t_k}$ be the mapping that carries the point x of C to the point $(x(t_1), \ldots, x(t_k))$ of $\mathbb{R}^k$. The finite-dimensional distributions of a probability measure P on $(C, \mathcal{C})$ are defined as the measures $\pi_{t_1 \ldots t_k}(P)$. Since these projections are continuous, the weak convergence of probability measures on $(C, \mathcal{C})$ implies the weak convergence of the corresponding finite-dimensional distributions, but the converse fails (perhaps in the presence of singular measures).

Definition B.70. A sequence (X_n) of random variables with values in a common measurable space $(S, \mathcal{S})$ converges in distribution to the random variable X (notation $X_n \xrightarrow{\mathcal{D}} X$), if the probability laws P_n of the X_n converge weakly to the probability law P of X:

$$P_n \xrightarrow{\mathcal{W}} P.$$

B.2 Prohorov's Theorem

Prohorov's theorem, gives, under suitable hypotheses, equivalence among relative compactness and tightness of families of probability measures.

Theorem B.71. (Prohorov). *Let Π be a family of probability measures on the probability space $(S, \mathcal{S})$. Then*

1. *if Π is tight, then it is relatively compact;*
2. *suppose S is separable and complete; if Π is relatively compact, then it is tight.*

Proof: See, e.g., Billingsley (1968). □

B.3 Donsker's Theorem

Weak Convergence and Tightness in $C([0,1])$

Consider the space $C := C([0,1])$ of continuous functions on $[0,1]$. Weak convergence of finite-dimensional distributions of a sequence of probability measures on C is not a sufficient condition for weak convergence of the sequence itself in C. One can prove (see, e.g., Billingsley (1968)) that an additional condition is needed, i.e., relative compactness of the sequence. Since C is a Polish space, i.e., a separable and complete metric space, by Prohorov's theorem we have the following result.

Theorem B.72. *Let (P_n) and P be probability measures on $(C, \mathcal{C})$. If the finite-dimensional distributions of P_n converge weakly to those of P, and if $\{P_n\}$ is tight, then $P_n \xrightarrow{\mathcal{W}} P$.*

To use this theorem we provide here some characterization of tightness. Given a $\delta \in]0,1]$, a δ-continuity modulus of an element x of C is defined by

$$w_x(\delta) = w(x, \delta) = \sup_{|s-t|<\delta} |x(s) - x(t)|, \qquad 0 < \delta \leq 1.$$

Let (P_n) be a sequence of probability measures on $(C, \mathcal{C})$.

Theorem B.73. *The sequence (P_n) is tight if and only if these two conditions hold:*

1. *For each positive η, there exists an a such that*
$$P_n(x||x(0)| > a) \leq \eta, \qquad n \geq 1.$$
2. *For each positive ϵ and η, there exists a δ, with $0 < \delta < 1$, and an integer n_0 such that*
$$P_n(x|w_x(\delta) \geq \epsilon) \leq \eta, \quad n \geq n_0.$$

The following theorem gives a sufficient condition for compactness.

Theorem B.74. *If the following two conditions are satisfied:*

1. *For each positive* η*, there exists an* a *such that*

$$P_n(x||x(0)| > a) \leq \eta \qquad n \geq 1.$$

2. *For each positive* ϵ *and* η*, there exists a* δ*, with* $0 < \delta < 1$*, and an integer* n_0 *such that*

$$\frac{1}{\delta} P_n \left(x \middle| \sup_{t \leq s \leq t+\delta} |x(s) - x(t)| \geq \epsilon \right) \leq \eta, \qquad n \geq n_0,$$

for all t*, then the sequence* (P_n) *is tight.*

Let X be a mapping from $(\Omega, \mathcal{F}, P)$ into $(C, \mathcal{C})$. For all $\omega \in \Omega$, $X(\omega)$ is an element of C, i.e., a continuous function on $[0,1]$, whose value at t we denote by $X(t,\omega)$. For fixed t, let $X(t)$ denote the real function on Ω with value $X(t,\omega)$ at ω. Then $X(t)$ is the projection $\pi_t X$.

Similarly, let $(X(t_1), X(t_2), \ldots, X(t_k))$ denote the mapping from Ω into $\mathbb{R}^k$ with values $(X(t_1,\omega), X(t_2,\omega), \ldots, X(t_k,\omega))$ at ω. If each $X(t)$ is a random variable, X is said to be a random function. Suppose now that (X_n) is a sequence of random functions. According to Theorem B.73, (X_n) is tight if and only if the sequence $(X_n(0))$ is tight, and for any positive real numbers ϵ and η there exists δ, $(0 < \delta < 1)$ and an integer n_0 such that

$$P\left(w_{X_n}(\delta) \geq \epsilon\right) \leq \eta, \qquad n \geq n_0.$$

This condition states that the random functions X_n do not oscillate too much. Theorem B.74 can be restated in the same way: (X_n) is tight if $(X_n(0))$ is tight and if for any positive ϵ and η there exists a δ, $0 < \delta < 1$, and an integer n_0 such that

$$\frac{1}{\delta} P\left(\sup_{t \leq s \leq t+\delta} |X_n(s) - X_n(t)| \geq \epsilon \right) \leq \eta \tag{B.1}$$

for $n \geq n_0$ and $0 \leq t \leq 1$. Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables on $(\Omega, \mathcal{F}, P)$ with mean 0 and variance σ^2. We define the sequence of partial sums $S_n = \xi_1 + \cdots + \xi_n$, with $S_0 = 0$. Let us construct the sequence of random variables X_n from the sequence (S_n) by means of rescaling and linear interpolation, as follows:

$$X_n\left(\frac{i}{n}, \omega\right) = \frac{1}{\sigma\sqrt{n}} S_i(\omega) \qquad \text{for} \qquad \frac{i}{n} \in [0,1[; \tag{B.2}$$

$$\frac{X_n(t) - X_n\left(\frac{i-1}{n}\right)}{X_n\left(\frac{i}{n}\right) - X_n\left(\frac{i-1}{n}\right)} - \frac{t - \frac{i-1}{n}}{\frac{1}{n}} = 0 \qquad \text{for} \qquad t \in \left[\frac{(i-1)}{n}, \frac{i}{n}\right]. \tag{B.3}$$

With a little algebra, we obtain

$$\begin{aligned}
X_n(t) &= X_n\left(\frac{i-1}{n}\right) + \frac{t-\frac{i-1}{n}}{\frac{1}{n}}\left(X_n\left(\frac{i}{n}\right) - X_n\left(\frac{i-1}{n}\right)\right)\\
&= \frac{t-\frac{i-1}{n}}{\frac{1}{n}} X_n\left(\frac{i}{n}\right) + \left(\frac{\frac{i}{n}-t}{\frac{1}{n}}\right)\\
&= \frac{1}{\sigma\sqrt{n}} S_{i-1}(\omega)\frac{\frac{i}{n}-t}{\frac{1}{n}} + \frac{t-\frac{(i-1)}{n}}{\frac{1}{n}}\frac{1}{\sigma\sqrt{n}} S_i(\omega)\\
&= \frac{1}{\sigma\sqrt{n}} S_{i-1}(\omega)\left(\frac{\frac{i}{n}-t}{\frac{1}{n}} + \frac{t-\frac{i}{n}+\frac{1}{n}}{\frac{1}{n}}\right) + \frac{1}{\sigma\sqrt{n}}\frac{t-\frac{(i-1)}{n}}{\frac{1}{n}}\xi_i(\omega)\\
&= \frac{1}{\sigma\sqrt{n}} S_{i-1}(\omega) + n\left(t-\frac{i-1}{n}\right)\frac{1}{\sigma\sqrt{n}}\xi_i(\omega).
\end{aligned}$$

Since $i-1=[nt]$, if $t \in [\frac{(i-1)}{n}, \frac{i}{n}]$, we may rewrite equation (B.3) as follows:

$$X_n(t,\omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt-[nt])\frac{1}{\sigma\sqrt{n}}\xi_{[nt]+1}(\omega). \tag{B.4}$$

For any fixed ω, $X_n(.,\omega)$ is a piecewise linear function whose pieces' amplitude decreases as n increases. Since the ξ_i and hence the S_i are random variables it follows by (B.4) that $X_n(t)$ is a random variable for each t. Therefore, the X_n are random functions.

The following theorem provides a sufficient condition for (X_n) to be a tight sequence.

Theorem B.75. *Suppose (X_n) is defined by (B.4). The sequence (X_n) is tight if for each positive ϵ there exists a λ, with $\lambda > 1$, and an integer n_0 such that, if $n \geq n_0$, then*

$$P\left(\max_{i\leq n}|S_{k+i}-S_k| \geq \lambda\sigma\sqrt{n}\right) \leq \frac{\epsilon}{\lambda^2} \tag{B.5}$$

holds for all k.

Let us denote by P_W the probability measure of the Wiener process as defined in Definition 2.134 and whose existence is a consequence of Theorem 2.54. We will refer here to its restriction to $t \in [0,1]$, so that its trajectories are almost surely elements of $C([0,1])$.

