Introduction to Stochastic Analysis

Introduction to Stochastic Analysis

Integrals and Differential Equations

Vigirdas Mackevičius

First published 2011 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Cataloging-in-Publication Data

Mackevičius, Vigirdas.
Introduction to stochastic analysis, integrals, and differential equations / Vigirdas Mackevičius.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-84821-311-1
1. Stochastic analysis. I. Title.
QA274.2.M33 2011
519.2'2--dc22

2011012249

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-84821-311-1

Printed and bound in Great Britain by CPI Antony Rowe, Chippenham and Eastbourne

FSC www.fsc.org MIX Paper from responsible sources FSC® C013604

Contents

Preface

Initially, I wanted to entitle the textbook "Stochastic Analysis for *All*" or "Stochastic Analysis *without Tears*", keeping in mind that it will be accessible not only to students of mathematics but also to physicists, chemists, biologists, financiers, actuaries, etc. However, though aiming for as wide a readability as possible, finally, I rejected such titles regarding them as too ambitious.

Most people have an intuitive concept of probability based on their own life experience. However, efforts to precisely define probabilistic notions meet serious difficulties; this is seen looking at the history of probability theory—from elementary combinatorial calculations in hazard games to a rigorous axiomatic theory, having a store of applications in various practical and scientific areas. Possibly, as in no other area of mathematics, in probability theory, there is a huge distance from the beginning and elements to the precise and rigorous theory. This is firstly related to the fact that the "palace" of probability theory is built on the substructure of the rather subtle and abstract measure theory. For example, for a mathematician, a random variable is a real measurable function defined on the space of elementary events, while for practitioners—physicists, chemists, biologists, actuaries, etc.—it is some quantity depending on chance. For a mathematician, the randomness is externalized by a probability measure on the measure space (space of elementary events with a σ-algebra of its subsets), which, together with the latter, constitutes a probability space, the primary notion of probability theory. On its basis, all the notions of probability theory, such as random variable, independence, expectation, variance, etc., are defined. For a practitioner, randomness is represented by a distribution function, which is well understood intuitively and can be used to define many of the above-mentioned notions, though with some loss of strictness and generality. Thus, every author of a book on probability theory has to look for a middle ground between strict abstract theory and accessibility for researchers in other sciences and even for mathematicians working in other areas of mathematics. The author of this book is no an exception.

A relatively recent area of probability theory, stochastic differential equations are receiving increasing attention by researchers and practitioners in various natural and applied sciences. First of all, this is related to the fact that "ordinary" (deterministic) differential equations, when modeling various real-world phenomena, usually describe only the average behavior of one system or another. However, real-world systems are most often influenced by many different random factors, also called perturbations. It appears that such perturbations, when they are sufficiently intensive, do not only "disorganize" the system forcing it to oscillate about the average behavior, but also *qualitatively* change the average behavior itself. It is clear that, in such a situation, deterministic equations cannot, in principle, adequately describe the system.

However, in stochastic analysis (theory of stochastic integration and stochastic differential equations) the problem of strictness-to-simplicity ratio arises much more strongly. Modern stochastic analysis is closely related with rather abstract areas of probability theory, such as general theory of random processes and theory of martingales. How can we avoid them when teaching the basics of stochastic differential equations? The author tries to present them in the spirit of "naive" stochastic integration. Just as in applications, the notion of a random variable is imperceptibly replaced by its distribution function, here we replace the very important notion of *filtration* (an increasing system of σ-algebras of events) by the intuitively easier notion of *history*, or *past*. Correspondingly, the notion of an *adapted* random process[1] becomes easier to understand if by an adapted process we mean the one with values that, at every time moment, "belong to the history", or, in other words, "depend only on the past".

A result of all these methodical searches and "inventions" is that the textbook is written on two levels. The main level, which is devoted "for everybody", contains a simplified theory of stochastic integration and stochastic differential equations, with some lack of rigidity and preciseness (though without any "cheating"). The author hopes that it will be understandable to everybody who is acquainted with the basics of probability theory in the scope of a standard elementary course, together with a minimal mathematical "ear". The second level is devoted to a rigorous introduction to stochastic analysis for students of mathematics. The comments, definitions, detailed proofs, or their revisions written in this level are marked by the symbol Ⓜ.

To apply stochastic analysis in a treatment of real-world random processes, we have to construct one or another theoretical model and to know how to simulate it. Therefore, in the book (Chapter 13) much attention is paid to numerical solution of stochastic differential equations or, in other words, to their computer simulation.

1. As a process which, at every time moment, is measurable with respect to the corresponding σ-algebra from the given filtration. . . .

We essentially restrict ourselves to the one-dimensional case. Passing to the multidimensional case is often related not with principal difficulties, but rather with technical inconvenience in using, for a beginner, complicated notation and formulas. However, for the reader's convenience, in the last chapter, we present an overview of the main definitions and statements in the multidimensional case.

We use the double numbering of statements (theorems, propositions, etc.): the first number denotes the number of the chapter, and the second indicates the number of the statement within the chapter.

Though an enormous number of books and papers have influenced the contents of this book, the short reference list includes only those directly used by the author.

Almost all the graphs in the book were drawn by using the TEX macro package PICTEX, while the graph points were calculated by using Pascal programs in Free-Pascal environment.

The book essentially is a translation of the author's book *Stochastic Analysis: Stochastic Integrals and Stochastic Differential Equations* (Vilnius University Press, 2005) from Lithuanian, complemented by the chapter with an application example in finance. The author would like to thank a large number of students of the Department of Mathematics and Informatics of Vilnius University and, especially, Kęstutis Gadeikis. Thanks to them, the book contains significantly fewer mistakes and misprints.

Vigirdas Mackevičius
Vilnius, May 2011

Notation

$\mathbb{N}$	The set of positive integers $\{1, 2, \ldots\}$
$\overline{\mathbb{N}}$	$\mathbb{N} \cup \{+\infty\}$
$\mathbb{N}_+$	$\mathbb{N} \cup \{0\}$
$\mathbb{R}$	Real line $(-\infty, +\infty)$
$\overline{\mathbb{R}}$	Extended real line $\mathbb{R} \cup \{-\infty, +\infty\}$
$\mathbb{R}_+$	The set of non-negative real numbers $[0, +\infty)$
$\forall$	"for all", "for each", "for every"
$\exists \ldots : \ldots$	There exists ... such that...
$:=$	"Denote", "is equal by definition to"
$\equiv$	Identically equal
$\implies \Rightarrow$	"implies"
$\iff, \Leftrightarrow$	"if and only if", "necessary and sufficient", "equivalent"
$\emptyset$	Empty set
$x \in A$	An element x belongs to a set A
$x \notin A$	An element x does not belong to a set A
A^c	The complement of a set A; the contrary event to an event A
$A \subset B, B \supset A$	A set A is a subset of a set B; an event A implies an event B
$\{x_n\} \subset A$	A sequence $\{x_n,\, n \in \mathbb{N}\}$ of elements of a set A
$A \cup B$	The union of sets (events) A and B
$A \cap B$	The intersection of sets (events) A and B
$\bigcup\limits_{n=1}^{\infty} A_n, \bigcup_{n=1}^{\infty} A_n$	The union of a sequence of sets (events) $A_n, n \in \mathbb{N}$
$\bigcap\limits_{n=1}^{\infty} A_n, \bigcap_{n=1}^{\infty} A_n$	The intersection of a sequence of sets (events) $A_n, n \in \mathbb{N}$

$f\colon X \to Y$	A function (mapping) from X to Y
$[x]$	The integer part of a number x (the maximal integer not exceeding x)
$x \vee y$	$\max\{x, y\}$
$x \wedge y$	$\min\{x, y\}$
$\mathbb{1}_A$	The indicator of a set (an event) A: $\mathbb{1}_A(x) = 1$ for $x \in A$; $\mathbb{1}_A(x) = 0$ for $x \in A^c$
$C(I)$	The set of continuous functions $f\colon I \to \mathbb{R}$ ($I = \mathbb{R}$ or $[a, b]$)
$C^k(I)$	The set of k times continuously differentiable functions $f\colon I \to \mathbb{R}$
$C_b^k(I)$	The set of bounded functions $f\colon I \to \mathbb{R}$ having bounded derivatives up to order k
$C_0^k(\mathbb{R})$	The set of k times continuously differentiable functions $f\colon \mathbb{R} \to \mathbb{R}$ with compact support (that is, vanishing outside a finite interval)
$L^2[a, b]$	The set of (measurable) functions $f\colon [a, b] \to \mathbb{R}$ such that $\|f\| = \|f\|_{L^2} := (\int_a^b f^2(x)\,\mathrm{d}x)^{1/2} < +\infty$
$\mathbf{E}(X)$, $\mathbf{E}\,X$	The expectation (or mean) of a random variable X
$\mathbf{D}(X)$, $\mathbf{D}\,X$	The variance of a random variable X
$N(a, \sigma^2)$	Normal distribution with expectation a and variance σ^2
$X \sim N(a, \sigma^2)$	A random variable X with distribution $N(a, \sigma^2)$
$X \overset{d}{=} Y$	Random variables X and Y are identically distributed
$X \perp\!\!\!\perp Y$	Random variables X and Y are independent
$X_n \to X$ (a.s.)	A sequence $\{X_n\}$ of random variables converges almost surely (or with probability one) to a random variable X (section 1.8)
$X_n \xrightarrow{\mathbf{P}} X$	A sequence $\{X_n\}$ of random variables converges in probability to a random variable X (section 1.8)
$X_n \xrightarrow{L^2} X$	A sequence $\{X_n\}$ of random variables converges in mean square (or in the L^2 sense) to a random variable X (section 1.8)
$X_n \xrightarrow{w} X$	A sequence $\{X_n\}$ of random variables converges weakly (or in distribution) to a random variable X (section 1.8)
$\mathcal{H}_t = \mathcal{H}_t^B$	The history (or the past) of Brownian motion B up to moment t (Chapter 2)
$H^2[0, T]$	The set of adapted processes $X = \{X_t,\, t \in [0, T]\}$ with $\|X\|_{H^2} = (\mathbf{E}\int_0^T X_s^2\,\mathrm{d}s)^{1/2} < +\infty$ (Chapter 4)

$\widehat{H}^2[0,T]$	The set of adapted processes $X = \{X_t,\, t \in [0,T]\}$ with $\mathbf{P}\{\int_0^T X_s^2 \,\mathrm{d}s < +\infty\} = 1$ (Chapter 4)
$Y \bullet X_t = \int_0^t Y_s \,\mathrm{d}X_s$	The stochastic (Itô) integral of a process Y with respect to an Itô process X (Chapters 4 and 7)
$Y \circ X_t = \int_0^t Y_s \circ \mathrm{d}X_s$	The Stratonovich integral of a process Y with respect to an Itô process X (Chapter 8)
$\langle X, Y \rangle$	The (quadratic) covariation of Itô processes X and Y (Chapter 7)
$\langle X \rangle = \langle X, X \rangle$	The quadratic variation of an Itô process X (Chapter 7)

Chapter 1

Introduction: Basic Notions of Probability Theory

1.1. Probability space

The main notion of probability theory is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ consisting of any set of elementary events (or outcomes) Ω, a system of events $\mathcal{F}$, and probability measure $\mathbf{P}$. Though these objects form an unanimous whole, we shall try to consider them separately.

Sample space Ω is any non-empty set. Its elements are interpreted as all possible outcomes of an experiment (test, monitoring, phenomenon, and so on) and are called outcomes or elementary events. They are often denoted by letter ω (possibly with some index(es)). Let us consider some examples.

EXAMPLE 1A. Suppose that our experiment involves throwing a die once. Usually, we are only interested in the number of dots, and so all possible outcomes can be described by the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$. Naturally, the outcome of the experiment "the number of dots that appeared on top is five" is represented by the simple event $\omega = 5$.

EXAMPLE 1B. Consider the more complex experiment of throwing a die thrice. It can be described by the sample space consisting of all triples (i, j, k) of the natural numbers from 1 to 6 (the number of such triples is $6^3 = 216$):

$$\begin{aligned}\Omega &= \big\{(i, j, k)\colon i, j, k = 1, 2, 3, 4, 5, 6\big\} \\ &= \big\{(1, 1, 1), (1, 1, 2), (1, 1, 3), \dots, (6, 6, 5), (6, 6, 6)\big\}.\end{aligned}$$

EXAMPLE 1C. If a die is thrown an unknown (in advance) number of times (for example, until six dots appear three consecutive times or until we are tired of throwing), it is convenient to consider the abstract model with unlimited number of dice throws. An outcome of such an "experiment" can be described by a sequence $\omega = \{\omega_n\} = \{\omega_1, \omega_2, \dots\}$, the elements ω_n of which are arbitrary natural numbers from 1 to 6. Thus, in this case, the sample space is the set of all such sequences:

$$\Omega = \bigl\{\omega = \{\omega_n,\, n \in \mathbb{N}\} \colon \omega_n \in \{1,2,3,4,5,6\}, n \in \mathbb{N}\bigr\}.$$

EXAMPLE 2A. Suppose that we decided to measure the outdoor temperature. An outcome of the "experiment" could be the temperature at a fixed moment in time. Since, practically, the temperature cannot be lower than (say) $-60°$C and higher than $+60°$C, for this experiment, we can consider the sample space $\Omega = [-60, 60]$. However, it is often more convenient not to define any reasonable bounds for the temperature (perhaps we measure the temperature in Mars or in the center of the Sun) and take $\Omega = \mathbb{R}$.

EXAMPLE 2B. Let us measure the outdoor temperature each hour for the whole day. Then the outcome of the experiment is a set of 24 numbers meaning the temperatures at time moments $t = 1, 2, \dots, 24$, and all possible outcomes are described by the sample space

$$\Omega = [-40, 40]^{24} = \bigl\{(\omega_1, \omega_2, \dots, \omega_{24}) \colon \omega_i \in [-40, 40], i = 1, \dots, 24\bigr\}$$

or, without bounding the possible temperature range,

$$\Omega = \mathbb{R}^{24} = \bigl\{(\omega_1, \omega_2, \dots, \omega_{24}) \colon \omega_i \in \mathbb{R}, i = 1, \dots, 24\bigr\}.$$

EXAMPLE 2C. The devices of meteorological stations usually measure the temperature continuously, drawing the temperature curves on the paper or saving them in computer files. Such a curve can be considered as a graph of a continuous function, and the latter can be interpreted as an outcome of measurement (experiment). Thus, in this case, as a sample space, we can consider the set $\Omega = C[0, T]$ (or $\Omega = C[0, +\infty)$ when the measurement time is unlimited) of continuous real functions on the interval $[0, T]$ (or $[0, +\infty)$). Such sample spaces are typical in the theory of random processes, since every element may be interpreted as a possible outcome of a continuous experiment. Of course, a temperature curve is only one of numerous interpretations. A function may show, for example, the dependence on time of the strength or frequency of a radio signal, or of sugar concentration in blood, or the oscillation of stock price, etc. To describe some random processes, the continuous functions are insufficient. For example, modeling the intensity of telephone calls in a service station, we have to introduce the sample spaces containing discontinuous functions, since the number of telephone calls varies not continuously but with unit jumps.

Events. The second object, $\mathcal{F}$, of a probability space is a system of subsets of the sample space Ω. The sets $A \in \mathcal{F}$ are called *events*. In an experiment, the observer is often interested not in the outcome ω itself but rather in whether it belongs to some set $A \subset \Omega$. If it appears that $\omega \in A$, we say that the event occurred; otherwise, we say that the event A did not occur (or that the opposite event $A^c = \Omega \setminus A$ occurred, since then $\omega \in A^c$).

For example, throwing a die (Example 1a), we may be interested not in the number of dots but only in whether the number of dots is greater than three. To this event described in words, we correspond the subset $A = \{4, 5, 6\}$ of the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$. While measuring the temperature (each hour or continuously), we may only be interested in whether the average temperature of the day is positive. The latter event described in words is represented by the subset $\{\omega \in \Omega\colon \frac{1}{24}\sum_{i=1}^{24}\omega_i > 0\}$ of the sample space $\Omega = \mathbb{R}^{24}$ (Example 2b) or by the subset $\{\omega \in \Omega\colon \frac{1}{24}\int_0^{24}\omega(s)\,\mathrm{d}s > 0\}$ of $\Omega = C[0, T]$ (Example 2c).

It is natural to try to consider *all* subsets of Ω as events. However, this is only reasonable in the case where Ω is finite or countable, i.e. where the elementary events can be enumerated by natural numbers: $\Omega = \{\omega_1, \omega_2, \ldots\}$. If Ω is not countable (for example, $C[0, T]$), then this approach is almost always doomed to fail. The reason is rather technical: when we try to further define the probabilities of all subsets (that is, events), it appears that this is impossible (except in trivial or uninteresting cases). Therefore, we have to restrict ourselves to "good" subsets of Ω that can be reasonably called events. Usually, as events, we have to consider some relatively simple subsets of Ω and all subsets that can be obtained from them by the set operations, such as complement, (countable) union, and intersection. For example, in the case of $\Omega = \mathbb{R}$ (Example 2a), as simple subsets (events), we may consider the intervals. When Ω is a function space, say, $C[0, T]$, the role of simple subsets (events) is usually played by the so-called *cylindric* sets of the form

$$A = \bigl\{\omega \in C[0, T]\colon a_i < \omega(t_i) < b_i,\ i = 1, 2, \ldots, k\bigr\}$$

with arbitrary $k \in \mathbb{N}$, $-\infty \leqslant a_i < b_i \leqslant +\infty$, $i = 1, 2, \ldots, k$, $0 \leqslant t_1 < t_2 < \cdots < t_k \leqslant T$.

Thus, it becomes clear why the system $\mathcal{F}$ of subsets (i.e. events) of Ω is always assumed to be a σ-algebra, i.e. a system of subsets satisfying the following conditions:

1) $\Omega \in \mathcal{F}$;

2) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$;

3) $A_n \in \mathcal{F}, n \in \mathbb{N} \implies \displaystyle\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$.

These three conditions yield the closedness of σ-algebra with respect to all other set operations, for example:

4) $\emptyset \in \mathcal{F}$;

5) $A, B \in \mathcal{F} \implies A \cap B,\ A \cup B,\ A \setminus B \in \mathcal{F}$;

6) $A_n \in \mathcal{F}, n \in \mathbb{N} \implies \bigcap_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

Probability. The third object of a probability space is a *probability* (or *probability measure*) $\mathbf{P}$ in the σ-algebra $\mathcal{F}$. This is a function $\mathbf{P}\colon \mathcal{F} \to [0,1]$ satisfying the axioms:

1) $\mathbf{P}(\Omega) = 1$;

2) if $A_n \in \mathcal{F}$, $n \in \mathbb{N}$, is a sequence of pairwise non-intersecting events, then

$$\mathbf{P}\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n).$$

The probability $\mathbf{P}(A)$ of an event A intuitively is understood as subjective or objective degree of its possibility of occurring. For example, if $\mathbf{P}(A) = 1/2$, then we say that there are equal possibilities that event A will or will not occur. In the extreme cases where $\mathbf{P}(A) = 1$ or $\mathbf{P}(A) = 0$, we say, respectively, that an event A is certain or impossible. Thus, the first axiom can be interpreted as the requirement that "something must certainly happen", i.e. the sample space Ω is a certain event. Besides, here we want to mention the probabilistic notion "almost surely", which is related to certain events: if A is a certain event, then we say that it occurs *almost surely* (shortly, *a.s.*) or *with probability one*. Sometimes, it is convenient to use a more general phrase, *almost surely in the event A*, which means *in the event A, except in an event of zero probability*.

If $\Omega = \{\omega_1, \omega_2, \ldots\}$ is an at most countable (finite or countable) set, the probability is completely determined by the probabilities of all elementary events $p_i := \mathbf{P}(\omega_i) = \mathbf{P}(\{\omega_i\})$. Indeed, by the second axiom of probability, then the probability an arbitrary event $A \subset \Omega$ can be calculated by summing the probabilities of all of the elementary events it includes:

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

In the general case, the situation is more complicated: probabilities are constructed by the methods of measure theory. Usually, the probability is first defined for "simple" events, and then it is extended to the whole σ-algebra of events. In this book, we do not go too deeply into these rather subtle details, since this is a separate topic of the basics and justification of probability theory. We shall always assume that we "live" in the already defined or constructed probability space $(\Omega, \mathcal{F}, \mathbf{P})$, though in practice, we will often be able to "forget" it—similarly, knowing that we are part of the Earth or Universe, in everyday life, we rarely think of this.

1.2. Random variables

Often, theoretically and practically, the information about the sample space is obtained only in the form of some of its numerical characteristics. Such numerical characteristics are called *random variables*. More precisely, a random variable (in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$) is a function $X : \Omega \to \mathbb{R}$ (or $X : \Omega \to \overline{\mathbb{R}} = [-\infty, +\infty]$ when infinite values $\pm\infty$ are allowed). In order to have the possibility to consider the events related to a random variable X, it is always required that X be measurable, that is, that

$$\{X < x\} := \big\{\omega \in \Omega \colon X(\omega) < x\big\} \in \mathcal{F} \quad \text{for all} \quad x \in \mathbb{R}. \qquad [1.1]$$

Without this condition being satisfied, we would not be able to consider the events "the value of a random variable X is less than x" or "the value of X is in the interval $[a, b]$", since the corresponding sets of elementary events $\{\omega \in \Omega \colon X(\omega) < x\}$ and $\{\omega \in \Omega \colon a \leqslant X(\omega) \leqslant b\}$ would not be events.

On the other hand, condition [1.1] allows us to consider much more complicated events. The smallest σ-algebra $\mathcal{B}$ of subsets of $\mathbb{R}$ containing all the intervals is called the *Borel* σ-algebra, and its sets are called *Borel sets* or *measurable sets*. From condition [1.1] it follows that for all sets $B \in \mathcal{B}$, $\{X \in B\} := \{\omega \in \Omega \colon X(\omega) \in B\} \in \mathcal{F}$, i.e. the fact that "the value of a random variable X is in a Borel set B", can be reasonably considered an event.

REMARK.– We would like to encourage the reader who falls into despair when facing concepts such as σ-algebra, Borel sets, measurable function. Practically any common sense imaginable subsets of real line are measurable, and common sense imaginable subsets of a sample space Ω are events. Finally, any common sense imaginable real function in a sample space (with sufficiently rich σ-algebra of events) is measurable and, thus, a random variable. Moreover, in many cases, intuitively, it suffices to comprehend random variables simply as quantities depending on chance. Their properties and characteristics are rather well described by the concepts of *distribution function, average, variance*, etc. For these reasons, to widen the circle of readers, subtle issues related to measurability in this book are often deliberately circumvented or suppressed. In the cases where, speaking about functions and sets, the term *measurable* is unavoidable, the reader, without risk of being deceived, may replace it by something like *good*. Of course, here we do not have in mind the mathematicians, for whom complete precision and rigidity are their "daily bread".

1.3. Characteristics of a random variable

The distribution of a random variable X is the probability $\mathbf{P}_X$ on the Borel σ-algebra $\mathcal{B}$ of subsets of $\mathbb{R}$ defined by

$$\mathbf{P}_X(B) := \mathbf{P}\{X \in B\} = \mathbf{P}\big\{\omega \in \Omega \colon X(\omega) \in B\big\}, \quad B \in \mathcal{B}.$$

The distribution $\mathbf{P}_X$ is uniquely determined by the *distribution function* $F = F_X$ of X defined by

$$F(x) = \mathbf{P}\{X < x\} = \mathbf{P}\big\{\omega \in \Omega\colon X(\omega) < x\big\}, \quad x \in \mathbb{R}.$$

The distribution function possesses the following properties:

1) $F(-\infty) := \lim_{x\to-\infty} F(x) = 0$;
2) $F(+\infty) := \lim_{x\to+\infty} F(x) = 1$;
3) F is an increasing left-continuous function, i.e. $\lim_{y\uparrow x} F(y) = F(x)$, $x \in \mathbb{R}$.

The converse is also true: every function F possessing these three properties is a distribution function of some random variable.

In practical applications, the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is often unknown and unspecified, and sufficient information is provided by distribution functions of random variables. On the other hand, theoretically, the use even of an unknown probability space gives some advantages, especially when we need to consider, at the same time, several random variables, for example when considering their convergence. Then we can treat the distributions or distribution functions of different random variables as a manifestation of a single probability $\mathbf{P}$.

Two random variables X and Y are said to be identically distributed if $\mathbf{P}_X = \mathbf{P}_Y$, that is, if $\mathbf{P}\{X \in B\} = \mathbf{P}\{Y \in B\}$ for all good (measurable) sets $B \subset \mathbb{R}$. In such a case, we will write $X \stackrel{d}{=} Y$. For this, the coincidence of the distribution functions is sufficient: $X \stackrel{d}{=} Y$ if $F_X(x) = F_Y(x)$ for all $x \in \mathbb{R}$.

An important characteristic of a random variable X is its *expectation* $\mathbf{E}X = \mathbf{E}(X)$. It can be determined in various ways depending on the mathematical readiness of the reader. For those experienced in general measure theory, the expectation of a random variable X is the integral of X in the sample space Ω with respect to probability measure $\mathbf{P}$:

$$\mathbf{E}X := \int_\Omega X\,\mathrm{d}\mathbf{P} = \int_\Omega X(\omega)\,\mathbf{P}(d\omega).$$

For those who know the Stieltjes integral, the expectation of X can be defined as the integral with respect to the distribution function:

$$\mathbf{E}X := \int_{-\infty}^{+\infty} x\,\mathrm{d}F(x).$$

In the two most frequent cases, the expectation can be defined even more simply (see section 1.4).

If we know the distribution function F of a random variable X and $Y = f(X)$ with a (non-random) function f, then the expectation of Y can be calculated without knowing the distribution function of Y:

$$\mathbf{E}Y = \int_{-\infty}^{+\infty} f(x)\,\mathrm{d}F(x).$$

The expectation is a linear functional: if X and Y are random variables and α, β are real numbers, then $\mathbf{E}(\alpha X + \beta Y) = \alpha\mathbf{E}X + \beta\mathbf{E}Y$. This directly follows from the linearity of integrals (in our case, of that with respect to probability).[1]

The *variance* $\mathbf{D}X = \mathbf{D}(X)$ of a random variable X characterizes its spread around its expectation and is defined by

$$\mathbf{D}X = \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}(X^2) - (\mathbf{E}X)^2$$

or

$$\mathbf{D}X = \int_{-\infty}^{+\infty} (x - \mathbf{E}X)^2\,\mathrm{d}F(x) = \int_{-\infty}^{+\infty} x^2\,\mathrm{d}F(x) - (\mathbf{E}X)^2.$$

1.4. Types of random variables

For the practitioner, the most important (but not exhaustive) types of random variables are discrete and continuous random variables.

Discrete random variables. A random variable X is called discrete if it can take at most a countable (finite or countable) number of values. The distribution function of a discrete random variable X is fully determined by the probabilities of its possible values $p_i = \mathbf{P}\{X = x_i\}$, $i = 1, 2, \dots$ ($\sum_i p_i = 1$). It varies by jumps at the points of the range set of X: $F_X(x) = \sum_{i: x_i < x} p_i$. The expectation and variance of such a random variable can be calculated (or defined) by the formulas

$$\mathbf{E}X = \sum_i x_i p_i, \qquad \mathbf{D}X = \sum_i (x_i - \mathbf{E}X)^2 p_i = \sum_i x_i^2 p_i - (\mathbf{E}X)^2.$$

1. On the other hand, it is rather difficult to prove this using distribution functions—this is yet one argument in favor of considering probabilistic objects in a single probability space.

Examples:

1) A uniform discrete random variable X taking values $1, 2, \ldots, n$ with equal probabilities, i.e. $\mathbf{P}\{X = k\} = \frac{1}{n}$, $k = 1, 2, \ldots, n$.

2) A binomial random variable X: $\mathbf{P}\{X = k\} = C_n^k p^k (1-p)^{n-k}$, $k = 0, 1, \ldots, n$ ($n \in \mathbb{N}$, $p \in (0,1)$). In this case, we write $X \sim B(n, p)$.

3) A Poissonian random variable X: $\mathbf{P}\{X = k\} = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{N}_+$ ($\lambda > 0$). In this case, we write $X \sim P(\lambda)$. A Poissonian random variable is in a certain sense limiting random variable for binomial random variables: $B(n, p_n) \to P(\lambda)$ as $np_n \to \lambda$, $n \to \infty$. The latter relation is understood as follows: if $X_n \sim B(n, p_n)$, $n \in \mathbb{N}$, and $X \in P(\lambda)$, then, for all $k \in \mathbb{N}_+$, we have $\mathbf{P}\{X_n = k\} \to \mathbf{P}\{X = k\}$, provided that $np_n \to \lambda$ as $n \to \infty$.

Continuous random variables. A random variable X is called continuous[2] if its distribution function can be expressed as the integral $F_X(x) = \int_{-\infty}^{x} p(x)\,\mathrm{d}x$, $x \in \mathbb{R}$. In this case, the function p is called the *density* of X. A random variable X is continuous if, for example, its distribution function $F = F_X$ is continuous and is differentiable everywhere, except a finite number of points; then its density $p = F'$ (p can be defined arbitrarily at the points where F' does not exist). The probabilities of taking value in a set, expectation, and variance of such a random variable can be calculated (or defined) by the formulas:

$$\mathbf{P}\{X \in B\} = \int_B p(x)\,\mathrm{d}x, \qquad \mathbf{E}X = \int_{-\infty}^{\infty} xp(x)\,\mathrm{d}x,$$

$$\mathbf{D}X = \int_{-\infty}^{\infty} (x - \mathbf{E}X)^2 p(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x^2 p(x)\,\mathrm{d}x - (\mathbf{E}X)^2.$$

Examples:

1) A random variable uniformly distributed in the interval $[a, b]$. Its density is defined by

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

2) A normal (or normally distributed) or Gaussian random variable X with the density

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(x-a)^2}{2\sigma^2}}, \quad x \in \mathbb{R} \quad (a \in \mathbb{R}, \quad \sigma^2 > 0).$$

2. Often, *absolutely continuous*.

In this case, we write $X \sim N(a, \sigma^2)$. The expectation of such a random variable $\mathbf{E}X = a$ and variance $\mathbf{D}X = \sigma^2$. When $\mathbf{E}X = 0$ and $\mathbf{D}X = 1$, X is called a *standard normal* random variable.

3) An exponential random variable X with the density $p(x) = \alpha \mathrm{e}^{-\alpha x}$, $x \geqslant 0$ $(\alpha > 0)$.

When we consider several random variables, usually, it is not sufficient to know the distribution functions of them all, since they may depend on each other in various ways. Therefore, the probabilistic behavior of a system is characterized by their joint characteristics. The (joint) distribution function of two random variables X and Y is defined as

$$F(x,y) = F_{XY}(x,y) := \mathbf{P}\big\{\omega \in \Omega \colon X(\omega) < x \text{ and } Y(\omega) < y\big\}, \quad x, y \in \mathbb{R}.$$

A function $p = p(x,y) = p_{XY}(x,y)$, $x, y \in \mathbb{R}$, is called the (joint) density of X and Y if

$$F_{XY}(x,y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} p(u,v) \,\mathrm{d}u \,\mathrm{d}v, \quad x, y \in \mathbb{R}.$$

If $F_{XY} \in C^2$, then the random variables X and Y have a joint density, which equals $p(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$.

If the joint density $p = p(x,y)$, $x, y \in \mathbb{R}$, of random variables X and Y is known, then, as before, the expectation of a random variable of the form $Z = f(X,Y)$ can be calculated without knowing the distribution function or density of Z:

$$\mathbf{E}Z = \mathbf{E}f(X,Y) = \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} f(x,y) p(x,y) \,\mathrm{d}x \,\mathrm{d}y.$$

Moreover, they are continuous, and their densities are expressed by the integrals

$$p_X(x) = \int\limits_{-\infty}^{+\infty} p(x,y) \,\mathrm{d}y, \quad x \in \mathbb{R}, \quad \text{and} \quad p_Y(y) = \int\limits_{-\infty}^{+\infty} p(x,y) \,\mathrm{d}x, \quad y \in \mathbb{R}.$$

The joint distribution functions and densities of more than two random variables are similarly defined.

1.5. Conditional probabilities and distributions

Let A and B be two events with $\mathbf{P}(B) > 0$. The conditional probability of event A, given the (occurrence of) event B is defined as

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

In applications, from the context of the problem we often know namely conditional probabilities. In the simplest case, this allows us to calculate the (unconditional) probability:

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

In more complicated cases, the following formulas may be useful: the *multiplication rule*

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

and the *formula of full probability*

$$\mathbf{P}(A) = \sum_k \mathbf{P}(H_k)\mathbf{P}(A|H_k),$$

where $\{H_k\}$ is a finite or countable system of events (interpreted, in this formula, as "hypotheses") such that $H_k \cap H_m = \emptyset$ for $k \neq m$ and $\bigcup_k H_k = \Omega$ (that is, exactly one of the hypotheses H_k certainly occurs).

Especially important is the notion of the conditional probability of an event given a random variable. Suppose first that Y is a discrete random variable taking values in a finite or countable set $\{y_j\}$ (i.e. $\mathbf{P}\{Y = y_j\} > 0$ for all j, and $\sum_j \mathbf{P}\{Y = y_j\} = 1$), then we can calculate the conditional probabilities

$$\mathbf{P}\{A|Y = y_j\} = \frac{\mathbf{P}\{A \cap \{Y = y_j\}\}}{\mathbf{P}\{Y = y_j\}}.$$

For every fixed y such that $\mathbf{P}\{Y = y\} > 0$, the conditional probability $\mathbf{P}\{A|Y = y\}$, as a function of an event A, is a "usual" probability. Therefore, all notions related to probabilities can be redefined to conditional probabilities (with the addition of the word *conditional*). In particular, we can consider the *conditional distribution* of another random variable X, given $Y = y$,

$$\mathbf{P}\{X \in B|Y = y\} = \frac{\mathbf{P}\big\{\{X \in B\} \cap \{Y = y\}\big\}}{\mathbf{P}\{Y = y\}}.$$

The conditional distribution function of X, given $Y = y$, is

$$F(x|Y = y_j) := \mathbf{P}\{X < x|Y = y_j\} = \frac{\mathbf{P}\big\{\{X < x\} \cap \{Y = y_j\}\big\}}{\mathbf{P}\{Y = y_j\}}, \quad x \in \mathbb{R};$$

and its *conditional expectation*, given $Y = y$, equals

$$\mathbf{E}(X|Y = y_j) = \int\limits_{-\infty}^{+\infty} x \, \mathrm{d}F(x|Y = y_j), \quad \text{etc.}$$

It is more complicated to define the conditional probabilities with respect to continuous random variables because of the zero probability of taking, by such a random variable, some fixed value. Suppose that random variables X and Y have a joint density $p = p_{XY}(x, y)$, $x, y \in \mathbb{R}$, and $p_Y(y) = \int_{\mathbb{R}} p_{XY}(x, y) \, \mathrm{d}x > 0$, $y \in \mathbb{R}$. Denote

$$p(x|y) = p_{X|Y}(x|y) := \frac{p_{XY}(x, y)}{p_Y(y)}, \quad x, y \in \mathbb{R}.$$

This function is called the *conditional density* of X, given Y, and is a basis for other *conditional* characteristics of X with respect to Y. For example, the conditional distribution function of X, given Y is defined by

$$F(x|Y = y) = F_{X|Y}(x|y) := \int\limits_{-\infty}^{x} p(u|y) \, \mathrm{d}u, \quad x, y \in \mathbb{R};$$

the conditional probability that X will take a value in the interval $[a, b]$, given $Y = y$, is equal to $\mathbf{P}\{X \in [a, b]|Y = y\} = F(b|Y = y) - F(a|Y = y)$; and the *conditional expectation* of X, given Y, is the function

$$\mathbf{E}(X|Y = y) = \int\limits_{-\infty}^{+\infty} x \, \mathrm{d}F(x|y) = \int\limits_{-\infty}^{+\infty} x p(x|y) \, \mathrm{d}x, \quad y \in \mathbb{R}.$$

1.6. Conditional expectations as random variables

Assuming further that X and Y have a joint density, denote $e(y) = \mathbf{E}(X|Y = y)$, $y \in \mathbb{R}$. Then the random variable $e(Y)$ is also called the conditional expectation of X, given Y (or with respect to Y), and is denoted by $\mathbf{E}(X|Y)$. Then, for every bounded (measurable) function $f\colon \mathbb{R} \to \mathbb{R}$, we have

$$\begin{aligned}
\mathbf{E}\big(f(Y)X\big) &= \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(y)xp(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(y)xp(x|y)p(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int\limits_{-\infty}^{+\infty} f(y)\Bigg(\int\limits_{-\infty}^{+\infty} xp(x|y)\,\mathrm{d}x\Bigg)p(y)\,\mathrm{d}y \\
&= \int\limits_{-\infty}^{+\infty} f(y)\mathbf{E}(X|Y=y)p(y)\,\mathrm{d}y = \mathbf{E}\big(f(Y)\mathbf{E}(X|Y)\big).
\end{aligned}$$

The latter equality can be used to define the conditional expectation of an *arbitrary* random variable X with respect to an *arbitrary* Y. Precisely, by definition, the conditional expectation of X with respect to Y is the random variable $e(Y)$, denoted by $\mathbf{E}(X|Y)$, satisfying the equality

$$\mathbf{E}\big(f(Y)e(Y)\big) = \mathbf{E}\big(f(Y)X\big)$$

for all bounded (measurable) functions $f\colon \mathbb{R} \to \mathbb{R}$. In particular, taking $f = 1$, we get an important corollary, the so-called iteration rule

$$\mathbf{E}\big(\mathbf{E}(X|Y)\big) = \mathbf{E}(X).$$

$\mathcal{M}$ The notion of conditional expectation can be further significantly generalized. Suppose that, in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, we are given another σ-algebra $\mathcal{G} \subset \mathcal{F}$. The conditional expectation of a random variable X with respect to σ-algebra $\mathcal{G}$ is a $\mathcal{G}$-measurable random variable $\mathbf{E}(X|\mathcal{G})$ satisfying the condition

$$\mathbf{E}\big(Z\mathbf{E}(X|\mathcal{G})\big) = \mathbf{E}(ZX)$$

for all bounded $\mathcal{G}$-measurable random variables Z. The expectation $\mathbf{E}(X|\mathcal{G})$ exists if there exists finite $\mathbf{E}(X)$. If $\mathcal{G} = \sigma\{Y\}$ (the σ-algebra generated by Y), then $\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X|Y)$. The main property of the conditional expectation with respect to σ-algebra is as follows: *if* $\mathcal{G}_0 \subset \mathcal{G} \subset \mathcal{F}$ *are* σ*-algebras, and* $\mathbf{E}|X| < +\infty$*, then* $\mathbf{E}(\mathbf{E}(X|\mathcal{G})|\mathcal{G}_0) = \mathbf{E}(X|\mathcal{G}_0)$. The above (rather formal!) definition can be motivated by the following observation: if σ-algebra $\mathcal{G}$ is generated by a *finite or countable* system of incompatible events $\{G_i,\, i \in I\}$ (with $\mathbf{P}(G_i) > 0$ for all i), then

$$\mathbf{E}(X|\mathcal{G})(\omega) = \mathbf{E}(X|G_i) \quad \text{for} \quad \omega \in G_i;$$

in other words, the conditional expectation $\mathbf{E}(X|\mathcal{G})$ equals $\mathbf{E}(X|G_i)$ if we know that the event G_i has occurred.

1.7. Independent events and random variables

Two events A and B are said to be *independent* if $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. When $\mathbf{P}(B) > 0$, this property is equivalent to the equality $\mathbf{P}(A|B) = \mathbf{P}(A)$, intuitively understood that the probability of the event A is independent of whether event B has occurred or not. Random variables X and Y are said to be independent if, for all (measurable) sets B_1 and B_2, the events $\{X \in B_1\}$ and $\{Y \in B_2\}$ are independent. In such a case, we shall write $X \perp\!\!\!\perp Y$. For the independence of random variables X and Y, the following equality for distribution functions is sufficient:

$$F_{XY}(x, y) = F_X(x)F_Y(y), \quad x, y \in \mathbb{R}.$$

If random variables X and Y have a joint density p, then the equality $p(x, y) = p_X(x)p_Y(y)$, $x, y \in \mathbb{R}$, is sufficient for their independence.

Properties of independent random variables:

(1) if $X_1 \perp\!\!\!\perp Y$, then $f(X) \perp\!\!\!\perp g(Y)$ for all (measurable) functions $f \colon \mathbb{R} \to \mathbb{R}$ and $g \colon \mathbb{R} \to \mathbb{R}$;

(2) if $X_n \perp\!\!\!\perp Y$, $n \in \mathbb{N}$, and $X_n \to X$, then $X \perp\!\!\!\perp Y$;

(3) if $X \perp\!\!\!\perp Y$, then $\mathbf{E}(XY) = \mathbf{E}X \cdot \mathbf{E}Y$, provided that the expectations $\mathbf{E}X$ and $\mathbf{E}Y$ are finite.

(4) if $X \perp\!\!\!\perp Y$, then $\mathbf{D}(X + Y) = \mathbf{D}X + \mathbf{D}Y$, provided that the variances $\mathbf{D}X$ and $\mathbf{D}Y$ are finite.

Generalization: systems of independent random variables. Events A_i, $i \in I$, are said to be independent if $\mathbf{P}\{\bigcap_{i \in J} A_i\} = \prod_{i \in J} \mathbf{P}(A_i)$ for every *finite* $J \subset I$. Random variables X_i, $i \in I$, are said to be independent if for any (measurable) sets $B_i \in \mathbb{R}$, $i \in I$, the events $\{X_i \in B_i\}$, $i \in I$, are independent. For systems of independent random variables, analogs of properties (1)–(4) hold (the formulations are left to the reader).

1.8. Convergence of random variables

Let X and X_n, $n \in \mathbb{N}$, be random variables. The sequence $\{X_n\}$ converges to X:

(a) *almost surely* (or *with probability one*) if

$$\mathbf{P}\Big\{\lim_{n\to\infty} X_n = X\Big\} = \mathbf{P}\Big\{\omega \in \Omega \colon \lim_{n\to\infty} X_n(\omega) = X(\omega)\Big\} = 1$$

(abbreviated to $X_n \to X$ a.s. or $\lim\limits_{n\to\infty} X_n = X$ a.s.);

(b) *in probability*, if, for all $\varepsilon > 0$,

$$\mathbf{P}\big\{|X_n - X| > \varepsilon\big\} \to 0, \quad n \to \infty$$

(abbreviated to $X_n \xrightarrow{\mathbf{P}} X$ or $\mathbf{P}\text{-}\lim_{n\to\infty} X_n = X$);

(c) *in mean square* (or in the L^2 *sense*) if

$$\mathbf{E}(X_n - X)^2 \to 0, \quad n \to \infty$$

(abbreviated to $X_n \xrightarrow{L^2} X$ or $L^2\text{-}\lim_{n\to\infty} X_n = X$);

(d) *weakly* (or *in distribution*), if, for all $f \in C_b(\mathbb{R})$ (continuous bounded),

$$\mathbf{E}f(X_n) \to \mathbf{E}f(X), \quad n \to \infty$$

(abbreviated to $X_n \xrightarrow{w} X$); for this, it is necessary and sufficient that

$$F_{X_n}(x) \to F_X(x), \quad n \to \infty,$$

at all continuity points $x \in \mathbb{R}$ of the limit distribution function F_X.

Both almost sure convergence and mean square convergence imply convergence in probability, and the latter implies weak convergence. In the particular case where the limit random variable is, in fact, a constant, say C, we have the converse:

If $X_n \xrightarrow{w} C$, *then* $X_n \xrightarrow{\mathbf{P}} C$.

Note that in this case (where the limit is a constant), convergence in probability also has sense for sequences of random variables defined in *different* probability spaces. On the other hand, in the general case, we can get convergence in probability from weak convergence by "embedding" a weakly convergent sequence of random variables into a single probability space. This fact is often used in proofs since it is more convenient to "live" in one probability space rather than in different spaces. More precisely, from (d) it follows that:

(b′) There exist a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbf{P}})$ and random variables $\widetilde{X}$ and $\widetilde{X}_n$, $n \in \mathbb{N}$, defined in this space such that $\widetilde{X} \stackrel{d}{=} X$, $\widetilde{X}_n \stackrel{d}{=} X_n$, $n \in \mathbb{N}$, and $\widetilde{X}_n \xrightarrow{\mathbf{P}} \widetilde{X}$.

Below, we will need the following properties of the limits of sums and products of sequences of random variables:

– *If* $X_n \to X$ *and* $Y_n \to Y$ *almost surely, in probability, and in mean square, then* $X_n + Y_n \to X + Y$ *in the same sense*.

– *If $X_n \to X$ and $Y_n \to Y$ almost surely or in probability, then $X_n Y_n \to XY$ in the same sense*.

It is convenient to consider the notions related to the convergence in mean square in terms of the so-called L^2 norms. The L^2 norm of a random variable X is the number $\|X\| = \|X\|_{L^2} := \sqrt{\mathbf{E}(X^2)}$ (provided that it is finite). Then the mean square convergence can be defined as follows: $X_n \xrightarrow{L^2} X \iff \|X_n - X\| \to 0$. This norm possesses the following properties that are well known from functional analysis:

(1) $\|\alpha X\| = |\alpha| \, \|X\|$, $\alpha \in \mathbb{R}$;

(2) $\|X + Y\| \leqslant \|X\| + \|Y\|$;

(2′) $\big|\|X\| - \|Y\|\big| \leqslant \|X - Y\|$.