Lemma B.76. *Let $\xi_1,\ldots,\xi_m$ be independent random variables with mean 0 and finite variance σ_i^2; put $S_i = \xi_1+\cdots+\xi_i$ and $s_i^2 = \sigma_1^2+\cdots+\sigma_i^2$. Then*

$$P\left(\max_{i\leq m}|S_i| \geq \lambda s_m\right) \leq 2P\left(|S_m| \geq (\lambda-\sqrt{2})s_m\right). \tag{B.6}$$

Theorem B.77. (Donsker). *Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent identically distributed random variables defined on $(\Omega, \mathcal{F}, P)$ with mean 0 and finite, positive variance σ^2:*

$$E[\xi_n] = 0, \qquad E[\xi_n^2] = \sigma^2.$$

Let $S_n = \xi_1 + \xi_2 + \cdots + \xi_n$. Then the random functions

$$X_n(t,\omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}(\omega) + (nt - [nt]) \frac{1}{\sigma\sqrt{n}} \xi_{[nt]+1}(\omega)$$

satisfy $X_n \xrightarrow{\mathcal{D}} W$.

Proof: We first show that the finite-dimensional distributions of $\{X_n\}$ converge to those of W. Consider first a single time point s; we need to prove that

$$X_n(s) \xrightarrow{\mathcal{W}} W_s.$$

Since

$$\left| X_n(s) - \frac{1}{\sigma\sqrt{n}} S_{[ns]} \right| = (ns - [ns]) \left| \frac{1}{\sigma\sqrt{n}} \xi_{[ns+1]} \right|$$

and since, by Chebyshev's inequality,

$$\begin{aligned} P\left(\left| \frac{1}{\sigma\sqrt{n}} \xi_{[ns]+1} \right| \geq 1 \right) &\leq \frac{E\left[\left| \frac{1}{\sigma\sqrt{n}} \xi_{[ns]+1} \right|^2 \right]}{\epsilon^2} \\ &= \frac{1}{\sigma n \epsilon^2} E\left[\xi_{[ns]+1}^2 \right] = \frac{1}{\sigma n \epsilon} \sigma^2 \\ &= \frac{\sigma}{n\epsilon} \to 0, \qquad n \to \infty, \end{aligned}$$

we obtain

$$\left| X_n(s) - \frac{1}{\sigma\sqrt{n}} S_{[ns]} \right| \xrightarrow{P} 0. \tag{B.7}$$

Since $\lim_{n\to\infty} \frac{[ns]}{ns} = 1$, by the Lindeberg Theorem 1.92

$$\frac{1}{\sigma\sqrt{ns}} \sum_{k=1}^{[ns]} \xi_k \xrightarrow{\mathcal{D}} N(0,1),$$

so that

$$\frac{1}{\sigma\sqrt{n}} S_{[ns]} \xrightarrow{\mathcal{D}} W_s.$$

Therefore, by Theorem B.66, $X_n(s) \xrightarrow{\mathcal{D}} W_s$. Consider now two time points s and t with $s < t$. We must prove

$$(X_n(s), X_n(t)) \xrightarrow{\mathcal{D}} (W_s, W_t).$$

Since

$$\left|X_n(t) - \frac{1}{\sigma\sqrt{n}} S_{[nt]}\right| \xrightarrow{P} 0 \qquad \text{and} \qquad \left|X_n(s) - \frac{1}{\sigma\sqrt{n}} S_{[ns]}\right| \xrightarrow{P} 0$$

by Chebyshev's inequality, so that

$$\left\|(X_n(s), X_n(t)) - \left(\frac{1}{\sigma\sqrt{n}} S_{[ns]}, \frac{1}{\sigma\sqrt{n}} S_{[nt]}\right)\right\|_{\mathbb{R}^2} \xrightarrow{P} 0,$$

and by Theorem B.66, it is sufficient to prove that

$$\frac{1}{\sigma\sqrt{n}} \left(S_{[ns]}, S_{[nt]}\right) \xrightarrow{\mathcal{D}} (W_s, W_t).$$

By Corollary B.68 of Theorem B.67 this is equivalent to proving

$$\frac{1}{\sigma\sqrt{n}} \left(S_{[ns]}, S_{[nt]} - S_{[ns]}\right) \xrightarrow{\mathcal{D}} (W_s, W_t - W_s).$$

For independence of the random variables ξ_i, $i = 1, 2, \ldots, n$, the random variables $S_{[ns]}$ and $S_{[nt]} - S_{[ns]}$ are independent, so that

$$\begin{aligned} &\lim_{n\to\infty} E\left[e^{\frac{\mathrm{i}u}{\sigma\sqrt{n}} \sum_{j=1}^{[ns]} \xi_j + \frac{\mathrm{i}v}{\sigma\sqrt{n}} \sum_{j=[ns]+1}^{[nt]} \xi_j}\right] \\ &= \lim_{n\to\infty} E\left[e^{\frac{\mathrm{i}u}{\sigma\sqrt{n}} \sum_{j=1}^{[ns]} \xi_j}\right] \cdot \lim_{n\to\infty} E\left[e^{\frac{\mathrm{i}v}{\sigma\sqrt{n}} \sum_{j=[ns]+1}^{[nt]} \xi_j}\right]. \end{aligned} \tag{B.8}$$

Since $\lim_{n\to\infty} \frac{[ns]}{ns} = 1$, by the Lindeberg Theorem 1.92

$$\frac{1}{\sigma\sqrt{n}} S_{[ns]} \xrightarrow{\mathcal{D}} N(0, s)$$

and for the same reason

$$\frac{1}{\sigma\sqrt{n}} (S_{[nt]} - S_{[ns]}) \xrightarrow{\mathcal{D}} N(0, t - s),$$

so that

$$\lim_{n\to\infty} E\left[e^{\frac{\mathrm{i}u}{\sigma\sqrt{n}} S_{[ns]}}\right] = e^{-\frac{u^2 s}{2}}$$

and

$$\lim_{n\to\infty} E\left[e^{\frac{\mathrm{i}v}{\sigma\sqrt{n}} S_{[nt]} - S_{[ns]}}\right] = e^{-\frac{u^2 s}{2}}.$$

Substitution of these two last equations into (B.8) gives

$$\frac{1}{\sigma\sqrt{n}} \left(S_{[ns]}, S_{[nt]} - S_{[ns]}\right) \xrightarrow{\mathcal{D}} (W_s, W_t - W_s),$$

and consequently

$$(X_n(s), X_n(t)) \xrightarrow{\mathcal{D}} (W_s, W_t).$$

A set of three or more time points can be treated in the same way, and hence the finite-dimensional distributions converge properly. Applying Lemma B.76 to the random variables $\xi_1, \xi_2, \ldots, \xi_n$, we have

$$P\left(\max_{i \le n} |S_i| \ge \lambda\sqrt{n}\sigma\right) \le 2P\left(|S_n| \ge (\lambda - \sqrt{2})\sqrt{n}\sigma\right).$$

For $\frac{\lambda}{2} > \sqrt{2}$ we have

$$P\left(\max_{i \le n} |S_i| \ge \lambda\sqrt{n}\sigma\right) \le 2P\left(|S_n| \ge \frac{\lambda}{2}\sqrt{n}\sigma\right).$$

By the Central Limit Theorem,

$$P\left(|S_n| \ge \frac{1}{2}\lambda\sigma\sqrt{n}\right) \to P\left(|N| \ge \frac{1}{2}\lambda\right) < \frac{8}{\lambda^3} E\left[|N|^3\right],$$

where the last inequality follows by Chebyshev's and $N \sim N(0, 1)$. Therefore, if ϵ is positive, there exists a λ such that

$$\limsup_{n\to\infty} P\left(\max_{i \le n} |S_i| \ge \lambda\sigma\sqrt{n}\right) < \frac{\epsilon}{\lambda^2}$$

and then, by Theorem B.75, the family of the distribution functions of X_n is tight. Since C is separable and complete, by Prohorov's theorem this family is relatively compact and then $X_n \xrightarrow{\mathcal{D}} X$. □

An Application of Donsker's Theorem

Donsker's theorem has the following qualitative interpretation: $X_n \xrightarrow{\mathcal{D}} W$ implies that, if τ is small, then a particle subject to independent displacements $\xi_1, \xi_2, \ldots$ at successive times $\tau_1, \tau_2, \ldots$ appears to follow approximately a Brownian motion.

More important than this qualitative interpretation is the use of Donsker's theorem to prove limit theorems for various functions of the partial sums S_n. By using Donsker's theorem it is possible to use the relation $X_n \xrightarrow{\mathcal{D}} W$ to derive the limiting distribution of $\max_{i \le n} S_i$.

Since $h(x) = \sup_t x(t)$ is a continuous function on C, $X_n \xrightarrow{\mathcal{D}} W$ implies, by Corollary B.68, that

$$\sup_{0 \le t \le 1} X_n(t) \xrightarrow{\mathcal{D}} \sup_{0 \le t \le 1} W_t.$$

The obvious relation

$$\sup_{0\le t\le 1} X_n(t) = \max_{i\le n} \frac{1}{\sigma\sqrt{n}} S_i$$

implies

$$\frac{1}{\sigma\sqrt{n}} \max_{i\le n} S_i \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1} W_t. \tag{B.9}$$

Thus, under the hypotheses of Donsker's theorem, if we knew the distribution of $\sup_t W_t$ we would have the limiting distribution of $\max_{i\le n} S_i$. The technique we shall use to obtain the distribution of $\sup_t W_t$ is to compute the limit distribution of $\max_{i\le n} S_i$ in a simple special case and then using $h(X_n) \xrightarrow{\mathcal{D}} h(W)$, where h is continuous on C or continuous except at points forming a set of Wiener measure 0, we obtain the distribution of $\sup_t W_t$ in the general case.

Suppose that $S_0, S_1, \dots$ are the random variables for a symmetric random walk starting from the origin; this is equivalent to supposing that ξ_n are independent and satisfy

$$P(\xi_n = 1) = P(\xi_n = -1) = \frac{1}{2}. \tag{B.10}$$

Let us show that if a is a nonnegative integer, then

$$P\left(\max_{0\le i\le n} S_i \ge a\right) = 2P(S_n > a) + P(S_n = a). \tag{B.11}$$

If $a = 0$ the previous relation is obvious; in fact, since $S_0 = 0$,

$$P\left(\max_{0\le i\le n} S_i \ge 0\right) = 1$$

and obviously, by symmetry of S_n

$$2P(S_n > 0) + P(S_n = 0) = P(S_n > 0) + P(S_n < 0) + P(S_n = 0) = 1.$$

Suppose now that $a > 0$ and put $M_i = \max_{0\le j\le i} S_j$. Since

$$\{S_n = a\} \subset \{M_n \ge a\}$$

and

$$\{S_n > a\} \subset \{M_n \ge a\},$$

we have

$$P(M_n \ge a) - P(S_n = a) = P(M_n \ge a, S_n < a) + P(M_n \ge a, S_n > a)$$

and

$$P(M_n \ge a, S_n > a) = P(S_n > a).$$

Hence we have to show that

$$P(M_n \geq a, S_n < a) = P(M_n \geq a, S_n > a). \tag{B.12}$$

Because of (B.10), all 2^n possible paths $(S_1, S_2, \ldots, S_n)$ have the same probability 2^{-n}. Therefore, (B.12) will follow, if we show that the number of paths contributing to the left-hand event is the same as the number of paths contributing to the right-hand event. To show this it suffices to find a one-to-one correspondence between the paths contributing to the right-hand event and the paths contributing to the left-hand event.

Given a path $(S_1, S_2, \ldots, S_n)$ contributing to the left-hand event in (B.12), match it with the path obtained by reflecting through a all the partial sums after the first one that achieves the height a. Since the correspondence is one-to-one, (B.12) follows. This argument is an example of the reflection principle. See also Lemma 2.144.

Let α be an arbitrary nonnegative number, and let $a_n = -[-\alpha n^{\frac{1}{2}}]$. By (B.12) we have

$$P\left(\max_{i \leq n} \frac{1}{\sqrt{n}} S_i \geq a_n\right) = 2P(S_n > a_n) + P(S_n = a_n).$$

Since S_i can assume only integer values and since a_n is the smallest integer greater than or equal to $\alpha n^{\frac{1}{2}}$,

$$P\left(\max_{i \leq n} \frac{1}{\sqrt{n}} S_i \geq \alpha\right) = 2P(S_n < a_n) + P(S_n = a_n). \tag{B.13}$$

By the central limit theorem

$$P(S_n \geq a_n) \to P(N \geq \alpha),$$

where $N \sim N(0,1)$ and $\sigma^2 = 1$ by (B.10).

Since in the symmetric binomial distribution $S_n \to 0$ almost certainly, the term $P(S_n = a_n)$ is negligible. Thus

$$P\left(\max_{i \leq n} \frac{1}{\sqrt{n}} S_i \geq \alpha\right) \to 2P\left(N \geq \alpha\right), \qquad \alpha \geq 0. \tag{B.14}$$

By (B.14), (B.9), and (B.10), we conclude that

$$P\left(\sup_{0 \leq t \leq 1} W_t \leq \alpha\right) = \frac{2}{\sqrt{2\pi}} \int_0^\alpha e^{-\frac{1}{2}u^2} du, \qquad \alpha \geq 0. \tag{B.15}$$

If we drop the assumption (B.10) and suppose that the random variables ξ_n are independent and identically distributed and satisfy the hypothesis of Donsker's theorem, then (B.9) holds and from (B.15) we obtain

$$P\left(\frac{1}{\sigma\sqrt{n}} \max_{i \leq n} S_i \leq \alpha\right) \to \frac{2}{\sqrt{2\pi}} \int_0^\alpha e^{-\frac{1}{2}u^2} du, \qquad \alpha \geq 0. \tag{B.16}$$

Thus we have derived the limiting distribution of $\max_{i \le n} S_i$ by Lindeberg's theorem. Therefore, if the ξ_n are independent and identically distributed with $E[\xi_n] = 0$ and $E[\xi_n^2] = \sigma^2$, then the limit distribution of $h(X_n)$ does not depend on any further properties of the ξ_n. For this reason, Donsker's theorem is often called an invariance principle.

C

Maximum Principles of Elliptic and Parabolic Operators

The maximum principle is a generalization of the fact that if a function $f : [a,b] \to \mathbb{R}$, endowed with a first and second derivative, has $f'' > 0$ ($f'' < 0$) in $[a,b]$, then it attains its maximum (minimum) at the limits of the interval it is defined on.

In fact, if a function, as the solution of a certain differential equation, attains its maximum on the boundary of the domain Ω on which it is defined, then it is said to underlie a maximum principle. The latter is a remarkable instrument for the study of partial differential equations (e.g., uniqueness of solutions, comparison of solutions, etc.).

C.1 Maximum Principles of Elliptic Equations

Let $\Omega \subset \mathbb{R}$ be open bounded and let a, b, c, be real-valued functions defined on Ω. We consider the partial differential operator

$$L[u] = \frac{1}{2}a(x)u_{xx} + b(x)u_x + c(x)u. \tag{C.1}$$

L is said to be *elliptic in a point* $x_0 \in \Omega$ if $ax_0 > 0$. If for all $x \in \Omega\colon\ a(x) > 0$, then L is said to be *uniformly elliptic.*

Lemma C.1. *For* $a(x) > 0$, $c(x) \leq 0$, *for all* $x \in \Omega$, *if*

$$\exists \max_{x\in\Omega} u(x) = u(x_0) > 0, \qquad x_0 \in \Omega,$$

and $u \in C^2(\Omega)$. *Then* $L[u](x_0) \leq 0$.

Proof:

$$L[u](x_0) = \frac{1}{2}a(x_0)u_{xx}(x_0) + b(x_0)u_x(x_0) + c(x_0)u(x_0),$$

where $c(x_0)u(x_0) \leq 0$, $b(x_0)u_x(x_0) = 0$ and $a(x_0)u_{xx}(x_0) \leq 0$ with x_0 being the maximum point of u. $\square$

Theorem C.2. *Let $a(x) > 0$, $c(x) \leq 0$ for all $x \in \Omega$ and there exists a $\lambda > 0$ such that $\frac{1}{2}\lambda^2 a(x) + \lambda b(x) > 0$ for all $x \in \Omega$. If $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$, $L[u] \geq 0$ in Ω and if $\max_{\bar{\Omega}} u(x) > 0$, then $\sup_{\Omega} u(x) \leq \max_{\partial\Omega} u(x)$, where $\partial\Omega$ is the boundary of Ω.*

Proof: See, e.g., Friedman (1963). □

Corollary C.3. Under the assumptions of the preceding theorem, if $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$, $L[u] \leq 0$ in Ω, and if $\min_{\bar{\Omega}} u(x) < 0$, then $\inf_{\Omega} u \geq \min_{\partial\Omega} u$.

Proof: See, e.g., Friedman (1963). □

Theorem C.4. (Strong maximum principle). *Let L be a uniformly elliptic operator ($a(x) > 0$ in Ω) with bounded coefficients a, b, c on compact sets of Ω and let $c(x) \leq 0$ in Ω. If $u \in C^2(\Omega)$, $L[u] \geq 0$ ($L[u] \leq 0$) in Ω, and if $u \neq$ constant, then u cannot attain a positive maximum (negative minimum) in Ω.*

Proof: See, e.g., Friedman (1963). □

Remark C.5. The boundedness of the coefficients a, b, c is essential, as the following example demonstrates:

$$u_{xx} + b(x)u_x = 0, \tag{C.2}$$

where

$$b(x) = \begin{cases} -\frac{3}{x} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0. \end{cases}$$

It is easily verified that $u = 1 - x^4$ is the solution of (C.2), and moreover $\max_{[-1,1]} u(x) = u(0) = 1$. In fact, b is not bounded in compact neighborhoods of zero.

The First Boundary Value or Dirichlet Problem

The Dirichlet problem consists of finding a solution u of the system

$$\begin{cases} L[u](x) = f(x) & \text{in } \Omega, \\ u(x) = \phi(x) & \text{in } \partial\Omega. \end{cases} \tag{C.3}$$

Theorem C.6. *Let $a(x) > 0$, $c(x) \leq 0$; a, b, c, f uniformly Hölder continuous with exponent α in $\bar{\Omega}$. Then there exists a unique $u \in C^2(\Omega) \cap C^0(\bar{\Omega})$, solution of the Dirichlet problem.*

Proof: See, e.g., Friedman (1963) or (1964). □

C.2 Maximum Principles of Parabolic Equations

Let $Q \subset \mathbb{R}^2$ be open bounded and a, b, c be real-valued functions defined on Q. We consider the partial differential operator

$$M[u] = a(x,t)u_{xx} + b(x,t)u_x + c(x,t)u - u_t. \tag{C.4}$$

M is of *parabolic type* in $(x_0, t_0) \in Q$ if $a(x_0, t_0) > 0$. If $a(x,t) > 0$ in Q, then M is said to be *uniformly parabolic.*

We suppose $Q \subset \mathbb{R}\times]0, T[$ and define

$$\begin{aligned} D_T &= \{(x,T)|(x, t-\delta) \in Q, \forall 0 < \delta < \delta_0, \delta_0 \text{ independent of } x\}, \\ Q_0 &= Q \cup D_T, \\ \partial_0 Q &= \partial Q \setminus D_T. \end{aligned}$$

$\partial_0 Q$ is a closed set of $\mathbb{R}^2$ and is called a *parabolic boundary.*

Example C.7. $Q = \Omega\times]t_0, T[$.