1.9. Cauchy criterion

To check whether a sequence converges in some sense, the corresponding Cauchy[3] criterion is often used. Roughly speaking, the Cauchy criterion says that if $X_n - X_m \to 0$ in some sense as $n, m \to \infty$, then the sequence has a limit in the same sense. The Cauchy criterion holds in cases (a), (b), and (c). Let us formulate more precisely the Cauchy criterion for the mean-square convergence. A sequence of random variables $\{X_n\}$ is called a *fundamental* (or *Cauchy*) in mean square if $\|X_n\| < \infty$, $n \in \mathbb{N}$, and

$$\forall \varepsilon > 0,\ \exists N \in \mathbb{N}\colon \|X_n - X_m\| = \sqrt{\mathbf{E}(X_n - X_m)^2} < \varepsilon \quad \text{for} \quad n, m > N$$

(abbreviated to $X_n - X_m \xrightarrow{L^2} 0$, $n, m \to \infty$). The Cauchy criterion states that:

Every fundamental in mean square sequence of random variables $\{X_n\}$ has a limit in mean square, i.e. there is a random variable X such that $X_n \xrightarrow{L^2} X$ as $n \to \infty$.

To check the convergence in probability, the following criterion is also often used:

A sequence of random variables $\{X_n\}$ converges in probability if and only if every subsequence has a subsubsequence that converges almost surely.

1.10. Series of random variables

For every type of convergence of random variables (except weak convergence), there is the corresponding type of convergence of a series of random variables. We

3. Augustin Louis Cauchy.

say that a series $\sum_{n=1}^{\infty} X_n$ converges almost surely (in probability, in mean square) if its partial-sum sequence $S_n := \sum_{k=1}^{n} X_k$, $n \in \mathbb{N}$, converges almost surely (resp. in probability, in mean square) to some random variable X. Then we write $X = \sum_{n=1}^{\infty} X_n$ almost surely (in probability, in mean square). A sufficient condition for a series $\sum_{n=1}^{\infty} X_n$ to converge in mean square is $\sum_{n=1}^{\infty} \|X_n\|_{L^2} < \infty$.

1.11. Lebesgue theorem

To prove the convergence of integrals and, in particular, expectations, a rather simple but powerful tool is the Lebesgue[4] theorem on dominated convergence. Its essence can be briefly formulated as follows: *given the convergence of functions $f_n \to f$, for the convergence of integrals $\int f_n \to \int f$, it suffices that there is an integrable function g such that $|f_n| \leqslant g$ for all $n \in \mathbb{N}$.* Of course, all functions f_n are defined (and measurable) in a space with a measure, with respect to which the integrals are taken; the function g is said to dominate the sequence $\{f_n\}$, hence the name of the theorem. We formulate it more precisely in two important cases:

– *If a sequence of random variables $\{X_n\}$ converges (almost surely) to a random variable X and there is a random variable Y with finite $\mathbf{E}Y$ such that $|X_n| \leqslant Y$ for all $n \in \mathbb{N}$, then $\mathbf{E}X_n \to \mathbf{E}X$.*

– *If a sequence of functions $\{f_n\}$, integrable in an interval $[a, b]$, converges (almost everywhere) in that interval to a function f and there is a function g, integrable in $[a, b]$, such that $|f_n(x)| \leqslant g(x)$, $x \in [a, b]$, for all $n \in \mathbb{N}$, then $\int_a^b f_n(x)\,\mathrm{d}x \to \int_a^b f(x)\,\mathrm{d}x$, $n \to \infty$.*

1.12. Fubini theorem

The Fubini[5] theorem allows us to reduce the calculation of the integrals of functions of several variables to the calculation of the (iterated) integrals of single-variable functions. For example, the integral of a two-variable function in a rectangle can be calculated by the formula

$$\int_a^b \int_c^d f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_a^b \left(\int_c^d f(x,y)\,\mathrm{d}x \right) \mathrm{d}y = \int_c^d \left(\int_a^b f(x,y)\,\mathrm{d}y \right) \mathrm{d}x,$$

which holds under rather general conditions; for this, it suffices that the (measurable) integrand function f satisfies one of the following three conditions:

– $\int_a^b \int_c^d |f(x,y)|\,\mathrm{d}x\,\mathrm{d}y < +\infty$, i.e. f is integrable as a two-variable function;

4. Henri Leon Lebesgue.

5. Guido Fubini.

– $f \geqslant 0$ (in this case, all integrals in the equality may be equal to $+\infty$);

– $\int_a^b (\int_c^d |f(x,y)| \,\mathrm{d}x) \,\mathrm{d}y < +\infty$ or $\int_c^d (\int_a^b |f(x,y)| \,\mathrm{d}y) \,\mathrm{d}x < +\infty$.

The second equality in the Fubini theorem is equally important to the first one, since it allows us to interchange the integration order, and it frequently happens that the calculation of the (iterated) integral in one order is simpler than in another. Moreover, the Fubini theorem is valid in a much more general situation where the intervals are replaced by arbitrary sets with arbitrary measures. In this book, we will often face the case where one of such measures is a probability. Namely, if $X(t) = X(t,\omega)$, $t \in [a,b]$, is family of random variables, the general Fubini theorem guarantees the equality

$$\mathbf{E}\int_a^b X(t)\,\mathrm{d}t = \int_a^b \mathbf{E}X(t)\,\mathrm{d}t$$

if $X(t) \geqslant 0$ a.s., $t \in [a,b]$, or at least one of the (equal) integrals $\mathbf{E}\int_a^b |X(t)|\,\mathrm{d}t$ and $\int_a^b \mathbf{E}|X(t)|\,\mathrm{d}t$ is finite (of course, $X = X(t,\omega)$, $t \in [a,b]$, $\omega \in \Omega$, must be rather "good" (that is, measurable) as a two-variable function).

1.13. Random processes

Generalizing the notion of a random variable, we can consider mappings (functions) defined in the sample space Ω and taking not real values but rather values in any set E. Such mappings are called *random elements* with values in E. An especially important case is where E is a space of functions on an interval $I \subset \mathbb{R}$, and then a random element is called a random function. Most often, $E = C(I)$ (continuous function on I) or $E = D(I)$ (right-continuous functions with left limits on I). Since I is usually interpreted as a time interval, random functions are also called random (or stochastic) processes.

Thus, a *random process* (or a *random function*) is a mapping X which maps every elementary event $\omega \in \Omega$ to a function $X(\omega) = \{X_t(\omega) = X(t,\omega),\, t \in I\}$, called a *path* (or *trajectory*) of the random process. If we fix $t \in I$, we get a random variable $X_t = X_t(\omega)$, $\omega \in \Omega$. Therefore, a random process is often alternatively defined as a family of random variables X_t, $t \in I$. Usually, it is required that a random process $X = X(t,\omega)$ is a sufficiently "good" (measurable) function with respect to both variables (t,ω) (for example, in order to be able to change the integration order in the integral $\mathbf{E}\int_I X_t\,\mathrm{d}t = \int_I \mathbf{E}X_t\,\mathrm{d}t$; see section 1.12). It is often convenient to take some function space as the sample space itself, for example, $\Omega = C(I)$. Then the random process $X_t(\omega) := \omega(t)$, $t \in I$, is called the *canonical* process.

1.14. Kolmogorov theorem

In this book, we mostly consider continuous processes. Therefore, it is important to have a simple condition to check the continuity of a process. Such a rather universal condition is provided by the Kolmogorov[6] theorem:

If, for a random process $X = \{X_t,\ t \in [0,T]\}$, *there are constants* $C > 0$, $\alpha > 0$, *and* $\beta > 0$ *such that*

$$\mathbf{E}|X_t - X_s|^{\alpha} \leqslant C|t-s|^{\beta+1}, \quad t,s \in [0,T],$$

then X *is continuous almost surely.*

REMARK.– To be precise, we have to specify the statement of the theorem. A random process $\widetilde{X}$ is called a *modification* of X if, for all $t \geqslant 0$, $\widetilde{X}_t = X_t$ a.s. So, in a precise formulation, "X is continuous almost surely" must be replaced by "there exists a continuous modification of X". In fact, specialists of random process theory always understand the continuity of a random process in the sense of the existence of a continuous modification. Similarly, in functional analysis, "a continuous function $f \in L^2[a,b]$" is understood so that there exists $\tilde{f} \in C[a,b]$ coinciding with f almost everywhere (i.e. at all points of $[a,b]$, except a zero-measure set).

6. Andrei Nikolaevich Kolmogorov.

Chapter 2

Brownian Motion

2.1. Definition and properties

Brownian motion plays a great role in mathematics, physics, biology, chemistry, and finance. First, it was used to describe the random chaotic movement of particles suspended in a liquid under bombardment by a tremendous number of atoms of the liquid.[1] The mass of the particle is much greater than that of a molecule, and therefore the influence of a single molecule blow is practically negligible. However, the number of blows is tremendous (about 10^{21} per second), and we can watch by microscope the uninterrupted chaotic movement of the particle. It is also important that all such blows are independent of each other. These facts lead to a mathematical model of Brownian motion. Mathematically, "chaotic" means that the trajectories of Brownian motion, though continuous, are nowhere differentiable. For this reason, a number of mathematical difficulties and "exotic" phenomena appear while considering questions related to Brownian motion.

Eventually it became clear that the Brownian motion is a much more universal random process. Many real life events are accompanied by a number of small independent factors (blows of "molecules"), each of which individually is probably completely irrelevant, but the overall sum of their effects can be quite perceptible. For example, consider stock fluctuations. We do not always see the reasons or shocks ("molecules") that lead to chaotic changes of stock prices: political events and scandals, investment legislation changes, natural disasters, crimes, bankruptcy

1. Scottish botanist Robert Brown was the first who described such a movement in 1987, hence the name.

of banks, deliberately or unconsciously disseminated rumors, moon phases acting on finance brokers' feelings, prejudices, the mood of the risk, and so on.

Those who have a background in probability theory well know the so-called central limit theorem (CLT), which claims that, under very general conditions, a sum of independent small-size random variables is approximately distributed as a normal random variable. Considering a sequence of such sums as a random process, we approximately get a Brownian motion.

Let us strictly formulate the central limit theorem in the simplest form.

THEOREM 2.1 (CLT).– *Let* $\{\xi_n\}$ *be a sequence of independent identically distributed random variables with* $\xi_n = 0$ *and* $\mathbf{D}\xi_n = \mathbf{E}\xi_n^2 = 1$. *Denote* $X_0 = 0$, $X_n = \xi_1 + \xi_2 + \cdots + \xi_n$, $n \in \mathbb{N}$. *Then*

$$\frac{X_n}{\sqrt{n}} \xrightarrow{w} \xi \sim N(0,1), \quad n \to \infty. \qquad [2.1]$$

We see that, for large n, the sums X_n, multiplied by an appropriate normalizing factor $1/\sqrt{n}$, are approximately distributed as a normal random variable $\xi \sim N(0,1)$. How can we characterize the behavior of the whole sequence $\{X_n\}$? In view of relation [2.1], it is natural to try shrinking the sequence by the factor $1/\sqrt{n}$, i.e. consider the sequence $\{Y_k = X_k/\sqrt{n},\ k = 0,1,2,\ldots\}$. It appears that it is also convenient to shrink the time scale by the factor $1/n$ in order that, for example, the value Y_n would be taken at the time moment 1, rather than at moment n.

Let us extend the random sequence $\{X_n\}$ to a random process $\{X_t,\ t \geqslant 0\}$ by taking

$$X_t := X_{[t]} = X_k, \quad t \in [k, k+1),\ k \in \mathbb{N}_+ = \mathbb{N} \cup \{0\},$$

where $[t]$ denotes the integer part of a number t. A typical trajectory of the process X is as follows:

From the CLT we also have that

$$X_t^n := \frac{X_{nt}}{\sqrt{n}} \xrightarrow{w} B_t \sim N(0,t), \quad n \to \infty,\ t \geqslant 0.$$

Indeed, for all $t > 0$,

$$\frac{X_{nt}}{\sqrt{n}} = \frac{X_{[nt]}}{\sqrt{[nt]}}\sqrt{\frac{[nt]}{n}} \xrightarrow{w} \xi\sqrt{t} \sim N(0,t), \quad n \to \infty,$$

where we used the fact that $[nt]/n \to t$ as $n \to \infty$ for all $t \geqslant 0$. This is seen from the inequalities $(nt-1)/n < [nt]/n \leqslant t$ with the left-hand side tending to t.

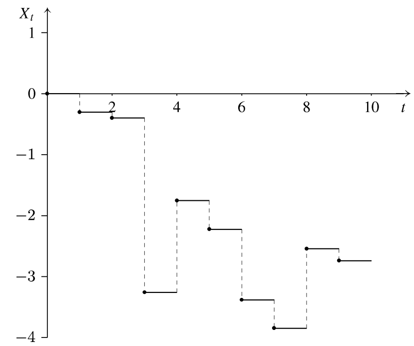


Figure 2.1. *A trajectory of X*

Thus, we get that, for each *fixed* t, the values X_t^n of the processes X^n weakly converge to a random variable $B_t \sim N(0,t)$. However, this does not mean that the sequence $\{X^n\}$ converges, in some sense, to some random process. In other words, is possible to construct, on a single probability space $(\Omega, \mathcal{F}, \mathbf{P})$, a random process B_t, $t \geqslant 0$, such that, *for all* $t \geqslant 0$, we should have

$$X_t^n \xrightarrow{w} B_t, \quad n \to \infty?$$

Before we answer the question, let us plot the graphs of X_t^n, $t \geqslant 0$, with various n, observing the behavior of the trajectories as n increases. The trajectory of X^2 (with the same sequence $\{\xi_n\}$) is shown in Figure 2.2.

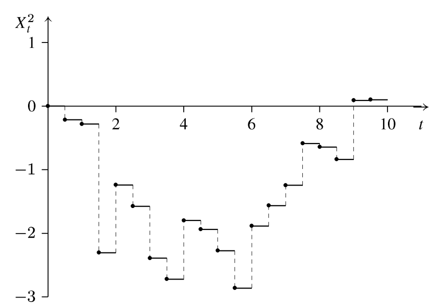


Figure 2.2. *A trajectory of* X^2

Note that the trajectory of the process X^2 in the time interval $[0,5]$ repeats the trajectory of the process $X = X^1$ in the interval $[0,10]$, but with the time scale two times smaller and with the values of the process $\sqrt{2}$ times smaller. Further doubling n

we get the trajectories of the process X^n, $n = 2^k$, $k = 2, 3, \ldots, 9$, that we see in Figures 2.3 to 2.6. In all the figures, we use *the same* values of the random sequence $\{\xi_n\}$ generated in advance. Therefore, we further see that, in each figure half of the trajectory repeats, with appropriately changed time and space scales the previous trajectory. However, it is much more important that, as n increases, step-like trajectories of the processes merge into rather irregular but *continuous* trajectories. Therefore, we may expect not only the positive answer to the question stated but also the continuity of the limit process.

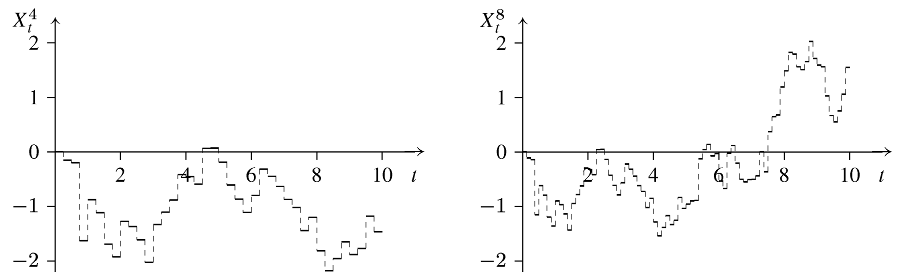


Figure 2.3. *Trajectories of X^4 and X^8*

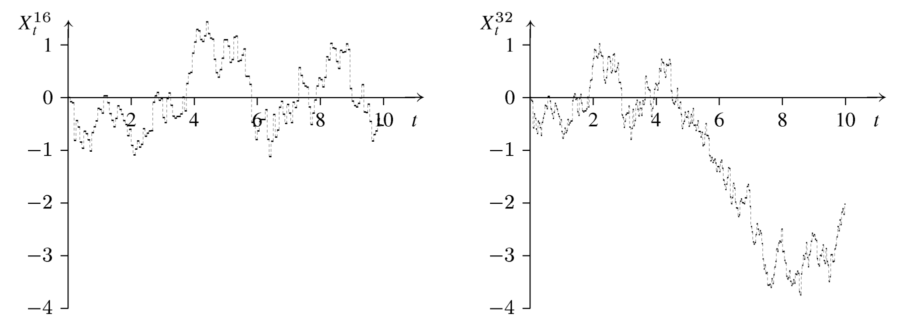


Figure 2.4. *Trajectories of X^{16} and X^{32}*

THEOREM 2.2.– *There exists a random process $B = \{B_t,\, t \geqslant 0\}$ such that*

$$\sum_{i=1}^{k} \lambda_i X_{t_i}^n \xrightarrow{w} \sum_{i=1}^{k} \lambda_i B_{t_i}, \quad n \to \infty,$$

with all collections $0 \leqslant t_1 < t_2 < \cdots < t_k$, $\lambda_i \in \mathbb{R}$, $i = 1, 2, \ldots k$, $k \in \mathbb{N}$. We denote this convergence by $X^n \xrightarrow{w} B$ and say that the sequence of the processes $\{X^n\}$ weakly converges to B in the sense of finite-dimensional distributions.

The process B is called a Brownian motion (or a standard Wiener[2] process).

2. Norbert Wiener.

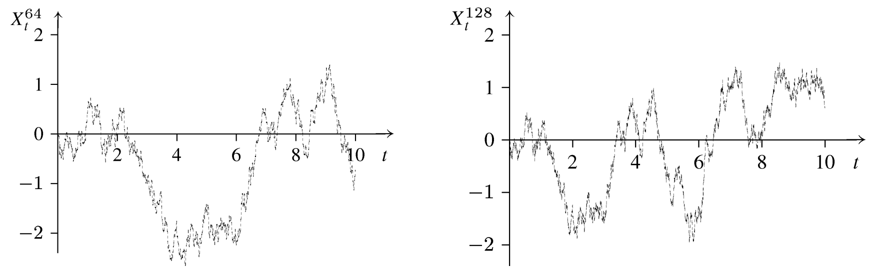


Figure 2.5. *Trajectories of* X^{64} *and* X^{128}

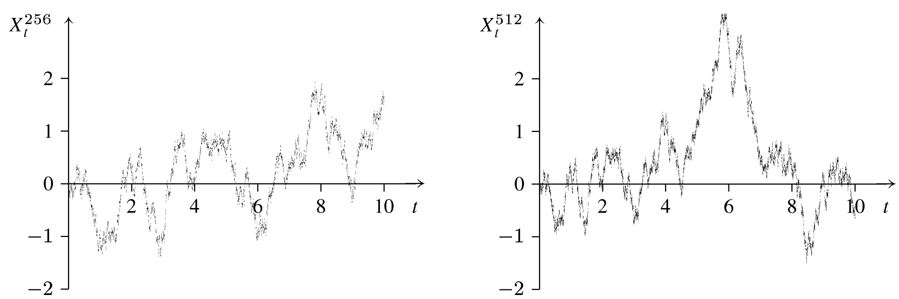


Figure 2.6. *Trajectories of* X^{256} *and* X^{512}

We shall not prove this theorem, although it is said that *the main result of the Brownian motion theory is... the existence of a Brownian motion*. However, we will use the theorem to prove some characteristic properties of a Brownian motion.

THEOREM 2.3 (Properties of a Brownian motion).–

1) $B_0 = 0$ *and* $B_t - B_s \sim N(0, t-s)$, $0 \leqslant s < t$;

2) *the random variables* $B_{t_1} - B_{t_0}$, $B_{t_2} - B_{t_1}$, $B_{t_3} - B_{t_2}, \dots, B_{t_k} - B_{t_{k-1}}$ *are independent with arbitrary* $0 \leqslant t_0 < t_1 < t_2 < \cdots < t_k < \infty$. *(Processes with this property are called processes with independent increments);*

3) B *is a continuous random process, i.e.* B_t, $t \geqslant 0$, *is a continuous function with probability one.*

Proof. 1) The equality $B(0) = 0$ is obvious. By Theorem 2.2,

$$X_t^n - X_s^n \xrightarrow{w} B_t - B_s, \quad n \to \infty.$$

On the other hand, by the CLT,

$$\begin{aligned}X_t^n - X_s^n &= \frac{X_{nt} - X_{ns}}{\sqrt{n}} = \frac{\xi_{[ns]+1} + \xi_{[ns]+2} + \cdots + \xi_{[nt]}}{\sqrt{n}} \\ &\stackrel{d}{=} \frac{\xi_1 + \xi_2 + \cdots + \xi_{[nt]-[ns]}}{\sqrt{n}} \\ &= \frac{\xi_1 + \xi_2 + \cdots + \xi_{[nt]-[ns]}}{\sqrt{[nt]-[ns]}} \cdot \sqrt{\frac{[nt]}{n} - \frac{[ns]}{n}} \\ &\xrightarrow{w} \xi\sqrt{t-s} \sim N(0, t-s), \quad n \to \infty,\end{aligned}$$

where $\xi \sim N(0,1)$.

2) For every $n \in \mathbb{N}$, the random variables

$$X_{t_i}^n - X_{t_{i-1}}^n = \frac{\sum_{j=[nt_{i-1}]+1}^{[nt_i]} \xi_j}{\sqrt{n}}, \quad i = 1, 2, \ldots, k,$$

are independent, since in the numerator, for different i, we have sums of different random variables. Therefore, their limits $B_{t_i} - B_{t_{i-1}}$, $i = 1, 2, \ldots, k$, are also independent.

3) Since $B_t - B_s \sim N(0, t-s)$ for $t > s$, we have $\mathbf{E}|B_t - B_s|^4 = 3|t-s|^2$, $t, s \geqslant 0$. Therefore, the continuity of the Brownian follows from the Kolmogorov theorem (section 1.14) by taking $\alpha = 4$, $\beta = 1$, $C = 3$, and any T. △

We now present a pair of more subtle properties of Brownian motion.

THEOREM 2.4 (Additional properties of Brownian motion).–

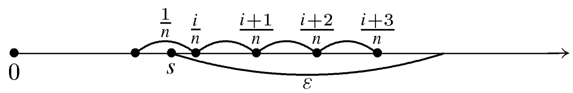
1) *With probability one, the trajectories of Brownian motion are nowhere differentiable.*

2) *Let* $M_t := \sup_{s \leqslant t} B_s$ *and* $m_t := \inf_{s \leqslant t} B_s$. *Then* $M_t \stackrel{d}{=} |B_t|$ *ir* $m_t \stackrel{d}{=} -|B_t|$ *for all* $t \geqslant 0$.

Proof. 1) We first give an idea of the proof. For any $h > 0$, the increment $B_{t+h} - B_t \sim N(0, h)$, and therefore $(B_{t+h} - B_t)/h \sim N(0, h^{-1})$. Since the variance of the ratio $(B_{t+h} - B_t)/h$ equals h^{-1} and tends to ∞ as $h \to 0$, this ratio cannot have a limit as $h \to 0$ (the limit of a sequence of normal random variables may only be a normal random variable or a constant).

$\mathcal{M}$ It suffices to prove the non-differentiability of a trajectory of Brownian motion on any finite time interval $I = [0, T]$. If a trajectory of B is differentiable at some point $s \in I$, then the finite limit $|B'_s| = \lim_{t \to s} |B_t - B_s|/|t-s|$ exists. For an arbitrary (random) positive integer $k > |B'_s|$, there exists (random) $\varepsilon > 0$, such that

$$|B_t - B_s| < k(t-s) \quad \text{for} \quad 0 < t-s < \varepsilon.$$



Now, taking (random) $n \in \mathbb{N}$ such that $4/n < \varepsilon$ and $i = [ns] + 1$ (see the figure above) , we get $0 < j/n - s < \varepsilon$, $j = i, i+1, i+2, i+3$, and therefore

$$|B_{j/n} - B_s| < k\Big(\frac{j}{n} - s\Big) < k\frac{4}{n}, \quad j = i, i+1, i+2, i+3.$$

Hence,

$$\begin{aligned} |B_{j/n} - B_{(j-1)/n}| &\leqslant |B_{j/n} - B_s| + |B_{(j-1)/n} - B_s| \\ &< k\frac{4}{n} + k\frac{4}{n} = \frac{8k}{n}, \quad j = i+1, i+2, i+3. \end{aligned}$$

Define the events

$$A^j_{k,n} := \Big\{ |B_{j/n} - B_{(j-1)/n}| < \frac{8k}{n} \Big\}$$

and consider the event

$$A := \bigcup_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \bigcup_{i=1}^{[nT]} \bigcap_{j=i+1}^{i+3} A^j_{k,n}.$$

It occurs when there is a positive integer k such that, for all sufficiently large $n \in \mathbb{N}$ and some i, we have the equalities $|B_{j/n} - B_{(j-1)/n}| < 8k/n$, $j = i+1, i+2, i+3$, that were obtained from the assumption that B is differentiable at some point $s \in [0, T]$. Therefore, A contains the event that a trajectory of Brownian motion is differentiable at some point $s \in [0, T]$. Thus, the statement will be proved if we show that $\mathbf{P}(A) = 0$. The event A is a countable union of the events

$$B_{k,m} := \bigcap_{n=m}^{\infty} \bigcup_{i=1}^{[nT]} \bigcap_{j=i+1}^{i+3} A^j_{k,n}, \quad k, m \in \mathbb{N}.$$

Therefore, it suffices to show that $\mathbf{P}(B_{k,m}) = 0$ for all $k, m \in \mathbb{N}$. Since

$$\mathbf{P}(B_{k,m}) \leqslant \mathbf{P}\Big(\bigcup_{i=1}^{[nT]} \bigcap_{j=i+1}^{i+3} A^j_{k,n} \Big)$$

for $n \geqslant m$, in turn it suffices to show that, for all $k \in \mathbb{N}$,

$$\mathbf{P}\Big(\bigcup_{i=1}^{[nT]} \bigcap_{j=i+1}^{i+3} A_{k,n}^j\Big) \to 0, \quad n \to \infty.$$

Since the events $A_{k,n}^j$, $j = i+1, i+2, i+3$, are independent and

$$|B_{j/n} - B_{(j-1)/n}| \overset{d}{=} |B(1/n)| \overset{d}{=} |B(1)|/\sqrt{n},$$

for all $i, k, n \in \mathbb{N}$, we have

$$\mathbf{P}\Big(\bigcap_{j=i+1}^{i+3} A_{k,n}^j\Big) = \prod_{j=i+1}^{i+3} \mathbf{P}\big(A_{k,n}^j\big) = \Big(\mathbf{P}\Big\{\frac{|B(1)|}{\sqrt{n}} < \frac{8k}{n}\Big\}\Big)^3$$
$$= \Big(\mathbf{P}\Big\{|B(1)| < \frac{8k}{\sqrt{n}}\Big\}\Big)^3.$$

Using the simple inequality

$$\mathbf{P}\big\{|B(1)| < x\big\} = \frac{1}{\sqrt{2\pi}} \int_{-x}^{x} \mathrm{e}^{-y^2/2}\, \mathrm{d}y \leqslant \frac{2x}{\sqrt{2\pi}}, \quad x > 0,$$

we get:

$$\mathbf{P}\Big(\bigcap_{j=i+1}^{i+3} A_{k,n}^j\Big) \leqslant \left(\frac{2 \cdot \frac{8k}{\sqrt{n}}}{\sqrt{2\pi}}\right)^3 = \frac{C_k}{n^{3/2}}$$

with a constant C_k depending only on k. From this we finally obtain

$$\mathbf{P}\Big(\bigcap_{j=i+1}^{i+3} A_{k,n}^j\Big) \leqslant \sum_{i=1}^{[nT]} \mathbf{P}\Big(\bigcap_{j=i+1}^{i+3} A_{k,n}^j\Big)$$
$$\leqslant [nT]\frac{C_k}{n^{3/2}} \leqslant \frac{TC_k}{\sqrt{n}} \to 0, \quad n \to \infty.$$

$\mathcal{M}$ 2) We will apply the so-called *symmetry principle* of Brownian motion. Random variable X is called symmetric if $X \overset{d}{=} -X$. The increments of Brownian motion $B_t - B_\tau$, $t > \tau$, have the distributions $N(0, t-\tau)$ and therefore are symmetric. The symmetry principle claims that these increments are also symmetric when τ is not a fixed time moment but rather the first (random) moment when Brownian motion hits some level $x \in \mathbb{R}$, i.e.

$$\tau = \tau_x = \min\{s \geqslant 0\colon B_s = x\}.$$

(Intuitively, this fact is obvious, though not trivial, and is a partial case of a general strong Markov property of a Brownian motion.) More precisely, the random variables

$$\big(B_t - B_{\tau_x}\big)1\!\!1_{\{\tau_x \leqslant t\}} = \begin{cases} B_t - x & \text{for } \tau_x \leqslant t, \\ 0 & \text{for } \tau_x > t, \end{cases}$$

are symmetric. In particular, for all $x > 0$ and $t \geqslant 0$, we get:

$$\begin{aligned} \mathbf{P}\{\tau_x \leqslant t,\, B_t > x\} &= \mathbf{P}\big\{\big(B_t - B_{\tau_x}\big)1\!\!1_{\{\tau_x \leqslant t\}} > 0\big\} \\ &= \mathbf{P}\big\{\big(B_t - B_{\tau_x}\big)1\!\!1_{\{\tau_x \leqslant t\}} < 0\big\} \\ &= \mathbf{P}\{\tau_x \leqslant t,\, B_t < x\}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{P}\{\tau_x \leqslant t\} &= \mathbf{P}\{\tau_x \leqslant t,\, B_t > x\} + \mathbf{P}\{\tau_x \leqslant t,\, B_t < x\} \\ &= 2\mathbf{P}\{\tau_x \leqslant t,\, B_t > x\} = 2\mathbf{P}\{B_t > x\} = \mathbf{P}\{|B_t| > x\}. \end{aligned}$$

Since the events $\{M_t \geqslant x\}$ and $\{\tau_x \leqslant t\}$ coincide, we have

$$\mathbf{P}\{M_t \geqslant x\} = \mathbf{P}\{\tau_x \leqslant t\} = \mathbf{P}\{|B_t| > x\} = \mathbf{P}\{|B_t| \geqslant x\}, \quad x > 0,$$

or

$$F_{M_t}(x) = 1 - \mathbf{P}\{M_t \geqslant x\} = 1 - \mathbf{P}\{|B_t| \geqslant x\} = \mathbf{P}\{|B_t| < x\} = F_{|B_t|}(x), \quad x \in \mathbb{R}$$

($F_{M_t}(x) = F_{|B_t|}(x) = 0$ for $x \leqslant 0$).

The equality $m_t \stackrel{d}{=} -|B_t|$ can be proved similarly or by using the property just proved: since the random process $\widetilde{B}_t := -B_t$, $t \geqslant 0$, is also a Brownian motion, we have

$$m_t = \inf_{s \leqslant t} B_s = -\sup_{s \leqslant t} \widetilde{B}_s \stackrel{d}{=} -\big|\widetilde{B}_t\big| = -|B_t|. \qquad \triangle$$

REMARK.– It is interesting to note that, as we have just proved, $M_t = \sup_{s \leqslant t} B_t \stackrel{d}{=} |B_t|$ for every *fixed* t, the processes $M = \{M_t,\, t \geqslant 0\}$ and $|B| = \{|B_t|,\, t \geqslant 0\}$ are not identically distributed; moreover, their behavior is completely different!

DEFINITION 2.5.– *We say that a random variable X belongs to the class $\mathcal{H}_t = \mathcal{H}_t^B$ if it is completely determined by the values of a Brownian motion B in the time interval $[0, t]$ or, in other words, if X is a functional of the trajectory B_s, $s \in [0, t]$. The class $\mathcal{H}_t$ of all such random variables is called the* past *or* history *of a Brownian motion up to moment t. We also write that an* event $A \in \mathcal{H}_t$ *if* $1\!\!1_A \in \mathcal{H}_t$.

PROPOSITION 2.6.– *If* $X \in \mathcal{H}_t$, *then* $X \perp\!\!\!\perp B_s - B_u$, $s \geqslant u \geqslant t$.

Proof. We defined the class $\mathcal{H}_t$ rather intuitively. To prove the proposition, we have to characterize random variables belonging to the class $\mathcal{H}_t$ more precisely. First of all, this class obviously contains the random variables of the form

$$X = f\big(B_{t_1}, B_{t_2}, \dots, B_{t_k}\big),$$

where $f: \mathbb{R}^k \to \mathbb{R}$ is a good (measurable) function, and $0 \leqslant t_1 < t_2 < \cdots < t_k \leqslant t$, $k \in \mathbb{N}$. Such a random variable can be expressed by increments of Brownian motion up to moment t:

$$X = f\left(B_{t_1} - B_0, \big(B_{t_2} - B_{t_1}\big) + \big(B_{t_1} - B_0\big), \dots, \sum_{i=1}^{k} \big(B_{t_i} - B_{t_{i-1}}\big)\right), \quad t_0 = 0.$$

This shows (see section 1.7, property 1) that $X \perp\!\!\!\perp B_s - B_u$, $s \geqslant u \geqslant t$. It is also natural that if $X^n \in \mathcal{H}_t$, $n \in \mathbb{N}$, and $X^n \to X$ a.s., then the limit X also belongs to the class $\mathcal{H}_t$. From the properties of independent random variables (section 1.7, property 2) we again have that $X \perp\!\!\!\perp B_s - B_u$, $s \geqslant u \geqslant t$. Continuing, it is also natural to attribute to the class $\mathcal{H}_t$ all the limits of sequences of random variables already attributed to $\mathcal{H}_t$ since the independence (from the increments $B_s - B_u$, $s \geqslant u \geqslant t$) is preserved when passing to the limit (in fact, such limits do not give any new random variables; however, this fact is not important for us). △

EXAMPLES 2.7.– Here are some examples of random variables belonging to the class $\mathcal{H}_t$:

$$X_1 = f(B_t), \quad \text{where } f \in C(\mathbb{R});$$

$$X_2 = \int_0^t B_s \,\mathrm{d}s = \lim_{n\to\infty} \frac{t}{n} \sum_{k=1}^{n} B\Big(\frac{k}{n}t\Big);$$

$$X_3 = \sup_{s\in[0,t]} B_s = \lim_{n\to\infty} \max_{0\leqslant k\leqslant n} B\Big(\frac{k}{n}t\Big);$$

$$X_4 = \min\big\{s \geqslant 0 \colon B(s) > 1\big\} \wedge t = \lim_{n\to\infty} \min\big\{k2^{-n} \colon B(k2^{-n}) > 1\big\} \wedge t.$$

All these rather simple examples can be expressed in the form $X = F(B_s,\, s \leqslant t)$ with a continuous functional $F: C[0,t] \to \mathbb{R}$; however, the class $\mathcal{H}_t$ is incomparably wider.

REMARKS.– 1. If needed, the history $\mathcal{H}_t = \mathcal{H}_t^B$ may be complemented by random variables that are independent of the Brownian motion. For further consideration, it is only important to preserve the independence of $\mathcal{H}_t$ and the increments $B_s - B_u$, $s \geqslant u \geqslant t$. For example, we can take $\mathcal{H}_t$ as all "simple" random variables of the form

$$X = f\big(\xi_1, \xi_2, \ldots, \xi_m, B_{t_1}, B_{t_2}, \ldots, B_{t_k}\big)$$

and the limits of sequences of such variables; here ξ_i, $i = 1, 2, \ldots$, are arbitrary random variables independent of the Brownian motion B.

$\mathcal{M}$ 2. The class $\mathcal{H}_t$ can be strictly defined as follows. Denote by $\mathcal{F}_t = \mathcal{F}_t^B$ the σ-algebra $\sigma\{B_s,\ s \leqslant t\}$ generated by all random variables B_s, $s \leqslant t$, and by all zero-probability events. In other words, $\mathcal{F}_t$ is the minimal σ-algebra with respect to which all these random variables are measurable and which contains all zero-probability events. Then the history of Brownian motion up to moment t is the class $\mathcal{H}_t$ of all $\mathcal{F}_t$-measurable random variables. In view of the previous remark, we can take $\mathcal{F}_t$ to be the σ-algebra generated not only by B_s, $s \leqslant t$, but also by any random variables independent of B. Moreover, in full generality, the history $\mathcal{H}_t$ can be defined by extending the σ-algebras $\mathcal{F}_t$, $t \geqslant 0$, so that the following requirements remain satisfied:

(i) $\mathbb{F} = \{\mathcal{F}_t,\ t \geqslant 0\}$ is an increasing family of σ-algebras, i.e. $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leqslant t$;

(ii) σ-algebras $\mathcal{F}_t$ contain all zero-probability events;

(iii) for all $t \geqslant 0$, B_t are $\mathcal{F}_t$-measurable, and the increments $B_s - B_u$, $s \geqslant u \geqslant t$, of the σ-algebra $\mathcal{F}_t$, i.e., of all $\mathcal{F}_t$-measurable random variables.

Any system $\mathbb{F}$ of σ-algebras satisfying conditions (i)–(ii) is called *filtration*. If, moreover, condition (iii) is satisfied, we say that the Brownian motion B is adapted to the filtration $\mathbb{F}$ or that B is a Brownian motion with respect to the filtration $\mathbb{F}$.

2.2. White noise and Brownian motion

Let us return to the sequence of random processes $\{X^n\}$ weakly converging to B and modify them so that they become continuous by taking "polygonal" processes instead of "step" processes,. More precisely, we take $B_t^n = X_t^n$ for $t = \frac{k}{n}$, $k = 0, 1, 2, \ldots$, and extend B^n linearly in the intervals $[k/n, (k+1)/n]$, i.e.

$$\begin{aligned} B_t^n &= X_{k/n}^n + \big(X_{(k+1)/n}^n - X_{k/n}^n\big)\frac{t - k/n}{1/n} \\ &= X_{k/n}^n + \sqrt{n}\,\xi_{k+1}(t - k/n), \quad t \in \big[k/n, (k+1)/n\big), \quad k = 0, 1, 2, \ldots. \end{aligned}$$

Then, for this sequence, we also have $B^n \xrightarrow{w} B$.

To compare typical trajectories of polygonal processes B^n with those of step processes X^n (see Figures 2.1 to 2.6), we plot their graphs with the same n and the same sequences $\{\xi_n\}$ generated in advance (Figures 2.7 to 2.11).

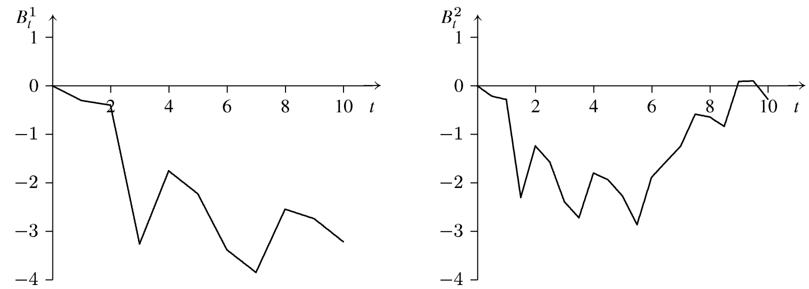


Figure 2.7. *Trajectories of* B^1 *and* B^2

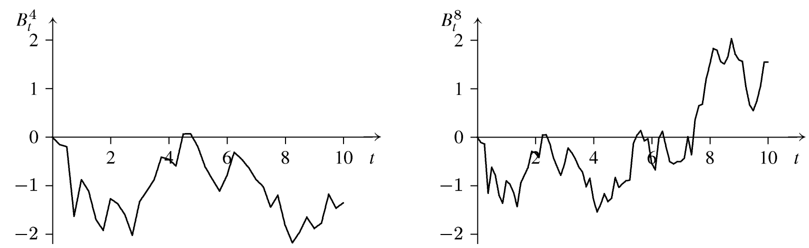


Figure 2.8. *Trajectories of* B^4 *and* B^8

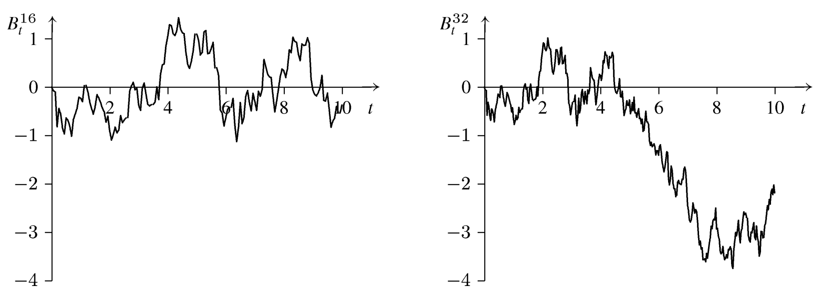


Figure 2.9. *Trajectories of* B^{16} *and* B^{32}

Note that the processes B^n are piecewise differentiable and

$$B_t^n = \int_0^t \eta_s^n \,\mathrm{d}s, \quad \eta_s^n = \sqrt{n}\,\xi_{k+1}, \quad s \in \big[k/n, (k+1)/n\big), \quad k = 0, 1, 2, \ldots.$$

For better understanding of the behavior of random processes η^n, we also give a pair of their graphs together with the graphs of B^n. True, it is difficult to do this with large n, since the jumps of η^n increase with n; therefore, we restrict ourselves to the cases $n = 4$ and $n = 8$ (Figure 2.12).

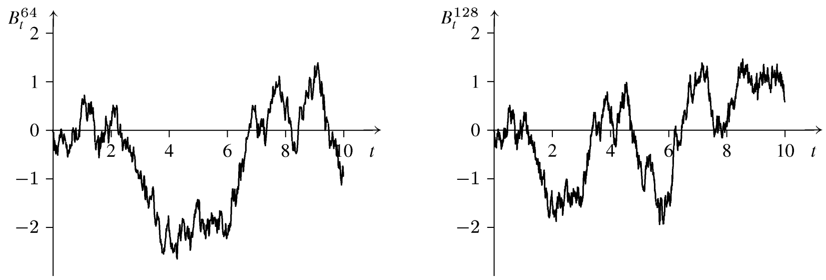


Figure 2.10. *Trajectories of* B^{64} *and* B^{128}

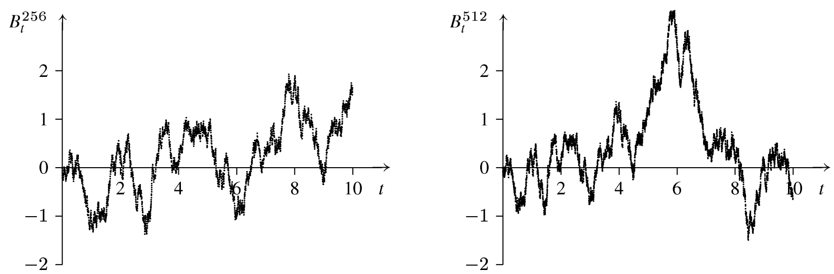


Figure 2.11. *Trajectories of* B^{256} *and* B^{512}

We can interpret processes η^n as chaotic processes with large jumps and short memory: $\eta_t^n \perp\!\!\!\perp \eta_s^n$ if $|t-s| > \frac{1}{n}$. Such processes are called *noises*. As $n \to \infty$, in the limits, we gate the "process" η with infinite values and without any memory. Such a process does not exist in a usual sense. However, it is used as an idealized model of a real noise with short memory. The "process" η is called *white noise*, while η^n and similar processes are called *colored noises*. Since $B_t^n = \int_0^t \eta_s^n \, \mathrm{d}s \xrightarrow{w} B_t$ (and $\frac{\mathrm{d}B_t^n}{\mathrm{d}t} = \eta_t^n$), the derivative of a Brownian motion is said to be a white noise (and one writes $\frac{\mathrm{d}B_t}{\mathrm{d}t} = \eta_t$ or $B_t = \int_0^t \eta_s \, \mathrm{d}s$), though, as is proved in Theorem 2.4, almost all trajectories of a Brownian motion are nowhere differentiable!