Theorem C.8. *Let* $a(x,t) \geq 0$ *and* $c(x,t) \leq 0$ *in* Q*. If* $u \in C^0(\bar{Q})$*,* u_x, u_{xx}, u_t *belong to* $C^0(Q_0)$ *and if* $M[u] \geq 0$ *in* Q_0 *and* $\max_{\bar{Q}} u > 0$*, then* $\sup_{Q_0} u(x,t) \leq \max_{\partial_0 Q} u(x,t)$*.*

Proof: See, e.g., Friedman (1963) or (1964). □

Definition C.9. For every $P_0 = (x_0, t_0) \in Q$, let $S(P_0) = \{P \in Q|$ a simple continuous curve γ_{P_0} exists that is contained within Q and does not decrease along t passing from P to P_0 connecting P to $P_0\}$ and let $C(P_0)$ be the connecting component at $t = t_0$ of $Q \cap \{t = t_0\}$ that contains P_0. Clearly $C(P_0) \subset S(P_0)$.

Theorem C.10. (Strong maximum principle). *Let* M *be uniformly parabolic in* Q *with bounded coefficients and let* $c(x,t) \leq 0$*. If* u, u_x, u_{xx}, u_t *are continuous in* Q *and* $M[u] \geq 0$ *in* Q *and if* u *attains a positive maximum in the point* $P_0 = (x_0, t_0) \in Q$*, then*

$$u(P) = u(P_0) \qquad \forall P \in S(P_0).$$

Proof: See, e.g., Friedman (1963) or (1964). □

The First Boundary Value Problem

Let Q be a domain bounded in $\mathbb{R}$, $Q \subset \mathbb{R}\times]0, T[$ and define

$$\begin{aligned} \tilde{B}_T &= \bar{Q} \cap \{t = T\}, \\ \tilde{B} &= \tilde{Q} \cap \{t = 0\}, \\ B_T = \tilde{B}_T, &\quad B = \tilde{B}, \\ S_0 &= \{(x,t) \in \partial Q, 0 < t \leq T\}, \\ S &= S_0 \setminus B_T, \\ \partial_0 Q &= B \cup S \text{ parabolic boundary of } Q. \end{aligned}$$

The first boundary value problem consists of finding a solution u of the system

$$\begin{cases} M[u](x,t) = f(x,t) & \text{in } Q \cup B_T, \\ u(x,0) = \phi(x) & \text{in } B \text{ (initial condition)}, \\ u(x,t) = g(x,t) & \text{in } S \text{ (boundary condition)},, \end{cases} \tag{C.5}$$

where f, ϕ, g are appropriately chosen functions. If $g = \phi$ in $\bar{B} \cap \bar{S}$, then the solution u is always understood to be continuous in $\bar{Q}$.

Definition C.11. $\omega_R(P)$ is a *barrier function* in $R \in \bar{B} \cup S$ if

1. $\omega_R(P)$ is continuous in $\bar{Q}$,
2. $\omega_R(P) > 0$ for $P \in \bar{Q}, P \neq R, \omega_R(R) = 0$,
3. $M[\omega_R] \leq -1$ in $Q \cup B_T$.

Remark C.12. If $Q = \Omega \times (0,T)$, then there always exists a barrier function in any point $P_0 = (x_0, t_0)$ of S $(0 < t_0 \leq T)$ that is given by

$$\omega_{P_0} = K e^{\gamma t} \left(\frac{1}{R_0^p} - \frac{1}{R} \right),$$

where K and p are positive constants, $\gamma \geq c(x,t), R_0 = |x - x_0|, R = (|x - \bar{x}|^2 + (t - t_0)^2)^{\frac{1}{2}}$ with $\bar{x} > x_0$.

Theorem C.13. *Let M be uniformly parabolic in Q; the functions a, b, c, and f uniformly Hölder continuous in $\bar{Q}$; ϕ continuous in $\bar{B}$; and g continuous in $\bar{S}$ with $\phi = g$ in $\bar{B} \cap \bar{S}$. If, for all $R \in S$, there exists ω_R a barrier function in R, then there exists a unique u, solution of (C.5), with u_x, u_{xx}, u_t Hölder continuous.*

Proof: See, e.g., Friedman (1963) or (1964). □

The Cauchy Problem

Let $L[u] = a u_{xx} + b u_x + c u$ be an elliptic operator in $\mathbb{R}$ for all $t \in [0,T]$ and let $f : \mathbb{R} \times [0,T] \to \mathbb{R}$, $\phi : \mathbb{R} \to \mathbb{R}$ be two appropriately assigned functions. The Cauchy problem consists of finding a solution u of

$$\begin{cases} M[u] \equiv L[u] - u_t = f(x,t) & \text{in } \mathbb{R} \times]0,T], \\ u(x,0) = \phi(x) & \text{in } \mathbb{R}. \end{cases} \tag{C.6}$$

The solution is understood to be continuous in $\mathbb{R} \times [0,T]$, with the derivatives u_x, u_{xx}, u_t continuous in $\mathbb{R} \times]0,T]$.

Theorem C.14. *Let*

$$0 \leq a(x,t) \leq C, |b(x,t)| \leq C(|x| + 1), c(x,t) \leq C(|x|^2 + 1), \tag{C.7}$$

where C is a constant. If $M[u] \leq 0$ in $\mathbb{R} \times]0,T]$, $u(x,t) \geq -B \exp\{\beta |x|^2\}$ in $\mathbb{R} \times [0,T]$ (B, β positive constants), $u(x,0) \geq 0$ in $\mathbb{R}$, then $u(x,t) \geq 0$ in $\mathbb{R} \times [0,T]$.

Proof: See, e.g., Friedman (1963) or (1964). □

Corollary C.15. If $a(x,t) \geq 0$, satisfying (C.7), then there exists at least one solution u of the Cauchy problem with

$$|u(x,t)| \leq Be^{\beta|x|^2},$$

where b, β are positive constants.

Proof: See, e.g., Friedman (1963) or (1964). □

Theorem C.16. *Let*

$$a(x,t) \geq 0, |a(x,t)| \leq C(|x|^2+1), |b(x,t)| \leq C(|x|+1), c \leq C, \tag{C.8}$$

where C is a constant. If $M[u] \leq 0$ in $\mathbb{R}\times]0,T]$, $u(x,t) \geq -N(|x|^q+1)$ in $\mathbb{R}\times[0,T]$ (N, q positive constants), $u(x,0) \geq 0$ in $\mathbb{R}$, then $u(x,t) \geq 0$ in $\mathbb{R}\times[0,T]$.

Proof: See, e.g., Friedman (1964). □

Corollary C.17. If $a(x,t) \geq 0$, satisfying (C.8), then there exists at least one solution u of the Cauchy problem with

$$|u(x,t)| \leq N(1+|x|^q),$$

where N, q are positive constants.

Proof: See, e.g., Friedman (1964). □

Definition C.18. A *fundamental solution of the parabolic operator* $L - \frac{\partial}{\partial t}$ in $\mathbb{R}\times[0,T]$ is a function $\Gamma(x,t;\xi,r)$, defined, for all $(x,t) \in \mathbb{R}\times[0,T]$ and all $(\xi,t) \in \mathbb{R}\times[0,T]$, $t > r$, such that, for all f with compact support[14], the function

$$u(x,t) = \int_{\mathbb{R}} \Gamma(x,t;\xi,r) f(\xi) d\xi$$

satisfies

1. $L[u] - u_t = 0$ if $x \in \mathbb{R}, r < t \leq T$,
2. $u(x,t) \to f(x)$ if $t \downarrow r$.

We impose the following conditions:
(A_1) there exists a $\mu > 0$ such that $a(x,t) \geq \mu$ for all $(x,t) \in \mathbb{R}\times[0,T]$;
(A_2) the coefficients of L are continuous functions, bounded in $\mathbb{R}\times[0,T]$, and the coefficient $a(x,t)$ is continuous in t uniformly with respect to $(x,t) \in \mathbb{R}\times[0,T]$;
(A_3) the coefficients of L are Hölder continuous functions (with exponent α) in x, uniformly with respect to the variables (x,t) in compacts of $\mathbb{R}\times[0,T]$, and the coefficient $a(x,t)$ is Hölder continuous (with exponent α) in x, uniformly with respect to $(x,t) \in \mathbb{R}\times[0,T]$.

[14] The support of a function $f : \mathbb{R} \to \mathbb{R}$ is the set $\{x \in \mathbb{R} | f(x) \neq 0\}$.

Theorem C.19. *If* (A_1), (A_2), *and* (A_3) *are satisfied, then there exists* $\Gamma(x,t;\xi,r)$, *a fundamental solution of* $L - \frac{\partial}{\partial t}$, *with*

$$|D_x^m \Gamma(x,t;\xi,r)| \leq c_1(t-r)^{-\frac{m+1}{2}} \exp\left\{-c_2 \frac{(x-\xi)^2}{t-r}\right\}, \qquad m = 0,1,$$

where c_1 *and* c_2 *are positive constants. The functions* $D_x^m \Gamma$, $m = 0,1,2$, *and* $D_t\Gamma$ *are continuous in* $(x,t;\xi,r) \in \mathbb{R} \times [0,T] \times \mathbb{R} \times [0,T]$, $t > r$, *and* $L[\Gamma] - \Gamma_t = 0$, *as function of* (x,t). *Finally, for all* f *bounded continuous, we have*

$$\int_{\mathbb{R}} \Gamma(x,t;\xi,r) f(x) dx \to f(\xi) \text{ for } t \downarrow r.$$

Proof: See, e.g., Friedman (1963). □

Theorem C.20. *Let* (A_1), (A_2), (A_3) *be satisfied,* $f(x,t)$ *be a continuous function in* $\mathbb{R} \times [0,T]$, *Hölder continuous in* x, *uniformly with respect to* (x,t) *in compacts of* $\mathbb{R} \times [0,T]$, *and let* ϕ *be a continuous function in* $\mathbb{R}$. *Moreover, we suppose that*

$$|f(x,t)| \leq Ae^{a_1|x|^2} \text{ in } \mathbb{R} \times [0,T],$$
$$|\phi(x,t)| \leq Ae^{a_1|x|^2} \text{ in } \mathbb{R},$$

where A, a_1 *are positive constants. There exists a solution of the Cauchy problem in* $0 \leq t \leq T^*$, *where* $T^* = \min\{T, \frac{\bar{c}}{a_1}\}$ *and* $\bar{c}$ *is a constant, that only depends on the coefficients of* L *and*

$$|u(x,t)| \leq A' e^{a_1'|x|^2} \text{ in } \mathbb{R} \times [0,T^*],$$

with positive constants A', a_1'. *The solution is given by*

$$u(x,t) = \int_{\mathbb{R}} \Gamma(x,t;\xi,0)\phi(\xi)d\xi - \int_0^t \int_{\mathbb{R}} \Gamma(x,t;\xi,r) f(\xi,r) d\xi dr.$$

The operator M^*, as a supplement to $M = L - \frac{\partial}{\partial t}$, is given by

$$M^*[v] = L^*[v] + \frac{\partial v}{\partial t},$$
$$L^*[v] = \frac{1}{2} a v_{xx} + b^* v_x + c^* v,$$

where $b^* = -b + a_x$, $c^* = c - b_x + \frac{1}{2} a_{xx}$, assuming that a_x, a_{xx}, a_t exist and are bounded.

Remark C.21.

$$M^*[v] = \frac{1}{2} a v_{xx} + a_x v_x - b v_x + \frac{1}{2} a_{xx} v - b_x v + cv + v_t$$
$$= \frac{1}{2}(av)_{xx} - (bv)_x + cv + v_t,$$

from which follows *Green's formula*:

$$\begin{aligned}
&vM[u] - uM^*[v] \\
&= v\left(\frac{1}{2}au_{xx} + bu_x + cu - u_t\right) - u\left(\frac{1}{2}(av)_{xx} - (bv)_x + cv + v_t\right) \\
&= \frac{1}{2}(va)u_{xx} - \frac{1}{2}(av)_{xx} + vbu_x - u(bv)_x + vcu - vcu - vu_t - uv_t \\
&= \frac{1}{2}((va)u_x - (ua)v_x - va_x)_x + (vbu)_x - (uv)_t.
\end{aligned}$$

Therefore, if u and v have compact support in a domain G, we have that

$$\int\int_G (vMu - uM^*v)dxdt = 0.$$

Definition C.22. A *fundamental solution of the operator* $L^* + \frac{\partial}{\partial t}$ in $\mathbb{R} \times [0, T]$ is a function $\Gamma^*(x, t; \xi, r)$, defined, for all $(x, t) \in \mathbb{R} \times [0, T]$ and all $(\xi, t) \in \mathbb{R} \times [0, T]$, $t > r$, such that, for all g continuous with compact support, the function

$$v(x, t) = \int_{\mathbb{R}} \Gamma^*(x, t; \xi, r)g(\xi)d\xi$$

satisfies

1. $L^*[v] + v_t = 0$ if $x \in \mathbb{R}, 0 \leq t \leq r$;
2. $v(x, t) \to g(x)$ if $t \uparrow r$.

We consider the following additional condition.
(A_4) The functions a, a_x, a_{xx}, b, b_x, c are bounded and the coefficients of L^* satisfy the conditions (A_2) and (A_3).

Theorem C.23. *If* (A_1), (A_1), (A_3), *and* (A_4) *are satisfied, then there exists a fundamental solution* $\Gamma^*(x, t; \xi, r)$ *of* $L^* + \frac{\partial}{\partial t}$ *such that*

$$\Gamma(x, t; \xi, r) = \Gamma^*(\xi, r; x, t), \qquad t > r.$$

Proof: See, e.g., Friedman (1963). □

D

Stability of Ordinary Differential Equations

We consider the system of ordinary differential equations

$$\begin{cases} \frac{d}{dt}\mathbf{u}(t) = \mathbf{f}(t, \mathbf{u}(t)),\ t > t_0, \\ \mathbf{u}(t_0) = \mathbf{c} \end{cases} \tag{D.1}$$

in $\mathbb{R}^d$ and we suppose that, for all $\mathbf{c} \in \mathbb{R}^d$, there exists a unique general solution $\mathbf{u}(t, t_0, \mathbf{c})$ in $[t_0, +\infty[$. We further suppose that $\mathbf{f}$ is continuous in $[t_0, +\infty[\times\mathbb{R}^d$ and that $\mathbf{0}$ is the equilibrium solution of $\mathbf{f}$. Thus $\mathbf{f}(t, \mathbf{0}) = \mathbf{0}$ for all $t \geq t_0$.

Definition D.1. The equilibrium solution $\mathbf{0}$ is *stable* if, for all $\epsilon > 0$:

$$\exists\delta = \delta(\epsilon, t_0) > 0 \text{ such that } \forall\mathbf{c} \in \mathbb{R}^d, |\mathbf{c}| < \delta \Rightarrow \sup_{t_0 \leq t \leq +\infty} |\mathbf{u}(t, t_0, \mathbf{c})| < \epsilon. \tag{D.2}$$

If the condition (D.2) is not verified, then the equilibrium solution is *unstable.* The position of the equilibrium is said to be *asymptotically stable* if it is stable and *attractive*, namely, if along with (D.2), it can also be verified that

$$\lim_{t \to +\infty} \mathbf{u}(t, t_0, \mathbf{c}) = \mathbf{0} \qquad \forall\mathbf{c} \in \mathbb{R}^d, |\mathbf{c}| < \delta \text{ (chosen suitably).} \tag{D.3}$$

Remark D.2. There may be attraction without stability.

Remark D.3. If $\mathbf{x}^* \in \mathbb{R}^d$ is the equilibrium solution of $\mathbf{f}$, then the position $\mathbf{y}(t) = \mathbf{u}(t) - \mathbf{x}^*$ tends towards $\mathbf{0}$.

Definition D.4. We consider the ball $B_h \equiv \bar{B}_h(0) = \{\mathbf{x} \in \mathbb{R}^d | |\mathbf{x}| \leq h\}, h > 0$, which contains the origin. The continuous function $v : B_h \to \mathbb{R}_+$ is *positive definite* (in the Lyapunov sense) if

$$\begin{cases} v(\mathbf{0}) = 0, \\ v(\mathbf{x}) > 0\ \ \forall\mathbf{x} \in B_h \setminus \{\mathbf{0}\}. \end{cases} \tag{D.4}$$

The continuous function $v : [t_0, +\infty[\times B_h \to \mathbb{R}_+$ is *positive definite* if

$$\begin{cases} v(t, \mathbf{0}) = 0 & \forall t \in [t_0, +\infty[, \\ \exists \omega : B_h \to \mathbb{R}_+ \text{ positive definite such that } v(t, \mathbf{x}) \geq \omega(\mathbf{x}) & \forall t \in [t_0, +\infty[. \end{cases} \tag{D.5}$$

v is negative definite if $-v$ is positive definite.

Now let $v : [t_0, +\infty[\times B_h \to \mathbb{R}_+$ be a positive definite function endowed with continuous first partial derivatives with respect to t and x_i, $i = 1, \ldots, d$. We consider the function

$$V(t) = v(t, \mathbf{u}(t, t_0, \mathbf{c})) : [t_0, +\infty[\to \mathbb{R}_+,$$

where $\mathbf{u}(t, t_0, \mathbf{c})$ is the solution of (D.1). V is differentiable with respect to t and we have

$$\frac{d}{dt}V(t) = \frac{\partial v}{\partial t} + \sum_{i=1}^{d} \frac{\partial v}{\partial x_i}\frac{du_i}{dt}.$$

But $\frac{du_i}{dt} = f_i(t, \mathbf{u}(t, t_0, \mathbf{c}))$, therefore

$$\dot{v} \equiv \frac{d}{dt}V(t) = \frac{\partial v}{\partial t} + \sum_{i=1}^{d} \frac{\partial v}{\partial x_i} f_i(t, \mathbf{u}(t, t_0, \mathbf{c})),$$

and this is the derivative of v with respect to time "along the trajectory" of the system. If $\frac{d}{dt}V(t) \leq 0$ for all $t \in (t_0, +\infty[$, then $\mathbf{u}(t, t_0, \mathbf{c})$ does not increase the value v, which measures by how much $\mathbf{u}$ moves away from $\mathbf{0}$. Through this observation, the required stability of the Lyapunov criterion for the stability of $\mathbf{0}$ has been formulated.