Thus, white noise has no real meaning, though it makes sense to consider processes that are close to white noise. The phrase "a process η_t is close to white noise" means that the process $B_t = \int_0^t \eta_s \, \mathrm{d}s$ is close to a Brownian motion. If we want to create an idealized mathematical model of a real physical phenomenon, we first have to consider a model with real, colored noises η^n, and then pass to the limit as $B_t^n = \int_0^t \eta_s^n \, \mathrm{d}s \xrightarrow{w} B_t$. There are a number of random processes (more precisely, their sequences) η^n that can be considered colored noises that are close to white noise. There many theorems on sufficient conditions that guarantee the weak convergence of

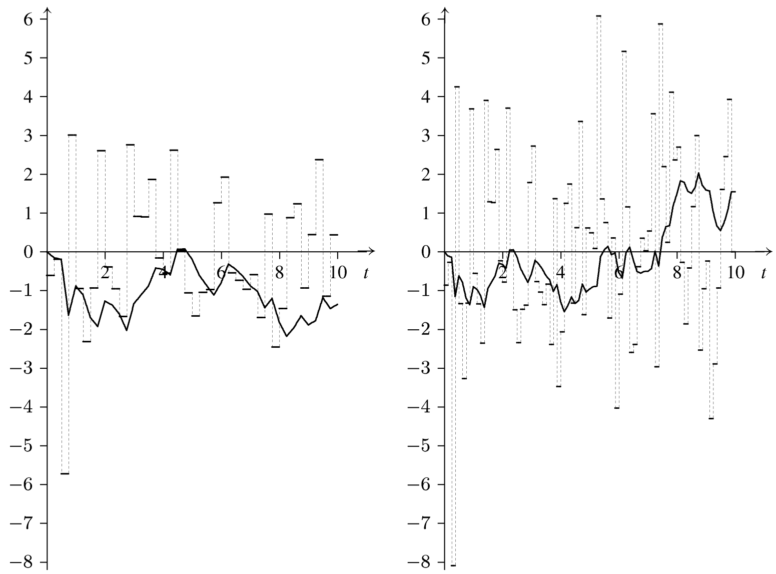


Figure 2.12. *Trajectories of B^n ir η^n, $n = 2$; 4*

the corresponding integrals B^n to a Brownian motion. For example, $B^n \xrightarrow{w} B$ when $B_t^n = \int_0^t \eta_s^n \,\mathrm{d}s$, $\eta_s^n = \tilde{\xi}_{nk}$, $s \in [t_k^n, t_{k+1}^n)$, $0 = t_0^n < t_1^n < t_2^n < \cdots \to \infty$, $\max_k \Delta t_k^n = \max_k (t_{k+1}^n - t_k^n) \to 0$, $n \to \infty$, and, for every fixed $n \in \mathbb{N}$, random variables $\tilde{\xi}_{nk}$, $k \in \mathbb{N}$, are independent and identically distributed with means $\mathbf{E}\tilde{\xi}_{nk} = 0$ and variances $\mathbf{E}\tilde{\xi}_{nk}^2 = \frac{1}{\Delta t_k^n}$.

REMARK.– In another interpretation, a white noise is a *generalized stationary process with zero mean and covariance function equal to the Dirac function*. Let us comment on the above-mentioned notions. Random process X_t, $t \geqslant 0$, is called a *stationary process* (in the wide sense) if the expectation $E(t) := \mathbf{E}X_t$ is constant and the *covariance function* $C(s,t) := \mathbf{E}(X_s X_t)$ depends only on the difference $t - s$, that is $C(s,t) = C(t-s)$. The Dirac function is defined as the "function" $\delta : \mathbb{R} \to [0, +\infty]$ possessing the following properties: $\int_{\mathbb{R}} \delta(s)\,\mathrm{d}s = 1$, $\delta(0) = +\infty$, and $\delta(s) = 0$, $s \in \mathbb{R} \setminus \{0\}$. It is not a function in the usual sense, but it can be understood as the limit of a sequence of non-negative functions $\{\delta_n\}$ such that, for all n, $\int_{\mathbb{R}} \delta_n(s)\,\mathrm{d}s = 1$ and $\delta_n(s) = 0$ for $s \notin [-1/n, 1/n]$. Rigorously, the Dirac function is defined and considered in the theory of generalized functions (also called distributions). Let us find the covariance function C of white noise. By the preceding, it is naturally defined as the limit of the covariance functions C_n of colored noises η^n tending to white noise. We calculate $C_n(s,t) = \mathbf{E}(\eta_s^n \eta_t^n) = n$ when s and t are in the same interval $[k/n, (k+1)/n)$ and $C_n(s,t) = 0$ otherwise; moreover, $\int_{\mathbb{R}} C_n(s,t)\,\mathrm{d}s = 1$ for all t. Passing to the limit as $n \to \infty$, we find that the limit "covariance function" $C(s,t) = \lim_{n\to\infty} C_n(s,t)$ equals $+\infty$ for $s = t$ and 0 for $s \neq t$. It is also natural to suppose that $\int_{\mathbb{R}} C(s,t)\,\mathrm{d}s = \lim_{n\to\infty} \int_{\mathbb{R}} C_n(s,t)\,\mathrm{d}s = 1$. These facts are interpreted as the equality $C(s,t) = C(t-s) = \delta(t-s)$.

2.3. Exercises

2.1. Find $\mathbf{E}\displaystyle\int_0^t B_s\,ds, \quad \mathbf{E}\Big(\int_0^t B_s\,ds\Big)^2, \quad \mathbf{E}\Big(\int_0^t B_s\,ds\Big)^3, \quad t \geqslant 0.$

2.2. Find $\mathbf{E}B_tB_s, \quad \mathbf{E}B_t^2B_s^2, \quad t,s \geqslant 0; \quad \mathbf{E}B_tB_sB_u, \quad t \geqslant s \geqslant u \geqslant 0.$

2.3. Prove that $\mathbf{P}\{B_t \leqslant 0\ \forall\, t \in [0,T]\} = 0$ for every $T > 0$.

2.4. By the general Doob inequality (see Chapter 4), $\mathbf{E}\sup_{t\leqslant T} B_t^2 \leqslant 4T$. Show that, actually, $\mathbf{E}\sup_{t\leqslant T} B_t^2 \leqslant 2T$.

2.5. Give an example of a stochastic process X_t, $t \geqslant 0$, such that

1) $X_t \overset{d}{=\!=} B_t$ for all $t \geqslant 0$;

2) X_t is differentiable at all $t > 0$.

2.6. Can a trajectory of a Brownian motion be a Lipschitz function on some finite interval with positive probability? (Note that the non-differentiability of a function does not imply that it is Lipschitz.)
Reminder: A function $f : [a,b] \to \mathbb{R}$ is a Lipschitz function in an interval $[a,b]$ if there is a constant C such that $|f(t) - f(s)| \leqslant C|t-s|$ for all $t,s \in [a,b]$.

2.7. For $c > 0$, denote $W_t := \sqrt{c}B_{t/c}$, $t \geqslant 0$ (a scaled Brownian motion). Show that W is a Brownian motion.

2.8. Let B and $\tilde{B}$ be two independent Brownian motions. For $\rho \in [-1,1]$, denote

$$W_t := \rho B_t + \sqrt{1-\rho^2}\,\tilde{B}_t, \quad t \geqslant 0.$$

Show that W also is a Brownian motion and find $\operatorname{cov}(W_t, B_t)$.

2.9. The process $Z_t := B_t - tB_1$, $t \in [0,1]$, is called a Brownian bridge. Find $\operatorname{cov}(Z_t, Z_s)$, $t,s \in [0,1]$.

2.10. Let $\Delta^n = \{0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ such that $|\Delta^n| = \max_i |t_{i+1}^n - t_i^n| \to 0$ as $n \to \infty$. Let X and

Y be two *independent* Brownian motions. Show that

$$\sum_i \Delta X_i \Delta Y_i = \sum_{i=0}^{k_n-1} \big(X(t_{i+1}^n) - X(t_i^n)\big)\big(Y(t_{i+1}^n) - Y(t_i^n)\big) \to 0, \qquad n \to \infty,$$

in probability.

2.11. Let $\{\Delta^n\}$ be as in the previous problem. Let X and Y be two, not necessarily independent, Brownian motions. Find all possible limits in probability, as $n \to \infty$, of the sums

$$\sum_i \Delta X_i \Delta Y_i = \sum_{i=0}^{k_n-1} \big(X(t_{i+1}^n) - X(t_i^n)\big)\big(Y(t_{i+1}^n) - Y(t_i^n)\big).$$

Chapter 3

Stochastic Models with Brownian Motion and White Noise

In the previous chapter, we have learnt that a Brownian motion can be used as:

1) an approximation of a discrete-time process (for example, a process of sums of independent random variables) as space and time scales infinitely increase with an appropriate ratio;

2) a mathematical idealization of (the integral of) a high-intensity and short-memory continuous-time random process (noise).

These two cases of applications of Brownian motion can be essentially extended.

3.1. Discrete time

Discrete-time (non-random) sequences X_{t_n}, $n \in \mathbb{N}_+$, describing, for example, biological, physical, chemical or financial processes are often defined by recurrent equations:

$$X_{t_0} = x, \qquad X_{t_{n+1}} = X_{t_n} + b\big(X_{t_n}, t_n\big)\,\Delta t_n, \qquad \Delta t_n = t_{n+1} - t_n. \quad [3.1]$$

Consider, for example, the following simplest model describing the dynamics of the population growth:

$$X_{t_0} = x, \qquad X_{t_{n+1}} = X_{t_n} + \lambda X_{t_n}\,\Delta t_n, \quad [3.2]$$

where λ is the coefficient showing the reproduction intensity of population. In this model, we do not take into account the fact that a population cannot grow infinitely, and its reproduction intensity decreases for various reasons (limited food resources, diseases, predators, and the like). The following model is more realistic:

$$X_{t_0} = x, \qquad X_{t_{n+1}} = X_{t_n} + \lambda X_{t_n}\Big(1 - \frac{X_{t_n}}{M}\Big)\,\Delta t_n, \tag{3.3}$$

where M is a number showing a natural threshold that cannot be exceeded by a population since the reproduction intensity appropriately decreases as the population size approaches this threshold. The more we take into account the factors influencing the population, the more complicated the expression of coefficient b becomes. For example, to take into account the influence of predators, we have to introduce an increasing positive function h reflecting this influence (the larger the population, the more active the predators),[1] and then the equation may be as follows:

$$X_{t_0} = x, \qquad X_{t_{n+1}} = X_{t_n} + \Big[\lambda X_{t_n}\Big(1 - \frac{X_{t_n}}{M}\Big) - h(X_{t_n})\Big]\,\Delta t_n.$$

Other possible interpretations include: proliferation of tumor cells (with the immune system being the "predator" destroying these cells); the evolution of a bank account (with λ as the dividend intensity and taxes, deductions, and expenses as "predators").

In the first case, the solution of the equation is

$$X_{t_0} = x, \qquad X_{t_n} = x\prod_{k=0}^{n-1}(1 + \lambda\Delta t_k), \quad n \in \mathbb{N}.$$

In general, finding the solution is more complicated. Therefore, a continuous-time approximation of equation [3.1] is often used. It can be rewritten as

$$X_{t_{n+1}} = x + \sum_{k=0}^{n} b\big(X_{t_k}, t_k\big)\,\Delta t_k.$$

Note that, on the right-hand side of the equation, we have a Riemann integral sum of the function $b(X_t, t)$. If the time intervals Δt_k are small in comparison with the considered time interval (say, $[t_0, T] = [0, T]$), then, with a small error, the sequence $\{X_{t_n}\}$ can be replaced in this interval by the function X_t, $t \in [0, T]$, satisfying the integral equation

$$X_t = x + \int_0^t b(X_s, s)\,\mathrm{d}s, \quad t \in [0, T],$$

1. Hunters can also be interpreted as predators. Then the function h is not necessarily positive. For example, if the population significantly decreases until some critical value x_0, smart hunters may start to take care of the population. This means that $h(x) > 0$ for $x > x_0$ and $h(x) < 0$ for $x < x_0$.

or by the equivalent differential equation with initial value

$$X_t' = b(X_t, t), \qquad X_0 = x.$$

Solving such an approximating equation is usually easier than the initial discrete-time one. For example, for equation [3.2], the corresponding differential equation (which is called growth equation) is

$$X_t' = \lambda X_t, \qquad X_0 = x, \tag{3.2'}$$

which has the solution $X_t = x\mathrm{e}^{\lambda t}$, $t \geqslant 0$. The approximate equation for equation [3.3] is the Verhulst[2] (or logistic) equation

$$X_t' = \lambda X_t\Big(1 - \frac{X_t}{M}\Big), \quad X_0 = x; \tag{3.3'}$$

If $0 \leqslant x < M$, its solution is

$$X_t = \frac{Mx}{x + (M - x)\,\mathrm{e}^{-\lambda t}}, \quad t \geqslant 0.$$

Though, historically, the Verhulst equation was introduced to describe population growth, it further appeared in various other areas, such as physics, chemistry (protein dynamics), genetics, etc.

In reality, such a deterministic model describes only the average behavior of a phenomenon and does not provide its detailed image since it does not take into account random perturbations. One often obtains rather realistic models assuming that the perturbations that influence physical, biological, chemical, financial systems are:

– short in time;

– independent;

– zero-mean;

– have an intensity ($=$ second moment) proportional to their action time.

In view of these assumptions, in the simplest case the perturbed equation [3.1] is as follows:

$$X_{t_0} = x, \qquad X_{t_{n+1}} = X_{t_n} + b\big(X_{t_n}, t_n\big)\,\Delta t_n + \xi_n,$$

where ξ_n, $n \in \mathbb{N}_+$, are independent identically distributed (most often normal) random variables with $\mathbf{E}\xi_n = 0$ and $\mathbf{E}\xi_n^2 = C\Delta t_n$. Denoting $\eta_{t_n} = C^{-1/2}\xi_n/\Delta t_n$, we have

$$X_{t_0} = x, \qquad X_{t_{n+1}} = X_{t_n} + b\big(X_{t_n}, t_n\big)\,\Delta t_n + \sqrt{C}\eta_{t_n}\Delta t_n,$$

2. Pierre François Verhulst.

where $\mathbf{E}\eta_{t_n} = 0$, $\mathbf{E}\eta_{t_n}^2 = 1/\Delta t_n$. We rewrite this equation in the form

$$X_{t_{n+1}} = x + \sum_{i=0}^{n} b\big(X_{t_i}, t_i\big)\,\Delta t_i + \sqrt{C}\sum_{i=0}^{n}\eta_{t_i}\,\Delta t_i.$$

As in the deterministic case, if the time intervals Δt_k are small in comparison with the time interval considered, then the sequence $\{X_{t_n}\}$ may be replaced in this interval by (this time random) function X_t, $t \in [0, T]$, satisfying the integral equation

$$X_t = x + \int_0^t b(X_s, s)\,\mathrm{d}s + \sqrt{C}\int_0^t \eta_s\,\mathrm{d}s,$$

where $\eta_s = \eta_{t_i}$ for $s \in [t_i, t_{i+1})$. Since $\mathbf{E}(\eta_{t_i}\Delta t_i) = 0$ and $\mathbf{E}(\eta_{t_i}\Delta t_i)^2 = \Delta t_i$, for small Δt_i, we get[3] $\int_0^t \eta_s\,\mathrm{d}s \approx \sum_{t_i<t}\eta_{t_i}\Delta t_i \xrightarrow{w} B_t$, i.e. η is a process close to a white noise. Therefore, it is natural to replace the second integral by a Brownian motion B_t. We get the integral equation

$$X_t = x + \int_0^t b(X_s, s)\,\mathrm{d}s + \sqrt{C}B_t. \qquad [3.4a]$$

This equation cannot be replaced by a differential equation, since the trajectories of a Brownian motion are nowhere differentiable (Theorem 2.4). However, this equation makes sense as a Volterra-type[4] equation, which has a unique solution for every (continuous) trajectory B_t, $t \geqslant 0$, of Brownian motion, provided that the coefficient b satisfies the Lipschitz condition.

In a more general case, the intensity of noise is proportional to a quantity depending on the state of the system at every time moment:

$$X_{t_{n+1}} = x + \sum_{i=0}^{n} b\big(X_{t_i}, t_i\big)\,\Delta t_i + \sum_{i=0}^{n}\sigma\big(X_{t_i}, t_i\big)\eta_{t_i}\,\Delta t_i.$$

By similar arguments, from this equation we get the integral equation

$$X_t = x + \int_0^t b(X_s, s)\,\mathrm{d}s + \int_0^t \sigma(X_s, s)\eta_s\,\mathrm{d}s.$$

This time we can only formally replace the differential $\eta_s\,\mathrm{d}s$ by the differential of the Brownian motions $\mathrm{d}B_s$:

$$X_t = x + \int_0^t b(X_s, s)\,\mathrm{d}s + \int_0^t \sigma(X_s, s)\,\mathrm{d}B_s. \qquad [3.4m]$$

3. See section 2.2.

4. Vito Volterra.

Again, recalling the non-differentiability of trajectories of Brownian motion, we have unfortunately to realize that the second integral does not exist in the usual (Stieltjes) sense. Thus, we cannot give any sense to this equation by means of usual “deterministic” integrals. Fortunately, there is a way to define the integrals $\int_0^t Y_s \,\mathrm{d}B_s$ of random processes with respect to Brownian motion and, at the same time, to make sense of equation [3.4m]. However, the situation is a little more complicated, since, as we will see, the stochastic integral can be defined in various ways. Moreover, it is not always clear which known definition—Itô or Stratonovich—is to be used. Note that here there is no analogy with Riemann and Lebesgue integrals that coincide when they both exist. The Itô and Stratonovich integrals are different even if they both exist! After rather long discussions in the mathematical and physical literature on the question of which integral is better, a general consensus was reached that both integrals are needed, with their pluses and minuses, and both have their “spheres of influence”. Usually, the Itô integral is used when discrete-time models are approximated by continuous processes, while the Stratonovich integral is used when continuous-time processes influenced by perturbations close to white noise are considered.

Equation [3.4m] is often formally written in the differential form

$$\mathrm{d}X_t = b(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t, \qquad X_0 = x,$$

and called a *stochastic differential equation* (SDE), though, in fact, it is an *integral* equation. This can be partially explained by tradition and history. However, the main reason is possibly that, in various applications, heuristic arguments and intuition often lead to the differential form of an equation.

3.2. Continuous time

The same equations can be obtained while considering some continuous-time models. Let us start with the differential equation

$$X_t' = b(X_t, t), \qquad X_0 = x,$$

which is equivalent to the integral equation

$$X_t = x + \int_0^t b(X_s, s)\,\mathrm{d}s.$$

Such equations are often used to describe the motion of a particle in a force field, or the dynamics of a signal in an electrical circuit. In reality, the motion of a particle may be influenced by chaotic “kicks” of much smaller particles, e.g. molecules. Although in comparison with the influence of a force field these kicks are very irregular, they partially compensate each other. In an electric field, for various reasons, noises appear that perturb the pure signal. In another financial market interpretation regular changes (dividends) in the values of assets are perturbed by “noises”, random and relatively chaotic fluctuations of stock prices. Adding the noise (denoted by η_t), multiplied by a proportionality coefficient C, to the right-hand side of the differential equation, we obtain the equation

$$X_t' = b(X_t, t) + C\eta_t, \qquad X_0 = x.$$

If the proportionality coefficient depends on the state of the a solution, we obtain a more general equation

$$X_t' = b(X_t, t) + \sigma(X_t, t)\eta_t, \qquad X_0 = x.$$

If the noise η_t is close to a white noise, we again get equations [3.4a] and [3.4m].[5]

Another way to get a stochastic version of an ordinary differential equation is randomization of a parameter (or several parameters) assuming that the latter is perturbed by a random process proportional to white noise (more precisely, to a process which is close to white noise). Suppose that we have an equation describing some phenomenon,

$$X_t' = b_\lambda(X_t, t) = b(X_t, t) + \lambda h(X_t, t), \qquad X_0 = x,$$

with the right-hand side depending on a real parameter λ corresponding to the averaged state of a system. Assuming that parameter λ is perturbed by a process proportional to a white noise, we can replace it by the perturbed process $\lambda_t = \lambda + \sigma\eta_t$, $t \geqslant 0$. We get the equation

$$X_t' = b(X_t, t) + \lambda h(X_t, t) + \sigma h(X_t, t)\eta_t, \qquad X_0 = x,$$

or

$$X_t' = b_\lambda(X_t, t) + \tilde{\sigma}(X_t, t)\eta_t, \qquad X_0 = x.$$

Passing to the integral notation, we get the equation

$$X_t = x + \int_0^t b_\lambda(X_s, s)\,\mathrm{d}s + \int_0^t \tilde{\sigma}(X_s, s)\,\mathrm{d}B_s.$$

Let us consider several examples.

---

5. The letters a and m are not accidental. Equations [3.4a] and [3.4m] are said to be equations with *additive* and *multiplicative* noise, respectively.

Growth equation. The above-mentioned equation $X_t' = \lambda X_t$ with the initial condition $X_0 = x$ depends on the intensity parameter λ. By randomizing it we get the stochastic equation

$$X_t = x + \lambda \int_0^t X_s \, \mathrm{d}s + \sigma \int_0^t X_s \, \mathrm{d}B_s.$$

It is interesting that this equation describes not only a simple (and, therefore, not very realistic) model of population growth under perturbations but also a rather popular model of fluctuations in stock price.

Verhulst equation. Consider equation [3.3′], taking, for simplicity, $M = 1$, that is

$$X_t' = \lambda X_t(1 - X_t), \qquad X_0 = x.$$

Randomizing the parameter λ, we get the stochastic differential equation

$$X_t = x + \lambda \int_0^t X_s(1 - X_s) \, \mathrm{d}s + \sigma \int_0^t X_s(1 - X_s) \, \mathrm{d}B_s.$$

Taking $M = \lambda$, we get the equation $X_t' = \lambda X_t - X_t^2$, and the randomization of the parameter λ results in the stochastic differential equation

$$X_t = x + \int_0^t \big(\lambda X_s - X_s^2\big) \, \mathrm{d}s + \sigma \int_0^t X_s \, \mathrm{d}B_s.$$

A generalization of the Verhulst equation, the equation

$$X_t' = a - X_t + \lambda X_t(1 - X_t), \qquad X_0 = x,$$

is used in genetics and chemistry. Its stochastic version (called the stochastic genetic model) obtained by randomization of the parameter λ is

$$X_t = x + \int_0^t \big(a - X_s + \lambda X_s(1 - X_s)\big) \, \mathrm{d}s + \sigma \int_0^t X_s(1 - X_s) \, \mathrm{d}B_s.$$

Chapter 4

Stochastic Integral with Respect to Brownian Motion

4.1. Preliminaries. Stochastic integral with respect to a step process

In the previous chapter, we discussed that in order to make sense of stochastic differential equations, it is important to define the integrals $\int_0^t Y_s \, \mathrm{d}B_s$, where the integrand function Y is random, and the integrating function (also random) is a Brownian motion B.

Before defining the stochastic integral, recall the definition of the Stieltjes integral. Roughly speaking, the Stieltjes (or Riemann–Stieltjes) integral of a function f in the interval $[a, b]$ with respect to a function g is the limit

$$\int_a^b f(t) \, \mathrm{d}g(t) = \lim_{\max_i |\Delta t_i| \to 0} \sum_i f(\xi_i)\big(g(t_{i+1}) - g(t_i)\big),$$

where $a = t_0 < t_1 < \cdots < t_k = b$ is a partition of the interval $[a, b]$, and $\xi = \{\xi_i\}$ is its intermediate partition, i.e. $\xi_i \in [t_i, t_{i+1}]$, $\Delta t_i := t_{i+1} - t_i$.

$\mathcal{M}$ To rigorously define the Stieltjes integral $\int_a^b f(t) \, \mathrm{d}g(t)$, consider any sequence $\Delta^n = \{a = t_0^n < t_1^n < t_2^n < \cdots < t_{k_n}^n = b\}$, $n \in \mathbb{N}$, of partitions of the interval $[a, b]$ such that $|\Delta^n| := \max_i \Delta t_i^n = \max_i |t_{i+1}^n - t_i^n| \to 0$, and any sequence of intermediate partitions $\xi^n = \{\xi_i^n \in [t_i^n, t_{i+1}^n],\ i = 0, 1, \ldots, k_n - 1\}$, $n \in \mathbb{N}$. The Stieltjes integral of a function f in the interval $[a, b]$ with respect to a function g is the limit

$$\int_a^b f(t)\,\mathrm{d}g(t) := \lim_{n\to\infty} \sum_{i=0}^{k_n-1} f(\xi_i^n)\big(g(t_{i+1}^n) - g(t_i^n)\big),$$

provided that this limit exists and does not depend on the choice of the sequences of partitions Δ^n, $n \in \mathbb{N}$, and $\{\xi^n\}$, $n \in \mathbb{N}$.

For the existence of such a integral, it suffices that $f \in C[a,b]$ and

$$V(g) = V(g;a,b) := \sup\left\{\sum_i \big|g(t_{i+1}) - g(t_i)\big|\right\} < +\infty,$$

where the supremum is taken over all partitions $a = t_0 < t_1 < t_2 < \cdots < t_k = b$, $k \in \mathbb{N}$, of the interval $[a,b]$. The number $V(g)$ is called the *variation* of function g. If $V(g) < +\infty$, then g is said to be a *finite-variation function*. If $g \in C^1[a,b]$, then $V(g) = \int_a^b |g'(t)|\,\mathrm{d}t$, and $\int_a^b f(t)\,\mathrm{d}g(t) = \int_a^b f(t)g'(t)\,\mathrm{d}t$.[1] Thus, if $f \in C[a,b]$ and g has a continuous derivative, then the Stieltjes integral $\int_a^b f(t)\,\mathrm{d}g(t)$ exists, and the differential $\mathrm{d}g(t)$ may be replaced by its formal expression $g'(t)\,\mathrm{d}t$. However, this is unacceptable for us, recalling that, unfortunately, the trajectories of Brownian motion are nowhere differentiable. Moreover, the finite-variation functions are the only functions that integrate *all* continuous functions. More precisely, if the integral $\int_a^b f(t)\,\mathrm{d}g(t)$ exists for all $f \in C[a,b]$, then g is a finite-variation function! Unfortunately, trajectories of Brownian motion have finite variation in every interval (Corollary 4.2). Nevertheless, Brownian motion has so-called quadratic variation.

THEOREM 4.1.– *Let B_t, $t \geqslant 0$, be a Brownian motion. Let*

$$\Delta^n = \{s = t_0^n < t_1^n < \cdots < t_{k_n}^n = t\}, \quad n \in \mathbb{N},$$

be a sequence of partitions of the interval $[s,t]$ such that $|\Delta^n| := \max_i \Delta t_i^n = \max_i |t_{i+1}^n - t_i^n| \to 0$, $n \to \infty$. Then

$$\sum_{i=0}^{k_n-1} \big(\Delta B_i^n\big)^2 = \sum_{i=0}^{k_n-1} \big(B(t_{i+1}^n) - B(t_i^n)\big)^2 \xrightarrow{L^2} t - s, \quad n \to \infty.$$

REMARKS.– 1. In view of this theorem, a Brownian motion is said to have, in every interval $[s,t]$, the *quadratic variation* equal to $t - s$. This fact is often symbolically written as $(\mathrm{d}B_t)^2 = \mathrm{d}t$.

---

1. Or, a little more generally: if $g(t) = g(a) + \int_a^t h(s)\,\mathrm{d}s$ and h is an integrable function in $[a,b]$, then $V(g) = \int_a^b |h(t)|\,\mathrm{d}t$, and $\int_a^b f(t)\,\mathrm{d}g(t) = \int_a^b f(t)h(t)\,\mathrm{d}t$.

2. If $|\Delta^n| \to 0$ sufficiently "rapidly", then convergence in the theorem is also almost sure.

Proof. Omitting the indices n (though keeping them in mind), we denote:

$$\Delta B_i = B(t_{i+1}^n) - B(t_i^n), \qquad \Delta t_i = t_{i+1}^n - t_i^n.$$

Then

$$\begin{aligned}\varepsilon_n &:= \mathbf{E}\Big(\sum_i \Delta B_i^2 - (t-s)\Big)^2 = \mathbf{E}\Big(\sum_i \Delta B_i^2 - \sum_i \Delta t_i\Big)^2 \\ &= \mathbf{E}\Big(\sum_i \big(\Delta B_i^2 - \Delta t_i\big)\Big)^2.\end{aligned}$$

Since the increments of Brownian motion ΔB_i are independent, the random variables $\Delta B_i^2 - \Delta t_i$, as their increments, are also independent. Using this and the equality $\mathbf{E}(\Delta B_i^2 - \Delta t_i) = 0$, we have

$$\begin{aligned}\varepsilon_n &= \mathbf{D}\Big(\sum_i \big(\Delta B_i^2 - \Delta t_i\big)\Big) = \sum_i \mathbf{D}\big(\Delta B_i^2 - \Delta t_i\big) \\ &= \sum_i \mathbf{E}\big(\Delta B_i^2 - \Delta t_i\big)^2 = \sum_i \big(\mathbf{E}\Delta B_i^4 - 2\Delta t_i \mathbf{E}\Delta B_i^2 + \Delta t_i^2\big) \\ &= \sum_i \big(3\Delta t_i^2 - 2\Delta t_i^2 + \Delta t_i^2\big) = 2\sum_i \Delta t_i^2 \\ &\leqslant 2|\Delta^n| \sum_i \Delta t_i = 2(t-s)|\Delta^n| \to 0.\end{aligned}$$

△

COROLLARY 4.2.– *Brownian motion has infinite variation in every interval $[s,t]$ almost surely.*

Proof. Using the same notation, we have

$$\sum_i |\Delta B_i| = \sum_i \frac{\Delta B_i^2}{|\Delta B_i|} \geqslant \sum_i \frac{\Delta B_i^2}{\max\limits_j |\Delta B_j|} = \frac{1}{\max\limits_j |\Delta B_j|} \sum_i \Delta B_i^2.$$

By Theorem 4.1, $\sum_i \Delta B_i^2 \xrightarrow{L^2} t-s$, and hence $\sum_i \Delta B_i^2 \xrightarrow{\mathbf{P}} t-s$ as $n \to \infty$. Passing, if necessary, to an appropriate subsequence of the sequence of partitions $\{\Delta^n\}$, we may assume that $\sum_i \Delta B_i^2 \to t-s$ as $n \to \infty$ almost surely.

Since the trajectories of Brownian motion are uniformly continuous functions on $[s,t]$, $\max_j |\Delta B_j| \to 0$ (almost surely). Therefore,

$$\sum_i |\Delta B_i| = \sum_i \left|B(t_{i+1}^n) - B(t_i^n)\right| \to \infty, \quad n \to \infty,$$

and thus $V(B;s,t) = \sup\{\sum_i |\Delta B_i|\} = +\infty$ (where the supremum is taken over all partitions of $[s,t]$). △

DEFINITION 4.3.– *A random process $X = \{X_t,\, t \in [0,T]\}$ is called adapted*[2] *(with respect to Brownian motion B) in the interval $[0,T]$ if $X_t \in \mathcal{H}_t = \mathcal{H}_t^B$ for all $t \in [0,T]$.*

We denote by $H^2[0,T]$ the set of adapted processes X in the interval $[0,T]$ such that

$$\|X\| = \|X\|_{H^2} := \left(\mathbf{E}\int_0^T X_t^2\,\mathrm{d}t\right)^{1/2} < +\infty.$$

We shall say that a sequence $\{X^n\}$ of random processes converges to a random process X in the space $H^2[0,T]$ and write $X^n \xrightarrow{H^2} X$ if $\|X^n - X\| \to 0$ as $n \to \infty$.

The function $\|\cdot\|$ has the usual properties of a norm:

1. $\|\alpha X\| = |\alpha|\,\|X\|$, $\alpha \in \mathbb{R}$;
2. $\|X + Y\| \leqslant \|X\| + \|Y\|$;

2′. $\big|\|X\| - \|Y\|\big| \leqslant \|X - Y\|$.

$\mathcal{M}$ The set $H^2[0,T]$ is a normed and even a Banach space with norm $\|\cdot\|$ if we consider two random processes $X, \widetilde{X} \in H^2[0,T]$ coinciding if $\|X - \widetilde{X}\| = 0$. It is also a Hilbert space with scalar product $(X,Y) := \mathbf{E}\int_0^T X_t Y_t\,\mathrm{d}t$.

DEFINITION 4.4.– *An adapted random process $X = \{X_t,\, t \in [0,T]\}$ is called a step process if there is a partition $0 = t_0 < t_1 < \cdots < t_k = T$ of the interval $[0,T]$ such that*

$$X_t = X_{t_i}, \quad \textit{for } t \in [t_i, t_{i+1}),\ i = 0,1,2,\dots,k-1,$$

or $X = \sum_{i=0}^{k-1} X_{t_i} 1\!\!1_{[t_i,t_{i+1})}$ (i.e. $X_t(\omega) = \sum_{i=0}^{k-1} X_{t_i}(\omega) 1\!\!1_{[t_i,t_{i+1})}(t)$). The stochastic integral s(or Itô[3] *integral) with respect to a Brownian motion B in the interval $[0,T]$*

2. Or *non-anticipating.*

3. Kyoshi Itô.

is the sum

$$\int_0^T X_t \, \mathrm{d}B_t := \sum_{i=0}^{k-1} X_{t_i} \big(B_{t_{i+1}} - B_{t_i} \big).$$

We denote by $S^2[0,T]$ the class of all step processes belonging to $H^2[0,T]$ and by $S_b[0,T]$ all bounded step processes.

The stochastic integral of a process $X \in S[0,T]$ with respect to a Brownian motion in a subinterval $[T_1, T_2] \subset [0,T]$ is defined as

$$\int_{T_1}^{T_2} X_t \, \mathrm{d}B_t := \int_0^T X_t \mathbb{1}_{[T_1,T_2]}(t) \, \mathrm{d}B_t.$$

REMARKS.– 1. To define the stochastic integral in an interval $[T_1, T_2] \subset [0,T]$, we may also proceed as follows. Without loss of generality, suppose that the endpoints of $[T_1, T_2]$ are also endpoints of constancy intervals of the process $X \in S[0,T]$, say, $t_n = T_1$ and $t_m = T_2$. Then

$$\int_{T_1}^{T_2} X_t \, \mathrm{d}B_t := \sum_{i=n}^{m-1} X_{t_i} \big(B_{t_{i+1}} - B_{t_i} \big).$$

Such a definition can be also used when the step process is defined in the time interval $[T_1, T_2]$ only.

2. In the definition of the stochastic integral of a step process, we do not formally need the process to be adapted. However, without this assumption, the basic properties of the stochastic integrals proved below would not hold!

PROPOSITION 4.5.– *The stochastic integral in the class $S_b = S_b[0,T]$ possesses the following properties:*

1) $X, Y \in S_b$, $\alpha, \beta \in \mathbb{R} \implies \int_0^T (\alpha X + \beta Y)_t \, \mathrm{d}B_t = \alpha \int_0^T X_t \, \mathrm{d}B_t + \beta \int_0^T Y_t \, \mathrm{d}B_t$.

2) $\mathbf{E} \int_0^T X_t \, \mathrm{d}B_t = 0$.

2'). If $Z \in \mathcal{H}_s$ is a bounded random variable, then $Z \int_s^T X_t \, \mathrm{d}B_t = \int_s^T Z X_t \, \mathrm{d}B_t$, and, in particular, $\mathbf{E}(Z \int_s^T X_t \, \mathrm{d}B_t) = 0$.

3) $\mathbf{E}(\int_0^T X_t \,\mathrm{d}B_t)^2 = \mathbf{E}\int_0^T X_t^2 \,\mathrm{d}t$, *i.e.,* $\|\int_0^T X_t \,\mathrm{d}B_t\|_{L^2} = \|X\|_{H^2}$. *In other words, the mapping associating a step process X with its stochastic integral preserves the norm.*[4]

4) $\mathbf{E}(\int_0^T X_t \,\mathrm{d}B_t \int_0^T Y_t \,\mathrm{d}B_t) = \mathbf{E}\int_0^T X_t Y_t \,\mathrm{d}t$.

5) $\mathbf{E}(\int_0^T X_t \,\mathrm{d}B_t)^4 \leqslant 36\,\mathbf{E}(\int_0^T X_t^2 \,\mathrm{d}t)^2$.

Proof. 1) First, note that, without loss of generality, we may assume that the step processes X and Y are constant on the same intervals of constancy. Otherwise, joining the partitions corresponding to the processes X and Y would give a new partition with common intervals of constancy for both processes. Thus, let us take a partition $0 = t_0 < t_1 < \cdots < t_k = T$ of $[0, T]$ such that

$$X_t = X_{t_i} \quad \text{and} \quad Y_t = Y_{t_i} \quad \text{for } t \in [t_i, t_{i+1}),\ i = 0, 1, 2, \ldots, k-1.$$

Then

$$\begin{aligned}
\int_0^T (\alpha X + \beta Y)_t \,\mathrm{d}B_t &= \sum_{i=0}^{k-1} (\alpha X + \beta Y)_{t_i} \big(B_{t_{i+1}} - B_{t_i}\big) \\
&= \sum_{i=0}^{k-1} \big(\alpha X_{t_i} + \beta Y_{t_i}\big)\big(B_{t_{i+1}} - B_{t_i}\big) \\
&= \alpha \sum_{i=0}^{k-1} X_{t_i} \big(B_{t_{i+1}} - B_{t_i}\big) + \beta \sum_{i=0}^{k-1} Y_{t_i} \big(B_{t_{i+1}} - B_{t_i}\big) \\
&= \alpha \int_0^T X_t \,\mathrm{d}B_t + \beta \int_0^T Y_t \,\mathrm{d}B_t.
\end{aligned}$$

2) Let the step process X be as in Definition 4.4. Denote $\Delta B_i = B_{t_{i+1}} - B_{t_i}$. Since $X_{t_i} \in \mathcal{H}_{t_i}$, we have $X_{t_i} \perp\!\!\!\perp \Delta B_i$. Therefore,

$$\begin{aligned}
\mathbf{E}\int_0^T X_t \,\mathrm{d}B_t &= \mathbf{E}\sum_i X_{t_i} \Delta B_i = \sum_i \mathbf{E}\big(X_{t_i} \Delta B_i\big) \\
&= \sum_i \mathbf{E}\big(X_{t_i}\big)\mathbf{E}(\Delta B_i) = 0,
\end{aligned}$$

4. Such mappings are called isometries. As we will see in section 4.2, this is the main property that will enable us to extend the stochastic integral to the processes from $H^2[0, T]$.

where the last equality follows from $\mathbf{E}(\Delta B_i) = 0$.

2′) Suppose that $t_n = s$ and $t_m = T$ are the ends of constancy intervals of the step process X. Then

$$Z\int_s^T X_t \,\mathrm{d}B_t = Z\sum_{i=n}^{m-1} X_{t_i}\,\Delta B_i = \sum_{i=n}^{m-1} ZX_{t_i}\,\Delta B_i.$$

Note that if $X \in S_b$ and if $Z \in \mathcal{H}_s$ is bounded, then the random process $\widetilde{X}_t := ZX_t 1\!\!1_{[s,T]}(t)$, $t \in [0,T]$, is also a bounded (adapted) step process. Therefore, the last sum can be written as a stochastic integral of this process, $\int_0^T ZX_t 1\!\!1_{[s,T]}(t)\,\mathrm{d}B_t = \int_s^T ZX_t\,\mathrm{d}B_t$. Thus,

$$Z\int_s^T X_t\,\mathrm{d}B_t = \int_s^T ZX_t\,\mathrm{d}B_t,$$

and it remains to apply Property 2 to this integral.

3) We have

$$\begin{aligned}
\mathbf{E}\bigg(\int_0^T X_t\,\mathrm{d}B_t\bigg)^2 &= \mathbf{E}\bigg(\sum_i X_{t_i}\Delta B_i\bigg)^2 = \mathbf{E}\bigg(\sum_i X_{t_i}\Delta B_i \cdot \sum_j X_{t_j}\Delta B_j\bigg)\\
&= \mathbf{E}\sum_{i,j}\big(X_{t_i}X_{t_j}\Delta B_i \Delta B_j\big)\\
&= \mathbf{E}\sum_i \big(X_{t_i}^2 \Delta B_i^2\big) + 2\mathbf{E}\sum_{i<j}\big(X_{t_i}X_{t_j}\Delta B_i\Delta B_j\big)\\
&= \sum_i \mathbf{E}\big(X_{t_i}^2\Delta B_i^2\big) + 2\sum_{i<j}\mathbf{E}\big(X_{t_i}X_{t_j}\Delta B_i\Delta B_j\big).
\end{aligned}$$

Since $X_{t_i} \perp\!\!\!\perp \Delta B_i$, we also have $X_{t_i}^2 \perp\!\!\!\perp \Delta B_i^2$. Moreover, $X_{t_i}X_{t_j}\Delta B_i \in \mathcal{H}_{t_j}$ for $i < j$ (i.e. $t_i < t_{i+1} \leqslant t_j$), and therefore $X_{t_i}X_{t_j}\Delta B_i \perp\!\!\!\perp \Delta B_j$ for $i < j$. Thus, continuing, we get:

$$\begin{aligned}
\mathbf{E}\Big(\int_0^T X_t\,\mathrm{d}B_t\Big)^2 &= \sum_i \big(\mathbf{E}X_{t_i}^2\mathbf{E}\Delta B_i^2\big) + 2\sum_{i<j}\Big(\mathbf{E}\big(X_{t_i}X_{t_j}\Delta B_i\big)\mathbf{E}\Delta B_j\Big)\\
&= \sum_i \mathbf{E}X_{t_i}^2\Delta t_i + 0 = \mathbf{E}\sum_i X_{t_i}^2\Delta t_i = \mathbf{E}\int_0^T X_t^2\,\mathrm{d}t.
\end{aligned}$$

4) Using twice the elementary identity $ab = ((a+b)^2 - (a-b)^2)/4$ and Properties 3 and 1, we have

$$
\begin{aligned}
&\mathbf{E}\bigg(\int\limits_0^T X_t \,\mathrm{d}B_t \int\limits_0^T Y_t \,\mathrm{d}B_t\bigg) \\
&\quad = \mathbf{E}\frac{\big(\int_0^T X_t \,\mathrm{d}B_t + \int_0^T Y_t \,\mathrm{d}B_t\big)^2 - \big(\int_0^T X_t \,\mathrm{d}B_t - \int_0^T Y_t \,\mathrm{d}B_t\big)^2}{4} \\
&\quad = \frac{\mathbf{E}\big(\int_0^T (X_t + Y_t) \,\mathrm{d}B_t\big)^2 - \mathbf{E}\big(\int_0^T (X_t - Y_t) \,\mathrm{d}B_t\big)^2}{4} \\
&\quad = \frac{\mathbf{E}\int_0^T (X_t + Y_t)^2 \,\mathrm{d}t - \mathbf{E}\int_0^T (X_t - Y_t)^2 \,\mathrm{d}t}{4} \\
&\quad = \mathbf{E}\int\limits_0^T \frac{(X_t + Y_t)^2 - (X_t - Y_t)^2}{4} \,\mathrm{d}t = \mathbf{E}\int\limits_0^T X_t Y_t \,\mathrm{d}t.
\end{aligned}
$$

5. Usually, this inequality is proved by using martingale theory. Here, we prefer to give a direct proof, similar to that of Property 3. In the latter, we wrote the square of the integral $\int_0^T X_t \,\mathrm{d}B_t$ as the product of *two* identical sums, then multiplied the summands one-by-one and, finally, after some grouping, used the independence of multiplicands. Let us try to consider the fourth power of the integral $\int_0^T X_t \,\mathrm{d}B_t$, writing it as the product of *four* identical sums.[5] Writing, for short, X_i instead of X_{t_i}, we have

$$
\begin{aligned}
\mathbf{E}\bigg(\int\limits_0^T X_t \,\mathrm{d}B_t\bigg)^4 &= \mathbf{E}\Big(\sum_i X_i \Delta B_i \cdot \sum_j X_j \Delta B_j \cdot \sum_k X_k \Delta B_k \cdot \sum_l X_l \Delta B_l\Big) \\
&= \mathbf{E}\sum_{i,j,k,l} X_i X_j X_k X_l \Delta B_i \Delta B_j \Delta B_k \Delta B_l \\
&= \sum_{i,j,k,l} \mathbf{E}(X_i X_j X_k X_l \Delta B_i \Delta B_j \Delta B_k \Delta B_l).
\end{aligned}
$$

Let us group the summands according to how many coinciding largest indices there are in a summand. Their number can vary from 1 to 4. Also, note that the terms with the indices (i, j, k, l) that differ only by their order (as, for example, the terms with

5. The idea was provided by the author, and the proof by Kęstutis Gadeikis.

indices $(1,1,3,5)$ and $(1,5,3,1)$) coincide. If the largest index is one, it can be in one of four possible positions. If there are two such indices, they can be placed among four positions in $\binom{4}{2} = 6$ ways; three coinciding maximal indices in $\binom{4}{3} = 4$ ways; and, finally, when all indices coincide, they are clearly placed in a unique way. Thus, we continue the equality:

$$\begin{aligned}
&\mathbf{E}\left(\int_0^T X_t \,\mathrm{d}B_t\right)^4 \\
&= \left(4\sum_{i,j,k<l} + 6\sum_{i,j<k=l} + 4\sum_{i<j=k=l} + \sum_{i=j=k=l}\right)\mathbf{E}(X_iX_jX_kX_l\Delta B_i\Delta B_j\Delta B_k\Delta B_l) \\
&\quad = 4\sum_{i,j,k<l}\mathbf{E}\big(X_iX_jX_kX_l\Delta B_i\Delta B_j\Delta B_k\big)\mathbf{E}\big(\Delta B_l\big) \\
&\qquad + 6\sum_{i,j<l}\mathbf{E}\big(X_iX_jX_l^2\Delta B_i\Delta B_j\big)\mathbf{E}\big(\Delta B_l^2\big) \\
&\qquad + 4\sum_{i<l}\mathbf{E}\big(X_iX_l^3\Delta B_i\big)\mathbf{E}\big(\Delta B_l^3\big) + \sum_l \mathbf{E}\big(X_l^4\big)\mathbf{E}\big(\Delta B_l^4\big).
\end{aligned}$$

Now, applying the equalities $\mathbf{E}\Delta B_l = \mathbf{E}(\Delta B_l)^3 = 0$, $\mathbf{E}(\Delta B_l)^2 = \Delta t_l$, and $\mathbf{E}(\Delta B_l)^4 = 3\Delta t_l^2$, we get:

$$\mathbf{E}\left(\int_0^T X_t \,\mathrm{d}B_t\right)^4 = 6\sum_{i,j<l}\mathbf{E}\big(X_iX_jX_l^2\Delta B_i\Delta B_j\big)\Delta t_l + 3\sum_l \mathbf{E}\big(X_l^4\big)\Delta t_l^2. \quad [4.1]$$

Let us rewrite the first sum by complementing it until the sum over all i, j, l:

$$\begin{aligned}
&\sum_{i,j<l}\mathbf{E}\big(X_iX_jX_l^2\Delta B_i\Delta B_j\big)\Delta t_l \\
&\quad = \left(\sum_{i,j,l} - 2\sum_{i<j=l} - \sum_{i=j=l} - \sum_{i=j>l} - 2\sum_{i>j,l}\right)\mathbf{E}\big(X_iX_jX_l^2\Delta B_i\Delta B_j\big)\Delta t_l \\
&\quad = \mathbf{E}\left(\sum_i X_i\Delta B_i \sum_j X_j\Delta B_j \sum_l X_l^2\Delta t_l\right) \\
&\qquad - 2\sum_{i<l}\mathbf{E}\big(X_i\Delta B_iX_l^3\big)\mathbf{E}(\Delta B_l)\Delta t_l \\
&\qquad - \sum_l \mathbf{E}\big(X_l^4\big)\mathbf{E}\big(\Delta B_l^2\big)\Delta t_l - \sum_{i>l}\mathbf{E}\big(X_i^2X_l^2\big)\mathbf{E}\big(\Delta B_i^2\big)\Delta t_l \\
&\qquad - 2\sum_{i>j,l}\mathbf{E}\big(X_iX_jX_l^2\Delta B_j\big)\mathbf{E}(\Delta B_i)\Delta t_l
\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{E}\left(\int_0^T X_t \,\mathrm{d}B_t \int_0^T X_t \,\mathrm{d}B_t \int_0^T X_t^2 \,\mathrm{d}t\right) - 2\cdot 0 \\
&\quad - \mathbf{E}\sum_l X_l^4 \Delta t_l^2 - \mathbf{E}\sum_{i>l} X_i^2 X_l^2 \Delta t_i \Delta t_l - 2\cdot 0 \\
&= \mathbf{E}\left[\left(\int_0^T X_t \,\mathrm{d}B_t\right)^2 \int_0^T X_t^2 \,\mathrm{d}t\right] \\
&\quad - \frac{1}{2}\mathbf{E}\sum_l X_l^4 \Delta t_l^2 - \frac{1}{2}\left(\mathbf{E}\sum_l X_l^4 \Delta t_l^2 + \mathbf{E}\sum_{i\neq l} X_i^2 X_l^2 \Delta t_i \Delta t_l\right) \\
&= \mathbf{E}\left[\left(\int_0^T X_t \,\mathrm{d}B_t\right)^2 \int_0^T X_t^2 \,\mathrm{d}t\right] \\
&\quad - \frac{1}{2}\mathbf{E}\sum_l X_l^4 \Delta t_l^2 - \frac{1}{2}\left(\mathbf{E}\sum_{i,l} X_i^2 X_l^2 \Delta t_i \Delta t_l\right) \\
&= \mathbf{E}\left[\left(\int_0^T X_t \,\mathrm{d}B_t\right)^2 \int_0^T X_t^2 \,\mathrm{d}t\right] - \frac{1}{2}\mathbf{E}\sum_l X_l^4 \Delta t_l^2 - \frac{1}{2}\mathbf{E}\left(\int_0^T X_t^2 \,\mathrm{d}t\right)^2.
\end{aligned}
$$

Substituting the equality obtained into equation [4.1], we get:

$$
\begin{aligned}
\mathbf{E}\left(\int_0^T X_t \,\mathrm{d}B_t\right)^4 &= 6\mathbf{E}\left[\left(\int_0^T X_t \,\mathrm{d}B_t\right)^2 \int_0^T X_t^2 \,\mathrm{d}t\right] - 3\mathbf{E}\left(\int_0^T X_t^2 \,\mathrm{d}t\right)^2 \\
&\leqslant 6\mathbf{E}\left[\left(\int_0^T X_t \,\mathrm{d}B_t\right)^2 \int_0^T X_t^2 \,\mathrm{d}t\right],
\end{aligned}
$$

hence, by the Cauchy inequality,

$$
\mathbf{E}\left(\int_0^T X_t \,\mathrm{d}B_t\right)^4 \leqslant 6\sqrt{\mathbf{E}\left(\int_0^T X_t \,\mathrm{d}B_t\right)^4 \cdot \mathbf{E}\left(\int_0^T X_t^2 \,\mathrm{d}t\right)^2}.
$$

Squaring the latter and then dividing by $\mathbf{E}(\int_0^T X_t \,\mathrm{d}B_t)^4$ yields the desired inequality. △

REMARK.– Without rejecting the term $3\mathbf{E}(\int_0^T X_t^2\,\mathrm{d}t)^2$, we obtain a slightly improved inequality,

$$\mathbf{E}\bigg(\int_0^T X_t\,\mathrm{d}B_t\bigg)^4 \leqslant (3+\sqrt{6})^2\mathbf{E}\Big(\int_0^T X_t^2\,\mathrm{d}t\Big)^2$$

with the constant $(3+\sqrt{6})^2 \approx 29.7$ instead of 36.