Definition D.5. Let $v : [t_0, +\infty[\times B_h \to \mathbb{R}_+$ be a positive definite function. v is said to be a *Lyapunov function for the system (D.1) relative to the equilibrium position* $\mathbf{0}$, if

1. v is endowed with first partial derivatives with respect to t and $x_i, i = 1, \ldots, d$;
2. for all $t \in]t_0, +\infty[$: $\dot{v}(t) \leq 0$ for all $c \in B_h$.

Theorem D.6. (Lyapunov).

1. *If there exists $v(t, \mathbf{x})$ a Lyapunov function for the system (D.1) relative to the equilibrium position $\mathbf{0}$, then $\mathbf{0}$ is stable;*
2. *if moreover the Lyapunov function $v(t, \mathbf{x})$ is such that, for all $t \in [t_0, +\infty[$: $v(t, \mathbf{x}) \leq \omega(\mathbf{x})$ with $\mathbf{u}$ being a positive definite function and $\dot{v}$ negative definite along the trajectory, then $\mathbf{0}$ is asymptotically stable.*

Example D.7. We consider the autonomous linear system

$$\begin{cases} \frac{d}{dt}\mathbf{u}(t) = A\mathbf{u}(t), \ t > t_0, \\ \mathbf{u}(t_0) = \mathbf{c}, \end{cases}$$

where A is a matrix that does not depend on time. A matrix P is said to be positive definite if, for all $\mathbf{x} \in \mathbb{R}^d, \mathbf{x} \neq \mathbf{0} : \mathbf{x}'P\mathbf{x} > 0$. Considering the function $v(\mathbf{x}) = \mathbf{x}'P\mathbf{x}$, we have

$$\dot{v} = \frac{d}{dt}v(\mathbf{u}(t)) = \sum_{i=1}^{d} \frac{\partial v}{\partial x_i}(A\mathbf{u}(t))_i = \mathbf{u}'(t)PA\mathbf{u}(t) + \mathbf{u}'(t)A'P\mathbf{u}(t).$$

Therefore, if P is such that $PA + A'P = -Q$, with Q being positive definite, then $\dot{v} = -\mathbf{u}'Q\mathbf{u} < 0$ and, by 2 of Lyapunov's theorem, $\mathbf{0}$ is asymptotically stable.

References

1. Aalen O. O., *Nonparametric inference for a family of counting processes*, Annals of Statistics, 701–726, 6; 1978.
2. Aletti G. & Capasso V., *Profitability in a multiple strategy market*, Decis. Econ. Finance, 145–152, 26; 2003.
3. Andersen P. K., Borgan Ø., Gill R.D. & Keiding N., *Statistical Models Based on Counting Processes*, Springer-Verlag, Heidelberg; 1993.
4. Anderson W. J., *Continuous-Time Markov Chains. An Application-Oriented Approach*, Springer–Verlag, New York; 1991.
5. Arnold L., *Stochastic Differential Equations*, Wiley, New York; 1974.
6. Ash R. B., *Real Analysis and Probability*, Academic Press, London; 1972.
7. Ash R. B. & Gardner M. F., *Topics in Stochastic Processes*, Academic Press, London; 1975.
8. Aubin J.-P., *Applied Abstract Analysis*, Wiley, New York; 1977.
9. Bachelier L., *Théorie de la spéculation*, Ann. Sci. École Norm. Sup., 21–86, 17; 1900.
10. Bailey N. T. J., *The Mathematical Theory of Infectious Diseases*, Griffin, London; 1975.
11. Baldi P., *Equazioni differenziali stocastiche*, UMI, Bologna; 1984.
12. Bartholomew D. J., *Continuous time diffusion models with random duration of interest*, J. of Math. Sociology, 187–199, 4; 1976.
13. Bauer H., *Probability Theory and Elements of Measure Theory*, Academic Press, London; 1981.
14. Becker N., *Analysis of Infectious Disease Data*, Chapman and Hall, London; 1989.
15. Bertoin J., *Lévy Processes*, Cambridge University Press, Cambridge, UK; 1996.
16. Bhattacharya R.N. & Waymire E.C., *Stochastic Processes with Applications*, Wiley, New York; 1990.
17. Billingsley P., *Convergence of Probability Measures*, Wiley, New York; 1968.
18. Black F. & Scholes M., *The pricing of options and corporate liabilities*, J. of Pol. Econ., 637–654, 81; 1973.
19. Borodin A. & Salminen P., *Handbook of Brownian Motion: Facts and Formulae*, Birkhä user, Boston; 1996.
20. Boyle P. & Tian Y., *Pricing lookback and barrier options under the CEV process*, J. of Fin. and Quant. Analysis, 241–264, 34; 1999.

21. Brace A., Gatarek D., & Musiela M., *The market model of interest rate dynamics*, Math. Fin., 127–154, 7; 1997.
22. Breiman L., *Probability*, Addison–Wesley, Reading, MA; 1968.
23. Bremaud P., *Point Processes and Queues: Martingale Dynamics*, Springer-Verlag, Heidelberg; 1981.
24. Burger M., Capasso V., & Morale D., *On an aggregation model with long and short range interaction*, MIRIAM reports, Milan; 2003.
25. Capasso V., *A counting process approach for stochastic age-dependent population dynamics*, in *Biomathematics and Related Computational Problems* (L. M. Ricciardi, Ed.), pp. 255–269, Kluwer, Dordrecht; 1988.
26. Capasso V., *A counting process approach for age-dependent epidemic systems*, in *Stochastic Processes in Epidemic Theory* (J. P. Gabriel et al., Eds.) pp. 118–128, Lecture Notes in Biomathematics, Vol. 86, Springer-Verlag, Heidelberg; 1990.
27. Capasso V., *Mathematical Structures of Epidemic Systems*, Springer-Verlag, Heidelberg; 1993.
28. Capasso V., Di Liddo A. & Maddalena L., *Asymptotic behaviour of a nonlinear model for the geographical diffusion of innovations*, Dynamic Systems and Applications, 207–220, 3; 1994.
29. Capasso V., Morale D. & Sioli F., *An agent-based model for "price herding", applied to the automobile market*, MIRIAM reports, Milan; 2003.
30. Carrillo J., *Entropy solutions for nonlinear degenerate problems*, Arch. Rat. Mech. Anal., 269–361, 147; 1999.
31. Cherny A. S. & Shiryaev A. N., *Vector stochastic integrals and the fundamental theorems of asset pricing*, French–Russian A. M. Liapunov Institute for Applied Mathematics and Computer Science Transactions, 5–37, 3; 2001.
32. Chiang C.L., *Introduction to Stochastic Processes in Biostatistics*, Wiley, New York; 1968.
33. Chung K. L., *A Course in Probability Theory*,Second Edition, Academic Press, New York; 1974.
34. Cox J. C., *The constant elasticity of variance option pricing model*, J. of Portfolio Management, 15–17, 22; 1996.
35. Cox J. C., Ross S. A., & Rubinstein M., *Option pricing: A simplified approach*, J. Fin. Econ., 229–263, 7; 1979.
36. Dalang R. C., Morton A., & Willinger W., *Equivalent martingale measures and non-arbitrage in stochastic securities market models*, Stochastics and Stochastic Reports, 185–201, 29; 1990.
37. Darling D. A. D. & Siegert A. J. F., *The first passage time problem for a continuum Markov process*, Ann. Math. Stat., 624–639, 24; 1953.
38. Davis M. H. A., *Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models*, Journal of the Royal Statistical Society, Series B, 353–388, 46; 1984.
39. Delbaen F. & Schachermeyer W., *A general version of the fundamental theorem of asset pricing*, Mathematische Annalen, 463–520, 300; 1994.
40. De Masi A. & Presutti E., *Mathematical Methods for Hydrodynamical Limits*, Springer-Verlag, Heidelberg; 1991.
41. Dieudonné J., *Foundations of Modern Analysis*, Academic Press, New York; 1960.
42. Donsker M. D. & Varadhan S. R. S., *Large deviations from a hydrodynamical scaling limit*, Comm. Pure Appl. Math., 243–270, 42; 1989.