4.2. Definition and properties

PROPOSITION 4.6.– *For every random process $X \in H^2[0,T]$, there is a sequence of step processes $\{X^n\}$ that converges to X in $H^2[0,T]$, i.e. such that*

$$\|X^n - X\|^2 = \mathbf{E}\int_0^T \big(X_t^n - X_t\big)^2\,\mathrm{d}t \to 0, \quad n\to\infty.$$

In other words, the class $S_b[0,T]$ is dense in $H^2[0,T]$.

REMARK.– One can "avoid" this proposition (together with a rather technical proof below) by defining $H^2[0,T]$ as the class of random processes X for which there exists a sequence $X^n \in S_b[0,T]$, $n\in\mathbb{N}$, such that $\|X^n - X\| \to 0$.

Proof. We proceed similarly to the proof of the density, in $L^2[a,b]$, of the class of (non-random) step functions on $[a,b]$.

Step 1. Take any $X \in H^2[0,T]$. We shall show that X can be approximated by *bounded* processes from $H^2[0,T]$. Define the processes Y^N, $N\in\mathbb{N}$, by

$$Y_t^N := \max\big\{-N, \min\{X_t, N\}\big\}.$$

It is clear that $|Y_t^N| \leqslant N$. Since $Y_t^N = X_t$ when $|X_t| \leqslant N$, we have $Y_t^N \to X_t$, $N\to\infty$, and thus $(Y_t^N - X_t)^2 \to 0$, $N\to\infty$. On the other hand,

$$\big(Y_t^N - X_t\big)^2 \leqslant 2\big[\big(Y_t^N\big)^2 + X_t^2\big] \leqslant 4X_t^2, \quad t\in[0,T],$$

and

$$\int_0^T \mathbf{E}X_t^2\,\mathrm{d}t = \mathbf{E}\int_0^T X_t^2\,\mathrm{d}t = \|X\|^2 < +\infty.$$

By the Lebesgue dominated convergence theorem (section 1.11), we get:

$$\left\|Y^N - X\right\|^2 = \int_0^T \mathbf{E}\big(Y_t^N - X_t\big)^2 \,\mathrm{d}t \to 0, \quad N \to \infty.$$

Step 2. Now let Y be any of the defined processes Y^N. We shall show that Y can be approximated by *bounded continuous* processes. Define the continuous processes Z^k by

$$Z_t^k := k \int_{t-1/k}^{t} Y_s \,\mathrm{d}s, \quad k \in \mathbb{N},$$

with the convention that (for example) $Y_s = 0$ for $s < 0$. Since Z_t^k depends only on the values of the adapted process Y until the moment t, the process Z^k is also adapted. Since $|Y_t| \leqslant N$, the processes Z^k are also all bounded ($|Z_t^k| \leqslant N$). Let us estimate the difference:

$$\begin{aligned}
\|Z^k - Y\|^2 &= \int_0^T \mathbf{E}\left(k \int_{t-1/k}^{t} Y_s \,\mathrm{d}s - Y_t\right)^2 \mathrm{d}t \\
&= \int_0^T \mathbf{E}\left(k \int_{t-1/k}^{t} \big(Y_s - Y_t\big) \,\mathrm{d}s\right)^2 \mathrm{d}t \\
&\leqslant \int_0^T \mathbf{E}\left(k \int_{t-1/k}^{t} \big(Y_s - Y_t\big)^2 \,\mathrm{d}s\right) \mathrm{d}t \quad \Big(\text{inequality } \big(\textstyle\int_a^b f\big)^2 \leqslant (b-a)\int_a^b f^2\Big) \\
&= \int_0^T \mathbf{E}\left(k \int_0^{1/k} \big(Y_{t-\tilde{s}} - Y_t\big)^2 \,\mathrm{d}\tilde{s}\right) \mathrm{d}t \quad (\text{change of variable } s = t - \tilde{s}) \\
&= k \int_0^{1/k} \mathbf{E}\left(\int_0^T \big(Y_{t-s} - Y_t\big)^2 \,\mathrm{d}t\right) \mathrm{d}s.
\end{aligned}$$

By the mean-square continuity of square-integrable functions,[6]

$$\int_0^T (Y_{t-s} - Y_t)^2 \,\mathrm{d}t \to 0 \quad \text{as } s \downarrow 0.$$

6. If $f \in L^2[a,b]$, then $\int_a^b (f(t+h) - f(t))^2 \,\mathrm{d}t \to 0$ as $h \to 0$; here $f(t) := 0$ for $t \notin [a,b]$.

By the boundedness of $(Y_{t-s}-Y_t)^2$ ($\leqslant 4N^2$) we find that $\mathbf{E}\int_0^T (Y_{t-s}-Y_t)^2\,\mathrm{d}t \to 0$ as $s\downarrow 0$. Therefore, for every $\varepsilon>0$, there is $k_0\in\mathbb{N}$ such that

$$\mathbf{E}\int_0^T (Y_{t-s}-Y_t)^2\,\mathrm{d}t<\varepsilon \quad \text{for} \quad 0<s<1/k_0,$$

and thus,

$$\left\|Z^k-Y\right\|^2 \leqslant k\int_0^{1/k}\varepsilon\,\mathrm{d}t=\varepsilon \quad \text{for } k>k_0.$$

Hence, $\|Z^k-Y\|\to 0$ as $k\to\infty$.

Step 3. Let Z be any process of the bounded continuous processes Z^k defined in Step 2. We shall show that Z can be approximated by *bounded step* processes. Take a sequence $\Delta^n=\{0=t_0^n<t_1^n<\cdots<t_{k_n}^n=T\}$, $n\in\mathbb{N}$, of partitions of the interval $[0,T]$ such that their diameters $|\Delta^n|=\max_i\Delta t_i^n=\max_i|t_{i+1}^n-t_i^n|\to 0$, $n\to\infty$. Define the sequence of step processes $\{X^n\}$ by

$$X_t^n:=Z\big(t_i^n\big), \quad t\in[t_i^n,t_{i+1}^n).$$

Because of the continuity of Z, $|X_t^n-Z_t|^2\to 0$ as $n\to\infty$. On the other hand, since Z is bounded, the sequence $\{X^n\}$ is uniformly bounded: if $|Z|\leqslant C$, then $|X^n|\leqslant C$, $n\in\mathbb{N}$. Therefore,

$$\left|X_t^n-Z_t\right|^2\leqslant 2\big((X_t^n)^2+Z_t^2\big)\leqslant 4C^2, \quad t\in[0,T].$$

Since

$$\mathbf{E}\int_0^T 4C^2\,\mathrm{d}t<\infty,$$

again by the Lebesgue theorem we get:

$$\left\|X^n-Z\right\|^2=\mathbf{E}\int_0^T\left|X_t^n-Z_t\right|^2\mathrm{d}t\to 0, \quad n\to\infty.$$

Step 4. Consider any $X\in H^2[0,T]$. By steps 1–3, for any $n\in\mathbb{N}$, we can sequentially choose a bounded process Y^N, a bounded continuous process Z^k, and step process $X^n\in S_b[0,T]$ such that

$$\left\|X-Y^N\right\|<\frac{1}{n}, \quad \left\|Y^N-Z^k\right\|<\frac{1}{n}, \quad \text{and} \quad \left\|Z^k-X^n\right\|<\frac{1}{n}.$$

Hence, we get:

$$\left\|X - X^n\right\| \leqslant \left\|X - Y^N\right\| + \left\|Y^N - Z^k\right\| + \left\|Z^k - X^n\right\| < \frac{3}{n} \to 0, \quad n \to \infty,$$

and we are done. △

DEFINITION 4.7.– *Suppose that* $X \in H^2[0,T]$ *and* $S_b[0,T] \ni X^n \xrightarrow{H^2} X$. *The stochastic* (*or Itô*) *integral of process* X *with respect to Brownian motion in the interval* $[0,T]$ *is the limit*

$$\int_0^T X_t \,\mathrm{d}B_t := L^2\text{-}\lim_{n\to\infty} \int_0^T X_t^n \,\mathrm{d}B_t.$$

Note that the latter limit always exists. Indeed,

$$\begin{aligned}
&\mathbf{E}\left(\int_0^T X_t^n \,\mathrm{d}B_t - \int_0^T X_t^m \,\mathrm{d}B_t\right)^2 \\
&\quad = \mathbf{E}\left(\int_0^T \left(X_t^n - X_t^m\right) \mathrm{d}B_t\right)^2 = \left\|X^n - X^m\right\|^2 \\
&\quad \leqslant \left(\left\|X^n - X\right\| + \left\|X - X^m\right\|\right)^2 \to 0, \quad n, m \to \infty.
\end{aligned}$$

Thus, the sequence of integrals $\{\int_0^T X_t^n \,\mathrm{d}B_t\}$ is a Cauchy sequence of random variables in the L^2 sense and therefore converges in the L^2 sense. Moreover, this limit does not depend on the choice of the sequence $S_b[0,T] \ni X^n \xrightarrow{H^2} X$. Indeed, let $\widetilde{X}^n \in S_b[0,T]$, $n \in \mathbb{N}$, be another sequence such that $\widetilde{X}^n \xrightarrow{H^2} X$. We define the new sequence

$$\{Y^n\} = \{X^1, \widetilde{X}^1, X^2, \widetilde{X}^2, \ldots, X^n, \widetilde{X}^n, \ldots\},$$

which clearly also converges to X in $H^2[0,T]$. As is shown, the corresponding sequence of the integrals $\{\int_0^T Y_t^n \,\mathrm{d}B_t\}$ converges in the L^2 sense. Therefore, its subsequences, $\{\int_0^T X_t^n \,\mathrm{d}B_t\}$ and $\{\int_0^T \widetilde{X}_t^n \,\mathrm{d}B_t\}$, have the same limit.

THEOREM 4.8.– *The stochastic integral in the class* $H^2 = H^2[0,T]$ *possesses all the properties 1–5 formulated in Proposition 4.5, that is*:

1) $X, Y \in H^2$, $\alpha, \beta \in \mathbb{R} \implies \int_0^T (\alpha X + \beta Y)_t \,\mathrm{d}B_t = \alpha \int_0^T X_t \,\mathrm{d}B_t + \beta \int_0^T Y_t \,\mathrm{d}B_t$;

2) $\mathbf{E}\int_0^T X_t\,\mathrm{d}B_t = 0$*;*

$2'$*) For bounded* $Z \in \mathcal{H}_s$*,* $Z\int_s^T X_t\,\mathrm{d}B_t = \int_s^T ZX_t\,\mathrm{d}B_t$*, and, in particular,* $\mathbf{E}(Z\int_s^T X_t\,\mathrm{d}B_t) = 0$.

3) $\mathbf{E}(\int_0^T X_t\,\mathrm{d}B_t)^2 = \mathbf{E}\int_0^T X_t^2\,\mathrm{d}t$*, i.e.,* $\|\int_0^T X_t\,\mathrm{d}B_t\|_{L^2} = \|X\|_{H^2}$*;*

4) $\mathbf{E}(\int_0^T X_t\,\mathrm{d}B_t\int_0^T Y_t\,\mathrm{d}B_t) = \mathbf{E}\int_0^T X_tY_t\,\mathrm{d}t$*;*

5) $\mathbf{E}(\int_0^T X_t\,\mathrm{d}B_t)^4 \leqslant 36\,\mathbf{E}(\int_0^T X_t^2\,\mathrm{d}t)^2$.

Moreover, the following Doob[7] *inequalities hold for the maximum of the stochastic integral*:

6) $\mathbf{P}\{\sup_{t\leqslant T}|\int_0^t X_s\,\mathrm{d}B_s| \geqslant \lambda\} \leqslant \frac{1}{\lambda^p}\mathbf{E}|\int_0^T X_s\,\mathrm{d}B_s|^p,\ \lambda > 0,\ p \geqslant 1$*; in particular,*
$\mathbf{P}\{\sup_{t\leqslant T}|\int_0^t X_s\,\mathrm{d}B_s| \geqslant \lambda\} \leqslant \frac{1}{\lambda^2}\mathbf{E}\int_0^T X_t^2\,\mathrm{d}t,\ \lambda > 0$*;*

$6'$*) (quadratic Doob inequality)* $\mathbf{E}\sup_{t\leqslant T}(\int_0^t X_s\,\mathrm{d}B_s)^2 \leqslant 4\mathbf{E}\int_0^T X_t^2\,\mathrm{d}t$.

Proof. The idea of proving Properties 1–3 is rather simple. We have to pass to the limit in the corresponding properties of Proposition 4.5 for arbitrary sequences $S_b \ni X^n \xrightarrow{H^2} X \in H^2$, $S_b \ni Y^n \xrightarrow{H^2} Y \in H^2$.

1) Since

$$\begin{aligned}\big\|(\alpha X^n + \beta Y^n) - (\alpha X + \beta Y)\big\| &= \big\|\alpha(X^n - X) + \beta(Y^n - Y)\big\| \\ &\leqslant |\alpha|\,\big\|X^n - X\big\| + |\beta|\,\big\|Y^n - Y\big\| \to 0, \quad n \to \infty,\end{aligned}$$

we have $\alpha X^n + \beta Y^n \xrightarrow{H^2} \alpha X + \beta Y$. Therefore, twice applying the definition of stochastic integral, we get:

$$\begin{aligned}\int_0^T (\alpha X + \beta Y)_t\,\mathrm{d}B_t &= L^2\text{-}\lim_{n\to\infty}\int_0^T \big(\alpha X^n + \beta Y^n\big)_t\,\mathrm{d}B_t \\ &= \alpha\,L^2\text{-}\lim_{n\to\infty}\int_0^T X_t^n\,\mathrm{d}B_t + \beta\,L^2\text{-}\lim_{n\to\infty}\int_0^T Y_t^n\,\mathrm{d}B_t \\ &= \alpha\int_0^T X_t\,\mathrm{d}B_t + \beta\int_0^T Y_t\,\mathrm{d}B_t.\end{aligned}$$

7. Joseph Leo Doob.

2) By the definition of the stochastic integral,

$$Y^n := \int_0^T X_t^n \,\mathrm{d}B_t \xrightarrow{L^2} Y := \int_0^T X_t \,\mathrm{d}B_t, \quad n \to \infty,$$

and $\mathbf{E}Y^n = 0$ for all $n \in \mathbb{N}$; therefore,

$$\mathbf{E}Y = \lim_{n\to\infty} \mathbf{E}Y^n = 0.$$

Indeed,

$$\big(\mathbf{E}Y^n - \mathbf{E}Y\big)^2 = \big(\mathbf{E}(Y^n - Y)\big)^2 \leqslant \mathbf{E}\big(Y^n - Y\big)^2 \to 0, \quad n \to \infty.$$

2′) This property is also proved by passing to the limit in mean-square sense in the corresponding property for step processes (for any bounded $Z \in \mathcal{H}_s$),

$$Z \int_s^T X_t^n \,\mathrm{d}B_t = \int_s^T Z X_t^n \,\mathrm{d}B_t$$

or, denoting $Y^n := X^n 1\!\!1_{[s,T]}$ and $Y := X 1\!\!1_{[s,T]}$,

$$Z \int_0^T Y_t^n \,\mathrm{d}B_t = \int_0^T Z Y_t^n \,\mathrm{d}B_t.$$

Note that all processes Y^n, Y, ZY^n, and ZY are adapted (and equal to zero in the interval $[0, s)$). Since

$$\begin{aligned}\big\|Y^n - Y\big\|^2 &= \mathbf{E}\int_0^T \big(Y_t^n - Y_t\big)^2 \,\mathrm{d}t = \mathbf{E}\int_s^T \big(X_t^n - X_t\big)^2 \,\mathrm{d}t \\ &\leqslant \mathbf{E}\int_0^T \big(X_t^n - X_t\big)^2 \,\mathrm{d}t = \big\|X^n - X\big\|^2 \to 0, \quad n \to \infty,\end{aligned}$$

and $|Z| \leqslant C$, we have

$$\begin{aligned}\mathbf{E}\bigg(Z\int_0^T Y_t^n \,\mathrm{d}B_t - Z\int_0^T Y_t \,\mathrm{d}B_t\bigg)^2 &= \mathbf{E}\bigg(Z\bigg[\int_0^T Y_t^n \,\mathrm{d}B_t - \int_0^T Y_t \,\mathrm{d}B_t\bigg]\bigg)^2 \\ &\leqslant C^2 \cdot \mathbf{E}\bigg(\int_0^T Y_t^n \,\mathrm{d}B_t - \int_0^T Y_t \,\mathrm{d}B_t\bigg)^2 \to 0, \quad n \to \infty,\end{aligned}$$

that is, the limit of the left-hand side (in the mean-square sense) is $Z\int_0^T Y_t\,\mathrm{d}B_t$. Since $|Z| \leqslant C$ and therefore

$$\big\|ZY^n - ZY\big\| = \big\|Z(Y^n - Y)\big\| \leqslant C\big\|Y^n - Y\big\| \to 0, \quad n \to \infty,$$

we have that the right-hand side

$$\int_0^T ZY_t^n\,\mathrm{d}B_t \xrightarrow{L^2} \int_0^T ZY_t\,\mathrm{d}B_t, \quad n \to \infty.$$

Thus,

$$Z\int_0^T Y_t\,\mathrm{d}B_t = \int_0^T ZY_t\,\mathrm{d}B_t,$$

as required.

3) We get the required equality by passing to the limit, as $n \to \infty$, in the equality

$$\left\|\int_0^T X_t^n\,\mathrm{d}B_t\right\|_{L^2}^2 = \mathbf{E}\left(\int_0^T X_t^n\,\mathrm{d}B_t\right)^2 = \big\|X^n\big\|^2, \quad n \in \mathbb{N}.$$

Indeed, $Y^n := \int_0^T X_t^n\,\mathrm{d}B_t \xrightarrow{L^2} Y := \int_0^T X_t\,\mathrm{d}B_t$; therefore, using the properties of a norm, we get:

$$\big|\|Y^n\| - \|Y\|\big|_{L^2} \leqslant \big\|Y^n - Y\big\|_{L^2} \to 0, \quad n \to \infty,$$

and

$$\big|\|X^n\| - \|X\|\big| \leqslant \big\|X^n - X\big\| \to 0, \quad n \to \infty.$$

Hence,

$$\|Y\|_{L^2} = \|X\|,$$

as required.

4) This property is shown in exactly the same way as in Proposition 4.5.

5) If $\mathbf{E}(\int_0^T X_t^2\,\mathrm{d}t)^2 = +\infty$, the statement is obvious. If $\mathbf{E}(\int_0^T X_t^2\,\mathrm{d}t)^2 < +\infty$, then, with a slight modification of Proposition 4.6, we can construct a sequence of step processes $\{X^n\}$ such that

$$\mathbf{E}\left(\int_0^T \big(X_t^n - X_t\big)^2\,\mathrm{d}t\right)^2 \to 0, \quad n \to \infty,$$

and

$$\mathbf{E}\left(\int_0^T X_t^n \,\mathrm{d}B_t - \int_0^T X_t \,\mathrm{d}B_t\right)^4 \to 0, \quad n \to \infty.$$

Then we obtain the desired inequality by passing to the limit, as $n \to \infty$, in the inequality

$$\mathbf{E}\left(\int_0^T X_t^n \,\mathrm{d}B_t\right)^4 \leqslant 36\,\mathbf{E}\left(\int_0^T \left(X_t^n\right)^2 \mathrm{d}t\right)^2.$$

6) We shall show the inequality in the most often used cases $p = 1$ and $p = 2$. Denote $M_t := \int_0^t X_s \,\mathrm{d}B_s$, $t \in [0, T]$. We first check that if $A \in \mathcal{H}_t$, i.e. $1\!\!1_A \in \mathcal{H}_t$, then

$$\mathbf{E}|M_t 1\!\!1_A|^p \leqslant \mathbf{E}|M_T 1\!\!1_A|^p, \quad p = 1, 2. \tag{4.2}$$

In the case $p = 1$, for all events $A \in \mathcal{H}_t$,

$$\begin{aligned}
\mathbf{E}\big(|M_T| 1\!\!1_A\big) = \mathbf{E}|M_T 1\!\!1_A| \geqslant \big|\mathbf{E}(M_T 1\!\!1_A)\big| &= \left|\mathbf{E}\left(\int_0^T X_s \,\mathrm{d}B_s \cdot 1\!\!1_A\right)\right| \\
&= \left|\mathbf{E}\left(\int_0^t X_s \,\mathrm{d}B_s \cdot 1\!\!1_A\right) + \mathbf{E}\left(\int_t^T X_s \,\mathrm{d}B_s \cdot 1\!\!1_A\right)\right| \\
&= \big|\mathbf{E}(M_t 1\!\!1_A)\big|.
\end{aligned}$$

Here we used the fact that $\mathbf{E}(\int_t^T X_s \,\mathrm{d}B_s \cdot 1\!\!1_A) = 0$ by property $2'$. Instead of $A \in \mathcal{H}_t$, considering the events

$$A_+ := A \cap \{M_t \geqslant 0\} \in \mathcal{H}_t \qquad \text{and} \qquad A_- := A \cap \{M_t < 0\} \in \mathcal{H}_t,$$

we get:

$$\mathbf{E}\big(|M_T| 1\!\!1_{A_+}\big) \geqslant \big|\mathbf{E}(M_t 1\!\!1_{A_+})\big| = \big|\mathbf{E}(|M_t| 1\!\!1_{A_+})\big| = \mathbf{E}\big(|M_t| 1\!\!1_{A_+}\big)$$

and

$$\mathbf{E}(|M_T| 1\!\!1_{A_-}) \geqslant |\mathbf{E}(M_t 1\!\!1_{A_-})| = \big|\mathbf{E}(-|M_t| 1\!\!1_{A_-})\big| = \mathbf{E}(|M_t| 1\!\!1_{A_-}).$$

Adding the inequalities and using the equality $\mathbb{1}_{A_+} + \mathbb{1}_{A_-} = \mathbb{1}_A$, we finally have

$$\mathbf{E}|M_T \mathbb{1}_A| = \mathbf{E}\big(|M_T|\mathbb{1}_A\big) \geqslant \mathbf{E}\big(|M_t|\mathbb{1}_A\big) = \mathbf{E}|M_t \mathbb{1}_A|,$$

that is, inequality [4.2] for $p = 1$.

In the case $p = 2$, we obtain inequality [4.2] in a simpler manner:

$$\begin{aligned}
\mathbf{E}(M_T \mathbb{1}_A)^2 &= \mathbf{E}\left\{\left[M_t^2 + 2M_t \cdot \int_t^T X_s \,\mathrm{d}B_s + \left(\int_t^T X_s \,\mathrm{d}B_s\right)^2\right]\mathbb{1}_A\right\} \\
&= \mathbf{E}(M_t \mathbb{1}_A)^2 + 2\mathbf{E}\left((M_t \mathbb{1}_A)\int_t^T X_s \,\mathrm{d}B_s\right) + \mathbf{E}\left(\mathbb{1}_A \int_t^T X_s \,\mathrm{d}B_s\right)^2 \\
&\geqslant \mathbf{E}(M_t \mathbb{1}_A)^2.
\end{aligned}$$

Here we used the fact that $\mathbf{E}((M_t \mathbb{1}_A)\int_t^T X_s \,\mathrm{d}B_s) = 0$ by property $2'$.

Now, denote $I(t) := |M_t| = |\int_0^t X_s \,\mathrm{d}B_s|$, $t \in [0, T]$. We shall prove the more general inequality

$$\mathbf{P}\Big\{\sup_{t \leqslant T} I(t) \geqslant \lambda\Big\} \leqslant \frac{1}{\lambda^p}\mathbf{E}I^p(T), \quad \lambda > 0,\ p \geqslant 1, \tag{4.3}$$

which, for $p = 2$, by Property 3 becomes

$$\mathbf{P}\Big\{\sup_{t \leqslant T} I(t) \geqslant \lambda\Big\} \leqslant \frac{1}{\lambda^2}\mathbf{E}I^2(T) = \frac{1}{\lambda^2}\mathbf{E}\int_0^T X_t^2 \,\mathrm{d}t,$$

as desired. The process $I(t)$, $t \in [0, T]$, is continuous.[8] Therefore, it suffices to check that

$$\mathbf{P}\Big\{\max_{0 \leqslant k \leqslant n} I(t_k) \geqslant \lambda\Big\} \leqslant \frac{1}{\lambda^p}\mathbf{E}I^p(T), \quad \lambda > 0, \tag{4.4}$$

for all $n \in \mathbb{N}$ and all partitions $0 = t_0 < t_1 < \cdots < t_n = T$. Denote

$$A_k := \Big\{I(t_k) \geqslant \lambda \quad \text{and} \quad \max_{0 \leqslant i \leqslant k-1} I(t_i) < \lambda\Big\}, \quad 1 \leqslant k \leqslant n.$$

8. This is proved in Proposition 4.14 (without using the Doob inequalities).

It is clear that $A_i \cap A_j = \emptyset$ for $i \neq j$, $B_\lambda := \{\max_{0 \leqslant k \leqslant n} I(t_k) \geqslant \lambda\} = \bigcup_{k=1}^n A_k$, and $A_k \in \mathcal{H}_{t_k}$ (i.e., $\mathbb{1}_{A_k} \in \mathcal{H}_{t_k}$). Therefore,

$$\begin{aligned}
\mathbf{P}(B_\lambda) &= \sum_{k=1}^n \mathbf{P}(A_k) = \sum_{k=1}^n \mathbf{E}(\mathbb{1}_{A_k}) \\
&\leqslant \sum_{k=1}^n \mathbf{E}\Big(\frac{I^p(t_k)}{\lambda^p}\mathbb{1}_{A_k}\Big) = \frac{1}{\lambda^p}\sum_{k=1}^n \mathbf{E}\big(I^p(t_k)\mathbb{1}_{A_k}\big) \\
&\leqslant \frac{1}{\lambda^p}\sum_{k=1}^n \mathbf{E}\big(I^p(T)\mathbb{1}_{A_k}\big) = \frac{1}{\lambda^p}\mathbf{E}\bigg(I^p(T)\sum_{k=1}^n \mathbb{1}_{A_k}\bigg) \\
&= \frac{1}{\lambda^p}\mathbf{E}\Big(I^p(T)\mathbb{1}_{\bigcup_{k=1}^n A_k}\Big) = \frac{1}{\lambda^p}\mathbf{E}\big(I^p(T)\mathbb{1}_{B_\lambda}\big).
\end{aligned}$$

In the second inequality, we have applied inequality [4.2]. In particular, we get inequality [4.4] (and, thus, inequality [4.3], since the last term does not exceed $\frac{1}{\lambda^p}\mathbf{E}I^p(T)$.

$\mathcal{M}$ 6′) Denote $Y := \sup_{t \leqslant T} I(t)$. Then, taking $p = 1$, we can rewrite the inequality just proved as

$$\mathbf{P}\{Y \geqslant \lambda\} \leqslant \frac{1}{\lambda}\mathbf{E}\big(I(T)\mathbb{1}_{\{Y \geqslant \lambda\}}\big), \quad \lambda > 0.$$

Therefore,

$$\begin{aligned}
\mathbf{E}Y^2 &= 2\mathbf{E}\bigg(\int_0^Y \lambda \,\mathrm{d}\lambda\bigg) = 2\mathbf{E}\bigg(\int_0^\infty \lambda \mathbb{1}_{\{Y \geqslant \lambda\}}\,\mathrm{d}\lambda\bigg) \\
&= 2\int_0^\infty \lambda \mathbf{E}(\mathbb{1}_{\{Y \geqslant \lambda\}})\,\mathrm{d}\lambda = 2\int_0^\infty \lambda \mathbf{P}\{Y \geqslant \lambda\}\,\mathrm{d}\lambda \\
&\leqslant 2\int_0^\infty \mathbf{E}\big(I(T)\mathbb{1}_{\{Y \geqslant \lambda\}}\big)\,\mathrm{d}\lambda = 2\mathbf{E}\bigg(I(T)\int_0^\infty \mathbb{1}_{\{Y \geqslant \lambda\}}\,\mathrm{d}\lambda\bigg) \\
&= 2\mathbf{E}\big(I(T)Y\big).
\end{aligned}$$

Applying the Cauchy inequality, we get:

$$\mathbf{E}Y^2 \leqslant 2\big(\mathbf{E}I^2(T)\big)^{1/2}\big(\mathbf{E}Y^2\big)^{1/2}.$$

Finally, dividing both sides of the latter by $(\mathbf{E}Y^2)^{1/2}$, we have

$$\big(\mathbf{E}Y^2\big)^{1/2} \leqslant 2\big(\mathbf{E}I^2(T)\big)^{1/2}$$

or

$$\mathbf{E}Y^2 \leqslant 4\mathbf{E}I^2(T).$$

$\triangle$

REMARK.– Strictly speaking, at the end of the proof, in order to divide the by $(\mathbf{E}Y^2)^{1/2}$, we must be sure that $\mathbf{E}Y^2 < \infty$. To this end, it suffices first to consider the truncated processes $I_n(t) := I(t) \wedge n$ instead of $I(t)$ and then pass to the limit as $n \to \infty$.

The stochastic integral of a continuous adapted process can be defined similarly to Stieltjes-type integrals. It is important that then, in the Stieltjes-type integral sums, we necessarily have to take values of the integrated process at the *left* points of the partition intervals.

PROPOSITION 4.9.– *Let $X \in H^2[0,T]$ be an adapted process which is L^2-continuous in the following sense: for all $t \in [0,T]$,*

$$\mathbf{E}X_t^2 < \infty \qquad \textit{and} \qquad \mathbf{E}|X_s - X_t|^2 \to 0 \quad \textit{as} \quad s \to t.$$

If $\Delta^n = \{0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T\}$, $n \in \mathbb{N}$, is a sequence of partitions of $[0,T]$ such that $|\Delta^n| = \max_i \Delta t_i = \max_i |t_{i+1}^n - t_i^n| \to 0$, then

$$\int_0^T X_t \,\mathrm{d}B_t = L^2\text{-}\lim_{n\to\infty} \sum_{i=0}^{k_n-1} X\big(t_i^n\big)\big(B(t_{i+1}^n) - B(t_i^n)\big).$$

Proof. First, note that from the L^2-continuity of X there follows the continuity of its second moment $\mathbf{E}X_t^2$ with respect to t:

$$\left|\sqrt{\mathbf{E}X_s^2} - \sqrt{\mathbf{E}X_t^2}\right| \leqslant \sqrt{\mathbf{E}(X_s - X_t)^2} \to 0, \quad s \to t.$$

In particular, the function $\mathbf{E}X_t^2$, $t \in [0,T]$, is bounded. Let $C := \sup_{t\in[0,T]} \mathbf{E}X_t^2$.

Similarly to the proof of Proposition 4.6 (step 3), define the sequence $\{X^n\} \subset S^2[0,T]$ by

$$X_t^n := X\big(t_i^n\big), \quad t \in \big[t_i^n, t_{i+1}^n\big).$$

Then

$$\sum_{i=0}^{k_n-1} X\big(t_i^n\big)\big(B(t_{i+1}^n) - B(t_i^n)\big) = \int_0^T X_t^n \,\mathrm{d}B_t.$$

Because of the L^2-continuity of X, $\mathbf{E}|X_t^n - X_t|^2 \to 0$, $n \to \infty$, for all $t \in [0,T]$. On the other hand, by the boundedness of $\mathbf{E}X_t^2$ ($\mathbf{E}X_t^2 \leqslant C$),

$$\mathbf{E}\big|X_t^n - X_t\big|^2 \leqslant 2\big(\mathbf{E}(X_t^n)^2 + \mathbf{E}X_t^2\big) \leqslant 2\Big(\max_i \mathbf{E}X_{t_i^n}^2 + \max_{t\in[0,T]} \mathbf{E}X_t^2\Big)$$

$$\leqslant 4 \max_{t\in[0,T]} \mathbf{E}X_t^2 \leqslant 4C, \quad t \in [0,T].$$

Since

$$\int_0^T (4C)\,\mathrm{d}t < \infty,$$

by the Lebesgue theorem we get:

$$\big\|X^n - X\big\|^2 = \int_0^T \mathbf{E}\big|X_t^n - X_t\big|^2\,\mathrm{d}t \to 0, \quad n \to \infty.$$

Thus, by the definition of stochastic integral, we may write

$$\int_0^T X_t\,\mathrm{d}B_t = L^2\text{-}\lim_{n\to\infty}\int_0^T X_t^n\,\mathrm{d}B_t = L^2\text{-}\lim_{n\to\infty}\sum_{i=0}^{k_n-1} X(t_i^n)\big(B(t_{i+1}^n) - B(t_i^n)\big).$$

△

EXAMPLE 4.10.– Let us calculate the stochastic integral $\int_0^T B_t\,\mathrm{d}B_t$ by applying Proposition 4.9. Consider the sequence of partitions Δ^n of $[0,T]$ as in the proposition ($|\Delta^n| \to 0$). For brevity, denote $B_i = B(t_i^n)$ $\Delta B_i = B_{i+1} - B_i$. Then, by the quadratic-variation property of Brownian motion (Theorem 4.1), we obtain

$$\begin{aligned}\sum_i B_i\Delta B_i &= \sum_i B_i(B_{i+1} - B_i)\\ &= \frac{1}{2}\sum_i \big(B_{i+1}^2 - B_i^2 - \Delta B_i^2\big)\\ &= \frac{1}{2}\sum_i \big(B_{i+1}^2 - B_i^2\big) - \frac{1}{2}\sum_i \Delta B_i^2\\ &= \frac{1}{2}B_T^2 - \frac{1}{2}\sum_i \Delta B_i^2 \xrightarrow{L^2} \frac{B_T^2}{2} - \frac{T}{2}, \quad n \to \infty,\end{aligned}$$

and, thus,

$$\int_0^T B_t\,\mathrm{d}B_t = \frac{B_T^2}{2} - \frac{T}{2}.$$

4.3. Extensions

The stochastic integral can be (slightly but very usefully) extended to the random processes X for which

$$\mathbf{P}\Big\{\int_0^T X_t^2\,\mathrm{d}t < +\infty\Big\} = 1. \qquad [4.5]$$

To this end, we will need the following local-behavior property of the stochastic integral.

LEMMA 4.11.– *Suppose that random processes $X, Y \in H^2[0,T]$ coincide on an event A, i.e.*

$$X_t(\omega) = Y_t(\omega), \quad t \in [0,T],\ \omega \in A.$$

Then, $\int_0^T X_t\,\mathrm{d}B_t = \int_0^T Y_t\,\mathrm{d}B_t$ almost surely on A.

Proof. Looking carefully through the proof of Proposition 4.6, where for any random process $X \in H^2[0,T]$ we constructed a sequence $S_b[0,T] \ni X^n \xrightarrow{H^2} X$, we can see that from the coincidence of X and Y on the event A there follows the coincidence, on A, of the corresponding step processes X^n and Y^n (for all $n \in \mathbb{N}$). By the definition of the stochastic integral of a step process, the integrals $\int_0^T X_t^n\,\mathrm{d}B_t$ and $\int_0^T Y_t^n\,\mathrm{d}B_t$ coincide on A (for all $n \in \mathbb{N}$). Passing, if necessary, to a subsequence, we can claim that $\int_0^T X_t^n\,\mathrm{d}B_t \to \int_0^T X_t\,\mathrm{d}B_t$ and $\int_0^T Y_t^n\,\mathrm{d}B_t \to \int_0^T Y_t\,\mathrm{d}B_t$ everywhere except an event of zero probability. Therefore, their limits, $\int_0^T X_t\,\mathrm{d}B_t$ and $\int_0^T Y_t\,\mathrm{d}B_t$, coincide where the integrals $\int_0^T X_t^n\,\mathrm{d}B_t$ and $\int_0^T Y_t^n\,\mathrm{d}B_t$ do, except in the same event of zero probability, that is, at least on A almost surely. △

DEFINITION 4.12.– *We denote by $\widehat{H}^2[0,T]$ the class of all adapted random processes X satisfying condition* [4.5].

For a process $X \in \widehat{H}^2[0,T]$, we denote

$$X_t^{(N)} = X_t 1\!\!1_{\{\int_0^t X_s^2\,\mathrm{d}s \leqslant N\}}, \quad t \in [0,T],\ N \in \mathbb{N}.$$

The stochastic (Itô) integral of X is defined by

$$\int_0^T X_t\,\mathrm{d}B_t := \lim_{N\to\infty} \int_0^T X_t^{(N)}\,\mathrm{d}B_t.$$

PROPOSITION 4.13.– *For every* $X \in \widehat{H}^2[0,T]$*, the latter limit exists* (*with probability* 1)*, i.e. the stochastic integral* $\int_0^T X_t \,\mathrm{d}B_t$ *is correctly defined.*

Proof. All processes $X^{(N)}$, $N \in \mathbb{N}$, belong to the class $H^2[0,T]$, since

$$\int_0^T \big(X_t^{(N)}\big)^2 \,\mathrm{d}t = \int_0^T X_t^2 \mathbb{1}_{\{\int_0^t X_s^2 \,\mathrm{d}s \leqslant N\}} \,\mathrm{d}t \leqslant N$$

and therefore $\mathbf{E} \int_0^T (X_t^{(N)})^2 \,\mathrm{d}t \leqslant N$. Hence, all the stochastic integrals $\int_0^T X_t^{(N)} \,\mathrm{d}B_t$, $N \in \mathbb{N}$, are well defined. Consider the events $\Omega_N := \{\int_0^T X_t^2 \,\mathrm{d}t \leqslant N\}$, $N \in \mathbb{N}$. Note that $X^{(N)} = X^{(M)} = X$ on the event Ω_M for all $N \geqslant M$. Therefore, $\int_0^T X_t^{(N)} \,\mathrm{d}B_t = \int_0^T X_t^{(M)} \,\mathrm{d}B_t$ (a.s.) on the event Ω_M for $N \geqslant M$. Hence, on every event Ω_M, all stochastic integrals $\{\int_0^T X_t^{(N)} \,\mathrm{d}B_t\}$ coincide (a.s.) for $N \geqslant M$. In particular, on every event Ω_M, their sequence $\{\int_0^T X_t^{(N)} \,\mathrm{d}B_t\}$, $N \in \mathbb{N}$, has a limit (a.s). Therefore, this sequence has a limit on the event $\Omega = \cup_M \Omega_M$, that is, almost surely, since the probability of Ω is one. △

REMARK.– In the general case, i.e. in the class $\widehat{H}^2[0,T]$, properties 2–5 of Theorem 4.8 may not hold.

PROPOSITION 4.14.– *For every process* $X \in \widehat{H}^2[0,T]$*, the stochastic integral*

$$I(t) := \int_0^t X_s \,\mathrm{d}B_s, \quad t \in [0,T],$$

is a continuous function of the upper bound t (*a.s.*).

Proof. We prove the continuity of the integral $I(t) = \int_0^t X_s \,\mathrm{d}B_s$, $t \in [0,T]$, by applying the Kolmogorov theorem (section 4.14) in the case where the trajectories of a process X are bounded (for example, continuous).[9] First, suppose that X is bounded by a *non-random* constant: $|X_t| \leqslant N$, $t \in [0,T]$. Then, by

9. It is worth emphasizing that the boundedness of a random process can be understood in two ways: 1) every trajectory is bounded (by a constant which depends on the trajectory and thus, in general, is a random variable); 2) the process itself is bounded by a (non-random) constant. For example, a Brownian motion is bounded (in every finite time interval) in the first sense and not in the second.

Property 5,

$$\mathbf{E}\big(I(t_2) - I(t_1)\big)^4 = \mathbf{E}\left(\int_0^{t_2} X_s \,\mathrm{d}B_s - \int_0^{t_1} X_s \,\mathrm{d}B_s\right)^4 = \mathbf{E}\left(\int_{t_1}^{t_2} X_s \,\mathrm{d}B_s\right)^4$$
$$\leqslant 36\,\mathbf{E}\left(\int_{t_1}^{t_2} X_s^2 \,\mathrm{d}s\right)^2 \leqslant 36N^4|t_2 - t_1|^2,$$

and the continuity of the process $I(t)$, $t \in [0,T]$, follows from the Kolmogorov theorem. In the general case where X is bounded by a possibly random constant, define the bounded processes $X^N = \max\{-N, \min\{N, X\}\}$. Then by Lemma 4.11 the stochastic integrals $I(t) = \int_0^t X_s \,\mathrm{d}B_s$, $t \in [0,T]$, (almost surely) coincide with the stochastic integrals $I^N(t) := \int_0^t X_s^N \,\mathrm{d}B_s$, $t \in [0,T]$, on the events $\Omega^N := \{|X_s| \leqslant N,\ \forall\, s \in [0,T]\}$ and therefore, on all these events, have continuous trajectories. Since the probability of the event $\cup_N \Omega^N$ is one, the process $I(t)$, $t \in [0,T]$, has continuous trajectories almost surely. △

We further give two more generalizations of the stochastic integrals, for *infinite* and *random* time intervals.

DEFINITION 4.15.– *We denote by $H^2[0,\infty)$ the class of all adapted random processes $X = \{X_t,\ t \geqslant 0\}$ for which*

$$\|X\| = \|X\|_{H^2} := \left(\mathbf{E}\int_0^\infty X_t^2 \,\mathrm{d}t\right)^{1/2} < +\infty.$$

We say that a sequence of adapted random processes $\{X^n\}$ converges to an adapted process X in the space $H^2[0,\infty)$ and write $X^n \xrightarrow{H^2} X$ if $\|X^n - X\| \to 0$ as $n \to \infty$.

The stochastic integral (or Itô integral) of a random process $X \in H^2[0,\infty)$ with respect to Brownian motion B in the interval $[0,\infty)$ is defined as the limit

$$\int_0^\infty X_t \,\mathrm{d}B_t := L^2\text{-}\lim_{n\to\infty} \int_0^n X_t \,\mathrm{d}B_t.$$

Such a limit always exists because $\{\int_0^n X_t\,\mathrm{d}B_t\}$ is a Cauchy sequence in the mean-square sense, i.e. if $n \geqslant m \to \infty$, then

$$\mathbf{E}\left(\int_0^n X_t\,\mathrm{d}B_t - \int_0^m X_t\,\mathrm{d}B_t\right)^2 = \mathbf{E}\left(\int_m^n X_t\,\mathrm{d}B_t\right)^2$$
$$= \mathbf{E}\int_m^n X_t^2\,\mathrm{d}t = \int_m^n \mathbf{E}X_t^2\,\mathrm{d}t \to 0,$$

since $\int_0^\infty \mathbf{E}X_t^2\,\mathrm{d}t = \mathbf{E}\int_0^\infty X_t^2\,\mathrm{d}t < \infty$.

THEOREM 4.8a. *All the properties of the stochastic integral stated in Theorem 4.8 remain true with $T = \infty$.*

Let τ be a non-negative random variable. When can we correctly define the stochastic integral $\int_0^\tau X_t\,\mathrm{d}B_t$? For a non-stochastic integral of a (deterministic or random) function $f(t)$, $t \geqslant 0$, we have the equality

$$\int_0^\tau f(t)\,\mathrm{d}t = \int_0^\infty f(t)\mathbb{1}_{[0,\tau]}(t)\,\mathrm{d}t,$$

provided that at least one of the integrals exists. Mimicking this equality, we may try to define a stochastic integral as

$$\int_0^\tau X_t\,\mathrm{d}B_t := \int_0^\infty X_t\mathbb{1}_{[0,\tau]}(t)\,\mathrm{d}B_t.$$

For the integral on the right-hand side to make sense, together with the random process X, we need the process $\mathbb{1}_{[0,\tau]}(t)$, $t \geqslant 0$, to be adapted. On the other hand, $\mathbb{1}_{[0,\tau]}(t) \in \mathcal{H}_t$ for all $t \geqslant 0$ if and only if $\tau \wedge t \in \mathcal{H}_t$ for all $t \geqslant 0$. Random variables with the latter property are called *stopping times* or *Markov moments* (with respect to Brownian motion B). Intuitively, a non-negative[10] random variable τ is a stopping time if, at every time moment $t \geqslant 0$, we know whether $\tau < t$, i.e. whether the stopping took place before moment t. A typical stopping time is the first moment when an adapted process (in particular, a Brownian motion) X hits a (measurable) set A:

$$\tau := \inf\{t \geqslant 0 \colon X_t \in A\}.$$

10. With the value $+\infty$ allowed.

A completely rigorous definition is as follows.

DEFINITION 4.16.– *Denote by* $\mathcal{F}_t = \mathcal{F}_t^B := \sigma\{B_s,\ s \leqslant t\}$ *the* σ*-algebra generated by random variables* B_s*,* $s \leqslant t$*, and all events of zero probability. A random variable* τ *taking values in the interval* $[0, +\infty]$ *is called stopping time or Markov moment (with respect to Brownian motion* B*) if the events* $\{\tau \leqslant t\} \in \mathcal{F}_t$ *for all* $t \geqslant 0$*.*

REMARK.– In the theory of random processes, the general notion of stopping time is widely used. Let $\mathbb{F} = \{\mathcal{F}_t,\ t \geqslant 0\}$ be any increasing family of σ-algebras; such a family is called a *filtration*. A random variable τ taking values in the interval $[0, +\infty]$, is called a *stopping time* or *Markov moment* (with respect to the filtration $\mathbb{F}$) if $\{\tau \leqslant t\} \in \mathcal{F}_t$ for all $t \geqslant 0$. Taking, in the latter definition, instead of $\{\tau \leqslant t\}$, the events $\{\tau < t\}$, we get a wider notion of *optional* moment. Note that when defining a stopping time τ with respect to Brownian motion as a random variable for which $\tau \wedge t \in \mathcal{F}_t$, $t \geqslant 0$, we actually define an optional moment. However, for a filtration generated by a Brownian motion, the notions of stopping time and optional moment *coincide*. In general, they coincide for a right-continuous filtration, i.e. such that

$$\mathcal{F}_{t+0} := \bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t, \quad t \geqslant 0.$$

DEFINITION 4.17.– *If* τ *is a stopping time with respect to a Brownian motion* B*, then we say that an adapted random process* X *belongs to the class* $H^2[0,\tau]$ *if*

$$\mathbf{E}\int_0^\tau X_t^2\,\mathrm{d}t = \mathbf{E}\int_0^\infty X_t^2 1\!\!1_{[0,\tau]}(t)\,\mathrm{d}t < +\infty.$$

The stochastic integral of such a process in the interval $[0,\tau]$ *is defined by*

$$X_t\,\mathrm{d}B_t := \int_0^\infty X_t 1\!\!1_{[0,\tau]}(t)\,\mathrm{d}B_t.$$

THEOREM 4.8b. *If* τ *is a stopping time with respect to a Brownian motion* B*, then all the properties of the stochastic integral stated in Theorem 4.8 remain true with* $T = \tau$*.*

4.4. Exercises

4.1. By Proposition 4.5 (property 5), for any step process $X \in S_b[0,T]$, we have the inequality

$$\mathbf{E}\Big(\int_0^T X_s\ \mathrm{d}B_s\Big)^4 \leqslant C\,\mathbf{E}\Big(\int_0^T X_s^2\ \mathrm{d}s\Big)^2$$

with $C = 36$. Find the smallest C for which the inequality holds with any *non-random* step function X.

4.2. Let $f \in L^2[0, \infty)$, i.e. $f : [0, \infty) \to \mathbb{R}$ is a measurable (non-random) function with $\int_0^\infty f^2(t)\, dt < +\infty$. Without using Theorem 4.8a, show that

$$\mathbf{E}\left(B_t \int_0^\infty f(s)\, dB_s\right) = \int_0^t f(s)\, ds, \qquad t \geqslant 0.$$

4.3. For (non-random) functions $f \in C^1(\mathbb{R})$, prove the "usual" integration-by-parts formula

$$\int_0^t f(s)\, \mathrm{d}B_s = f(t)B_t - \int_0^t B_s f'(s)\, \mathrm{d}s.$$

4.4. Denote $\tau := \min\{t \geqslant 0 : B_t = 1\}$. Then $B_\tau = 1$ a.s., and thus $\mathbf{E}B_\tau = 1$. On the other hand,

$$B_\tau = \int_0^\tau dB_t = \int_0^\infty \mathbb{1}_{[0,\tau]}(t)\, dB_t,$$

and thus $\mathbf{E}B_\tau = 0$ (Theorem 4.8a or 4.8b). Is this a contradiction?

4.5. Give an example of an adapted random process X such that $\mathbf{E} \int_0^T X_t\, \mathrm{d}B_t \neq 0$.

4.6. Explain what is wrong in the equalities

$$1 = \mathbf{E}(B_1^2) = \mathbf{E}\Big(B_1 \int_0^1 dB_s\Big) = \mathbf{E}\Big(\int_0^1 B_1\, dB_s\Big) = 0.$$

4.7. Find the expectation and variance of the random variable

$$X = \textstyle\int_0^T (B_t + B_t^3)\, \mathrm{d}B_t.$$

4.8. Let $X_t = \displaystyle\int_0^t B_s^2\, \mathrm{d}B_s$ and $Y_t = \displaystyle\int_0^t X_s\, \mathrm{d}B_s$. Find $\mathbf{E}Y_t^2$ and $\mathbf{E}Y_t Y_s$.

4.9. Find the limit (the "backward" stochastic integral $\int_0^1 B_t\, \mathrm{d}B_t$)

$$I := L^2\text{-}\lim_{n\to\infty} \sum_{k=1}^n B\Big(\frac{k}{n}\Big)\Big[B\Big(\frac{k}{n}\Big) - B\Big(\frac{k-1}{n}\Big)\Big].$$

4.10. Prove the inequality

$$\mathbf{E}\Big|\int_0^t B_s^2\, dB_s\Big| \leqslant t^{3/2}, \quad t \geqslant 0.$$

Chapter 5

Itô's Formula

For finite-variation continuous processes X and functions $F \in C^1(\mathbb{R})$, recall the main formula of integration[1]

$$F\big(X_T\big) - F\big(X_0\big) = \int_0^T F'\big(X_t\big)\,\mathrm{d}X_t.$$

From example 4.10 we see that it does not hold for stochastic integrals. Its stochastic counterpart is the Itô formula which is proved below. For this, we need the Taylor[2] formula with the remainder in some special Peano-type[3] form.

LEMMA 5.1 (Taylor formula).– *Let $F \in C^2(\mathbb{R})$ have a uniformly continuous second derivative (for example, F has a bounded third derivative F'''). Then*

$$F(y) - F(x) = F'(x)(y - x) + \frac{1}{2}F''(x)(y - x)^2 + R(x, y),$$

where, the remainder term satisfies the estimate $|R(x, y)| \leqslant r(|y - x|)(y - x)^2$ with an increasing function $r \colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{h\downarrow 0} r(h) = 0$.

Proof. We use the Taylor formula with the well-known Lagrange remainder term

$$F(y) - F(x) = F'(x)(y - x) + \frac{1}{2}F''(\xi)(y - x)^2,$$

1. Also called the Newton–Leibnitz formula (Isaac Newton, Gottfried Wilhelm Leibnitz).
2. Brook Taylor.
3. Giuseppe Peano.

where the point $\xi = \xi(x, y)$ is between x and y ($x < \xi < y$ or $y < \xi < x$). From this we have

$$F(y) - F(x) = F'(x)(y-x) + \frac{1}{2}F''(x)(y-x)^2 + \frac{1}{2}\big(F''(\xi) - F''(x)\big)(y-x)^2.$$

Clearly,

$$\big|F''(\xi) - F''(x)\big| \leqslant \sup_{|\tilde{y}-\tilde{x}|\leqslant h} \big|F''(\tilde{y}) - F''(\tilde{x})\big|$$

if $|y - x| \leqslant h$ (and thus $|\xi - x| \leqslant h$). On the other hand, because of the uniform continuity of F'',

$$\sup_{|\tilde{y}-\tilde{x}|\leqslant h} \big|F''(\tilde{y}) - F''(\tilde{x})\big| \to 0, \quad h \to 0.$$

Therefore, the desired formula holds with

$$r(h) := \frac{1}{2} \sup_{|x-y|\leqslant h} \big|F''(y) - F''(x)\big|, \quad h > 0. \qquad \triangle$$

THEOREM 5.2 (Itô's formula for Brownian motion).– *If* $F \in C^2(\mathbb{R})$*, then*

$$F\big(B_T\big) - F\big(B_0\big) = \int_0^T F'\big(B_t\big)\,\mathrm{d}B_t + \frac{1}{2}\int_0^T F''\big(B_t\big)\,\mathrm{d}t.$$

REMARK.– The formula is often written in the formal differential form

$$\mathrm{d}F(B_t) = F'(B_t)\,\mathrm{d}B_t + \frac{1}{2}F''(B_t)\,\mathrm{d}t.$$

It can be considered as the second-order Taylor decomposition

$$\mathrm{d}F(B_t) = F'(B_t)\,\mathrm{d}B_t + \frac{1}{2}F''(B_t)\,(\mathrm{d}B_t)^2,$$

where the second-order differential $(\mathrm{d}B_t)^2$ is interpreted as $\mathrm{d}t$.

Proof. We first suppose that $F \in C_b^2(\mathbb{R})$ and that, moreover, the second derivative F'' is uniformly continuous. We shall use the just proved Taylor formula.

Let us take a sequence $\Delta^n = \{0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T\}$, $n \in \mathbb{N}$, of partitions of the interval $[0, T]$ such that $|\Delta^n| = \max_i \Delta t_i^n = \max_i |t_{i+1}^n - t_i^n| \to 0$,

$n \to \infty$. For short, we denote $B_i = B(t_i^n)$, $\Delta B_i = B_{i+1} - B_i$, $\Delta t_i = \Delta t_i^n$. By the Taylor formula from Lemma 5.1, we have

$$F(B_T) - F(B_0) = \sum_i \big(F(B_{i+1}) - F(B_i)\big)$$
$$= \sum_i F'(B_i)\Delta B_i + \frac{1}{2}\sum_i F''(B_i)\Delta B_i^2 + \sum_i R(B_i, B_{i+1}).$$

The random process $F'(B_t)$, $t \in [0,T]$, is L^2-continuous since (for example) by the Lagrange mean value theorem,[4] $\mathbf{E}(F'(B_s) - F'(B_t))^2 \leqslant \mathbf{E}[C(B_s - B_t)]^2 = C^2(s-t) \to 0$, $s \to t$, where $C := \sup_{x\in\mathbb{R}} |F''(x)| < \infty$. Therefore, applying Proposition 4.9 to this process, we get:

$$\sum_i F'(B_i)\Delta B_i \xrightarrow{L^2} \int_0^T F'(B_t)\,\mathrm{d}B_t, \quad n \to \infty.$$

Let us now show that the sum $\sum_i F''(B_i)\Delta B_i^2$ converges to $\int_0^T F''(B_t)\,\mathrm{d}t$. Decompose it into two sums:

$$\sum_i F''(B_i)\Delta B_i^2 = \sum_i F''(B_i)\Delta t_i + \sum_i F''(B_i)\big(\Delta B_i^2 - \Delta t_i\big).$$

The first sum, as a Riemann sum of a continuous function $F''(B_t)$, $t \in [0,T]$, converges to its integral $\int_0^T F''(B_t)\,\mathrm{d}t$ (almost surely and, thus, in probability). Let us check that the second sum converges to zero in mean square (and, thus, in probability):

$$\mathbf{E}\bigg(\sum_i F''(B_i)(\Delta B_i^2 - \Delta t_i)\bigg)^2$$
$$= \mathbf{E}\bigg(\sum_i F''(B_i)\big(\Delta B_i^2 - \Delta t_i\big)\sum_j F''(B_j)\big(\Delta B_j^2 - \Delta t_j\big)\bigg)$$
$$= \sum_i \mathbf{E}\Big(F''(B_i)\big(\Delta B_i^2 - \Delta t_i\big)\Big)^2$$
$$+ 2\sum_{i<j} \mathbf{E}\Big(F''(B_i)\big(\Delta B_i^2 - \Delta t_i)F''(B_j)\big(\Delta B_j^2 - \Delta t_j\big)\Big).$$

Since $F''(B_i) \in \mathcal{H}_{t_i^n}$, we have $F''(B_i) \perp\!\!\!\perp \Delta B_i^2 - \Delta t_i$, and since $F''(B_i) \in \mathcal{H}_{t_i^n} \subset \mathcal{H}_{t_j^n}$, $F''(B_j) \in \mathcal{H}_{t_j^n}$, and $\Delta B_i^2 - \Delta t_i \in \mathcal{H}_{t_{i+1}^n} \subset \mathcal{H}_{t_j^n}$ for $i < j$, we also have

$$F''(B_i)F''(B_j)(\Delta B_i^2 - \Delta t_i) \in \mathcal{H}_{t_j^n}, \quad i < j;$$

---

4. *If $f \in C^1[a,b]$, then there is $c \in [a,b]$ such that $f(b) - f(a) = f'(c)(b-a)$.*

hence,

$$F''(B_i)F''(B_j)(\Delta B_i^2 - \Delta t_i) \perp\!\!\!\perp \Delta B_j^2 - \Delta t_j, \quad i < j.$$

Therefore,

$$\begin{aligned}
&\mathbf{E}\Big(\sum_i F''(B_i)\big(\Delta B_i^2 - \Delta t_i\big)\Big)^2 \\
&\quad = \sum_i \mathbf{E}\big(F''(B_i)\big)^2 \mathbf{E}\big(\Delta B_i^2 - \Delta t_i\big)^2 \\
&\qquad + 2\sum_{i<j} \mathbf{E}\Big(F''(B_i)F''(B_j)\big(\Delta B_i^2 - \Delta t_i\big)\Big)\mathbf{E}\big(\Delta B_j^2 - \Delta t_j\big) \\
&\quad \leqslant C^2 \sum_i \mathbf{E}\big(\Delta B_i^2 - \Delta t_i\big)^2 + 0 = 2C^2 \sum_i \Delta t_i^2 \\
&\quad \leqslant 2C^2|\Delta^n| \sum_i \Delta t_i = 2TC^2|\Delta^n| \to 0, \quad n \to \infty.
\end{aligned}$$

It remains to check that $\sum_i R(B_i, B_{i+1}) \xrightarrow{\mathbf{P}} 0$:

$$\begin{aligned}
\Big|\sum_i R(B_i, B_{i+1})\Big| &\leqslant \sum_i r\big(|\Delta B_i|\big)\Delta B_i^2 \leqslant \max_i r\big(|\Delta B_i|\big) \sum_i \Delta B_i^2 \\
&\leqslant r\big(\max_i |\Delta B_i|\big) \sum_i \Delta B_i^2 \xrightarrow{\mathbf{P}} 0 \cdot T = 0, \quad n \to \infty.
\end{aligned}$$

Here we have used that, because of the continuity of paths of Brownian motion, $\max_i |\Delta B_i| \to 0$, $n \to \infty$, almost surely (and, thus, in probability) and $\sum_i \Delta B_i^2 \xrightarrow{\mathbf{P}} T$ (Theorem 4.1).

Now we pass to the general case, without the boundedness restrictions on $F \in C^2(\mathbb{R})$ and its derivatives. Below, we shall show that we can construct a sequence of functions $F_n \in C_b^2(\mathbb{R})$, $n \in \mathbb{N}$, that have uniformly continuous second derivatives and (most important) coincide with F in the interval $[-n, n]$, i.e. $F_n(x) = F(x)$ for $x \in [-n, n]$. For such functions, the Itô formula is already proved, and thus, we may write:

$$F_n\big(B_T\big) - F_n\big(B_0\big) = \int_0^T F_n'\big(B_t\big)\,\mathrm{d}B_t + \frac{1}{2}\int_0^T F_n''\big(B_t\big)\,\mathrm{d}t, \quad n \in \mathbb{N}.$$

Since $F_n(B_t) = F(B_t)$, $t \in [0,T]$, on the event $\Omega_n = \{\sup_{t\in[0,T]} |B_t| \leqslant n\}$, by Lemma 4.11 it follows that, on this event,

$$F\big(B_T\big) - F\big(B_0\big) = \int_0^T F'\big(B_t\big)\,\mathrm{d}B_t + \frac{1}{2}\int_0^T F''\big(B_t\big)\,\mathrm{d}t.$$

Therefore, the Itô formula holds on the event $\widetilde{\Omega} = \bigcup_{n\in\mathbb{N}} \Omega_n$. It remains for us to note that $\widetilde{\Omega}$ is a certain event and thus the Itô formula holds with probability one.

$\mathcal{M}$ To complete the proof, it remains to us to construct a promised sequence of functions $\{F_n\}$. For every $n \in \mathbb{N}$, we define:

$$\varphi_n(x) = \frac{F(n) + F'(n)(x-n) + \frac{1}{2}F''(n)(x-n)^2}{1+(x-n)^2},$$

$$\psi_n(x) = \frac{F(-n) + F'(-n)(x+n) + \frac{1}{2}F''(-n)(x+n)^2}{1+(x+n)^2}.$$

The functions φ_n and ψ_n are chosen so that they belong to $C_b^\infty(\mathbb{R})$ (i.e. they have bounded derivatives of all orders) and that their values and the first- and second-order derivatives coincide with those of F at the ends of the intervals $[-n,n]$, i.e.

$$\begin{aligned} &\varphi_n(n) = F(n), && \varphi_n'(n) = F'(n), && \varphi_n''(n) = F''(n),\\ &\psi_n(-n) = F(-n), && \psi_n'(-n) = F'(-n), && \psi_n''(-n) = F''(-n). \end{aligned}$$

These values coincide since, in the numerators, we have taken the corresponding Taylor second-order polynomials of F at the points $\pm n$. To ensure the boundedness conditions, we divide them by square polynomials that do not change their values and the first two derivatives at the points $x = \pm n$.

Now, let us "glue" the function F at the points $x = \pm n$ with the functions φ_n ir ψ_n (see Figure 5.1, where $F(x) = (x^3+1)/10$, $n = 4$):

$$F_n(x) := \begin{cases} F(x), & x \in [-n,n],\\ \varphi_n(x), & x > n,\\ \psi_n(x), & x < -n. \end{cases}$$

Clearly, the functions F_n satisfy all the required conditions. △

EXAMPLE 5.3.– Let us apply Itô's formula to the function $F(x) = x^{n+1}$ ($n \in \mathbb{N}$):

$$B_T^{n+1} = (n+1)\int_0^T B_t^n\,\mathrm{d}B_t + \frac{1}{2}(n+1)n\int_0^T B_t^{n-1}\,\mathrm{d}t.$$

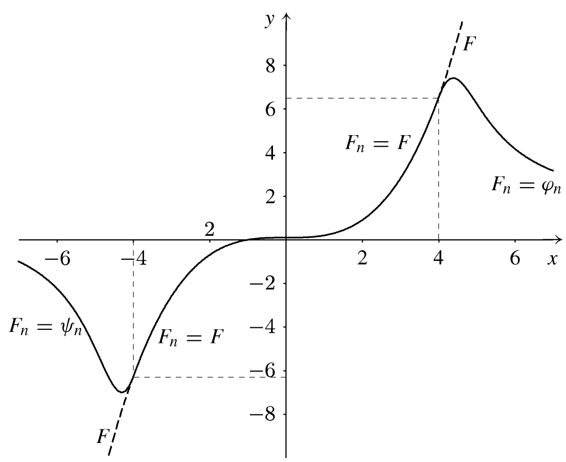


Figure 5.1. *Approximation of* $F \in C^2(\mathbb{R})$ *by* $F_n \in C_b^2(\mathbb{R})$

Hence,

$$\int_0^T B_t^n \,\mathrm{d}B_t = \frac{B_T^{n+1}}{n+1} - \frac{n}{2}\int_0^T B_t^{n-1}\,\mathrm{d}t, \quad n \in \mathbb{N}.$$

The Itô formula can easily be generalized to the case where $F = F(t,x)$, $t \in [0,T]$, $x \in \mathbb{R}$, also depends on time t.

THEOREM 5.4.– *If* $F \in C^2([0,T]\times\mathbb{R})$*, then*

$$F(T,B_T) - F(0,B_0) = \int_0^T F'_x(t,B_t)\,\mathrm{d}B_t + \int_0^T F'_t(t,B_t)\,\mathrm{d}t + \frac{1}{2}\int_0^T F''_{xx}(t,B_t)\,\mathrm{d}t.$$

REMARK.– As in Theorem 5.2, this formula can be obtained by "integrating" the formal Taylor decomposition

$$\begin{aligned}\mathrm{d}F(t,B_t) &= F'_x(t,B_t)\,\mathrm{d}B_t + F'_t(t,B_t)\,\mathrm{d}t \\ &\quad + \frac{1}{2}\big(F''_{xx}(t,B_t)\,(\mathrm{d}B_t)^2 + 2F''_{xt}(t,B_t)\,\mathrm{d}B_t\,\mathrm{d}t + F''_{tt}(t,B_t)\,(\mathrm{d}t)^2\big)\end{aligned}$$

if we settle to use the following multiplication rules for differentials:

$$(\mathrm{d}B_t)^2 = \mathrm{d}t, \quad \mathrm{d}B_t\,\mathrm{d}t = 0, \quad (\mathrm{d}t)^2 = 0.$$

Proof. Since the proof is similar to that of Theorem 5.2, we indicate only the main differences. As before, the proof can be reduced to the case of the function $F \in C^2([0,T] \times \mathbb{R})$ with *uniformly* continuous second-order derivatives. For such a function, the Taylor formula is as follows:

$$\begin{aligned}
&F(s,y) - F(t,x) \\
&\quad = F'_x(t,x)(y-x) + F'_t(t,x)(s-t) \\
&\qquad + \frac{1}{2}\big[F''_{xx}(t,x)(y-x)^2 + 2F''_{xt}(t,x)(y-x)(s-t) + F''_{tt}(t,x)(s-t)^2\big] \\
&\qquad + R(x,t,y,s),
\end{aligned}$$

where the remainder term satisfies the estimate

$$|R(x,t,y,s)| \leqslant r(|y-x| + |s-t|)((y-x)^2 + (s-t)^2)$$

with an increasing function $r\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{h\downarrow 0} r(h) = 0$. Now, the decomposition of the difference $F(T,B_T) - F(0,B_0)$ by a "telescopic" sum is as follows:

$$\begin{aligned}
&F(T,B_T) - F(0,B_0) \\
&\quad = \sum_i \big(F(t_{i+1},B_{i+1}) - F(t_i,B_i)\big) \\
&\quad = \sum_i F'_x(t_i,B_i)\Delta B_i + \sum_i F'_t(t_i,B_i)\Delta t_i \\
&\qquad + \frac{1}{2}\Big[\sum_i F''_{xx}(t_i,B_i)\Delta B_i^2 + 2\sum_i F''_{xt}(t_i,B_i)\Delta B_i \Delta t_i \\
&\qquad\qquad + \sum_i F''_{tt}(t_i,B_i)\Delta t_i^2\Big] + \sum_i R(B_i,t_i,B_{i+1},t_{i+1}).
\end{aligned}$$

As in the proof of Theorem 5.2, we have

$$\sum_i F'_x(t_i,B_i)\Delta B_i \xrightarrow{L^2} \int_0^T F'_x(t,B_t)\,\mathrm{d}B_t,$$

$$\sum_i F''_{xx}(t_i,B_i)\Delta B_i^2 \xrightarrow{\mathbf{P}} \int_0^T F''_{xx}(t,B_t)\,\mathrm{d}t,$$

$$\sum_i R(B_i,t_i,B_{i+1},t_{i+1}) \xrightarrow{\mathbf{P}} 0.$$

It remains to check the convergence of the remaining three sums that additionally appear in the decomposition:

$$\sum_i F_t'(t_i, B_i)\Delta t_i \to \int_0^T F_t'(t, B_t)\,\mathrm{d}t,$$

$$\frac{1}{2}\sum_i F_{tt}''(t_i, B_i)\Delta t_i^2 \to 0,$$

$$\sum_i F_{tx}''(t_i, B_i)\Delta B_i \Delta t_i \to 0.$$

The convergence of the first sum is clear. The second can be easily estimated:

$$\left|\sum_i F_{tt}''(t_i, B_i)\Delta t_i^2\right| \leqslant \sum_i \left|F_{tt}''(t_i, B_i)\right|\Delta t_i \cdot |\Delta^n| \to \int_0^T \left|F_{tt}''(t, B_t)\right|\mathrm{d}t \cdot 0 = 0.$$

The third sum can be similarly estimated:

$$\left|\sum_i F_{tx}''(t_i, B_i)\Delta B_i \Delta t_i\right| \leqslant \sum_i \left|F_{tx}''(t_i, B_i)\right|\Delta t_i \cdot \max_i |\Delta B_i|$$

$$\to \int_0^T \left|F_{tx}''(t, B_t)\right|\mathrm{d}t \cdot 0 = 0. \qquad \triangle$$

5.1. Exercises

5.1. Find the stochastic integrals

1) $\int_0^t \cos B_s\,dB_s,\ t \geqslant 0;$

2) $\int_0^t B_s e^{B_s}\,dB_s;$

3) $\int_0^t \sin(B_s + 2s)\,dB_s,\ t \geqslant 0.$

5.2. Find $\mathbf{E}\cos(2B_t),\ t \geqslant 0.$

5.3. Let $X \in H^2[0,T]$ with $\mathbf{E}\int_0^T X_t^2\,\mathrm{d}t > 0$. Can the process $\int_0^t X_s\,\mathrm{d}B_s$, $t \in [0,T]$, be bounded (i.e. $\left|\int_0^t X_s\,\mathrm{d}B_s\right| \leqslant C$, $t \in [0,T]$, with a *non-random* constant C)?

5.4. Let

$$X_t := \int_0^t \frac{\mathrm{d}B_s}{1+B_s^2} - \int_0^t \frac{B_s\,\mathrm{d}s}{(1+B_s^2)^2}, \qquad t \geqslant 0.$$

Show that $\mathbf{E}|X_t| < \pi/2$, $t \geqslant 0$.

Chapter 6

Stochastic Differential Equations

Now, we can already make sense of the integral equation

$$X_t = x_0 + \int_0^t b(X_s, s)\,\mathrm{d}s + \int_0^t \sigma(X_s, s)\,\mathrm{d}B_s, \quad t \in I, \qquad [6.1]$$

derived in Chapter 3 (most often, $I = [0,T]$ or $[0,\infty)$). Though partially for historical reasons, but rather for practical convenience, it is usually written in the formal differential form

$$\mathrm{d}X_t = b(X_t, t)\,\mathrm{d}t + \sigma(X_t, t)\,\mathrm{d}B_t, \quad X_0 = x_0. \qquad [6.2]$$

Both equations are called *stochastic differential equations*, though here the word *integral* would fit better. A formal definition is as follows.

DEFINITION 6.1.– *A continuous random process X_t, $t \in I$, is a solution of the stochastic differential equation* [6.2] *in an interval I if, for all $t \in I$, it satisfies equation* [6.1] *with probability one.*

EXAMPLE 6.2.– Consider the stochastic differential equation

$$\mathrm{d}X_t = \mu\, X_t\,\mathrm{d}t + \sigma\, X_t\,\mathrm{d}B_t, \quad X_0 = x_0.$$

This equation describes, for example, the dynamics of stock price in financial mathematics, or the development of some populations in biology. Let us check that the random process $X_t = x_0 \exp\{(\mu - \sigma^2/2)t + \sigma B_t\}$, $t \geqslant 0$, is its solution. Applying the Itô rule (Theorem 5.2) to the function $F(t,x) = x_0 \exp\{(\mu - \sigma^2/2)t + \sigma x\}$, we have:

$$
\begin{aligned}
X_t &= F(t, B_t) \\
&= F(0,0) + \int_0^t F'_x(s, B_s)\,\mathrm{d}B_s + \int_0^t F'_s(s, B_s)\,\mathrm{d}s + \frac{1}{2}\int_0^t F''_{xx}(s, B_s)\,\mathrm{d}s \\
&= x_0 + \int_0^t x_0 \sigma \exp\{(\mu - \sigma^2/2)s + \sigma B_s\}\,\mathrm{d}B_s \\
&\quad + \int_0^t x_0 (\mu - \sigma^2/2) \exp\{(\mu - \sigma^2/2)s + \sigma B_s\}\,\mathrm{d}s \\
&\quad + \frac{1}{2}\int_0^t x_0 \sigma^2 \exp\{(\mu - \sigma^2/2)s + \sigma B_s\}\,\mathrm{d}s \\
&= x_0 + \int_0^t \mu x_0 \exp\{(\mu - \sigma^2/2)s + \sigma B_s\}\,\mathrm{d}s \\
&\quad + \int_0^t \sigma x_0 \exp\{(\mu - \sigma^2/2)s + \sigma B_s\}\,\mathrm{d}B_s \\
&= x_0 + \int_0^t \mu X_s\,\mathrm{d}s + \int_0^t \sigma X_s\,\mathrm{d}B_s, \quad t \geqslant 0.
\end{aligned}
$$

In particular, for $\mu = 0$ and $\sigma = 1$, we get that the random process

$$X_t = x_0 \exp\{B_t - t/2\}, \quad t \geqslant 0,$$

is a solution of the stochastic differential equation

$$\mathrm{d}X_t = X_t\,\mathrm{d}B_t, \qquad X_0 = x_0.$$

Therefore, the process X_t is sometimes called the *stochastic exponent* of a Brownian motion. (If B were not a Brownian motion but a continuously differentiable process, the exponent $X_t = x_0 \mathrm{e}^{B_t}$ would be a solution of the equation.)

We will further assume that the coefficients of the equations satisfy the following Lipschitz[1] and linear growth conditions:

(1) $|b(x,t)-b(y,t)|^2+|\sigma(x,t)-\sigma(y,t)|^2 \leqslant C|x-y|^2,\ x,y\in\mathbb{R},\ t\in\mathbb{R}_+$;

(2) $|b(x,t)|^2+|\sigma(x,t)|^2 \leqslant C(1+x^2),\ x\in\mathbb{R},\ t\in\mathbb{R}_+$.

REMARK.– If the coefficients b and σ do not depend on t, the second condition follows from the first one:

$$\begin{aligned}|b(x)|^2+|\sigma(x)|^2 &\leqslant 2\big(|b(x)-b(0)|^2+|b(0)|^2\big)+2\big(|\sigma(x)-\sigma(0)|^2+|\sigma(0)|^2\big)\\ &\leqslant 2|b(0)|^2+2|\sigma(0)|^2+2Cx^2\leqslant \widetilde{C}(1+x^2),\quad x\in\mathbb{R},\ t\in\mathbb{R}_+,\end{aligned}$$

where $\widetilde{C}=\max\{2C, 2|b(0)|^2+2|\sigma(0)|^2\}$.

In the proof of a general theorem on the existence and uniqueness of a solution of a stochastic differential equation, we will need the Gronwall[2] lemma, various versions of which are widely used in the theory of differential equations.

LEMMA 6.3 (Gronwall lemma).– *Let a non-negative function $\varphi\in C[0,T]$ satisfy the inequality*

$$\varphi(t)\leqslant C\int_0^t \varphi(s)\,\mathrm{d}s+\varepsilon,\quad t\in[0,T],$$

with some non-negative constants C and ε. Then

$$\varphi(t)\leqslant \varepsilon \mathrm{e}^{Ct},\quad t\in[0,T].$$

In particular, if $0\leqslant\varphi(t)\leqslant C\int_0^t\varphi(s)\,\mathrm{d}s$, $t\in[0,T]$, then $\varphi\equiv 0$.

Proof. First, suppose that $\varepsilon>0$. From the equality

$$\left[\ln\left(C\int_0^t\varphi(s)\,\mathrm{d}s+\varepsilon\right)\right]'=\frac{C\varphi(t)}{C\int_0^t\varphi(s)\,\mathrm{d}s+\varepsilon}\leqslant C,\quad t\in[0,T],$$

1. Rudolf Otto Sigismund Lipschitz.

2. Tomas Hakon Gronwall.

we have

$$\ln\left(C\int_0^t \varphi(s)\,\mathrm{d}s+\varepsilon\right)-\ln\varepsilon=\int_0^t\left[\ln\left(C\int_0^s\varphi(u)\,\mathrm{d}u+\varepsilon\right)\right]'\mathrm{d}s$$

$$\leqslant\int_0^t C\,\mathrm{d}s=Ct,\quad t\in[0,T].$$

Therefore,

$$\varphi(t)\leqslant C\int_0^t\varphi(s)\,\mathrm{d}s+\varepsilon=\mathrm{e}^{\ln(C\int_0^t\varphi(s)\,\mathrm{d}s+\varepsilon)}\leqslant\mathrm{e}^{Ct+\ln\varepsilon}=\varepsilon\mathrm{e}^{Ct},\quad t\in[0,T].$$

In the case $\varepsilon=0$, by the preceding we find that the last estimate holds for all $\varepsilon>0$. Passing to the limit as $\varepsilon\downarrow 0$, we get $\varphi(t)\leqslant 0$ and thus $\varphi(t)=0$ for all $t\in[0,T]$. △

THEOREM 6.4.– *If conditions (1) and (2) are satisfied, then there exists a unique continuous process X that satisfies the stochastic differential equation* [6.1] *in the interval* $[0,\infty)$.

Proof. Uniqueness. We begin with the proof of uniqueness. For every continuous random process X, we denote by AX the random process defined by the right-hand side of equation [6.1], i.e.

$$AX_t:=x_0+\int_0^t b(X_s,s)\,\mathrm{d}s+\int_0^t\sigma(X_s,s)\,\mathrm{d}B_s,\quad t\geqslant 0.$$

Clearly, a random process X is a solution of equation [6.1] in the interval $[0,\infty)$ if and only if $X=AX$, i.e. $X_t=AX_t$, $t\geqslant 0$ (a.s.).

Now, consider two arbitrary solutions X and Y of equation [6.1]. Then $X=AX$ and $Y=AY$. Let us estimate the difference between X and Y:

$$\begin{aligned}
&\mathbf{E}|X_t-Y_t|^2=\mathbf{E}|AX_t-AY_t|^2\\
&\quad=\mathbf{E}\left[\int_0^t\big(b(X_s,s)-b(Y_s,s)\big)\,\mathrm{d}s+\int_0^t\big(\sigma(X_s,s)-\sigma(Y_s,s)\big)\,\mathrm{d}B_s\right]^2\\
&\quad\leqslant 2\mathbf{E}\left[\int_0^t\big(b(X_s,s)-b(Y_s,s)\big)\,\mathrm{d}s\right]^2+2\mathbf{E}\left[\int_0^t\big(\sigma(X_s,s)-\sigma(Y_s,s)\big)\,\mathrm{d}B_s\right]^2
\end{aligned}$$

$$\leqslant 2t\mathbf{E}\int_0^t \big(b(X_s,s)-b(Y_s,s)\big)^2\,\mathrm{d}s + 2\mathbf{E}\int_0^t \big(\sigma(X_s,s)-\sigma(Y_s,s)\big)^2\,\mathrm{d}s$$

$$\leqslant 2Ct\mathbf{E}\int_0^t |X_s-Y_s|^2\,\mathrm{d}s + 2C\mathbf{E}\int_0^t |X_s-Y_s|^2\,\mathrm{d}s$$

$$= 2C(t+1)\mathbf{E}\int_0^t |X_s-Y_s|^2\,\mathrm{d}s = M_t\int_0^t \mathbf{E}|X_s-Y_s|^2\,\mathrm{d}s,\quad t\geqslant 0,$$

where $M_t := 2C(t+1)$.

Applying the Gronwall lemma to the non-negative function $\varphi(t) := \mathbf{E}|X_t-Y_t|^2$, $t\in[0,T]$, with $\varepsilon = 0$ and the constant M_T instead of C, we find that $\varphi(t) = 0$, $t\in[0,T]$, i.e. $X_t = Y_t$ for almost all $t\in[0,T]$. Since in this instance T is arbitrary, $X_t = Y_t$ for all $t\geqslant 0$.

REMARKS.– Here, we would like to insert a couple of important remarks:

1. In the proof of uniqueness, there is one “but”: we cannot be sure that the function φ takes finite values! To overcome this “obstacle”, we shall use the so-called stopping of random processes, a usual tool in similar situations. For any constant $c>0$, we define the stopping time $\tau_c := \min\{t\geqslant 0\colon |X_t-Y_t| = c\}$ and random processes $\widetilde{X}_t := X_{t\wedge\tau_c}$ and $\widetilde{Y}_t := Y_{t\wedge\tau_c}$ (that is, the processes X and Y that are stopped at the moment τ_c). Then, for the processes $\widetilde{X}$ and $\widetilde{Y}$, as in the inequality just proved, we obtain:

$$\begin{aligned}
\mathbf{E}\big|\widetilde{X}_t-\widetilde{Y}_t\big|^2 &= \mathbf{E}\Bigg[\int_0^{t\wedge\tau_c}\big(b(X_s,s)-b(Y_s,s)\big)\,\mathrm{d}s \\
&\qquad + \int_0^{t\wedge\tau_c}\big(\sigma(X_s,s)-\sigma(Y_s,s)\big)\,\mathrm{d}B_s\Bigg]^2 \\
&= \mathbf{E}\Bigg[\int_0^t \mathbb{1}_{\{s\leqslant\tau_c\}}\big(b(X_s,s)-b(Y_s,s)\big)\,\mathrm{d}s \\
&\qquad + \int_0^t \mathbb{1}_{\{s\leqslant\tau_c\}}\big(\sigma(X_s,s)-\sigma(Y_s,s)\big)\,\mathrm{d}B_s\Bigg]^2
\end{aligned}$$

$$\leqslant \cdots \leqslant M_t \int_0^t \mathbf{E}\big(\mathbb{1}_{\{s\leqslant\tau_c\}}|X_s - Y_s|^2\big)\,\mathrm{d}s$$

$$= M_t \int_0^t \mathbf{E}\big(\mathbb{1}_{\{s\leqslant\tau_c\}}|\widetilde{X}_s - \widetilde{Y}_s|^2\big)\,\mathrm{d}s$$

$$\leqslant M_t \int_0^t \mathbf{E}|\widetilde{X}_s - \widetilde{Y}_s|^2\,\mathrm{d}s, \quad t \geqslant 0.$$

The function $\tilde{\varphi}(t) := \mathbf{E}|\widetilde{X}_t - \widetilde{Y}_t|^2$, $t \geqslant 0$, analogous to the previously defined function φ, is already bounded since, by the definition of $\widetilde{X}$ and $\widetilde{Y}$, we have $\tilde{\varphi}(t) \leqslant c^2$, $t \geqslant 0$. Therefore, now we can legally apply the Gronwall lema and obtain $\tilde{\varphi}(t) = 0$ for all $t \geqslant 0$. Thus,

$$\mathbf{E}\big|\widetilde{X}_t - \widetilde{Y}_t\big|^2 = \mathbf{E}|X_{t\wedge\tau_c} - Y_{t\wedge\tau_c}|^2 = 0, \quad t \geqslant 0,$$

for every $c > 0$. Since $\tau_c \to \infty$ and thus $X_{t\wedge\tau_c} \to X_t$ and $Y_{t\wedge\tau_c} \to Y_t$ as $c \to \infty$, passing to the limit in the last equality as $c \to \infty$, we get:

$$\mathbf{E}|X_t - Y_t|^2 = 0, \quad t \geqslant 0,$$

i.e. $X_t = Y_t$ almost surely for all $t \geqslant 0$.

2. To be precise, there is a rather subtle difference between the equality $X_t = Y_t$ *almost surely for all* $t \in [0, T]$ and the equality $X_t = Y_t$ *for all* $t \in [0, T]$ *almost surely*. In the first case, the zero probability event on which $X_t \neq Y_t$ may depend on t, and the union of all such events over $t \in [0, T]$ is not necessarily a zero probability event (since $[0, T]$ is not a countable set); therefore, we may not claim that almost surely $X_t = Y_t$ for all $t \in [0, T]$. However, if the processes X and Y are continuous (as in the case considered), then the statements that $X_t = Y_t$ a.s., $t \in [0, T]$, and $X_t = Y_t$, $t \in [0, T]$, a.s. are equivalent, since for the coincidence of two continuous processes, their coincidence is sufficient on the set of rational t.