43. Doob J. L., *Stochastic Processes*, Wiley, New York; 1953.
44. Duffie D., *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, NJ; 1996.
45. Dupire B., *Pricing with a smile*, RISK, 18–20, 1; 1994.
46. Durrett R. & Levin S. A., *The importance of being discrete (and spatial)*, Theor. Pop. Biol., 363–394, 46; 1994.
47. Dynkin E. B., *Markov Processes*, Springer-Verlag, Berlin, Vols. 1–2; 1965.
48. Embrechts P., Klüppelberg C., & Mikosch T., *Modelling Extreme Events for Insurance and Finance*, Springer-Verlag, Berlin; 1997.
49. Epstein J. & Axtell R., *Growing Artificial Societies–Social Sciences from the Bottom Up*, Brooking Institution Press & MIT Press, Cambridge, MA; 1996.
50. Ethier S. N. & Kurtz T. G., *Markov Processes, Characterization and Convergence*, Wiley, New York; 1986.
51. Feller W., *An Introduction to Probability Theory and Its Applications*, Wiley, New York; 1971.
52. Flierl G., Grünbaum D., Levin S. A., & Olson D., *From individuals to agrregations: The interplay between behavior and Physics*, J. Theor. Biol., 397–454, 196; 1999.
53. Fournier N. & Méléard S., *A microscopic probabilistic description of a locally regulated population and macroscopic approximations*, Prépublications du Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 6/7, PMA-798; 2003.
54. Friedman A., *Partial Differential Equations*, Krieger, New York; 1963.
55. Friedman A., *Partial Differential Equations of Parabolic Type*, Prentice–Hall, London; 1964.
56. Friedman A., *Stochastic Differential Equations and Applications*, Academic Press, London; 1975.
57. Fristedt B. & Gray L., *A Modern Approach to Probability Theory*, Birkhäuser, Boston; 1997.
58. Gard T. C., *Introduction to Stochastic Differential Equations*, Marcel Dekker, New York; 1988.
59. Gihman I. I. & Skorohod A. V., *Stochastic Differential Equations*, Springer-Verlag, Berlin; 1972.
60. Gihman I. I. & Skorohod A. V., *The Theory of Random Processes*, Springer-Verlag, Berlin; 1974.
61. Gihman I. I. & Skorohod A. V., *Introduction à la théorie des processes aléatoires*, MIR, Moscow; 1980.
62. Gnedenko B. V., *The Theory of Probability*, Chelsea, New York; 1963.
63. Grigoriu M., *Stochastic Calculus. Applications to Science and Engineering*, Birkhäuser, Boston; 2002.
64. Grünbaum D. & Okubo A., *Modelling social animal aggregations*, in *Frontiers of Theoretical Biology* (S. A. Levin, Ed.) pp. 296–325, Lectures Notes in Biomathematics, Vol. 100, Springer-Verlag, New York; 1994.
65. Gueron S., Levin S. A., & Rubenstein D. I., *The dynamics of herds: From individuals to aggregations*, J. Theor. Biol., 85–98, 182; 1996.
66. Harrison J. M. & Kreps D. M., *Martingales and arbitrage in multiperiod securities markets*, J. Econ. Th., 381–408, 20; 1979.
67. Harrison J. M. & Pliska S. R., *Martingales and stochastic integrals in the theory of continuous trading*, Stochastic Processes Appl., 215–260, 11; 1981.

68. Has'minskii R. Z., *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, The Netherlands; 1980.
69. Heath D., Jarrow R., & Morton A., *Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation*, Econometrica, 77–105, 1; 1992.
70. Hunt P. J. & Kennedy J. E., *Financial Derivatives in Theory and Practice*, Wiley, New York; 2000.
71. Hull J. & White A., *Pricing interest rate derivative securities*, Rev. Fin. Stud., 573–592, 4; 1990.
72. Ikeda N. & Watanabe S., *Stochastic Differential Equations and Diffusion Processes*, North–Holland, Kodansha; 1989.
73. Jacobsen M., *Statistical Analysis of Counting Processes*, Springer-Verlag, Heidelberg; 1982.
74. Jacod J. & Shiryaev A. N., *Limit Theorems for Stochastic Processes*, Springer Lecture Notes in Math., Springer-Verlag, Berlin; 1987.
75. Jamshidian F., *LIBOR and swap market models and measures*, Finance and Stochastics, 43–67, 1; 1997.
76. Jelinski Z. & Moranda P., *Software reliability research*, in *Statistical Computer Performance Evaluation*, pp. 466–484, Academic Press, New York; 1972.
77. Kallenberg O., *Foundations of Modern Probability*, Springer-Verlag, Berlin; 1997.
78. Karatzas I. & Shreve S. E., *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York; 1991.
79. Karlin S. & Taylor H. M., *A First Course in Stochastic Processes*, Academic Press, New York; 1975.
80. Karlin S. & Taylor H. M., *A Second Course in Stochastic Processes*, Academic Press, New York; 1981.
81. Karr A. F., *Point Processes and Their Statistical Inference*, Marcel Dekker, New York; 1986.
82. Kolmogorov A. N., *Foundations of the Theory of Probability*, Chelsea, New York; 1956.
83. Kolmogorov A. N. & Fomin E. S. V., *Elements of Theory of Functions and Functional Analysis*, Groylock, Moscow; 1961.
84. Kou S., *A jump diffusion model for option pricing*, Man. Sc., 1086–1101, 48; 2002.
85. Lamberton D. & Lapeyre B., *Introduction to Stochastic Calculus applied to Finance*, Chapman & Hall, London; 1996.
86. Lamperti J., *Probability*, Benjamin, New York; 1966.
87. Last G. & Brandt A., *Marked Point Processes on the Real Line. The Dynamic Approach*, Springer-Verlag, Heidelberg; 1995.
88. Loève M., *Probability Theory*, Van Nostrand–Reinhold, Princeton, NJ; 1963.
89. Mahajan V. & Wind Y., *Innovation Diffusion Models of New Product Acceptance*, Ballinger, Cambridge, MA; 1986.
90. Méléard S., *Asymptotic behaviour of some interacting particle systems; McKean–Vlasov and Boltzmann models*, in *Probabilistic Models for Nonlinear Partial Differential Equations* (D. Talay, L. Tubaro, Eds.) pp. 42–95, Lecture Notes in Mathematics, Vol. 1627, CIME Subseries, Springer-Verlag, Heidelberg; 1996.
91. Merton R. C., *Theory of rational option pricing*, Bell J. Econ. Man. Sc., 141–183, 4; 1973.

92. Merton R. C., *Option pricing when underlying stock returns are discontinuous*, J. Fin. Econ., 125–144, 3; 1976.
93. Métivier M., *Notions fondamentales de la théorie des probabilités*, Dunod, Paris; 1968.
94. Meyer P. A., *Probabilités et Potentiel*, Ilermann, Paris; 1966.
95. Miltersen K. R., Sandmann K., & Sondermann D., *Closed form solutions for term structure derivatives with log-normal interest rates*, J. of Finance, 409–430, 52; 1997.
96. Morale D., Capasso V., & Oelschläger K., *An interacting particle system modelling aggregation behaviour: From individuals to populations*, J. Math. Biology; 2004.
97. Morale D., Capasso V., & Oelschläger K., *A rigorous derivation of the mean-field nonlinear integro-differential equation for a population of aggregating individuals subject to stochastic fluctuations*, Preprint 98–38 (SFB 359), IWR, Universität Heidelberg, Juni; 1998.
98. Musiela M. & Rutkowski M., *Martingale Methods in Financial Modelling*, Springer-Verlag, Berlin; 1998.
99. Nagai T. & Mimura M., *Some nonlinear degenerate diffusion equations related to population dynamics*, J. Math. Soc. Japan, 539–561, 35; 1983.
100. Neveu J., *Mathematical Foundations of the Calculus of Probability*, Holden-Day, San Francisco; 1965.
101. Oelschläger K., *A law of large numbers for moderately interacting diffusion processes*, Z. Wahrsch. Verw. Geb., 279–322, 69; 1985.
102. Oelschläger K., *Large systems of interacting particles and the porous medium equation*, J. Differential Equations, 294–346, 88; 1990.
103. Øksendal B., *Stochastic Differential Equations*, Springer-Verlag, Berlin; 1998.
104. Okubo A., *Dynamical aspects of animal grouping: Swarms, school, flocks and herds*, Adv. BioPhys., 1–94, 22; 1986.
105. Parzen E., *Stochastic Processes*, Holden-Day, San Francisco; 1962.
106. Pliska S. R., *Introduction to Mathematical Finance: Discrete-Time Models*, Blackwell Publishers, Oxford; 1997.
107. Protter P., *Stochastic Integration and Differential Equations*, Springer-Verlag, Berlin; 1990.
108. Rebolledo R., *Central limit theorems for local martingales*, Z. Wahrsch. verw. Geb., 269–286, 51; 1980.
109. Robert P., *Stochastic Networks and Queues*, Springer-Verlag, Heidelberg; 2003.
110. Rogers L. C. G. & Williams D., *Diffusions, Markov Processes and Martingales*, Vol. 1, Wiley, New York; 1994.
111. Rolski T., Schmidli H., Schmidt V., & Teugels J., *Stochastic Processes for Insurance and Finance*, Wiley, New York; 1999.
112. Roozen H., *Equilibrium and extinction in stochastic population dynamics*, Bull. Math. Biology, 671–696, 49; 1987.
113. Samorodnitsky G. & Taqqu M. S., *Stable Non-Gaussian Random Processes*, Chapman & Hall/CRC Press, Boca Ration, FL; 1994.
114. Sato K. I., *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge, UK; 1999.
115. Schuss Z., *Theory and Applications of Stochastic Differential Equations*, Wiley, New York; 1980.
116. Shiryaev A. N., *Probability*, Springer-Verlag, New York; 1995.