Existence. We shall apply the Picard[3] iteration method, which is well known in the theory of differential equations, and construct a sequence of random processes converging to a solution of the equation. Let $X_t^0 = x_0$, $t \geqslant 0$, and define the sequence $\{X^n\}$ recurrently by the equality $X^{n+1} = AX^n$, $n \in \mathbb{N}_+$. Then $X^{n+1} - X^n = AX^n - AX^{n-1}$. Using the inequality obtained in the proof of uniqueness, we have:

$$\mathbf{E}\big|X_t^{n+1} - X_t^n\big|^2 = \mathbf{E}\big|AX_t^n - AX_t^{n-1}\big|^2 \leqslant M_t \int_0^t \mathbf{E}\big|X_s^n - X_s^{n-1}\big|^2\,\mathrm{d}s, \quad t \geqslant 0.$$

3. Charles Émile Picard.

Using this inequality once more, we get:

$$\begin{aligned}\mathbf{E}\big|X_t^{n+1} - X_t^n\big|^2 &\leqslant M_t \int_0^t \mathbf{E}\big|X_s^n - X_s^{n-1}\big|^2 \,\mathrm{d}s \\ &\leqslant M_t^2 \int_0^t\int_0^s \mathbf{E}\big|X_u^{n-1} - X_u^{n-2}\big|^2 \,\mathrm{d}u\,\mathrm{d}s, \quad t \geqslant 0.\end{aligned}$$

Continuing, we have:

$$\mathbf{E}\big|X_t^{n+1} - X_t^n\big|^2 \leqslant M_t^n \underbrace{\int_0^t\int_0^{s_1}\dots\int_0^{s_{n-1}}}_{n} \mathbf{E}\big|X_{s_n}^1 - X_{s_n}^0\big|^2 \,\mathrm{d}s_n \dots \mathrm{d}s_2\,\mathrm{d}s_1, \quad n \in \mathbb{N}.$$

Now, by the linear growth condition, we have:

$$\begin{aligned}\mathbf{E}\big|X_t^1 - X_t^0\big|^2 &= \mathbf{E}\left(\int_0^t b(x_0, s)\,\mathrm{d}s + \int_0^t \sigma(x_0, s)\,\mathrm{d}B_s\right)^2 \\ &\leqslant 2\mathbf{E}\left(\int_0^t b(x_0, s)\,\mathrm{d}s\right)^2 + 2\mathbf{E}\left(\int_0^t \sigma(x_0, s)\,\mathrm{d}B_s\right)^2 \\ &\leqslant 2\left(t\int_0^t |b(x_0, s)|^2\,\mathrm{d}s + \int_0^t |\sigma(x_0, s)|^2\,\mathrm{d}s\right) \\ &\leqslant 2Ct(t+1)\big(1 + x_0^2\big) = \widetilde{M}_t\end{aligned}$$

with the constant $\widetilde{M}_t = tM_t(1 + x_0^2)$. Substituting this into the previous inequality, for all $n \in \mathbb{N}$, we have:

$$\begin{aligned}\mathbf{E}\big|X_t^{n+1} - X_t^n\big|^2 &\leqslant \widetilde{M}_t M_t^n \underbrace{\int_0^t\int_0^{s_1}\dots\int_0^{s_{n-1}}}_{n} \mathrm{d}s_n \dots \mathrm{d}s_2\,\mathrm{d}s_1 \\ &= \frac{\widetilde{M}_t M_t^n t^n}{n!}, \quad t \geqslant 0. \qquad [6.3]\end{aligned}$$

From the latter inequality we see that

$$\sum_{n=1}^{\infty} \|X_t^{n+1} - X_t^n\|_{L^2} = \sum_{n=1}^{\infty} \left(\mathbf{E}|X_t^{n+1} - X_t^n|^2\right)^{1/2}$$
$$\leqslant \sum_{n=1}^{\infty} \left(\frac{\widetilde{M}_t (M_t t)^n}{n!}\right)^{1/2} < \infty, \quad t \geqslant 0,$$

that is, the series $\sum_{n=1}^{\infty}(X_t^{n+1} - X_t^n)$ converges in the mean square sense (section 1.10). Denote $X_t = X_t^0 + \sum_{n=0}^{\infty}(X_t^{n+1} - X_t^n)$ (in the mean square sense). Then

$$X_t = L^2\text{-}\lim_{n\to\infty}\left(X_t^0 + \sum_{k=0}^{n-1}\left(X_t^{k+1} - X_t^k\right)\right) = L^2\text{-}\lim_{n\to\infty} X_t^n.$$

Thus, the sequence of Picard iterations $\{X_t^n\}$ has a limit X_t (in the mean square sense) at all $t \geqslant 0$. Passing formally to the limit in the equality $X_t^{n+1} = AX_t^n$ as $n \to \infty$, we get the equality $X_t = AX_t$, $t \geqslant 0$, which means that X is a solution of equation [6.1]. A rigorous justification of this passing to the limit is more complicated. Again from the inequality in the proof of uniqueness we have the inequality

$$\mathbf{E}\left[\int_0^t \big(b(X_s, s) - b(X_s^n, s)\big)\,\mathrm{d}s + \int_0^t \big(\sigma(X_s, s) - \sigma(X_s^n, s)\big)\,\mathrm{d}B_s\right]^2$$
$$\leqslant M_t \int_0^t \mathbf{E}\big|X_s - X_s^n\big|^2\,\mathrm{d}s, \quad t \geqslant 0.$$

Therefore,

$$\mathbf{E}\left[X_t - x_0 - \int_0^t b(X_s, s)\,\mathrm{d}s - \int_0^t \sigma(X_s, s)\,\mathrm{d}B_s\right]^2$$
$$= \mathbf{E}\Bigg[X_t - X_t^n + \int_0^t \Big(b\big(X_s^{n-1}, s\big) - b(X_s, s)\Big)\,\mathrm{d}s$$
$$+ \int_0^t \Big(\sigma\big(X_s^{n-1}, s\big) - \sigma(X_s, s)\Big)\,\mathrm{d}B_s\Bigg]^2$$
$$\leqslant 2\mathbf{E}\big|X_t - X_t^n\big|^2 + 2\mathbf{E}\Bigg[\int_0^t \Big(b\big(X_s^{n-1}, s\big) - b(X_s, s)\Big)\,\mathrm{d}s$$
$$+ \int_0^t \Big(\sigma\big(X_s^{n-1}, s\big) - \sigma(X_s, s)\Big)\,\mathrm{d}B_s\Bigg]^2$$

$$\leqslant 2\mathbf{E}|X_t - X_t^n|^2 + 2M_t \int_0^t \mathbf{E}\big|X_s - X_s^{n-1}\big|^2\,\mathrm{d}s \to 0, \quad t \geqslant 0,\ n \in \mathbb{N}.$$

The first term in the right-hand side of the inequality, $2\mathbf{E}|X_t - X_t^n|^2 \to 0$ as $n \to \infty$. The second term, the integral $2M_t \int_0^t \mathbf{E}|X_s - X_s^{n-1}|^2\,\mathrm{d}s$, also converges to zero, though this is not so obvious.[4] This follows from the estimate obtained by using inequality [6.3]:

$$\mathbf{E}\big|X_s - X_s^{n-1}\big|^2 = \big(\|X_s - X_s^{n-1}\|_{L^2}\big)^2 = \Bigg\|\sum_{k=n-1}^{\infty}\big(X_s^{k+1} - X_s^k\big)\Bigg\|_{L^2}^2$$

$$\leqslant \Bigg(\sum_{k=n-1}^{\infty}\big\|X_s^{k+1} - X_s^k\big\|_{L^2}\Bigg)^2 \leqslant \varepsilon_s^n := \Bigg[\sum_{k=n-1}^{\infty}\bigg(\frac{\widetilde{M}_s(M_s s)^k}{k!}\bigg)^{1/2}\Bigg]^2.$$

Indeed, since the term ε_s^n on the right-hand side is an increasing function of s, we have $\int_0^t \mathbf{E}|X_s - X_s^{n-1}|^2\,\mathrm{d}s \leqslant \int_0^t \varepsilon_s^n\,\mathrm{d}s \leqslant \int_0^t \varepsilon_t^n\,\mathrm{d}s = \varepsilon_t^n t \to 0$, $n \to \infty$, for all $t \geqslant 0$. Hence, $\mathbf{E}(X_t - x_0 - \int_0^t b(X_s, s)\,\mathrm{d}s - \int_0^t \sigma(X_s, s)\,\mathrm{d}B_s)^2 = 0$ for all $t \geqslant 0$, i.e. the random process X_t, $t \geqslant 0$, is a solution of equation [6.1]. △

6.1. Exercises

6.1. Check that the process $X_t := \operatorname{tg}(\pi/4 + B_t)$, $t \geqslant 0$, is a solution of the SDE $dX_t = X_t(1 + X_t^2)\,dt + (1 + X_t^2)\,dB_t$ with $X_0 = 1$.

6.2. Write an SDE for the process $X_t = \mathrm{e}^{B_t}$, $t \geqslant 0$.

6.3. Let X_t, $t \geqslant 0$, be the solution of the SDE

$$dX_t = 2X_t\,dt + \cos^2 X_t\,dB_t, \qquad X_0 = 1.$$

Find $\mathbf{E}X_t$, $t \geqslant 0$.

6.4. Denote by X_t^x, $t \geqslant 0$, the solution of an SDE

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t, \quad X_0 = x,$$

with Lipschitz coefficients b and σ. Show that, for all $x \in \mathbb{R}$, $X_t^x \to X_t^{x_0}$ in probability as $x \to x_0$ (continuity with respect to initial condition).

6.5. Write an SDE that is satisfied by the random process $X_t = B_t^3$, $t \geqslant 0$. Is it a unique solution of the equation?

4. Since for this, it does not suffice that the integrand function $\mathbf{E}|X_s - X_s^{n-1}|^2 \to 0$ for all $s \in [0, t]$!

Chapter 7

Itô Processes

DEFINITION 7.1.– *Adapted random process* $M = \{M_t, t \geqslant 0\}$ *is called a martingale if*

$$\mathbf{E}\big(Z(M_s - M_t)\big) = 0 \qquad [7.1]$$

for all bounded random variables $Z \in \mathcal{H}_t$ *and* $s \geqslant t \geqslant 0$.

REMARKS.– 1. Random variables Z_1 and Z_2 are said to be orthogonal if $\mathbf{E}(Z_1 Z_2) = 0$. Therefore, property [7.1] can be interpreted as follows: a random process M is a martingale if its increments $M_s - M_t$, $s \geqslant t \geqslant 0$, are orthogonal to the past $\mathcal{H}_t$ (i.e. to all bounded random variables $Z \in \mathcal{H}_t$). A Brownian motion B is a martingale: recall that its increments are not only *orthogonal to* but, moreover, *independent of* the past. Many properties of Brownian motion can be generalized to martingales since their proofs are actually based on the orthogonality of the increments rather than on their independence.

2. In equality [7.1], the requirement of the boundedness of Z often is unnecessary, provided that the product $Z(M_s - M_t)$ has a finite expectation. For example, if $\mathbf{E}M_t^2 < \infty$, $t \geqslant 0$, then from equality [7.1] with bounded $Z \in \mathcal{H}_t$ it follows also with $Z \in \mathcal{H}_t$ having a finite second moment ($\mathbf{E}Z^2 < \infty$).

3. Most often, a martingale is defined using the notion of the conditional expectation of a random variable with respect to a σ-algebra. Let X be a random variable, and $\mathcal{G} \subset \mathcal{F}$ a σ-algebra. Recall that (section 1.6) a random variable Y is called the expectation of X with respect to the σ-algebra $\mathcal{G}$ if it is $\mathcal{G}$-measurable and

$$\mathbf{E}(ZX) = \mathbf{E}(ZY)$$

for every bounded $\mathcal{G}$-measurable random variable Z. It is denoted by $\mathbf{E}(X|\mathcal{G})$. Then property [7.1] defining a martingale can be written as follows:

$$\mathbf{E}(M_s|\mathcal{F}_t) = M_t, \quad s \geqslant t \geqslant 0.$$

PROPOSITION 7.2.– *Let $H \in H^2[0,T]$. Denote $M_t = \int_0^t H_s \, \mathrm{d}B_s$ and $\langle M \rangle_t = \int_0^t H_s^2 \, \mathrm{d}s$, $t \in [0,T]$. Then:*

(1) *the random process M_t, $t \in [0,T]$, is a martingale;*

(2) $\mathbf{E}(Z(M_s^2 - M_t^2)) = \mathbf{E}(Z(M_s - M_t)^2)$ *for all bounded $Z \in \mathcal{H}_t$, $0 \leqslant t \leqslant s \leqslant T$;*

(3) *the random process $N_t := M_t^2 - \langle M \rangle_t$, $t \in [0,T]$, is a martingale.*

Proof. (1) This statement directly follows from Theorem 4.8, part 2′, which gives that, for every bounded random variable $Z \in \mathcal{H}_t$,

$$\mathbf{E}\big(Z(M_s - M_t)\big) = \mathbf{E}\left(Z \int_t^s H_u \, \mathrm{d}B_u\right) = \mathbf{E}\left(\int_t^s Z H_u \, \mathrm{d}B_u\right) = 0, \quad s \geqslant t.$$

(2) Since $ZM_t \in \mathcal{H}_t$, by Definition 7.1 (see also Remark 7.2), we have:

$$\begin{aligned}\mathbf{E}\big(Z(M_s - M_t)^2\big) &= \mathbf{E}\big(Z(M_s^2 + M_t^2 - 2M_sM_t)\big)\\ &= \mathbf{E}\big(Z(M_s^2 - M_t^2)\big) - 2\mathbf{E}\big((ZM_t)(M_s - M_t)\big)\\ &= \mathbf{E}\big(Z(M_s^2 - M_t^2)\big).\end{aligned}$$

(3) Using the property, for every *non-negative* random variable $Z \in \mathcal{H}_t$, we get:

$$\begin{aligned}&\mathbf{E}\big(Z(N_s - N_t)\big)\\ &\quad = \mathbf{E}\Big(Z\big(M_s^2 - \langle M \rangle_s - (M_t^2 - \langle M \rangle_t)\big)\Big)\\ &\quad = \mathbf{E}\big(Z(M_s^2 - M_t^2) - Z(\langle M \rangle_s - \langle M \rangle_t)\big)\\ &\quad = \mathbf{E}\big(Z(M_s - M_t)^2 - Z(\langle M \rangle_s - \langle M \rangle_t)\big) \qquad \text{(property 2)}\\ &\quad = \mathbf{E}\left\{Z\left(\int_t^s H_u \, \mathrm{d}B_u\right)^2 - Z\int_t^s H_u^2 \, \mathrm{d}u\right\}\\ &\quad = \mathbf{E}\left\{\left(\int_t^s \sqrt{Z} H_u \, \mathrm{d}B_u\right)^2 - \int_t^s \big(\sqrt{Z} H_u\big)^2 \, \mathrm{d}u\right\} \qquad \text{(Thm. 4.8.2}'\text{)}\end{aligned}$$

$$= \mathbf{E}\left(\int_t^s \sqrt{Z} H_u \, \mathrm{d}B_u\right)^2 - \mathbf{E}\int_t^s \left(\sqrt{Z}H_u\right)^2 \mathrm{d}u = 0. \qquad \text{(Thm. 4.8.3)}$$

Now, for an *arbitrary* bounded random variable $Z \in \mathcal{H}_t$, denote $Z^+ := \max\{Z, 0\}$ and $Z^- := \max\{-Z, 0\}$. Then $Z^+, Z^- \geqslant 0$ and $Z = Z^+ - Z^-$. Therefore, applying the just proved inequality for non-negative $Z \in \mathcal{H}_t$, we have:

$$\begin{aligned}\mathbf{E}\big(Z(N_s - N_t)\big) &= \mathbf{E}\big((Z^+ - Z^-)(N_s - N_t)\big) \\ &= \mathbf{E}\big((Z^+(N_s - N_t)\big) - \mathbf{E}\big(Z^-(N_s - N_t)\big) = 0 - 0 = 0. \end{aligned}$$

$\triangle$

DEFINITION 7.3.– *The stochastic integral of a random process Y_s, $s \in [0,T]$, with respect to the martingale M defined in Proposition 7.2(1) is the random variable*

$$\int_0^T Y_s \, \mathrm{d}M_s := \int_0^T Y_s H_s \, \mathrm{d}B_s,$$

provided that the integral on the right-hand side is defined.

The definition is justified by the statements on the convergence of Riemann-type integral sums. Let $\Delta^n = \{0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0,T]$ with $|\Delta^n| = \max_i |t_{i+1}^n - t_i^n| \to 0$, $n \to \infty$. To simplify the notation, we again omit the indices n and write $Y_i = Y(t_i^n)$ and $\Delta Y_i = Y_{i+1} - Y_i$.

PROPOSITION 7.4.– *Let $H \in H^2[0,T]$, $M_t := \int_0^t H_s \, \mathrm{d}B_s$, $t \in [0,T]$, and $\langle M\rangle_t := \int_0^t H_s^2 \, \mathrm{d}s$, $t \in [0,T]$. Let Y_t, $t \in [0,T]$, be a continuous adapted process. Then:*

(1) $\sum_i Y_i \Delta M_i \xrightarrow{\mathbf{P}} \int_0^T Y_t \, \mathrm{d}M_t = \int_0^T Y_t H_t \, \mathrm{d}B_t, \quad n \to \infty;$

(2) $\sum_i Y_i \Delta M_i^2 \xrightarrow{\mathbf{P}} \int_0^T Y_t \, \mathrm{d}\langle M\rangle_t = \int_0^T Y_t H_t^2 \, \mathrm{d}t, \quad n \to \infty.$

In particular, by taking $Y \equiv 1$, we have:

$$\sum_i \Delta M_i^2 \xrightarrow{\mathbf{P}} \langle M\rangle_T, \quad n \to \infty.$$

Therefore, the random process $\langle M\rangle_t$, $t \in [0,T]$, is called the quadratic variation of the martingale M.

Proof. (1) We first suppose that Y is a *bounded* random process, i.e. $|Y_t| \leqslant C$, $t \in [0,T]$ (with non-random constant C). We denote $Y_t^n = Y_i = Y(t_i^n)$, $t \in [t_i^n, t_{i+1}^n)$. Then we have:

$$\mathbf{E}\left(\sum_i Y_i \Delta M_i - \int_0^T Y_t H_t \, \mathrm{d}B_t\right)^2$$

$$= \mathbf{E}\Bigg(\sum_i Y_i \int_{t_i^n}^{t_{i+1}^n} H_t \,\mathrm{d}B_t - \int_0^T Y_t H_t \,\mathrm{d}B_t\Bigg)^2$$

$$= \mathbf{E}\Bigg(\sum_i \int_{t_i^n}^{t_{i+1}^n} Y_i H_t \,\mathrm{d}B_t - \int_0^T Y_t H_t \,\mathrm{d}B_t\Bigg)^2$$

$$= \mathbf{E}\Bigg(\sum_i \int_{t_i^n}^{t_{i+1}^n} Y_t^n H_t \,\mathrm{d}B_t - \int_0^T Y_t H_t \,\mathrm{d}B_t\Bigg)^2$$

$$= \mathbf{E}\Bigg(\int_0^T \big(Y_t^n - Y_t\big) H_t \,\mathrm{d}B_t\Bigg)^2 = \mathbf{E}\int_0^T \big(Y_t^n - Y_t\big)^2 H_t^2 \,\mathrm{d}t.$$

Because of the continuity of the paths of Y, we have $Y_t^n \to Y_t$, $n \to \infty$, for all $t \in [0,T]$, and therefore the integrand $(Y_t^n - Y_t)^2 H_t^2 \to 0$, $n \to \infty$, $t \in [0,T]$. On the other hand,

$$\big(Y_t^n - Y_t\big)^2 H_t^2 \leqslant 2\big((Y_t^n)^2 + Y_t^2\big) H_t^2 \leqslant 4C^2 H_t^2, \quad t \in [0,T],$$

and $\mathbf{E}\int_0^T 4C^2 H_t^2 \,\mathrm{d}t = 4C^2\|H\|^2 < \infty$. Now the Lebesgue theorem (section 1.11) yields $\mathbf{E}\int_0^T (Y_t^n - Y_t)^2 H_t^2 \,\mathrm{d}t \to 0$, $n \to \infty$, and, thus,

$$\mathbf{E}\Bigg(\sum_i Y_i \Delta M_i - \int_0^T Y_t H_t \,\mathrm{d}B_t\Bigg)^2 \to 0, \quad n \to \infty.$$

In the general case where Y is not bounded by a (non-random) constant, we denote

$$Y_t^{(N)} := \max\big\{-N, \min\{Y_t, N\}\big\} = \begin{cases} N & \text{for } Y_t > N, \\ Y_t & \text{for } |Y_t| \leqslant N, \\ -N & \text{for } Y_t < -N. \end{cases}$$

All the processes $Y^{(N)}$ are bounded ($|Y_t^{(N)}| \leqslant N$). As for such processes, the proposition is already proved, for all N,

$$\sum_i Y_i^{(N)} \Delta M_i \to \int_0^T Y_t^{(N)} \,\mathrm{d}M_t, \quad n \to \infty,$$

in the L^2 sense and, consequently, in probability. Let us now show that from this the required convergence follows. Take arbitrary $\varepsilon > 0$ and $\delta > 0$. Consider the events $\Omega_N := \{|Y_t| \leqslant N,\, t \in [0,T]\}$, $N \in \mathbb{N}$. Since $\bigcup_N \Omega_N$ is a certain event, there is N such that $\mathbf{P}(\Omega_N) > 1 - \delta/2$ or, equivalently, $\mathbf{P}(\Omega_N^c) < \delta/2$. Since $Y^{(N)} = Y$ on Ω_N, we have:

$$\sum_i Y_i^{(N)} \Delta M_i = \sum_i Y_i \Delta M_i \quad \text{and} \quad \int_0^T Y_t^{(N)} \,\mathrm{d}M_t = \int_0^T Y_t \,\mathrm{d}M_t$$

almost surely on Ω_N (Lemma 4.11). Therefore,

$$\begin{aligned}
&\mathbf{P}\left\{\left|\sum_i Y_i \Delta M_i - \int_0^T Y_t \,\mathrm{d}M_t\right| > \varepsilon\right\} \\
&\qquad = \mathbf{P}\left\{\left\{\left|\sum_i Y_i^{(N)} \Delta M_i - \int_0^T Y_t^{(N)} \,\mathrm{d}M_t\right| > \varepsilon\right\} \cap \Omega_N\right\} \\
&\qquad\quad + \mathbf{P}\left\{\left\{\left|\sum_i Y_i \Delta M_i - \int_0^T Y_t \,\mathrm{d}M_t\right| > \varepsilon\right\} \cap \Omega_N^c\right\} \\
&\qquad \leqslant \mathbf{P}\left\{\left|\sum_i Y_i^{(N)} \Delta M_i - \int_0^T Y_t^{(N)} \,\mathrm{d}M_t\right| > \varepsilon\right\} + \mathbf{P}\big(\Omega_N^c\big) \\
&\qquad < \mathbf{P}\left\{\left|\sum_i Y_i^{(N)} \Delta M_i - \int_0^T Y_t^{(N)} \,\mathrm{d}M_t\right| > \varepsilon\right\} + \frac{\delta}{2} \to \frac{\delta}{2}, \quad n \to \infty.
\end{aligned}$$

Hence, there is n_0 such that

$$\mathbf{P}\left\{\left|\sum_i Y_i \Delta M_i - \int_0^T Y_t \,\mathrm{d}M_t\right| > \varepsilon\right\} < \delta$$

for all $n \geqslant n_0$. Since $\varepsilon > 0$ and $\delta > 0$ were arbitrary, we obtain:

$$\forall \varepsilon > 0, \quad \mathbf{P}\left\{\left|\sum_i Y_i \Delta M_i - \int_0^T Y_t \,\mathrm{d}M_t\right| > \varepsilon\right\} \to 0, \quad n \to \infty,$$

as required.

(2) Since $\sum_i Y_i \Delta\langle M\rangle_i$ are integral sums of the Stieltjes integral $\int_0^T Y_t \,\mathrm{d}\langle M\rangle_t$, they converge to the latter (a.s.) as $n \to \infty$. Therefore, it suffices to show that

$$\sum_i Y_i \Delta M_i^2 - \sum_i Y_i \Delta\langle M\rangle_i \xrightarrow{\mathbf{P}} 0, \quad n \to \infty.$$

We first check this with some boundedness restrictions, assuming that the processes Y and H are bounded: $|Y_t| \leqslant C$ and $|H_t| \leqslant C$, $t \in [0,T]$. Since the process $N_t = M_t^2 - \langle M \rangle_t$, $t \in [0,T]$, is a martingale (Proposition 7.2(3)), we get:

$$\begin{aligned}
\mathbf{E}\Big(\sum_i Y_i \Delta M_i^2 - \sum_i Y_i \Delta\langle M\rangle_i \Big)^2 &= \mathbf{E}\Big(\sum_i Y_i \big(\Delta M_i^2 - \Delta\langle M\rangle_i\big) \Big)^2 \\
&= \mathbf{E}\Big(\sum_i Y_i \big(\Delta M_i^2 - \Delta\langle M\rangle_i\big) \sum_j Y_j \big(\Delta M_j^2 - \Delta\langle M\rangle_j\big) \Big) \\
&= \mathbf{E} \sum_i \big(Y_i(\Delta M_i^2 - \Delta\langle M\rangle_i)\big)^2 \\
&\quad + 2\mathbf{E} \sum_{i<j} Y_i \big(\Delta M_i^2 - \Delta\langle M\rangle_i\big) Y_j \big(\Delta M_j^2 - \Delta\langle M\rangle_j\big) \\
&= \sum_i \mathbf{E}\Big[Y_i^2 \big(\Delta M_i^2 - \Delta\langle M\rangle_i\big)^2 \Big] \\
&\quad + 2 \sum_{i<j} \mathbf{E}\Big[\big(Y_i Y_j (\Delta M_i^2 - \Delta\langle M\rangle_i)\big)\big(\Delta M_j^2 - \Delta\langle M\rangle_j\big) \Big] \\
&\leqslant C^2 \sum_i \mathbf{E}\big(\Delta M_i^2 - \Delta\langle M\rangle_i\big)^2 + 0 \leqslant 2C^2 \sum_i \mathbf{E}\big(\Delta M_i^4 + \Delta\langle M\rangle_i^2\big).
\end{aligned}$$

Here we used the fact that

$$\begin{aligned}
\mathbf{E}\big(Z(\Delta M_j^2 - \Delta\langle M\rangle_j)\big) &\overset{7.2(2)}{=\!=} \mathbf{E}\big[Z\big(\Delta(M^2)_j - \Delta\langle M\rangle_j\big)\big] \\
&= \mathbf{E}\big[Z \Delta(M^2 - \langle M\rangle)_j\big] \overset{7.2(3)}{=\!=} 0
\end{aligned}$$

with $Z = Y_i Y_j (\Delta M_i^2 - \Delta\langle M\rangle_i) \in \mathcal{H}_{t_j}$, $i < j$ (see also Remark 2 on page 107). It remains to check that $\sum_i \mathbf{E}(\Delta M_i^4 + \Delta\langle M\rangle_i^2) \to 0$, $n \to \infty$. Using the inequality

$$\mathbf{E}\bigg(\int_{t_1}^{t_2} H_s \, \mathrm{d}B_s \bigg)^4 \leqslant 36\mathbf{E}\bigg(\int_{t_1}^{t_2} H_s^2 \, \mathrm{d}s \bigg)^2$$

(see Theorem 4.8(5)), we get

$$\begin{aligned}
&\sum_i \mathbf{E}\big(\Delta M_i^4 + \Delta\langle M\rangle_i^2\big) \\
&\quad = \sum_i \Bigg[\mathbf{E}\bigg(\int_{t_i^n}^{t_{i+1}^n} H_s \, \mathrm{d}B_s \bigg)^4 + \mathbf{E}\bigg(\int_{t_i^n}^{t_{i+1}^n} H_s^2 \, \mathrm{d}s \bigg)^2 \Bigg]
\end{aligned}$$

$$\leqslant 37 \sum_i \mathbf{E}\left(\int\limits_{t_i^n}^{t_{i+1}^n} H_s^2 \, \mathrm{d}s \right)^2 \leqslant 37C^4 \sum_i \left(t_{i+1}^n - t_i^n\right)^2$$

$$\leqslant 37C^4 |\Delta^n| \sum_i \left(t_{i+1}^n - t_i^n\right) = 37C^4 T |\Delta^n| \to 0, \quad n \to \infty.$$

Finally, we suppress the assumptions of boundedness on Y and H similarly to the first part of the proposition. △

DEFINITION 7.5.– *An adapted random process X_t, $t \in [0,T]$, is called an Itô process (or diffusion-type process) if it can be written in the form*

$$X_t = X_0 + \int\limits_0^t K_s \, \mathrm{d}s + \int\limits_0^t H_s \, \mathrm{d}B_s, \quad t \in [0,T], \qquad [7.2]$$

where K and H are adapted random processes such that the integrals on the right-hand side exist (a.s.), i.e. $\int_0^T |K_s| \, \mathrm{d}s < +\infty$ and $\int_0^T H_s^2 \, \mathrm{d}s < +\infty$ (a.s.). In this case, we say that the process X admits the stochastic differential

$$\mathrm{d}X_t = K_t \, \mathrm{d}t + H_t \, \mathrm{d}B_t,$$

or, shortly,

$$\mathrm{d}X = K \, \mathrm{d}t + H \, \mathrm{d}B.$$

In particular, when $H = 0$, we shall say that the Itô process X is regular.

The stochastic integral of an adapted random process Z_t, $t \in [0,T]$, with respect to the Itô process X as in equation [7.2] *is the stochastic process*

$$Z \bullet X_t = \int\limits_0^t Z_s \, \mathrm{d}X_s := \int\limits_0^t Z_s K_s \, \mathrm{d}s + \int\limits_0^t Z_s H_s \, \mathrm{d}B_s, \quad t \in [0,T],$$

provided that the integrals on the right-hand side exist for all $t \in [0,T]$. Note that, in such a case, the stochastic integral $Z \bullet X$ is also an Itô process.

DEFINITION 7.6.– *The covariation of two Itô processes*

$$X_t = X_0 + \int\limits_0^t K_s \, \mathrm{d}s + \int\limits_0^t H_s \, \mathrm{d}B_s \quad \text{and} \quad Y_t = Y_0 + \int\limits_0^t \widetilde{K}_s \, \mathrm{d}s + \int\limits_0^t \widetilde{H}_s \, \mathrm{d}B_s$$

is the random process $\langle X, Y \rangle_t := \int_0^t H_s \widetilde{H}_s \, ds$. The process $\langle X \rangle_t := \langle X, X \rangle_t$ is called the quadratic variation of the process X. Note that the covariation $\langle X, Y \rangle = 0$ if at least one of the processes X and Y is regular.

PROPOSITION 7.7.– (Properties of the stochastic integral) *Let X and Y be Itô processes, and let Z and W be adapted random processes in the interval $[0,T]$. Then*:

(1) $Z \bullet (\alpha X + \beta Y) = \alpha Z \bullet X + \beta Z \bullet Y$, $\alpha, \beta \in \mathbb{R}$;

(2) $(\alpha Z + \beta W) \bullet X = \alpha Z \bullet X + \beta W \bullet X$, $\alpha, \beta \in \mathbb{R}$;

(3) $W \bullet (Z \bullet X) = (WZ) \bullet X$;

(4) $\langle Z \bullet X, W \bullet Y \rangle = (ZW) \bullet \langle X, Y \rangle$;

provided that the integrals on the right-hand sides of the first three equalities and on the left-hand side of the last one are well defined.

Proof. All properties are easily verified from the definition of the stochastic integral. Indeed, let Itô processes X and Y be as in Definition 7.6.

(1) By the linearity of the stochastic integral with respect to Brownian motion, we have:

$$(\alpha X+\beta Y)_t = \alpha X_0+\beta Y_0+\int_0^t \big(\alpha K_s+\beta \widetilde{K}_s\big)\,\mathrm{d}s+\int_0^t \big(\alpha H_s+\beta \widetilde{H}_s\big)\,\mathrm{d}B_s, \quad t \in [0,T].$$

Therefore,

$$\begin{aligned}
&Z \bullet (\alpha X + \beta Y)_t \\
&\quad = \int_0^t Z_s\big(\alpha K_s + \beta \widetilde{K}_s\big)\,\mathrm{d}s + \int_0^t Z_s\big(\alpha H_s + \beta \widetilde{H}_s\big)\,\mathrm{d}B_s \\
&\quad = \int_0^t \big(\alpha Z_s K_s + \beta Z_s \widetilde{K}_s\big)\,\mathrm{d}s + \int_0^t \big(\alpha Z_s H_s + \beta Z_s \widetilde{H}_s\big)\,\mathrm{d}B_s \\
&\quad = \alpha\bigg(\int_0^t Z_s K_s\,\mathrm{d}s + \int_0^t Z_s H_s\,\mathrm{d}B_s\bigg) + \beta\bigg(\int_0^t Z_s \widetilde{K}_s\,\mathrm{d}s + \int_0^t Z_s \widetilde{H}_s\,\mathrm{d}B_s\bigg) \\
&\quad = \alpha Z \bullet X_t + \beta Z \bullet Y_t, \quad t \in [0,T].
\end{aligned}$$

(2) Similarly,

$$(\alpha Z+\beta W)\bullet X_t=\int_0^t(\alpha Z_s+\beta W_s)K_s\,\mathrm{d}s+\int_0^t(\alpha Z_s+\beta W_s)H_s\,\mathrm{d}B_s$$

$$=\alpha\int_0^t Z_sK_s\,\mathrm{d}s+\beta\int_0^t W_sK_s\,\mathrm{d}s+\alpha\int_0^t Z_sH_s\,\mathrm{d}B_s+\beta\int_0^t W_sH_s\,\mathrm{d}B_s$$

$$=\alpha\left(\int_0^t Z_sK_s\,\mathrm{d}s+\int_0^t Z_sH_s\,\mathrm{d}B_s\right)+\beta\left(\int_0^t W_sH_s\,\mathrm{d}s+\int_0^t W_sH_s\,\mathrm{d}B_s\right)$$

$$=\alpha Z\bullet X_t+\beta W\bullet X_t,\quad t\in[0,T].$$

(3) Since, by Definition 7.5, $Z\bullet X_t=\int_0^t Z_sK_s\,\mathrm{d}s+\int_0^t Z_sH_s\,\mathrm{d}B_s$, we have:

$$W\bullet(Z\bullet X)_t=\int_0^t W_s(Z_sK_s)\,\mathrm{d}s+\int_0^t W_s(Z_sH_s)\,\mathrm{d}B_s$$

$$=\int_0^t(W_sZ_s)K_s\,\mathrm{d}s+\int_0^t(W_sZ_s)H_s\,\mathrm{d}B_s=(WZ)\bullet X_t,\quad t\in[0,T].$$

(4) From the equalities

$$Z\bullet X_t=\int_0^t Z_sK_s\,\mathrm{d}s+\int_0^t Z_sH_s\,\mathrm{d}B_s$$

and

$$W\bullet Y_t=\int_0^t W_s\widetilde{K}_s\,\mathrm{d}s+\int_0^t W_s\widetilde{H}_s\,\mathrm{d}B_s$$

by Definition 7.6 we have:

$$\langle Z\bullet X,W\bullet Y\rangle_t=\int_0^t(Z_sW_s)\big(H_s\widetilde{H}_s\big)\,\mathrm{d}s$$

$$=\int_0^t Z_sW_s d\langle X,Y\rangle_s=(ZW)\bullet\langle X,Y\rangle_t,\quad t\in[0,T].\qquad\triangle$$

REMARK.– When calculating stochastic integrals (and "ordinary" integrals as well), it is often convenient to apply formal rules of summation and multiplication of random processes and their differentials. We shall write $\mathrm{d}Z = Y\,\mathrm{d}X$ if $Z_t - Z_0 = \int_0^t Y_s\,\mathrm{d}X_s$, $t \in [0,T]$. In particular (when $Y = 1$), for two Itô processes Z and X, we shall write $\mathrm{d}Z = \mathrm{d}X$ if $Z_t - Z_0 = X_t - X_0$, $t \in [0,T]$. We also define the sum and the product of Itô processes X and Y by

$$\mathrm{d}X + \mathrm{d}Y := \mathrm{d}(X+Y), \qquad \mathrm{d}X \cdot \mathrm{d}Y := \mathrm{d}\langle X, Y\rangle.$$

Then the properties stated in Proposition 7.7 can be written in the differential form as follows:

(1) $Z\,\mathrm{d}(\alpha X + \beta Y) = \alpha Z\,\mathrm{d}X + \beta Z\,\mathrm{d}Y$;

(2) $(\alpha Z + \beta W)\,\mathrm{d}X = \alpha Z\,\mathrm{d}X + \beta W\,\mathrm{d}X$;

(3) $W(Z\,\mathrm{d}X) = (WZ)\,\mathrm{d}X$;

(4) $(Z\,\mathrm{d}X)\cdot(W\,\mathrm{d}Y) = (ZW)\,\mathrm{d}X \cdot \mathrm{d}Y$.

Also note the following properties that directly follow from the definition:

(5) $(\mathrm{d}X + \mathrm{d}Y)\cdot \mathrm{d}Z = \mathrm{d}X \cdot \mathrm{d}Z + \mathrm{d}Y \cdot \mathrm{d}Z$,

(6) $\mathrm{d}X \cdot \mathrm{d}Y = \mathrm{d}Y \cdot \mathrm{d}X$,

(7) $(\mathrm{d}X\cdot \mathrm{d}Y)\cdot \mathrm{d}Z = \mathrm{d}X\cdot(\mathrm{d}Y \cdot \mathrm{d}Z) = 0$, since the covariation of two Itô processes is always a regular process.

Consider, say, Property (3). Let $Y_t = Y_0 + \int_0^t Z_s\,\mathrm{d}X_s$ or, in the differential form, $\mathrm{d}Y = Z\,\mathrm{d}X$. By Property (3), we may formally multiply this equality by a random process W, thus obtaining the equality $W\,\mathrm{d}Y = WZ\,\mathrm{d}X$; then, formally integrating this, we get the correct equality $\int_0^t W_s\,\mathrm{d}Y_s = \int_0^t W_s Z_s\,\mathrm{d}X_s$.

We now prove a generalization of Proposition 7.4 for Itô processes. Once again, let

$$\Delta^n = \{0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T\}, \quad n \in \mathbb{N},$$

be a refining sequence of partitions of the interval $[0,T]$ (that is, $|\Delta^n| = \max_i |t_{i+1}^n - t_i^n| \to 0$). As before, for a random process Y, we denote $Y_i = Y(t_i^n)$, $\Delta Y_i = Y_{i+1} - Y_i$, omitting the indices n.

PROPOSITION 7.8.– *Let* $X_t = X_0 + A_t + M_t = X_0 + \int_0^t K_s\,\mathrm{d}s + \int_0^t H_s\,\mathrm{d}B_s$, $t \in [0,T]$*, be an Itô process, and let* Y_t*,* $t \in [0,T]$*, be a continuous adapted random process. Then we have*:

(1) $\sum_i Y_i \Delta X_i \xrightarrow{\mathbf{P}} \int_0^T Y_t\,\mathrm{d}X_t, \quad n \to \infty$;

(2) $\sum_i Y_i \Delta X_i^2 \xrightarrow{\mathbf{P}} \int_0^T Y_t\,\mathrm{d}\langle X\rangle_t, \quad n \to \infty$.

In particular, for $Y \equiv 1$, we have

$$\sum_i \Delta X_i^2 \xrightarrow{\mathbf{P}} \langle X \rangle_T,\ n \to \infty.$$

Therefore, the random process $\langle X \rangle_t$, $t \in [0,T]$, is called the quadratic variation of X.

Proof. (1) From the properties of the Stieltjes integral it follows that $\sum_i Y_i \Delta A_i \to \int_0^T Y_t \, \mathrm{d}A_t$, $n \to \infty$ (almost surely and, thus, in probability). Therefore, by Proposition 7.4.1 we get:

$$\begin{aligned}\sum_i Y_i \Delta X_i &= \sum_i Y_i \Delta A_i + \sum_i Y_i \Delta M_i \xrightarrow{\mathbf{P}} \int_0^T Y_t \, \mathrm{d}A_t + \int_0^T Y_t \, \mathrm{d}M_t \\ &= \int_0^T Y_t \, \mathrm{d}X_t, \quad n \to \infty.\end{aligned}$$

(2) We have:

$$\sum_i Y_i \Delta X_i^2 = \sum_i Y_i \Delta M_i^2 + 2 \sum_i Y_i \Delta A_i \Delta M_i + \sum_i Y_i \Delta A_i^2.$$

By Proposition 7.4.2,

$$\sum_i Y_i \Delta M_i^2 \to \int_0^T Y_t \, \mathrm{d}\langle M \rangle_t = \int_0^T Y_t \, \mathrm{d}\langle X \rangle_t, \quad n \to \infty.$$

Therefore, it suffices to show that

$$\sum_i Y_i \Delta A_i \Delta M_i \xrightarrow{\mathbf{P}} 0 \quad \text{and} \quad \sum_i Y_i \Delta A_i^2 \xrightarrow{\mathbf{P}} 0, \quad n \to \infty.$$

We have:

$$\left| \sum_i Y_i \Delta A_i \Delta M_i \right| \leqslant \max_i |\Delta M_i| \cdot \sum_i |Y_i||\Delta A_i|.$$

By the (uniform) continuity of paths of the martingale M, $\max_i |\Delta M_i| \to 0, n \to \infty$, and the latter sum can be estimated by a random variable independent of n:

$$\begin{aligned}\sum_i |Y_i||\Delta A_i| &\leqslant \max_i |Y_i| \sum_i |\Delta A_i| \leqslant \max_{t\in[0,T]} |Y_t| \sum_i \left| \int_{t_i^n}^{t_{i+1}^n} K_s \, \mathrm{d}s \right| \\ &\leqslant \max_{t\in[0,T]} |Y_t| \sum_i \int_{t_i^n}^{t_{i+1}^n} |K_s| \, \mathrm{d}s = \max_{t\in[0,T]} |Y_t| \int_0^T |K_s| \, \mathrm{d}s.\end{aligned}$$

Therefore, $|\sum_i Y_i \Delta A_i \Delta M_i| \to 0$, $n \to \infty$. Similarly,

$$\left|\sum_i Y_i \Delta A_i^2\right| \leqslant \max_i |\Delta A_i| \cdot \sum_i |Y_i||\Delta A_i| \to 0, \quad n \to \infty. \qquad \triangle$$

COROLLARY 7.9.– *If X_t and Y_t, $t \in [0,T]$, are two Itô processes, then*

$$\sum_i \Delta X_i \Delta Y_i \xrightarrow{\mathbf{P}} \langle X, Y\rangle_T, \quad n \to \infty.$$

Proof.

$$\begin{aligned}
\sum_i \Delta X_i \Delta Y_i &= \sum_i \frac{(\Delta X_i + \Delta Y_i)^2 - (\Delta X_i - \Delta Y_i)^2}{4} \\
&= \sum_i \frac{(\Delta(X+Y)_i)^2 - (\Delta(X-Y)_i)^2}{4} \\
&\to \frac{\langle X+Y\rangle_T - \langle X-Y\rangle_T}{4} \\
&= \frac{(\langle X\rangle_T + 2\langle X,Y\rangle_T + \langle Y\rangle_T) - (\langle X\rangle_T - 2\langle X,Y\rangle_T + \langle Y\rangle_T)}{4} \\
&= \langle X, Y\rangle_T, \quad n \to \infty. \qquad \triangle
\end{aligned}$$

Using Proposition 7.8, we can prove Itô's formula for Itô processes similarly to that for Brownian motion (Theorems 5.2 and 5.2.a).

THEOREM 7.10 (Itô's formula for Itô processes).– *If $X_t = X_0 + \int_0^t K_s \,\mathrm{d}s + \int_0^t H_s \,\mathrm{d}B_s$, $t \in [0,T]$, is an Itô process and $F \in C^2([0,T] \times \mathbb{R})$, then*

$$\begin{aligned}
&F\big(T, X_T\big) - F\big(0, X_0\big) \\
&\quad = \int_0^T F'_x\big(t, X_t\big)\,\mathrm{d}X_t + \int_0^T F'_t\big(t, X_t\big)\,\mathrm{d}t + \frac{1}{2}\int_0^T F''_{xx}\big(t, X_t\big)\,\mathrm{d}\langle X\rangle_t \\
&\quad = \int_0^T F'_x\big(t, X_t\big) H_t\,\mathrm{d}B_t + \int_0^T F'_x\big(t, X_t\big) K_t\,\mathrm{d}t \\
&\qquad + \int_0^T F'_t\big(t, X_t\big)\,\mathrm{d}t + \frac{1}{2}\int_0^T F''_{xx}\big(t, X_t\big) H_t^2\,\mathrm{d}t.
\end{aligned}$$

REMARK.– A more general result, Itô's formula for a function of several Itô processes, is formulated below in Theorem 13.5.