117. Skellam J. G., *Random dispersal in theoretical populations*,Biometrika, 196–218, 38; 1951.
118. Sobczyk K., *Stochastic Differential Equations. With Applications to Physics and Engineering*, Kluwer, Dordrecht; 1991.
119. Stein R. B., *A theoretical analysis of neuronal variability*, Biophys. J., 173–194, 5; 1965.
120. Stein R. B., *Some models of neuronal variability*, Biophys. J., 37–68, 7; 1967.
121. Taira K., *Diffusion Processes and Partial Differential Equations*, Academic Press, New York; 1988.
122. Tan W. Y., *Stochastic Models with Applications to Genetics, Cancers, AIDS and other Biomedical Systems*, World Scientific, Singapore; 2002.
123. Tuckwell H. C., *On the first exit time problem for temporarily homogeneous Markov process*, J. Appl. Prob., 39–48, 13; 1976.
124. Tuckwell H. C., *Stochastic Processes in the Neurosciences*, SIAM, Philadelphia; 1989.
125. Vasicek O., *An equilibrium characterisation of the term structure*, J. Fin. Econ., 177–188, 5; 1977.
126. Wang F. J. S., *Gaussian approximation of some closed stochastic epidemic models*, J. Appl. Prob., 221–231, 14; 1977.
127. Warburton K. & Lazarus J., *Tendency-distance models of social cohesion in animal groups*, J. Theor. Biol., 473–488, 150; 1991.
128. Wax N., *Selected Papers on Noise and Stochastic Processes*, Dover, New York; 1954.
129. Williams D., *Probability with Martingales*, Cambridge University Press, Cambridge, UK; 1991.
130. Wilmott P., Howison S. D., & Dewynne J. N., *Option Pricing: Mathematical Models and Computation*, Oxford Fin. Press, Oxford; 1993.
131. Yang G., *Stochastic epidemics as point processes*, in *Mathematics in Biology and Medicine* (V. Capasso, E. Grosso, S. L. Paveri-Fontana, Eds.), pp. 135–144, Lectures Notes in Biomathematics, Vol. 57, Springer-Verlag, Heidelberg; 1985.

Nomenclature

	"increasing" is used with the same meaning as "nondecreasing"; "decreasing" is used with the same meaning as "nonincreasing." In the strict cases "strictly increasing/strictly decreasing" is used.
$(\Omega, \mathcal{F}, P)$	probability space with Ω a set, $\mathcal{F}$ a σ-algebra of parts of Ω, and P a probability measure on $\mathcal{F}$
$(E, \mathcal{B}_E)$	measurable space with E a set and $\mathcal{B}_E$ a σ-algebra of parts of E
$:=$	equal by definition
$< f, g >$	scalar product of two elements f and g in an Hilbert space
$< M, N >$	predictable covariation of the martingales M and N
$< M >, < M, M >$	predictable variation of the martingale M
$[a, b[$	semiopen interval closed at extreme a and open at extreme b
$[a, b]$	closed interval of extremes a and b
$\bar{\mathbb{R}}$	extended set of real numbers; i.e., $\mathbb{R} \cup \{-\infty, +\infty\}$
$\bar{A}$	closure of a set A depending upon the context
$\bar{C}$	the complement of the set C depending upon the context
Δ	Laplace operator
δ_x	Dirac delta-function localized at x
δ_{ij}	Kronecker delta; i.e., $= 1$ for $i = j$, $= 0$ for $i \neq j$
$\emptyset$	the empty set
ϵ_x	Dirac delta-measure localized at x
$\equiv$	coincide
$\exp\{x\}$	exponential function e^x
$\int^*$	integral of a nonnegative measurable function, finite or not
$\lim_{s \downarrow t}$	limit for s decreasing while tending to t
$\lim_{s \uparrow t}$	limit for s increasing while tending to t
$\mathbb{C}$	the complex plane
$\mathbb{N}$	the set of natural nonnegative integers
$\mathbb{N}^*$	the set of natural (strictly) positive integers

$\mathbb{Q}$	the set of rational numbers
$\mathbb{R}^n$	n-dimensional Euclidean space
$\mathbb{R}_+$	the set of positive (nonnegative) real numbers
$\mathbb{R}_+^*$	the set of (strictly) positive real numbers
$\mathbb{Z}$	the set of all integers
$\mathcal{A}$	infinitesimal generator of a semigroup
$\mathcal{B}_{\mathbb{R}^n}$	σ-algebra of Borel sets on $\mathbb{R}^n$
$\mathcal{B}_E$	σ-algebra of Borel sets generated by the topology of E
$\mathcal{D}_A$	domain of definition of an operator A
$\mathcal{F}_t$ or $\mathcal{F}_t^X$	history of a process $(X_t)_{t\in\mathbb{R}_+}$ up to time t; i.e., the σ-algebra generated by $\{X_s, s \leq t\}$
$\mathcal{F}_{t+}$	$\bigcap_{s>t} \mathcal{F}_t$
$\mathcal{F}_{t-}$	σ-algebra generated by $\sigma(X_s, s < t)$
$\mathcal{F}_X$	σ-algebra generated by the random variable X
$\mathcal{L}(X)$	probability law of X
$\mathcal{L}^p(P)$	set of integrable functions with respect to the measure P
$\mathcal{M}(\mathcal{F}, \bar{\mathbb{R}}_+)$	set of all $\mathcal{F}$-measurable functions with values in $\bar{\mathbb{R}}_+$
$\mathcal{M}(E)$	set of all measures on E
$\mathfrak{P}(\Omega)$	the set of all parts of a set Ω
$\xrightarrow[n]{P}$ or P-lim	convergence in probability
$\xrightarrow[n]{\mathcal{W}}$	weak convergence
$\xrightarrow[n]{a.s.}$	almost sure convergence
$\xrightarrow[n]{d}$	convergence in distribution
$\xrightarrow[n]{P}$	convergence in probability
∇	gradient
Ω	the underlying sample space
ω	an element of the underlying sample space
$\otimes$	product of σ-algebras or product of measures
∂A	boundary of a set A
Φ	cumulative distribution function of a standard normal probability law
$\operatorname{sgn}\{x\}$	sign function; 1, if $x > 0$; 0, if $x = 0$; -1, if $x < 0$
$\sigma(\mathcal{R})$	σ-algebra generated by the family of events $\mathcal{R}$
$\square$	end of a proof
$\|a\|$	absolute value of a number a; or modulus of a complex number a
$\|A\|$ or $\sharp(A)$	cardinal number (number of elements) of a finite set A
$\|\|x\|\|$	the norm of a point x
$]a, b[$	open interval of extremes a, b
$]a, b]$	semiopen interval open at extreme a and closed at extreme b
$a \vee b$	maximum of two numbers

A'	transpose of a matrix A
$A \setminus B$	the set of elements of A that do not belong to B
$a \wedge b$	minimum of two numbers
$B(x,r)$ or $B_r(x)$	the open ball centered at x and having radius r
$C(A)$	set of continuous functions from A to $\mathbb{R}$
$C(A,B)$	set of continuous functions from A to B
$C^k(A)$	set of functions from A to $\mathbb{R}$ with continuous derivatives up to order k
$C^{k+\alpha}(A)$	set of functions from A to $\mathbb{R}$ whose k-th derivatives are Lipschitz continuous with exponent α
$C_0(A)$	set continuous functions on A with compact support
$C_b(A)$ or $BC(A)$	set of bounded continuous functions on A
$Cov[X,Y]$	the covariance of two random variables X and Y
$E[\cdot]$	expected value with respect to an underlying probability law clearly identifiable from the context
$E[Y\|\mathcal{F}]$	conditional expectation of a random variable Y with respect to the σ-algebra $\mathcal{F}$
$E_P[\cdot]$	expected value with respect to the probability law P
$E_x[\cdot]$	expected value conditional upon a given initial state x in a stochastic process
$f * g$	convolution of functions f and g
$f \circ X$ or $f(X)$	a function f composed with a function X
$f\|_A$	the restriction of a function f to the set A
f^-, f^+	negative (positive) part of f; i.e., $f^- = \max\{-f, 0\}$ ($f^+ = \max\{f, 0\}$)
$f^{-1}(B)$	the preimage of the set B by the function f
F_X	cumulative distribution function of a random variable X
$H \bullet X$	stochastic Stieltjes integral of the process H with respect to the stochastic process X
I_A	indicator function associated with a set A; i.e., $I_A(x) = 1$, if $x \in A$ otherwise $I_A(x) = 0$
$L^p(P)$	set of equivalence classes of a.e. equal integrable functions with respect to the measure P
$N(\mu, \sigma^2)$	normal (Gaussian) random variable with mean μ and variance σ^2
$O(\Delta)$	of the same order as Δ
$o(\delta)$	of higher order with respect to δ
P-a.s.	almost surely with respect to the measure P
$P(A\|B)$	conditional probability of an event A with respect to an event B
$P * Q$	convolution of measures P and Q
$P \ll Q$	the measure P is absolutely continuous with respect to the measure Q
$P \sim Q$	the measure P is equivalent to the measure Q
P_X	probability law of a random variable X

P_x	probability law conditional upon a given initial state x in a stochastic process
$Var[X]$	the variance of a random variable X
W_t	standard Brownian motion, Wiener process
$X \sim P$	the random variable X has probability law P
a.e.	almost everywhere
a.s.	almost surely

Index