Proof. The proof is similar to that of Theorems 5.2 and 5.4; therefore, we only give an outline here. Writing $F(T,X_T)-F(0,X_0)$ as a "telescopic" sum and applying Taylor's formula, we have:

$$\begin{aligned}
F\big(T,X_T\big)-F\big(0,X_0\big) &= \sum_i \big(F(t_{i+1},X_{i+1})-F(t_i,X_i)\big)\\
&= \sum_i F'_x(t_i,X_i)\Delta X_i + \sum_i F'_t(t_i,X_i)\Delta t_i\\
&\quad + \frac{1}{2}\bigg[\sum_i F''_{xx}(t_i,X_i)\Delta X_i^2 + 2\sum_i F''_{xt}(t_i,X_i)\Delta X_i\Delta t_i\\
&\quad + \sum_i F''_{tt}(t_i,X_i)\Delta t_i^2\bigg] + \sum_i R(X_i,t_i,X_{i+1},t_{i+1}),
\end{aligned}$$

where $|R(x,t,y,s)| \leqslant r(|y-x|+|s-t|)((y-x)^2+(s-t)^2)$ with $r\colon \mathbb{R}_+\to\mathbb{R}_+$ such that $\lim_{h\downarrow 0} r(h)=0$. Passing to the limit in the latter expression as $n\to\infty$, we get the formula stated. Indeed, by Proposition 7.8,

$$\sum_i F'_x(t_i,X_i)\Delta X_i \xrightarrow{\mathbf{P}} \int_0^T F'_x(t,X_t)\,\mathrm{d}X_t$$

and

$$\sum_i F''_{xx}(t_i,X_i)\Delta X_i^2 \xrightarrow{\mathbf{P}} \int_0^T F''_{xx}(t,X_t)\,\mathrm{d}\langle X\rangle_t.$$

Moreover, as in in the proofs of Theorems 5.2 and 5.4, we have:

$$\sum_i F'_t(t_i,X_i)\Delta t_i \to \int_0^T F'_t(t,X_t)\,\mathrm{d}t, \qquad \frac{1}{2}\sum_i F''_{tt}(t_i,X_i)\Delta t_i^2 \to 0,$$

$$\sum_i F''_{xt}(t_i,X_i)\Delta X_i\Delta t_i \xrightarrow{\mathbf{P}} 0, \qquad \sum_i R(X_i,t_i,X_{i+1},t_{i+1}) \xrightarrow{\mathbf{P}} 0. \quad \triangle$$

COROLLARY 7.11 (Integration by parts formula).– *If X and Y are Itô processes in the time interval $[0,T]$, then*

$$X_tY_t = X_0Y_0 + \int_0^t X_s\,\mathrm{d}Y_s + \int_0^t Y_s\,\mathrm{d}X_s + \langle X,Y\rangle_t, \quad t\in[0,T],$$

or, in a differential form,

$$\mathrm{d}(XY) = X\,\mathrm{d}Y + Y\,\mathrm{d}X + \mathrm{d}\langle X, Y\rangle.$$

In particular, if at least one of the processes X and Y is regular, these formulas become the usual integration by parts formulas

$$\int_0^t X_s\,\mathrm{d}Y_s = X_tY_t - X_0Y_0 - \int_0^t Y_s\,\mathrm{d}X_s, \qquad X\,\mathrm{d}Y = \mathrm{d}(XY) - Y\,\mathrm{d}X.$$

Proof. Let

$$X_t = X_0 + \int_0^t K_s\,\mathrm{d}s + \int_0^t H_s\,\mathrm{d}B_s \quad \text{and} \quad Y_t = Y_0 + \int_0^t \widetilde{K}_s\,\mathrm{d}s + \int_0^t \widetilde{H}_s\,\mathrm{d}B_s, \quad t \in [0, T].$$

Applying Itô's formula to the function $F(x) = x^2$, $x \in \mathbb{R}$, and the processes X, Y, and $X + Y$, we have:

$$X_t^2 = X_0^2 + 2\int_0^t X_s\,\mathrm{d}X_s + \int_0^t H_s^2\,\mathrm{d}s,$$

$$Y_t^2 = Y_0^2 + 2\int_0^t Y_s\,\mathrm{d}Y_s + \int_0^t \widetilde{H}_s^2\,\mathrm{d}s,$$

$$\begin{aligned}(X_t + Y_t)^2 &= (X_0 + Y_0)^2 + 2\int_0^t (X_s + Y_s)\,\mathrm{d}(X_s + Y_s) + \int_0^t \big(H_s + \widetilde{H}_s\big)^2\,\mathrm{d}s \\ &= X_0^2 + 2\int_0^t X_s\,\mathrm{d}X_s + \int_0^t H_s^2\,\mathrm{d}s + Y_0^2 + 2\int_0^t Y_s\,\mathrm{d}Y_s + \int_0^t \widetilde{H}_s^2\,\mathrm{d}s \\ &\quad + 2X_0Y_0 + 2\int_0^t X_s\,\mathrm{d}Y_s + 2\int_0^t Y_s\,\mathrm{d}X_s + 2\int_0^t H_s\widetilde{H}_s\,\mathrm{d}s.\end{aligned}$$

Subtracting the first two equalities from the last one and dividing by 2, we get:

$$X_tY_t = X_0Y_0 + \int_0^t X_s\,\mathrm{d}Y_s + \int_0^t Y_s\,\mathrm{d}X_s + \int_0^t H_s\widetilde{H}_s\,\mathrm{d}s. \qquad \triangle$$

Having defined the stochastic integrals with respect to Itô processes, we can also generalize the notion of a stochastic differential equation in which the stochastic integral is taken with respect to Itô process.

DEFINITION 7.12.– *Let Z_t, $t \geqslant 0$, be an Itô process, and A_t, $t \geqslant 0$, a regular Itô process. A random (Itô) process X_t, $t \geqslant 0$, is said to be a solution of the stochastic differential equation*

$$\mathrm{d}X_t = b(X_t, t)\,\mathrm{d}A_t + \sigma(X_t, t)\,\mathrm{d}Z_t, \qquad X_0 = x_0, \tag{7.3}$$

with coefficients $b, \sigma\colon \mathbb{R} \times [0, \infty) \to \mathbb{R}$ if it satisfies the equation

$$X_t = x_0 + \int_0^t b(X_s, s)\,\mathrm{d}A_s + \int_0^t \sigma(X_s, s)\,\mathrm{d}Z_s$$

for all $t \geqslant 0$.

The following analog of Theorem 6.4 holds.

THEOREM 7.13.– *If coefficients b and σ satisfy the Lipschitz and linear growth conditions, then equation* [7.3] *has a unique solution.*

7.1. Exercises

7.1. Itô processes X and Y satisfy the relations

$$X_t = x + \int_0^t Y_s\,\mathrm{d}B_s \quad \text{and} \quad Y_t = y - \int_0^t X_s\,\mathrm{d}B_s, \quad t \geqslant 0.$$

Show that $Z_t := X_t^2 + Y_t^2 = (x^2 + y^2)\,\mathrm{e}^t$, $t \geqslant 0$.

7.2. Give an example of a sequence of Itô processes $\{X^n\}$ and a non-regular Itô process X such that $X^n \to X$ uniformly in finite intervals, i.e. for all $T > 0$, almost surely,

$$\sup_{t\in[0,T]} |X_t^n - X_t| \to 0, \quad n \to \infty,$$

and

(1) $\langle X_n \rangle \not\to \langle X \rangle$, or, moreover,

(2) $\langle X_n \rangle_T \to \frac{1}{2}\langle X \rangle_T$.

7.3.

(1) Find a regular process A_t, $t \geqslant 0$, such that: the process $M_t = tB_t^2 - A_t$, $t \geqslant 0$, is a martingale.

(2) Find a function $f(s)$, $s \geqslant 0$, such that the random process $M_t = t^2 B_t + \int_0^t f(s) B_s \, ds$, $t \geqslant 0$, is a martingale.

(3) Find the drift coefficient $b(x)$ of the equation $dX_t = b(X_t)\, dt + X_t \, dB_t$ if the square of its solution $M_t = X_t^2$, $t \geqslant 0$, is a martingale.

(4) Let X_t, $t \geqslant 0$, be a solution of the stochastic differential equation

$$dX_t = X_t \, dt + X_t \, dB_t, \quad X_0 = 1.$$

Show that the random process

$$M_t := X_t^3 - 6 \int_0^t X_s^3 \, ds, \quad t \geqslant 0,$$

is a martingale.

7.4. Let $\Delta^n = \{0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = T\}$, $n \in \mathbb{N}$, be a sequence of partitions of the interval $[0, T]$ such that $|\Delta^n| := \max_i |t_{i+1}^n - t_i^n| \to 0$ as $n \to \infty$.

(1) Find

$$\mathbf{P}\text{–} \lim_{n \to \infty} \sum_i (\Delta X_i)^2 = \mathbf{P}\text{–} \lim_{n \to \infty} \sum_{i=0}^{k_n - 1} \big(X(t_{i+1}^n) - X(t_i^n) \big)^2$$

if $X_t = \exp\{2B_t - 3t\}$, $t \geqslant 0$.

(2) For any Itô processes X and Y, find the limit

$$\mathbf{P}\text{-} \lim_{n \to \infty} \sum_{i=0}^{k_n - 1} Y(t_i^n) \big(X(t_{i+1}^n) - X(t_i^n) \big)^3.$$

7.5. Find the limits

$$I_1 := L^2\text{-} \lim_{n \to \infty} \sum_{k=1}^{n} B\Big(\frac{k}{n}\Big) \Big[B^2\Big(\frac{k}{n}\Big) - B^2\Big(\frac{k-1}{n}\Big) \Big].$$

and

$$I_2 := L^2\text{-} \lim_{n \to \infty} \sum_{k=1}^{n} B^2\Big(\frac{k}{n}\Big) \Big[B\Big(\frac{k}{n}\Big) - B\Big(\frac{k-1}{n}\Big) \Big].$$

7.6. Let $X_t, t \geqslant 0$, be an Itô process, and let

$$Y_t := \int_0^t \frac{dX_s}{1+X_s^2} - \int_0^t \frac{X_s d\langle X\rangle_s}{(1+X_s^2)^2}, \qquad t \geqslant 0.$$

Show that $\mathbf{E}|Y_t| < \pi$, $t \geqslant 0$.

7.7. Find $\mathbf{E}\langle B^4, B^2\rangle_t$, $t \geqslant 0$.

7.8. Find $\mathbf{E}\langle X, Y\rangle_t$, $t \geqslant 0$, if $X_t := B_t^3$ and a random process Y satisfies the stochastic differential equation

$$dY_t = \cos^2 Y_t \, dt + dB_t, \qquad X_0 = 1.$$

7.9. Show that if X is an Itô process, $f \in C^2(\mathbb{R})$, and $Y_t := f(X_t)$, $t \geqslant 0$, then

$$\langle Y\rangle_t = \int_0^t \big(f'(X_s)\big)^2 \, d\langle X\rangle_s, \qquad t \geqslant 0.$$

7.10. Suppose that an Itô process $X_t = \int_0^t K_s \, ds + \int_0^t H_s \, dB_s = 0$, $t \geqslant 0$. Show that $K_t = H_t = 0$ a.s., $t \geqslant 0$.

7.11. Suppose that $X_t \geqslant 0$, $t \geqslant 0$, is a solution of the SDE $dX_t = \sqrt{X_t}\, dB_t$ with $X_0 = x \geqslant 0$. Find $\mathbf{E}X_t^2$ and $\mathbf{E}X_t^3$, $t \geqslant 0$.

7.12. Suppose that $X_t \geqslant 0$, $t \geqslant 0$, is a solution of the SDE $dX_t = (1 - X_t)\, dt + \sqrt{X_t}\, dB_t$ with $X_0 = x \geqslant 0$. Find $\mathbf{E}X_t^2$, $t \geqslant 0$.

Chapter 8

Stratonovich Integral and Equations

Non-differentiability of the trajectories of Brownian motion, stochastic integration theory, and accompanying "exotic" formulas of Itô, integration by parts, etc. is the price we pay for the mathematical idealization of various real-world models with perturbations close to white noise. However, our troubles do not end here. Suppose that, in some real experiment, we observe the integral $\int_0^T B_t \, \mathrm{d}B_t$ with "true" Brownian motion and wish to construct its realistic mathematical model. Although the trajectories of a "true" Brownian motion, as those of its mathematical model, are rather chaotic, they are necessarily differentiable.[1] To get an idealized model of the integral $\int_0^T B_t \, \mathrm{d}B_t$, we may try to take a sequence of continuously *differentiable* functions B^n, $n \in \mathbb{N}$, converging to a Brownian motion B and consider the corresponding sequence of integrals $\int_0^T B_t^n \, \mathrm{d}B_t^n$, $n \in \mathbb{N}$. Because of the differentiability of B^n, these integrals are $\int_0^T B_t^n \, \mathrm{d}B_t^n = (B_T^n)^2/2$. Since $\lim_{n\to\infty}(B_T^n)^2/2 = B_T^2/2$, in the real world, the above-mentioned integral $\int_0^T B_t \, \mathrm{d}B_t$, together with its mathematical model, should be equal to $B_T^2/2$, but not to $B_T^2/2 - T/2$, as we have obtained in Example 4.10. It looks like the stochastic integral incorrectly models the "real-world" integral!

As this happens with stochastic integrals, do not expect, a fortiori, anything good from stochastic differential equations. Suppose that by heuristic arguments we have decided that some phenomenon has to be modeled by the stochastic differential equation

$$X_t = x_0 + \int_0^t b(X_s, s) \, \mathrm{d}s + \int_0^t \sigma(X_s, s) \, \mathrm{d}B_s. \qquad [8.1]$$

1. Possibly with some risk, we can claim that, in nature, there are no continuous but non-differentiable functions.

To justify such a choice, we may start by taking a sequence $\{B^n,\, n \in \mathbb{N}\}$ (continuously differentiable) processes converging to Brownian motion B (say, uniformly on finite time intervals $[0, T]$). Consider the corresponding sequence of equations

$$X_t^n = x_0 + \int_0^t b\big(X_s^n, s\big)\,\mathrm{d}s + \int_0^t \sigma\big(X_s^n, s\big)\,\mathrm{d}B_s^n.$$

If the sequence of solutions $\{X^n\}$ has a limit X, it is natural to expect that this limit is a random process suitable for modeling the phenomenon. Unfortunately, as we will see, X is not a solution of equation [8.1]. Moreover, it satisfies another equation,

$$X_t = x_0 + \int_0^t \Big(b(X_s, s) + \frac{1}{2}\sigma\sigma_x'(X_s, s)\Big)\,\mathrm{d}s + \int_0^t \sigma(X_s, s)\,\mathrm{d}B_s, \qquad [8.2]$$

where we see the additional term $\frac{1}{2}\int_0^t \sigma\sigma_x'(X_s, s)\,\mathrm{d}s$. Thus, equation [8.1] that is intended to describe the phenomenon has to be complemented by an additional term, and instead we have to consider equation [8.2].

Similar "paradoxes" can be explained by a certain instability of the Itô integrals and stochastic differential equations. There are two possible outcomes from the current situation:

1. *Correcting terms in the stochastic differential equations*. We have to look critically at stochastic differential equations that model real-world phenomena. We have to keep in mind that the equations derived by heuristic arguments often need to be completed by correcting terms.

2. *Construction of another, more stable integral*. Instead of the Itô integral, we may try to construct another stochastic integral, free of the above-mentioned "paradoxes". Happily, such an integral is already constructed. Its essence can be described as follows. Recall Proposition 4.9, where the stochastic integral is characterized as the limit of Riemann-type integral sums:

$$\int_0^T X_t\,\mathrm{d}B_t = \lim_{n\to\infty} \sum_i X(t_i^n)\big(B(t_{i+1}^n) - B(t_i^n)\big).$$

We have already noted that, in order to use the adaptness of the integrated process X, the values of the latter are taken at the *left* ends t_i^n of the partition intervals $[t_i^n, t_{i+1}^n]$. And what about taking such values symmetrically, at the *mid-points* of the partition intervals, that is, at the points $\bar{t}_i^n := (t_i^n + t_{i+1}^n)/2$, that is, trying to define the integral as

$$\int_0^T X_t\,\mathrm{d}B_t = \lim_{n\to\infty} \sum_i X(\bar{t}_i^n)\big(B(t_{i+1}^n) - B(t_i^n)\big)?$$

It appears that if the integrated process X is sufficiently "good", we get the integral which is, in a sense, symmetric. Russian physicist and probabilist R.L. Stratonovich was the first who constructed and considered integrals of this type (in applications!). Later on, it appeared that a more cohesive theory is obtained when, instead of the values of the process at the mid-points of partition intervals, we take the averages of the values of the process at the endpoints of the latter. For a long time, in the literature, there was discussion about which integral is better, that of Itô or Stratonovich? We can say that the "competition" ended in a draw. Each integral has its "pluses" and "minuses". The Itô integral has better analytical properties (a wider class of integrable processes), while the Stratonovich integral has better geometrical properties (the usual change-of-variable formula). If both exist, they can be rather simply expressed one by another.[2] Most often, the Itô integrals and equations are preferred when *discrete-time* phenomena with random perturbations are approximated by continuous-time models (dynamics of populations, financial markets, etc.), while the Stratonovich integrals are more applicable for describing *continuous-time* phenomena with perturbations close to white noise (chemical reactions, radio signals with noise, etc.).

DEFINITION 8.1.– *Let X and Y be two Itô processes. The Stratonovich integral of the process Y in the interval $[0,t]$ with respect to X is the limit*

$$Y \circ X_t = \int_0^t Y_s \circ \mathrm{d}X_s := \mathbf{P}\text{-}\lim_n \sum_i \frac{Y_i + Y_{i+1}}{2} \Delta X_i \,,$$

where we continue using the usual simplified notation: $\Delta^n = \{0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t\}$, $n \in \mathbb{N}$, is a refining sequence of partitions of the interval $[0,t]$ ($|\Delta^n| := \max_i |t_{i+1}^n - t_i^n| \to 0$, $n \to \infty$), $Y_i := Y(t_i^n)$, $\Delta X_i := X_{i+1} - X_i = X(t_{i+1}^n) - X(t_i^n)$.

PROPOSITION 8.2.– *The above-defined integral exists and equals*

$$\int_0^t Y_s \circ \mathrm{d}X_s = \int_0^t Y_s \,\mathrm{d}X_s + \frac{1}{2}\langle X, Y\rangle_t,$$

or, in short,

$$Y \circ X = Y \bullet X + \frac{1}{2}\langle X, Y\rangle.$$

In particular, if at least one of the Itô processes X and Y is regular, the Stratonovich and Itô integrals coincide: $Y \circ X = Y \bullet X$.

2. The situation is essentially different from the "deterministic" Riemann and Lebesgue integrals that coincide when they both exist.

Proof. By applying Proposition 7.8 and Corollary 7.9 we get:

$$\begin{aligned}
\mathbf{P}\text{-}\lim_n \sum_i \frac{Y_i + Y_{i+1}}{2}\Delta X_i &= \mathbf{P}\text{-}\lim_n \left[\sum_i Y_i \Delta X_i + \sum_i \frac{Y_{i+1} - Y_i}{2}\Delta X_i\right] \\
&= \mathbf{P}\text{-}\lim_n \sum_i Y_i \Delta X_i + \frac{1}{2}\mathbf{P}\text{-}\lim_n \sum_i \Delta Y_i \Delta X_i \\
&= \int_0^t Y_s \,\mathrm{d}X_s + \frac{1}{2}\langle X, Y\rangle_t. \qquad \triangle
\end{aligned}$$

REMARK.– Note that the Stratonovich integral $Y \circ X$ exists if Y is an Itô process. Though this is not necessary for the integrability in the Stratonovich sense, the class of "Stratonovich-integrable" processes is narrower than that of "Itô-integrable" processes. This seems to be the most serious disadvantage of the Stratonovich integral.

PROPOSITION 8.3 (Properties of Stratonovich integral).– *Let X, Y, Z, and W be Itô processes in the interval $[0, T]$. Then*:

(1) $Z \circ (\alpha X + \beta Y) = \alpha Z \circ X + \beta Z \circ Y$, $\alpha, \beta \in \mathbb{R}$;

(2) $(\alpha Z + \beta W) \circ X = \alpha Z \circ X + \beta W \circ X$, $\alpha, \beta \in \mathbb{R}$;

(3) $W \circ (Z \circ X) = (WZ) \circ X$;

(4) $\langle Z \circ X, W \circ Y\rangle = \langle Z \bullet X, W \bullet Y\rangle = (ZW) \circ \langle X, Y\rangle = (ZW) \bullet \langle X, Y\rangle$.

Proof. (1) By the linearity of stochastic integrals (Proposition 7.7, property 1),

$$\begin{aligned}
Z \circ (\alpha X + \beta Y) &= Z \bullet (\alpha X + \beta Y) + \frac{1}{2}\langle Z, \alpha X + \beta Y\rangle \\
&= \alpha Z \bullet X + \beta Z \bullet Y + \frac{1}{2}\big(\alpha\langle Z, X\rangle + \beta\langle Z, Y\rangle\big) \\
&= \alpha\Big(Z \bullet X + \frac{1}{2}\langle Z, X\rangle\Big) + \beta\Big(Z \bullet Y + \frac{1}{2}\langle Z, Y\rangle\Big) \\
&= \alpha Z \circ X + \beta Z \circ Y.
\end{aligned}$$

(2) Similarly,

$$\begin{aligned}
(\alpha Z + \beta W) \circ X &= (\alpha Z + \beta W) \bullet X + \frac{1}{2}\big\langle(\alpha Z + \beta W), X\big\rangle \\
&= \alpha Z \bullet X + \beta W \bullet X + \frac{1}{2}\big(\alpha\langle Z, X\rangle + \beta\langle W, X\rangle\big) \\
&= \alpha\Big(Z \bullet X + \frac{1}{2}\langle Z, X\rangle\Big) + \beta\Big(W \bullet X + \frac{1}{2}\langle W, X\rangle\Big) \\
&= \alpha Z \circ X + \beta W \circ X.
\end{aligned}$$

(3) On the right-hand side, we have:

$$\begin{aligned}W\circ(Z\circ X) &= W\bullet(Z\circ X)+\frac{1}{2}\langle W, Z\circ X\rangle\\ &= W\bullet\Big(Z\bullet X+\frac{1}{2}\langle Z,X\rangle\Big)+\frac{1}{2}\Big\langle W, Z\bullet X+\frac{1}{2}\langle Z,X\rangle\Big\rangle\\ &= (WZ)\bullet X+\frac{1}{2}W\bullet\langle Z,X\rangle+\frac{1}{2}Z\bullet\langle W,X\rangle+\frac{1}{4}\big\langle W,\langle Z,X\rangle\big\rangle,\end{aligned}$$

while, on the right-hand side:

$$\begin{aligned}(WZ)\circ X &= (WZ)\bullet X+\frac{1}{2}\langle WZ,X\rangle\\ &\overset{7.11}{=\!=} (WZ)\bullet X+\frac{1}{2}\big\langle W\bullet Z+Z\bullet W+\langle W,Z\rangle, X\big\rangle\\ &= (WZ)\bullet X+\frac{1}{2}W\bullet\langle Z,X\rangle+\frac{1}{2}Z\bullet\langle W,X\rangle+\frac{1}{2}\big\langle\langle W,Z\rangle,X\big\rangle.\end{aligned}$$

The last terms of both equalities equal zero, and we are finished.

(4) We have

$$\begin{aligned}\langle Z\circ X, W\circ Y\rangle &= \Big\langle Z\bullet X+\frac{1}{2}\langle Z,X\rangle, W\bullet Y+\frac{1}{2}\langle W,Y\rangle\Big\rangle\\ &= \langle Z\bullet X, W\bullet Y\rangle+\frac{1}{2}\big\langle\langle Z,X\rangle, W\bullet Y\big\rangle+\frac{1}{2}\big\langle Z\bullet X,\langle W,Y\rangle\big\rangle\\ &\quad+\frac{1}{4}\big\langle\langle Z,X\rangle,\langle W,Y\rangle\big\rangle\\ &= \langle Z\bullet X, W\bullet Y\rangle=(ZW)\bullet\langle X,Y\rangle;\\ (ZW)\circ\langle X,Y\rangle &= (ZW)\bullet\langle X,Y\rangle+\frac{1}{2}\big\langle ZW,\langle X,Y\rangle\big\rangle\\ &= (ZW)\bullet\langle X,Y\rangle.\end{aligned}$$

Here we used the fact that the covariation of the covariation of two Itô processes (which is a regular process) and any third Itô process equals zero. △

THEOREM 8.4 (Itô's formula in the Stratonovich form).– *Let X_t, $t\geqslant 0$, be an Itô process. Then, for every $F\in C^{2,3}(\mathbb{R}_+\times\mathbb{R})$,*[3]

$$F\big(t,X_t\big)-F\big(0,X_0\big)=\int_0^t F'_x\big(s,X_s\big)\circ \mathrm{d}X_s+\int_0^t F'_s\big(s,X_s\big)\,\mathrm{d}s.$$

3. That is, F has two and three continuous derivatives in t and x, respectively.

In other words, we have the "usual" formula (*valid for continuously differentiable* X).

Proof. From Proposition 8.2 we have:

$$\int_0^t F_x'\big(s, X_s\big) \circ \mathrm{d}X_s = \int_0^t F_x'\big(s, X_s\big)\ \mathrm{d}X_s + \frac{1}{2}\big\langle F_x'(\cdot, X_\cdot), X\big\rangle_t. \qquad [8.3]$$

Now we apply Itô's formula to the process $F_x'(t, X_t)$, $t \geqslant 0$:

$$F_x'(t, X_t) = F_x'(0, X_0) + \int_0^t F_{xx}''(s, X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t F_{xxx}'''(s, X_s)\,\mathrm{d}\langle X\rangle_s + \int_0^t F_{sx}''(s, X_s)\,\mathrm{d}s, \quad t \geqslant 0.$$

The last two terms are regular processes. Therefore, their covariations with X equal zero. The covariation of the process $\int_0^t F_{xx}''(s, X_s)\,\mathrm{d}X_s$, $t \geqslant 0$, and X is

$$\int_0^t F_{xx}''(s, X_s)\,\mathrm{d}\langle X\rangle_s, \quad t \geqslant 0$$

(Proposition 7.7, property 4). Substituting this into equation [8.3] and applying Itô's formula (Theorem 7.10), we get:

$$\begin{aligned}\int_0^t F_x'\big(s, X_s\big) \circ \mathrm{d}X_s &= \int_0^t F_x'\big(s, X_s\big)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t F_{xx}''(s, X_s)\,\mathrm{d}\langle X\rangle_s \\ &= F\big(t, X_t\big) - F\big(0, X_0\big) - \int_0^t F_s'\big(s, X_s\big)\,\mathrm{d}s. \qquad \triangle\end{aligned}$$

PROPOSITION 8.5 (Wong–Zakai theorem for stochastic integrals[4]).– *Let* X, X^n, $n \in \mathbb{N}$, *be Itô processes. Suppose that a.s.* $X^n \to X$, $n \to \infty$, *uniformly in the interval* $[0, t]$. *Then, for every* $f \in C^2(\mathbb{R}_+ \times \mathbb{R})$,

$$\int_0^t f\big(s, X_s^n\big) \circ \mathrm{d}X_s^n \to \int_0^t f(s, X_s) \circ \mathrm{d}X_s, \quad n \to \infty.$$

Proof. Denote

$$F(t, x) = \int_0^x f(t, y)\,\mathrm{d}y, \quad (t, x) \in \mathbb{R}_+ \times \mathbb{R}.$$

4. Eugene Wong, Moshe Zakai.

Since $F \in C^{2,3}$ and $F'_x(t,x) = f(t,x)$, $(t,x) \in \mathbb{R}_+ \times \mathbb{R}$, by Theorem 8.4 we have:

$$\begin{aligned}\int_0^t f\big(s, X_s^n\big) \circ \mathrm{d}X_s^n &= F\big(t, X_t^n\big) - F\big(0, X_0^n\big) - \int_0^t F'_s\big(s, X_s^n\big)\,\mathrm{d}s \\ &\to F(t, X_t) - F(0, X_0) - \int_0^t F'_s(s, X_s)\,\mathrm{d}s \\ &= \int_0^t f(s, X_s) \circ \mathrm{d}X_s.\end{aligned}$$

Here we used the fact that the uniform convergence $X^n \to X$ in any finite interval $[0,t]$ and the continuity of F and F'_s implies the uniform convergence $F(s, X_s^n) \to F(s, X_s)$ and $F'_s(s, X_s^n) \to F'_s(s, X_s)$ for $s \in [0,t]$. △

The proved stability of the Stratonovich stochastic integral shows that we can also expect a certain stability for Stratonovich stochastic differential equations (where stochastic integrals are taken in the Stratonovich sense). First, let us derive a relation between Stratonovich and Itô stochastic differential equations. Consider the Stratonovich equation

$$X_t = x_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s) \circ \mathrm{d}B_s, \quad t \geqslant 0. \tag{8.4a}$$

It makes sense if, for any Itô process X, the process $\sigma(s, X_s)$, $s \geqslant 0$, again is an Itô process. In view of Itô's formula, for this, it suffices that the coefficient σ is of class C^2 (i.e. σ is twice continuously differentiable). Let us try to rewrite equation [8.4a] in the Itô form (as a stochastic differential equation with Itô's integral). First, note that by Proposition 8.2 any solution of equation [8.4a] satisfies the equation

$$X_t = x_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s + \frac{1}{2}\big\langle \sigma(\cdot,\, X.), B\big\rangle_t, \quad t \geqslant 0. \tag{8.4a'}$$

Therefore, $\langle X\rangle_t = \int_0^t \sigma^2(s, X_s)\,\mathrm{d}s$, and, thus, by Itô's formula (Theorem 7.10),

$$\begin{aligned}\sigma(t, X_t) = \sigma(0, X_0) &+ \int_0^t \sigma'_x(s, X_s)\,\mathrm{d}X_s + \int_0^t \sigma'_s(s, X_s)\,\mathrm{d}s \\ &+ \frac{1}{2}\int_0^t \sigma''_{xx}(s, X_s)\,\mathrm{d}\langle X\rangle_s\end{aligned}$$

$$
\begin{aligned}
&= \sigma(0,X_0) + \int_0^t b\sigma'_x(s,X_s)\,\mathrm{d}s + \int_0^t \sigma\sigma'_x(s,X_s)\,\mathrm{d}B_s \\
&\quad + \frac{1}{2}\int_0^t \sigma'_x(s,X_s)\,\mathrm{d}\big\langle \sigma(\cdot,X.),B\big\rangle_s \\
&\quad + \int_0^t \sigma'_s(s,X_s)\,\mathrm{d}s + \frac{1}{2}\int_0^t \sigma^2\sigma''_{xx}(s,X_s)\,\mathrm{d}s \\
&= \int_0^t \sigma\sigma'_x(s,X_s)\,\mathrm{d}B_s + \text{regular process.}
\end{aligned}
$$

Therefore,

$$
\big\langle \sigma(\cdot,X.),B\big\rangle_t = \int_0^t \sigma\sigma'_x(s,X_s)\,\mathrm{d}s. \tag{8.5}
$$

Substituting this into equation [8.4a′], we get the following Itô equation equivalent to equation [8.4a]:

$$
X_t = x_0 + \int_0^t \Big(b(s,X_s) + \frac{1}{2}\sigma\sigma'_x(s,X_s)\Big)\,\mathrm{d}s + \int_0^t \sigma(s,X_s)\,\mathrm{d}B_s, \quad t \geqslant 0. \tag{8.4b}
$$

Conversely, if a random process X satisfies equation [8.4b] and $\sigma \in C^2(\mathbb{R}_+ \times \mathbb{R})$, then we similarly obtain equality [8.5]. Therefore,

$$
\frac{1}{2}\int_0^t \sigma\sigma'_x(s,X_s)\,\mathrm{d}s + \int_0^t \sigma(s,X_s)\,\mathrm{d}B_s = \int_0^t \sigma(s,X_s)\circ\mathrm{d}B_s, \quad t \geqslant 0,
$$

and thus equation [8.4b] implies equation [8.4a]. Summarizing, we have the following:

PROPOSITION 8.6.– *If* $b\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ *and* $\sigma \in C^2(\mathbb{R}_+ \times \mathbb{R})$*, then each solution of equation* [8.4a] *is a solution of equation* [8.4b] *and vice versa.*

REMARK.– One easily notes that if, in equation [8.4a], B is an arbitrary Itô process, then equation [8.4b] must be replaced by the equation

$$
X_t = x_0 + \int_0^t b(s,X_s)\,\mathrm{d}s + \frac{1}{2}\int_0^t \sigma\sigma'_x(s,X_s)\,\mathrm{d}\langle B\rangle_s + \int_0^t \sigma(s,X_s)\,\mathrm{d}B_s, \quad t \geqslant 0.
$$

Let us now consider the sequence of equations

$$
X_t^n = x_0 + \int_0^t b\big(s,X_s^n\big)\,\mathrm{d}s + \int_0^t \sigma\big(s,X_s^n\big)\circ\mathrm{d}Z_s^n, \quad t \geqslant 0,\ n \in \overline{\mathbb{N}} = \mathbb{N}\cup\{\infty\},
$$

where Z^n, $n \in \mathbb{N}$, are arbitrary Itô processes converging, as $n \to \infty$, to an Itô process $Z := Z^\infty$. Without loss of generality, we may suppose that $Z_0^n = 0$, $n \in \overline{\mathbb{N}}$.

THEOREM 8.7 (Stability of Stratonovich SDEs).– *If $Z^n \to Z$, $n \to \infty$, uniformly on finite time intervals a.s., then $X^n \to X := X^\infty$, $n \to \infty$, uniformly on finite time intervals a.s., under the condition that the coefficients b and σ are good enough.*[5]

Proof. First, suppose that σ satisfies the following boundedness conditions:

$$0 < c \leqslant \sigma(t,x) \leqslant C, \qquad (t,x) \in \mathbb{R}_+ \times \mathbb{R}.$$

Rewrite the equation in the differential form:

$$\mathrm{d}X_t = b(t,X_t)\,\mathrm{d}t + \sigma(t,X_t) \circ \mathrm{d}Z_t.$$

By property 3 of Proposition 8.3, we can multiply the equation by $\sigma^{-1}(t,X_t)$:

$$\sigma^{-1}(t,X_t) \circ \mathrm{d}X_t = \sigma^{-1}b(t,X_t)\,\mathrm{d}t + \mathrm{d}Z_t.$$

Returning to the integral form, we get

$$\int_0^t \sigma^{-1}(s,X_s) \circ \mathrm{d}X_s = \int_0^t \sigma^{-1}b(s,X_s)\,\mathrm{d}s + Z_t.$$

Denote $G(t,x) = \int_0^x \sigma^{-1}(t,y)\,\mathrm{d}y$ (i.e. $G'_x(t,x) = \sigma^{-1}(t,x)$). Applying Itô's formula (Theorem 8.4) to the Stratonovich integral on the right-hand side, we get:

$$\begin{aligned} G(t,X_t) - G(0,X_0) &= \int_0^t G'_x(s,X_s) \circ \mathrm{d}X_s + \int_0^t G'_s(s,X_s)\,\mathrm{d}s \\ &= \int_0^t (\sigma^{-1}b + G'_s)(s,X_s)\,\mathrm{d}s + Z_t. \end{aligned}$$

For every fixed t, $G(t,\cdot)$ is a continuous and strictly increasing function, and

$$\lim_{x \to \pm\infty} G(t,x) = \pm\infty.$$

This follows from the inequality

$$\bigl|G(t,x)\bigr| = \bigl|G(t,x) - G(t,0)\bigr| \geqslant \inf_{y \in \mathbb{R}} \bigl|G'_y(t,y)\bigr| |x - 0| \geqslant C^{-1}|x|,$$

5. The precise meaning of "goodness" of the coefficients will become clear in the course of the proof.

which, in turn, follows from the Lagrange mean-value theorem applied to the function $G(t,\cdot)$. Therefore, for every fixed t, the function $G(t,\cdot)$ has the inverse defined on $\mathbb{R}$. Denote it by $H(t,\cdot)$. Then we have:

$$X_t = H\left(t, G(0,x_0) + \int_0^t \big(\sigma^{-1}b + G'_s\big)(s, X_s)\,\mathrm{d}s + Z_t\right). \qquad [8.6]$$

The same expression is valid for the approximating equations:

$$X^n_t = H\left(t, G(0,x_0) + \int_0^t \big(\sigma^{-1}b + G'_s\big)\big(s, X^n_s\big)\,\mathrm{d}s + Z^n_t\right). \qquad [8.6^n]$$

Since $C^{-1} \leqslant G'_x(t,x) \leqslant c^{-1}$, we have $c \leqslant H'_x(t,x) \leqslant C$, and thus $|H(t,x) - H(t,y)| \leqslant C|x-y|$, $x,y \in \mathbb{R}$, i.e. the function H satisfies the Lipschitz condition. Therefore, denoting $K := \sigma^{-1}b + G'_t$, we can estimate the difference of solutions:

$$\begin{aligned}\big|X^n_t - X_t\big| &\leqslant C\left|\int_0^t \big(K\big(s, X^n_s\big) - K(s, X_s)\big)\,\mathrm{d}s + Z^n_t - Z_t\right| \\ &\leqslant C\int_0^t \big|K\big(s, X^n_s\big) - K(s,X_s)\big|\,\mathrm{d}s + C\big|Z^n_t - Z_t\big|.\end{aligned}$$

We now see that it is natural to require the Lipschitz condition for the function K (this is an additional "goodness" requirement for the coefficients; it is satisfied if, for example, the derivatives $(b\sigma^{-1})'_x(t,x)$ and $(\sigma^{-1})'_t(t,x))$ are bounded). Then, in such a case, we should have:

$$\big|X^n_t - X_t\big| \leqslant C_1\int_0^t \big|X^n_s - X_s\big|\,\mathrm{d}s + C\big|Z^n_t - Z_t\big|.$$

Now we can apply the Gronwall lemma (Lemma 6.3) with any fixed T, $\varphi(t) := |X^n_t - X_t|$, and $\varepsilon := C\sup_{t\leqslant T}|Z^n_t - Z_t|$. This gives:

$$\big|X^n_t - X_t\big| \leqslant C\sup_{t\leqslant T}\big|Z^n_t - Z_t\big|\,\mathrm{e}^{C_1 t}, \quad t \leqslant T,$$

and hence,

$$\sup_{t\leqslant T}\big|X^n_t - X_t\big| \leqslant C\mathrm{e}^{C_1 T}\sup_{t\leqslant T}\big|Z^n_t - Z_t\big| \to 0, \quad n\to\infty.$$

△

REMARKS.– 1. The main point in the proof where we used the advantages of the Stratonovich integral is equations [8.6] and [8.6^n] where the limit process Z and its approximations Z^n "participate" without stochastic integrals. If we replaced the Stratonovich integrals by the Itô integrals, we should get the equations

$$X_t = H\left(t, G(0,x_0) + \int_0^t (\sigma^{-1}b + G'_s)(s, X_s)\,\mathrm{d}s + \frac{1}{2}\int_0^t G''_{xx}(s, X_s)\,\mathrm{d}\langle Z\rangle_s + Z_t\right)$$

and

$$X_t^n = H\left(t, G(0,x_0) + \int_0^t \left(\sigma^{-1}b + G'_s\right)\left(s, X_s^n\right)\mathrm{d}s + \frac{1}{2}\int_0^t G''_{xx}\left(s, X_s^n\right)\mathrm{d}\left\langle Z^n\right\rangle_s + Z_t^n\right).$$

Then, when trying, as in the proof of the theorem, to estimate the difference $|X_t^n - X_t|$, in the corresponding inequality, an extraneous term would appear:

$$\begin{aligned}\left|X_t^n - X_t\right| \leqslant C_1 \int_0^t \left|X_s^n - X_s\right|\mathrm{d}s + C\left|Z_t^n - Z_t\right| \\ + \frac{C}{2}\left|\int_0^t G''_{xx}\left(s, X_s^n\right)\mathrm{d}\left\langle Z^n\right\rangle_s - \int_0^t G''_{xx}(s, X_s)\,\mathrm{d}\left\langle Z\right\rangle_s\right|.\end{aligned}$$

This means that we cannot expect the convergence $|X_t^n - X_t| \to 0$, because the convergence of Z^n to Z does not imply that of quadratic variations. For a counterexample, it suffices to consider the case where $Z = B$ is a Brownian motion approximated by regular processes Z^n, for example, by polygonal lines; then $\langle Z\rangle_t = t$, while $\langle Z^n\rangle_t = 0$. Thus, to get a stability property for Itô equations, we have to require not only the convergence $Z^n \to Z$ but also the convergence $\langle Z^n\rangle \to \langle Z\rangle$, which is a too strong condition in view of applications.

2. The theorem is actually true when b is Lipschitz and σ has two bounded derivatives; however, in this case the proof is more complicated.

Now, let us consider the particular case where a sequence of continuous *regular* processes $\{B^n\}$ converges, uniformly in finite intervals, to a Brownian motion B. For example, we may approximate the Brownian motion B by polygonal lines B^n with values at the points $\frac{k}{n}$ coinciding with those of B. In this case, the Stratonovich equations (S^n) coincide with the Itô equations

$$X_t^n = X_0 + \int_0^t b\left(s, X_s^n\right)\mathrm{d}s + \int_0^t \sigma\left(s, X_s^n\right)\mathrm{d}B_s^n, \quad t \geqslant 0,\ n \in \mathbb{N}.$$

COROLLARY 8.8 (Wong–Zakai theorem for SDEs).– *Suppose that the coefficients b and σ are as in Theorem 8.7. If a sequence of regular processes $\{B^n\}$ converges, as*

$n \to \infty$, to a Brownian motion B uniformly in finite intervals, then the sequence of the corresponding solutions $\{X^n\}$ converges to the solution X of the equation

$$X_t = X_0 + \int_0^t \Big(b(s, X_s) + \frac{1}{2}\sigma\sigma'_x(s, X_s) \Big)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s, \quad t \geqslant 0,$$

uniformly in finite intervals.

Proof. It suffices to apply Theorem 8.7, recalling that the latter equation coincides with equation [8.4a] written in the Itô form (Proposition 8.6). △

8.1. Exercises

8.1. Find:

(1) $B \circ (B \circ B)_t$, $t \geqslant 0$;

(2) $\mathbf{E} \displaystyle\int_0^t B_s^3 \circ \mathrm{d}Y_s$ if $Y_t = B^2 \circ B_t$; and

(3) $\mathbf{E} \displaystyle\int_0^t X_s \circ \mathrm{d}Y_s$, $t \geqslant 0$, if $X_t := \displaystyle\int_0^t B_s\,\mathrm{d}s$ and $Y_t := \displaystyle\int_0^t s\,\mathrm{d}B_s$.

8.2. Find a positive coefficient $\sigma(x)$, $x \in \mathbb{R}$, in the Stratonovich SDE

$$\mathrm{d}X_t = -X_t\,\mathrm{d}t + \sigma(X_t) \circ \mathrm{d}B_t$$

such that the solution of the equation (with arbitrary initial condition) is a martingale.

8.3. Let $\{B^n\}$ be a sequence of regular processes converging to a Brownian motion B uniformly on compact intervals. Denote by X^n the solution of the SDE

$$dX_t^n = X_t^n\, d(B^n - B)_t, \qquad X_0^n = 1.$$

Find the limit

$$X_t := \lim_{n\to\infty} X_t^n, \qquad t \geqslant 0.$$

8.4. Can the probability

$$\mathbf{P}\Big\{ \int_0^T X_s \circ \mathrm{d}B_s > \int_0^T X_s\,\mathrm{d}B_s \Big\}$$

be equal to 0? 1? 1/2?

8.5. Solve the Stratonovich stochastic differential equation

$$\mathrm{d}X_t = 2X_t\,\mathrm{d}t + 3X_t \circ \mathrm{d}B_t, \qquad X_0 = 1.$$

Chapter 9

Linear Stochastic Differential Equations

9.1. Explicit solution of a linear SDE

DEFINITION 9.1.– *A linear stochastic differential equation* (LSDE) *is an equation of the form*

$$\mathrm{d}X_t = \big(a_1(t)X_t + a_2(t)\big)\,\mathrm{d}t + \big(b_1(t)X_t + b_2(t)\big)\,\mathrm{d}B_t, \qquad X_0 = x_0, \quad [9.1]$$

where a_i and b_i $(i = 1, 2)$ are non-random functions bounded on every finite interval $[0, T]$ (e.g. continuous); it is easy to check that the coefficients of an LSDE satisfy the Lipschitz conditions, and thus the equation has a unique solution. If a_i and b_i are constants, then the LSDE is called autonomous; if $a_2 \equiv b_2 \equiv 0$, then it is called homogeneous.

Denote

$$\Phi_t = \exp\left\{ \int_0^t \Big(a_1(s) - \frac{1}{2}b_1^2(s)\Big)\,\mathrm{d}s + \int_0^t b_1(s)\,\mathrm{d}B_s \right\}, \quad t \geqslant 0.$$

PROPOSITION 9.2.– *The random process Φ_t, $t \geqslant 0$, is a solution of the homogeneous LSDE*

$$\mathrm{d}Y_t = a_1(t)Y_t\,\mathrm{d}t + b_1(t)Y_t\,\mathrm{d}B_t \qquad [9.2]$$

with the initial condition $Y_0 = 1$. The process Φ is called the fundamental solution of equation [9.2].

Proof. Denote

$$Z_t = \int_0^t \Big(a_1(s) - \frac{1}{2}b_1^2(s)\Big)\,\mathrm{d}s + \int_0^t b_1(s)\,\mathrm{d}B_s.$$

Thus, $\Phi_t = \mathrm{e}^{Z_t}$, $t \geqslant 0$. Then, using Itô's formula, we get:

$$\begin{aligned}\mathrm{d}\Phi_t &= \mathrm{e}^{Z_t}\,\mathrm{d}Z_t + \frac{1}{2}\mathrm{e}^{Z_t}\,\mathrm{d}\langle Z\rangle_t \\ &= \mathrm{e}^{Z_t}\Big(a_1(t) - \frac{1}{2}b_1^2(t)\Big)\,\mathrm{d}t + \mathrm{e}^{Z_t}b_1(t)\,\mathrm{d}B_t + \frac{1}{2}\mathrm{e}^{Z_t}b_1^2(t)\,\mathrm{d}t \\ &= \mathrm{e}^{Z_t}a_1(t)\,\mathrm{d}t + \mathrm{e}^{Z_t}b_1(t)\,\mathrm{d}B_t \\ &= a_1(t)\Phi_t\,\mathrm{d}t + b_1(t)\Phi_t\,\mathrm{d}B_t.\end{aligned}$$

The initial condition $\Phi_0 = 1$ is clearly satisfied. △

PROPOSITION 9.3.– *The random process* Φ_t^{-1}*,* $t \geqslant 0$*, is a solution of the LSDE*

$$\mathrm{d}Y_t = \big(-a_1(t) + b_1^2(t)\big)Y_t\,\mathrm{d}t - b_1(t)Y_t\,\mathrm{d}B_t, \qquad Y_0 = 1.$$

Proof. A verification is similar to the previous one:

$$\begin{aligned}\mathrm{d}\Phi_t^{-1} &= \mathrm{d}\big(\mathrm{e}^{-Z_t}\big) = -\mathrm{e}^{-Z_t}\,\mathrm{d}Z_t + \frac{1}{2}\mathrm{e}^{-Z_t}\,\mathrm{d}\langle Z\rangle_t \\ &= -\Big[\mathrm{e}^{-Z_t}\Big(a_1(t) - \frac{1}{2}b_1^2(t)\Big)\,\mathrm{d}t + \mathrm{e}^{-Z_t}b_1(t)\,\mathrm{d}B_t\Big] + \frac{1}{2}\mathrm{e}^{-Z_t}b_1^2(t)\,\mathrm{d}t \\ &= \mathrm{e}^{-Z_t}\big(-a_1(t) + b_1^2(t)\big)\,\mathrm{d}t - \mathrm{e}^{-Z_t}b_1(t)\,\mathrm{d}B_t \\ &= \big(-a_1(t) + b_1^2(t)\big)\Phi_t^{-1}\,\mathrm{d}t - b_1(t)\Phi_t^{-1}\,\mathrm{d}B_t.\end{aligned}$$

The initial condition $\Phi_0^{-1} = 1$ again is clearly satisfied. △

THEOREM 9.4.– *The random process*

$$X_t = \Phi_t\left\{x_0 + \int_0^t \big(a_2(s) - b_1(s)b_2(s)\big)\Phi_s^{-1}\,\mathrm{d}s + \int_0^t b_2(s)\Phi_s^{-1}\,\mathrm{d}B_s\right\}, \quad t \geqslant 0,$$

is a solution of the LSDE [9.1].

Proof. Applying the integration-by-parts formula (Corollary 7.11) to the (unique) solution X_t of equation [9.1] and the process $Y_t = \Phi_t^{-1}$ and using Proposition 9.3, we get:

$$\begin{aligned}
\mathrm{d}\big(X_t\Phi_t^{-1}\big) &= X_t\,\mathrm{d}\Phi_t^{-1} + \Phi_t^{-1}\,\mathrm{d}X_t + \mathrm{d}\langle X,\Phi^{-1}\rangle_t \\
&= X_t\big(-a_1(t)+b_1^2(t)\big)\Phi_t^{-1}\,\mathrm{d}t + X_t\big(-b_1(t)\Phi_t^{-1}\big)\,\mathrm{d}B_t \\
&\quad + \Phi_t^{-1}\big(a_1(t)X_t + a_2(t)\big)\,\mathrm{d}t + \Phi_t^{-1}\big(b_1(t)X_t + b_2(t)\big)\,\mathrm{d}B_t \\
&\quad + \big(b_1(t)X_t + b_2(t)\big)\big(-b_1(t)\Phi_t^{-1}\big)\,\mathrm{d}t \\
&= \big(a_2(t) - b_1(t)b_2(t)\big)\Phi_t^{-1}\,\mathrm{d}t + b_2(t)\Phi_t^{-1}\,\mathrm{d}B_t.
\end{aligned}$$

Integrating the expression obtained and taking into account that $\Phi_0^{-1}X_0 = x_0$, we get:

$$\Phi_t^{-1}X_t = x_0 + \int_0^t \big(a_2(s) - b_1(s)b_2(s)\big)\Phi_s^{-1}\,\mathrm{d}s + \int_0^t b_2(s)\Phi_s^{-1}\,\mathrm{d}B_s, \quad t \geqslant 0.$$

It remains to multiply both sides by Φ_t. △

Now let us consider some particular cases.

9.5.– ADDITIVE NOISE. If $b_1 \equiv 0$, then equation [9.1] is called an LSDE with additive noise.

1. The *Langevin*[1] equation

$$\mathrm{d}X_t = -aX_t\,\mathrm{d}t + b\,\mathrm{d}B_t, \qquad X_0 = x_0,$$

describes the dynamics of velocity of a particle moving in viscous fluid influenced by chaotic "kicks" of molecules. The coefficient $a > 0$ characterizes the viscosity of a fluid, and $b > 0$ characterizes the intensity of the kicks. The solution of the Langevin equation can be taken for a model of velocity of "real" (physical) Brownian motion when the viscosity of medium is taken into account.

In this case $\Phi_t = \mathrm{e}^{-at}$, and therefore the solution of the equation is the process

$$\begin{aligned}
X_t &= \mathrm{e}^{-at}\left(x_0 + b\int_0^t \mathrm{e}^{as}\,\mathrm{d}B_s\right) \\
&= \mathrm{e}^{-at}\left(x_0 + b\mathrm{e}^{at}B_t - ab\int_0^t \mathrm{e}^{as}B_s\,\mathrm{d}s\right), \quad t \geqslant 0.
\end{aligned}$$

1. Paul Langevin.

Although the equation has the name of Langevin, its solution is usually called the *Ornstein–Uhlenbeck*[2] *process*.

2. The solution of the non-homogeneous equation with constant coefficients

$$\mathrm{d}X_t = (aX_t + b)\,\mathrm{d}t + c\,\mathrm{d}B_t, \quad X_0 = x_0,$$

is

$$X_t = \mathrm{e}^{at}\left(x_0 + \frac{b}{a}(1 - \mathrm{e}^{-at}) + c\int_0^t \mathrm{e}^{-as}\,\mathrm{d}B_s\right), \quad t \geqslant 0. \qquad [9.3]$$

3. The non-homogeneous equation with non-constant coefficients

$$\mathrm{d}X_t = \big(a(t)X_t + b(t)\big)\,\mathrm{d}t + c(t)\,\mathrm{d}B_t, \quad X_0 = x_0,$$

has the solution

$$X_t = \exp\left\{\int_0^t a(s)\,\mathrm{d}s\right\}\left[x_0 + \int_0^t b(s)\exp\left\{-\int_0^s a(u)\,\mathrm{d}u\right\}\mathrm{d}s + \int_0^t c(s)\exp\left\{-\int_0^s a(u)\,\mathrm{d}u\right\}\mathrm{d}B_s\right], \quad t \geqslant 0.$$

9.6.– MULTIPLICATIVE NOISE ($b_1 \not\equiv 0$).

1. The homogeneous equation with constant coefficients (growth equation)

$$\mathrm{d}X_t = aX_t\,\mathrm{d}t + bX_t\,\mathrm{d}B_t, \qquad X_0 = x_0,$$

has the solution

$$X_t = x_0\Phi_t = x_0\exp\left\{\Big(a - \frac{1}{2}b^2\Big)t + bB_t\right\}, \quad t \geqslant 0.$$

2. The non-homogeneous equation with constant coefficients

$$\mathrm{d}X_t = (aX_t + c)\,\mathrm{d}t + (bX_t + d)\,\mathrm{d}B_t, \qquad X_0 = x_0,$$

2. Leonard Salomon Ornstein, George Eugene Uhlenbeck.

has the solution

$$X_t = \Phi_t \left[x_0 + (c - bd) \int_0^t \Phi_s^{-1} \, \mathrm{d}s + d \int_0^t \Phi_s^{-1} \, \mathrm{d}B_s \right], \quad t \geqslant 0, \tag{9.4}$$

$$\Phi_t = \exp\left\{ \left(a - \frac{1}{2} b^2 \right) t + b B_t \right\}.$$

3. The homogeneous equation with non-constant coefficients

$$\mathrm{d}X_t = a(t) X_t \, \mathrm{d}t + b(t) X_t \, \mathrm{d}B_t, \qquad X_0 = x_0,$$

has the solution

$$X_t = x_0 \exp\left\{ \int_0^t \left(a(s) - \frac{1}{2} b^2(s) \right) \mathrm{d}s + \int_0^t b(s) \, \mathrm{d}B_s \right\}, \quad t \geqslant 0.$$

9.2. Expectation and variance of a solution of an LSDE

Important characteristics of a solution X of an LSDE—the expectation $E(t) = \mathbf{E}X_t$, second moment $Q(t) = \mathbf{E}X_t^2$, and variance $D(t) = \mathbf{D}X_t = Q(t) - E^2(t)$—can be found without solving the equation itself, only by using the properties of stochastic integrals. Taking the expectations of both sides of the equation

$$X_t = x_0 + \int_0^t \left(a_1(s) X_s + a_2(s) \right) \mathrm{d}s + \int_0^t \left(b_1(s) X_s + b_2(s) \right) \mathrm{d}B_s,$$

we get:

$$E(t) = x_0 + \int_0^t \left(a_1(s) E(s) + a_2(s) \right) \mathrm{d}s.$$

By differentiating the latter we get the ordinary differential equation

$$\frac{\mathrm{d}E(t)}{\mathrm{d}t} = a_1(t) E(t) + a_2(t), \qquad E(0) = x_0,$$

the solution of which is

$$E(t) = \exp\left\{ \int_0^t a_1(s) \, \mathrm{d}s \right\} \left(x_0 + \int_0^t a_2(s) \exp\left\{ - \int_0^s a_1(u) \, \mathrm{d}u \right\} \mathrm{d}s \right), \quad t \geqslant 0.$$

We also could get this formula by using the formula of Theorem 9.4 with $b_1 \equiv b_2 \equiv 0$.

Now, let us apply Itô's formula to the function $F(x) = x^2$ and the solution X:

$$\begin{aligned}
X_t^2 &= x_0^2 + 2\int_0^t X_s \,\mathrm{d}X_s + \int_0^t \mathrm{d}\langle X\rangle_s \\
&= x_0^2 + 2\int_0^t X_s\big[a_1(s)X_s + a_2(s)\big]\,\mathrm{d}s + 2\int_0^t X_s\big[b_1(s)X_s + b_2(s)\big]\,\mathrm{d}B_s \\
&\quad + \int_0^t \big[b_1(s)X_s + b_2(s)\big]^2\,\mathrm{d}s \\
&= x_0^2 + \int_0^t \Big[2a_1(s)X_s^2 + 2a_2(s)X_s + \big(b_1(s)X_s + b_2(s)\big)^2\Big]\,\mathrm{d}s \\
&\quad + 2\int_0^t \big[b_1(s)X_s^2 + b_2(s)X_s\big]\,\mathrm{d}B_s.
\end{aligned}$$

Taking the expectation, we get:

$$\begin{aligned}
Q(t) = \mathbf{E}X_t^2 &= x_0^2 + \int_0^t \mathbf{E}\Big[2\big(a_1(s)X_s + a_2(s)\big)X_s + \big(b_1(s)X_s + b_2(s)\big)^2\Big]\,\mathrm{d}s \\
&= x_0^2 + \int_0^t \mathbf{E}\big[2a_1(s)X_s^2 + 2a_2(s)X_s + b_1^2(s)X_s^2 \\
&\qquad\qquad + 2b_1(s)b_2(s)X_s + b_2^2(s)\big]\,\mathrm{d}s \\
&= x_0^2 + \int_0^t \big[2a_1(s)Q(s) + 2a_2(s)E(s) + b_1^2(s)Q(s) \\
&\qquad\qquad + 2b_1(s)b_2(s)E(s) + b_2^2(s)\big]\,\mathrm{d}s \\
&= x_0^2 + \int_0^t \Big[\big(2a_1(s) + b_1^2(s)\big)Q(s) \\
&\qquad\qquad + 2\big(a_2(s) + b_1(s)b_2(s)\big)E(s) + b_2^2(s)\Big]\,\mathrm{d}s.
\end{aligned}$$

Differentiation now gives the equation

$$\frac{\mathrm{d}Q(t)}{\mathrm{d}t} = \big(2a_1(t) + b_1^2(t)\big)Q(t) + 2\big(a_2(t) + b_1(t)b_2(t)\big)E(t) + b_2^2(t)$$

with the initial condition $Q(0) = x_0^2$. Its solution is

$$Q(t) = \exp\left\{\int_0^t A_1(s)\,\mathrm{d}s\right\}\left(x_0^2 + \int_0^t A_2(s)\exp\left\{-\int_0^s A_1(u)\,\mathrm{d}u\right\}\mathrm{d}s\right), \quad t \geqslant 0,$$

where we denoted

$$A_1(t) := 2a_1(t) + b_1^2(t), \qquad A_2(t) := 2\big(a_2(t) + b_1(t)b_2(t)\big)E(t) + b_2^2(t).$$

EXAMPLES 9.7.– In the case of *Langevin* equation $\mathrm{d}X_t = -aX_t\,\mathrm{d}t + b\,\mathrm{d}B_t$, we have $a_1(t) = -a$, $b_2(t) = b$, $a_2(t) = b_1(t) = 0$, $A_1(t) = -2a$, and $A_2(t) = b^2$. From this, using the expressions obtained for $E(t)$ and $Q(t)$, we get:

$$E(t) = x_0\mathrm{e}^{-at},$$

$$Q(t) = \mathrm{e}^{-2at}\left(x_0^2 + b^2\int_0^t \mathrm{e}^{2as}\,\mathrm{d}s\right) = \mathrm{e}^{-2at}\Big(x_0^2 + \frac{b^2}{2a}\big(\mathrm{e}^{2at} - 1\big)\Big)$$

$$= x_0^2\mathrm{e}^{-2at} + \frac{b^2}{2a}\big(1 - \mathrm{e}^{-2at}\big),$$

$$D(t) = Q(t) - E^2(t) = \frac{b^2}{2a}\big(1 - \mathrm{e}^{-2at}\big).$$

Note that, for any initial value $X_0 = x_0$, the expectation and variance of the Ornstein–Uhlenbeck process converge to constant values:

$$E(t) = x_0\mathrm{e}^{-at} \to 0, \qquad D(t) = \frac{b^2}{2a}\big(1 - \mathrm{e}^{-2at}\big) \to \frac{b^2}{2a}, \quad t \to \infty.$$

This can be interpreted as a certain stabilization of the behavior of the process after rather long time. This is also confirmed by Figure 9.1, where we see 11 trajectories of the Ornstein–Uhlenbeck process starting from different points.

We see that, eventually, the trajectories swarm not only around the time axis but also tend towards a narrow fiber of close trajectories. Such a phenomenon in physics is called *synchronization under identical noises*. It is specific not only for the Langevin equation but also for many other stochastic differential equations, the solutions of which starting at different points have the same limit distribution as $t \to \infty$. In the case of the Langevin equation, synchronization can be checked very simply. The

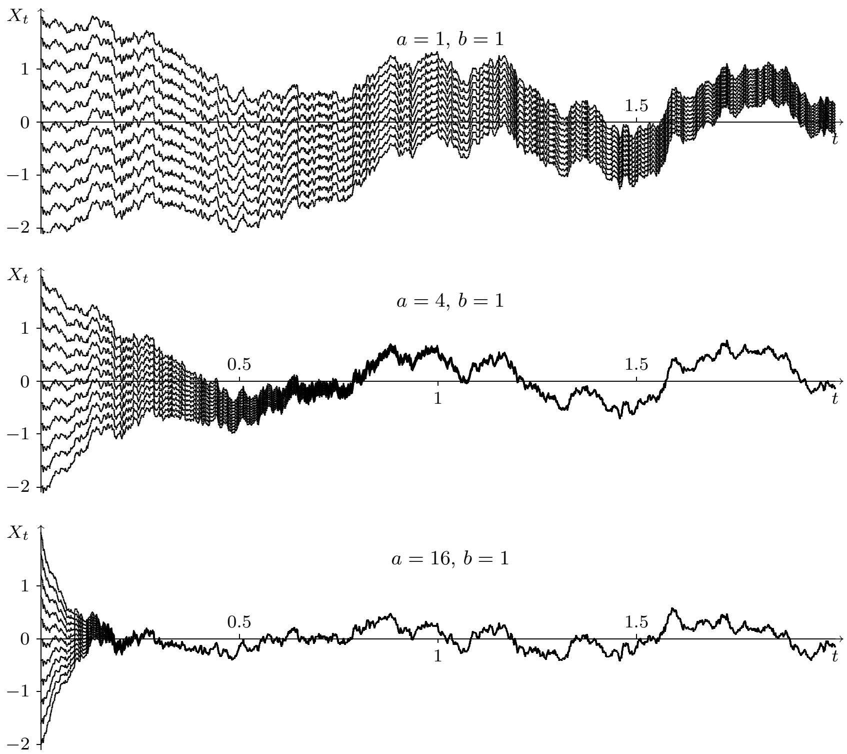


Figure 9.1. *Solutions of SDE* $\mathrm{d}X_t = -aX_t\,\mathrm{d}t + b\,\mathrm{d}B_t$, $X_0 = 0$, *for* $b = 1$ *and* $a = 1;\ 4;\ 16$

difference $Y_t := X_t^1 - X_t^2$ between two trajectories X_t^1 and X_t^2 with any initial points $X_0^1 = x_1$ and $X_0^2 = x_2$ satisfies the non-random equation

$$Y_t = x_1 - x_2 - a\int_0^t Y_s\,\mathrm{d}s;$$

its solution is

$$Y_t = (x_1 - x_2)\mathrm{e}^{-at} \to 0, \quad t \to \infty,$$

and the greater the coefficient a, the higher the convergence rate.

In Chapter 11, we shall see that, in fact, the distribution of the values of the Ornstein–Uhlenbeck process X_t have the limit normal distribution $N(0, b^2/2a)$.

In the case of the *growth* equation $\mathrm{d}X_t = aX_t\,\mathrm{d}t + bX_t\,\mathrm{d}B_t$, we have $a_1(t) = a$, $b_1(t) = b$, $a_2(t) = b_2(t) = 0$, $A_1(t) = 2a + b^2$, and $A_2(t) = 0$. Now we have:

$$E(t) = x_0\mathrm{e}^{at},$$

$$Q(t) = x_0^2\mathrm{e}^{(2a+b^2)t},$$

$$D(t) = x_0^2\mathrm{e}^{2at}\big(\mathrm{e}^{b^2t} - 1\big).$$

9.3. Other explicitly solvable equations

Consider an equation of the form

$$\mathrm{d}X_t = \frac{1}{2}\sigma(X_t)\sigma'(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \qquad X_0 = x_0,$$

with positive coefficient σ. Note that it can be rewritten in the Stratonovich form as

$$\mathrm{d}X_t = \sigma(X_t) \circ \mathrm{d}B_t, \qquad X_0 = x_0,$$

which, in fact, is already solved in the proof of Theorem 8.7:

$$X_t = H\big(G(x_0) + B_t\big),$$

where G is the antiderivative of σ^{-1} ($G' = \sigma^{-1}$), and H is the inverse function of G. For example, the solution of the equation

$$\mathrm{d}X_t = -\sin X_t \cos^3 X_t\,\mathrm{d}t + \cos^2 X_t\,\mathrm{d}B_t, \qquad X_0 = x_0 \in (-\pi/2, \pi/2),$$

is

$$X_t = \operatorname{arctg}\,\big(\operatorname{tg} x_0 + B_t\big).$$

As a slight generalization, we get that the equation

$$\mathrm{d}X_t = \Big(\alpha\sigma(X_t) + \frac{1}{2}\sigma(X_t)\sigma'(X_t)\Big)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \qquad X_0 = x_0,$$

or, equivalently,

$$\mathrm{d}X_t = \alpha\sigma(X_t)\,\mathrm{d}t + \sigma(X_t) \circ \mathrm{d}B_t, \qquad X_0 = x_0,$$

has the solution

$$X_t = H\big(G(x_0) + \alpha t + B_t\big).$$

Let us try to similarly solve an equation of the more general form

$$\mathrm{d}X_t = \Big(\alpha\sigma(X_t)h(X_t) + \frac{1}{2}\sigma(X_t)\sigma'(X_t)\Big)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t, \qquad X_0 = x_0.$$

We immediately get the equation

$$G(X_t) = G(x_0) + \alpha \int_0^t h(X_s)\,\mathrm{d}s + B_t, \quad t \geqslant 0.$$

Denote $Y_t := G(X_t)$. Then $X_t = H(Y_t)$, and we get a (nonlinear) equation with additive noise:

$$Y_t = Y_0 + \alpha \int_0^t h\big(H(Y_s)\big)\,\mathrm{d}s + B_t, \quad t \geqslant 0.$$

It can be interpreted as a Volterra integral equation with given function B_t, $t \geqslant 0$, that can be solved by functional analysis methods for every concrete observable trajectory of Brownian motion. If we are lucky, and the coefficient $l(y) = \alpha h(H(y))$, $y \in \mathbb{R}$, is, say, a linear function, we can solve the equation in an explicit form. Suppose that $l(y) = ay + b$. Then the solution (see equation [9.3]) is

$$Y_t = Y_0 \mathrm{e}^{at} + \frac{b}{a}(\mathrm{e}^{at} - 1) + \int_0^t \mathrm{e}^{a(t-s)}\,\mathrm{d}B_s, \quad t \geqslant 0,$$

and therefore

$$X_t = H\left(G(X_0)\mathrm{e}^{at} + \frac{b}{a}(\mathrm{e}^{at} - 1) + \int_0^t \mathrm{e}^{a(t-s)}\,\mathrm{d}B_s\right), \quad t \geqslant 0.$$

In the particular case where $h = G$, we get:

$$X_t = H\left(G(X_0)\mathrm{e}^{\alpha t} + \int_0^t \mathrm{e}^{\alpha(t-s)}\,\mathrm{d}B_s\right), \quad t \geqslant 0.$$

To finish, we shall solve the stochastic Verhulst equation, which is important in various areas,

$$\mathrm{d}X_t = \big(\lambda X_t - X_t^2\big)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t.$$

By change of variable $X_t = 1/Y_t$ it can reduced to a linear equation. Indeed, applying Itô's formula (Theorem 7.10) to the function $F(x) := 1/x$, we get:

$$\begin{aligned}
\mathrm{d}Y_t &= -\frac{1}{X_t^2}\,\mathrm{d}X_t + \frac{1}{2}\,\frac{2}{X_t^3}\,\mathrm{d}\langle X\rangle_t \\
&= -\frac{1}{X_t^2}\big(\lambda X_t - X_t^2\big)\,\mathrm{d}t - \frac{1}{X_t^2}\sigma X_t\,\mathrm{d}B_t + \frac{1}{X_t^3}\sigma^2 X_t^2\,\mathrm{d}t \\
&= \Big(-\frac{\lambda}{X_t} + 1 + \frac{\sigma^2}{X_t}\Big)\,\mathrm{d}t - \frac{\sigma}{X_t}\,\mathrm{d}B_t \\
&= \big((-\lambda + \sigma^2)Y_t + 1\big)\,\mathrm{d}t - \sigma Y_t\,\mathrm{d}B_t.
\end{aligned}$$

The solution of the latter is (see equation [9.4])

$$Y_t = \exp\Big\{-\Big(\lambda - \frac{1}{2}\sigma^2\Big)t - \sigma B_t\Big\}\bigg(Y_0 + \int_0^t \exp\Big\{\Big(\lambda - \frac{1}{2}\sigma^2\Big)s + \sigma B_s\Big\}\,\mathrm{d}s\bigg).$$

Hence,

$$X_t = \frac{1}{Y_t} = \frac{X_0 \exp\big\{\big(\lambda - \frac{1}{2}\sigma^2\big)t + \sigma B_t\big\}}{1 + X_0 \int_0^t \exp\big\{\big(\lambda - \frac{1}{2}\sigma^2\big)s + \sigma B_s\big\}\,\mathrm{d}s}.$$

Note that the use of Itô's formula was not well justified since the function $F(x) = 1/x$ does not belong to class $C^2(\mathbb{R})$. However, the game ended happily: if $X_0 \geqslant 0$, then in the obtained expression of X, the denominator is not zero; moreover, we can directly check that the process defined by the formula satisfies the Verhulst equation. If $X_0 < 0$, the situation is more complicated since the denominator can become zero at some (random) time moment. Roughly speaking, X is a solution of the equation until the moment at which the denominator becomes zero.

Similarly, the stochastic *Ginzburg–Landau* equation

$$\mathrm{d}X_t = \big(\lambda X_t - X_t^3\big)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t$$

has the solution

$$X_t = \frac{X_0 \exp\big\{\big(\lambda - \frac{1}{2}\sigma^2\big)t + \sigma B_t\big\}}{\sqrt{1 + 2X_0^2 \int_0^t \exp\big\{2\big(\lambda - \frac{1}{2}\sigma^2\big)s + 2\sigma B_s\big\}\,\mathrm{d}s}},$$

which is obtained by the change of variables $Y_t = 1/X_t^2$.

9.4. Stochastic exponential equation

A natural extension of the LSDEs [9.1] are equations where the "driving" Brownian motion B is replaced by any Itô process Y. We restrict ourselves to a *stochastic exponential equation*, which is extremely important in stochastic analysis,

$$X_t = 1 + \int_0^t X_s\,\mathrm{d}Y_s, \quad t \geqslant 0, \qquad [9.5]$$

and its generalization

$$X_t = H_t + \int_0^t X_s\,\mathrm{d}Y_s, \quad t \geqslant 0, \qquad [9.6]$$

where Y is an Itô process, and H is a continuous adapted process. The first equation is so called since when Y is a regular random process, its solution is the exponential $X_t = \exp\{Y_t - Y_0\}$, $t \geqslant 0$.

PROPOSITION 9.8.– *If $Y_0 = 0$, then the solution of equation* [9.5] *is*

$$X_t = \exp\left\{Y_t - \frac{1}{2}\langle Y\rangle_t\right\}, \quad t \geqslant 0.$$

It is called the stochastic exponent of Y and is denoted by $\mathcal{E}(Y)$.

Proof. Denote $Z_t = Y_t - \frac{1}{2}\langle Y\rangle_t$ and $F(x) = \mathrm{e}^x$. Then $X_t = F(Z_t)$ and $F'(Z_t) = F''(Z_t) = X_t$. Therefore, by applying Itô's formula (Theorem 7.10) we get:

$$\begin{aligned} X_t &= X_0 + \int_0^t F'(Z_s)\,\mathrm{d}Z_s + \frac{1}{2}\int_0^t F''(Z_s)\,\mathrm{d}\langle Z\rangle_s \\ &= 1 + \int_0^t X_s\,\mathrm{d}Z_s + \frac{1}{2}\int_0^t X_s\,\mathrm{d}\langle Z\rangle_s \\ &= 1 + \int_0^t X_s\,\mathrm{d}Y_s - \frac{1}{2}\int_0^t X_s\,\mathrm{d}\langle Y\rangle_s + \frac{1}{2}\int_0^t X_s\,\mathrm{d}\langle Y\rangle_s \\ &= 1 + \int_0^t X_s\,\mathrm{d}Y_s. \end{aligned}$$

△

PROPOSITION 9.9.– *If Y and Z are Itô processes and $Y_0 = Z_0 = 0$, then*

$$\mathcal{E}(Y)\mathcal{E}(Z) = \mathcal{E}(Y + Z + \langle Y, Z\rangle).$$

Proof. Denote $U_t = \mathcal{E}(Y)_t$ and $V_t = \mathcal{E}(Z)_t$. Then $\mathrm{d}U_t = U_t\,\mathrm{d}Y_t$, $\mathrm{d}V_t = V_t\,\mathrm{d}Z_t$, and $\mathrm{d}\langle U, V\rangle_t = U_tV_t\,\mathrm{d}\langle Y, Z\rangle_t$. Therefore, integrating by parts (Corollary 7.11), we get:

$$\begin{aligned} U_tV_t &= U_0V_0 + \int_0^t U_s\,\mathrm{d}V_s + \int_0^t V_s\,\mathrm{d}U_s + \langle U, V\rangle_t \\ &= 1 + \int_0^t U_sV_s\,\mathrm{d}Z_s + \int_0^t V_sU_s\,\mathrm{d}Y_s + \int_0^t U_sV_s\,\mathrm{d}\langle Y, Z\rangle_s \\ &= 1 + \int_0^t U_sV_s\,\mathrm{d}\big(Y + Z + \langle Y, Z\rangle\big)_s, \end{aligned}$$

that is, $UV = \mathcal{E}(Y + Z + \langle Y, Z\rangle)$. △

COROLLARY 9.10.– *If Y is an Itô process with $Y_0 = 0$, then*

$$\mathcal{E}(Y)^{-1} = \mathcal{E}(-Y + \langle Y\rangle).$$

Proof. By Proposition 9.9, we have:

$$\begin{aligned}\mathcal{E}(Y)\mathcal{E}\big(-Y + \langle Y\rangle\big) &= \mathcal{E}\big(Y + \big(-Y + \langle Y\rangle\big) + \big\langle Y, -Y + \langle Y\rangle\big\rangle\big)\\ &= \mathcal{E}\big(\langle Y\rangle - \langle Y\rangle + \big\langle Y, \langle Y\rangle\big\rangle\big) = \mathcal{E}(0) = 1.\end{aligned}$$ △

Now consider equation [9.6]. We denote its unique solution by $\mathcal{E}_H(Y)_t$.

THEOREM 9.11.– *Let H and Y be Itô processes and $Y_0 = 0$. Then*

$$X_t := \mathcal{E}_H(Y)_t = \mathcal{E}(Y)_t\left\{H_0 + \int_0^t \mathcal{E}(Y)_s^{-1}\,\mathrm{d}\big(H - \langle H, Y\rangle\big)_s\right\}, \quad t \geqslant 0.$$

Proof. We use the variation-of-constant method, well known in the theory of differential equations, and look for a solution in the form $X_t = C_t\mathcal{E}(Y)_t$. Denoting $E_t = \mathcal{E}(Y)_t$, we integrate by parts:

$$\begin{aligned}\mathrm{d}X_t &= C_t\,\mathrm{d}E_t + E_t\,\mathrm{d}C_t + \mathrm{d}\langle C, E\rangle_t\\ &= C_tE_t\,\mathrm{d}Y_t + E_t\,\mathrm{d}C_t + E_t\,\mathrm{d}\langle C, Y\rangle_t\\ &= X_t\,\mathrm{d}Y_t + E_t\,\mathrm{d}\big(C + \langle C, Y\rangle\big)_t.\end{aligned}$$

Comparing the equation obtained with equation [9.6], we have:

$$\mathrm{d}H_t + X_t\,\mathrm{d}Y_t = X_t\,\mathrm{d}Y_t + E_t\,\mathrm{d}\big(C + \langle C, Y\rangle\big)_t$$

or

$$\mathrm{d}H_t = E_t\,\mathrm{d}\big(C + \langle C, Y\rangle\big)_t.$$

Hence,

$$E_t^{-1}\,\mathrm{d}H_t = \mathrm{d}C_t + \mathrm{d}\langle C, Y\rangle_t.$$

By taking the covariation of both sides with Y and using the fact that $\langle\langle C, Y\rangle, Y\rangle = 0$, we get:

$$E_t^{-1}\,\mathrm{d}\langle H, Y\rangle_t = \mathrm{d}\langle C, Y\rangle_t.$$

Substituting the obtained expression for $\mathrm{d}\langle C, Y\rangle_t$ into the previous equality, we get:

$$E_t^{-1}\,\mathrm{d}H_t = \mathrm{d}C_t + E_t^{-1}\,\mathrm{d}\langle H, Y\rangle_t,$$

that is,

$$\mathrm{d}C_t = E_t^{-1}\,\mathrm{d}H_t - E_t^{-1}\,\mathrm{d}\langle H, Y\rangle_t, = E_t^{-1}\,\mathrm{d}\big(H - \langle H, Y\rangle\big)_t,$$

or, in the integral form,

$$C_t = C_0 + \int_0^t E_s^{-1}\,\mathrm{d}\big(H - \langle H, Y\rangle\big)_s.$$

It remains to note that $C_0 = H_0$ and recall that $X_t = \mathcal{E}(Y)_t C_t$. △

We now generalize the proposition to the case where H is not an Itô process.

COROLLARY 9.12.– *Let H be any continuous adapted process, and let Y be an Itô process with $Y_0 = 0$. Then the solution of equation* [9.6] *is*

$$X_t = \mathcal{E}_H(Y)_t = H_t + \mathcal{E}(Y)_t \int_0^t \mathcal{E}(Y)_s^{-1} H_s\,\mathrm{d}\big(Y - \langle Y\rangle\big)_s, \quad t \geqslant 0.$$

Proof. Denote $Z_t = X_t - H_t$. Then Z satisfies the equation

$$\begin{aligned} Z_t &= \int_0^t X_s\,\mathrm{d}Y_s = \int_0^t (Z_s + H_s)\,\mathrm{d}Y_s \\ &= \int_0^t Z_s\,\mathrm{d}Y_s + \int_0^t H_s\,\mathrm{d}Y_s = K_t + \int_0^t Z_s\,\mathrm{d}Y_s. \end{aligned}$$

Since here $K = H \bullet Y$ is an Itô process, applying Theorem 9.11, we have:

$$Z_t = \mathcal{E}(Y)_t \left\{ K_0 + \int_0^t \mathcal{E}(Y)_s^{-1}\,\mathrm{d}\big(K - \langle K, Y\rangle\big)_s \right\}.$$

Since $K_0 = 0$, $\mathrm{d}K_s = H_s\,\mathrm{d}Y_s$, and $\mathrm{d}\langle K, Y\rangle_s = H_s\,\mathrm{d}\langle Y\rangle_s$, we have:

$$Z_t = \mathcal{E}(Y)_t \int_0^t \mathcal{E}(Y)_s^{-1} H_s\,\mathrm{d}\big(Y - \langle Y\rangle\big)_s.$$

It remains to substitute this expression into the equality $X_t = H_t + Z_t$. △

Using the formula of Theorem 9.11, we can prove an important theorem on comparison of solutions of SDEs.

THEOREM 9.13 (Comparison of solutions of SDEs).– *Let X and Y be solutions of the stochastic differential equations*

$$X_t = x_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s$$

and

$$Y_t = y_0 + \int_0^t c(s, Y_s)\,\mathrm{d}s + \int_0^t \sigma(s, Y_s)\,\mathrm{d}B_s.$$

Suppose that the common diffusion coefficient σ satisfies the Lipschitz condition $|\sigma(t,x) - \sigma(t,y)| \leqslant M|x-y|$, $x, y \in \mathbb{R}$, $t \geqslant 0$, and that the drift coefficients of both equations are continuous functions. Then:

(1) *If $x_0 > y_0$ and $b(t,x) > c(t,x)$ for all $t \geqslant 0$, $x \in \mathbb{R}$, then $X_t > Y_t$ for all $t \geqslant 0$ a.s.*

(2) *If $x_0 \geqslant y_0$ and $b(t,x) \geqslant c(t,x)$ for all $t \geqslant 0$, $x \in \mathbb{R}$, and if the coefficients of at least one equation satisfy the uniqueness condition, then $X_t \geqslant Y_t$ for all $t \geqslant 0$ a.s.*

REMARK.– In the deterministic case (where $\sigma \equiv 0$), this theorem is well known in the theory of ordinary differential equations. Therefore, intuitively, the theorem can easily be understood when interpreting the solutions of stochastic differential equations as the solutions of deterministic differential equations that are perturbed by noise *identically* (because of the common diffusion coefficient σ). If one of the deterministic solutions is greater than another, it is natural to expect that the same remains true for the corresponding solutions of stochastic equations.

Proof. We first assume that $x_0 > y_0$, $b(t,x) > c(t,x)$. Denote $Z_t = X_t - Y_t$ and $K_t = b(t, X_t) - c(t, Y_t)$. Then we have:

$$\begin{aligned} Z_t &= x_0 - y_0 + \int_0^t K_s\,\mathrm{d}s + \int_0^t \big(\sigma(s, X_s) - \sigma(s, Y_s)\big)\,\mathrm{d}B_s \\ &= H_t + \int_0^t (X_s - Y_s)\frac{\sigma(s, X_s) - \sigma(s, Y_s)}{X_s - Y_s}\mathbb{1}_{\{X_s \neq Y_s\}}\,\mathrm{d}B_s \\ &= H_t + \int_0^t Z_s\,\mathrm{d}N_s, \end{aligned}$$

where

$$H_t = x_0 - y_0 + \int_0^t K_s \, \mathrm{d}s,$$

$$N_t = \int_0^t \frac{\sigma(s, X_s) - \sigma(s, Y_s)}{X_s - Y_s} \mathbb{1}_{\{X_s \neq Y_s\}} \, \mathrm{d}B_s$$

(in the second integral, the integrand process is bounded in absolute value by the constant M). By Theorem 9.11,

$$Z_t = \mathcal{E}(N)_t \left\{ x_0 - y_0 + \int_0^t \mathcal{E}(N)_s^{-1} \, \mathrm{d}H_s \right\}.$$

Rewrite the process in braces as follows:

$$\begin{aligned} V_t &:= x_0 - y_0 + \int_0^t \mathcal{E}(N)_s^{-1} \, \mathrm{d}H_s \\ &= x_0 - y_0 + \int_0^t \mathcal{E}(N)_s^{-1} K_s \, \mathrm{d}s \\ &= x_0 - y_0 + \int_0^t K_s \, \mathrm{d}A_s, \end{aligned}$$

where $A_t := \int_0^t \mathcal{E}(N)_s^{-1} \, \mathrm{d}s$, $t \geqslant 0$, is a non-decreasing process. Let us show that $V_t > 0$ for all $t \geqslant 0$ (almost surely). Suppose, on the contrary, that $V_t = 0$ for some (random) $t > 0$. Take the smallest such t, i.e. $t = \min\{s \colon V_s = 0\}$ (it does exist because of the continuity of the trajectories of V). Then $Z_t = X_t - Y_t = 0$, and thus $K_t = b(t, X_t) - c(t, Y_t) = b(t, X_t) - c(t, X_t) > 0$. Because of the continuity of K, there is a (random) $\varepsilon > 0$ such that $K_s > 0$ for all $s \in [t - \varepsilon, t]$. Therefore, $0 = V_t = V_{t-\varepsilon} + \int_{t-\varepsilon}^t K_s \, \mathrm{d}A_s \geqslant V_{t-\varepsilon} > 0$, a contradiction. Thus, $V_t > 0$ for all $t \geqslant 0$. Hence, $Z_t = \mathcal{E}(N)_t V_t > 0$, that is, $X_t > Y_t$, $t \geqslant 0$.

In the case of non-strict inequalities, suppose that the drift coefficient b of the *first* equation satisfies the uniqueness-of-solution condition. Consider the equations

$$X_t^\varepsilon = x_0 + \varepsilon + \int_0^t \big(b\big(s, X_s^\varepsilon\big) + \varepsilon\big) \, \mathrm{d}s + \int_0^t \sigma\big(s, X_s^\varepsilon\big) \, \mathrm{d}B_s, \quad \varepsilon > 0.$$

By the first part of the proof,

$$X_t^{\varepsilon_1} > X_t^{\varepsilon_2} > Y_t, \quad t \geqslant 0,$$

if $\varepsilon_1 > \varepsilon_2 > 0$. From this we have two facts. First, there exists the limit $X_t^0 := \lim_{\varepsilon \downarrow 0} X_t^\varepsilon \geqslant Y_t$. Second, passing to the limit in the last equation as $\varepsilon \downarrow 0$, we get:

$$X_t^0 = x_0 + \int_0^t b\big(s, X_s^0\big)\,\mathrm{d}s + \int_0^t \sigma\big(s, X_s^0\big)\,\mathrm{d}B_s,$$

that is, X^0 is a solution of the first equation. Because of the uniqueness of the solution, $X^0 = X$. Hence, $X_t = X_t^0 \geqslant Y_t$, $t \geqslant 0$. △

REMARK.– As can be seen from the proof, without the uniqueness-of-solution assumption, we could only claim that there *exists* a solution X^0 of the first equation such that $X_t^0 \geqslant X_t$, $t \geqslant 0$.

9.5. Exercises

9.1. Let $X_t \geqslant 0$, $t \geqslant 0$, be the solution of the SDE $\mathrm{d}X_t = -X_t\,\mathrm{d}t + X_t\,\mathrm{d}B_t$ with initial condition $X_0 = x \in \mathbb{R}$. Find $\mathbf{E}X_t^3$, $t \geqslant 0$.

9.2. Write an SDE satisfied by the process $Y_t = X_t^2$, $t \geqslant 0$, where X is the solution of the SDE $\mathrm{d}X_t = a(t)X_t\,\mathrm{d}t + b(t)X_t\,\mathrm{d}B_t$, $X_0 = 1$.

9.3. Solve the SDE

$$\mathrm{d}X_t = -a^2X_t^2(1 - X_t)\,\mathrm{d}t + aX_t(1 - X_t)\,\mathrm{d}B_t, \quad X_0 = x \in (0, 1),$$

where $a > 0$ is a constant.

9.4. Solve the SDE

$$X_t = 1 + \int_0^t (2 - 3X_s)\,\mathrm{d}s + 4\int_0^t X_s\,\mathrm{d}B_s.$$

9.5. Solve the SDE

$$X_t = 1 + \int_0^t (2X_s - 3X_s^4)\,\mathrm{d}s + 5\int_0^t X_s\,\mathrm{d}B_s.$$

9.6. Solve the SDE

$$X_t = x_0 + \int_0^t \cos X_s \sin^3 X_s \,\mathrm{d}s - \int_0^t \sin^2 X_s \,\mathrm{d}B_s, \quad t \geqslant 0,$$

when:

a) $x_0 = 0$,

b) $x_0 = \pi/2$,

c) $x_0 = 3\pi/2$.

9.7. Let X be an Itô process, and let Y and Z be the solutions of the equations $\mathrm{d}Y_t = Y_t \,\mathrm{d}X_t$ and $\mathrm{d}Z_t = -Z_t \,\mathrm{d}X_t$ with $Y_0 = 2$ and $Z_0 = 1/2$. Write an SDE for the random process $W_t := Y_t Z_t$, $t \geqslant 0$.

9.8. Show that the solution of the SDE

$$\mathrm{d}X_t = (X_t + \sin^2 X_t)\,\mathrm{d}t + X_t \,\mathrm{d}B_t$$

is positive for any initial condition $X_0 = x > 0$.

Chapter 10

Solutions of SDEs as Markov Diffusion Processes

10.1. Introduction

One consequence of the independence of increments of Brownian motion B is their orthogonality to the past or, in other words, that Brownian motion is a martingale. Another consequence is the so-called Markov property. Intuitively, the Markov property of a random process is described as follows: its probabilistic behavior *after* any moment s (in the future) depends on the value of the process at the time moment s (in the present) and does not depend on the values of the process until the moment s (in the past). There are many more or less rigorous ways to define the Markov property. It is relatively simple to do for random processes having densities. A process X is called a (real) *Markov process* if its conditional density satisfies the equality

$$p_{X_t|X_{t_1},X_{t_2},\dots,X_{t_k},X_s}(y|x_1,x_2,\dots,x_k,x) = p(s,x,t,y) := p_{X_t|X_s}(y|x) \quad [10.1]$$

for all $0 \leqslant t_1 < t_2 < \cdots < t_k < s < t$, $x_1, x_2, \dots, x_k, x, y \in \mathbb{R}$. Then the function $p = p(s,x,t,y)$, $0 < s < t$, $x,y \in \mathbb{R}$, is called a *transition density* of the Markov process X. Most often, the Markov property is applied when written in terms of conditional expectations From equation [10.1] it follows that, for all bounded (measurable) functions $f\colon \mathbb{R} \to \mathbb{R}$ and for all $0 \leqslant t_1 < t_2 < \cdots < t_k < s < t$, $x_1, x_2, \dots, x_k, x \in \mathbb{R}$, we have:

$$\begin{aligned}
&\mathbf{E}\big(f(X_t)|X_{t_1} = x_1,\, X_{t_2} = x_2, \dots, X_{t_k} = x_k,\, X_s = x\big) \\
&\quad = \mathbf{E}\big(f(X_t)|X_s = x\big) = \int_{-\infty}^{+\infty} f(y)\, p(s,x,t,y)\, \mathrm{d}y
\end{aligned}$$

or, shortly (see section 1.6), for all $0 \leqslant t_1 < \cdots < t_k < s < t$,

$$\begin{aligned}&\mathbf{E}\big(f(X_t)|X_{t_1}, X_{t_2}, \ldots, X_{t_k}, X_s\big)\\ &\quad = \mathbf{E}\big(f(X_t)|X_s\big) = \int_{-\infty}^{+\infty} f(y)p(s, X_s, t, y)\,\mathrm{d}y. \qquad [10.2]\end{aligned}$$

In the general case, a random process X is called a (real) *Markov process* with *transition probability* $P(s, x, t, B)$, $0 \leqslant s < t$, $x \in \mathbb{R}$, $B \in \mathcal{B}(\mathbb{R})$, if, almost surely,

$$\mathbf{P}\{X_t \in B|X_{t_1}, X_{t_2}, \ldots, X_{t_k}, X_s\} = P(s, X_s, t, B)$$

for all $0 \leqslant s < t$ and $B \in \mathcal{B}(\mathbb{R})$. In this case, the first equality in equation [10.2] also holds. When X has a transition density $p = p(s, x, t, y)$, the transition probability $P(s, x, t, B) = \int_B p(s, x, t, y)\,\mathrm{d}y$.

Let us check that Brownian motion is a Markov process and find its transition density. We first find the conditional density p defined by equation [10.1] and "pretending" to be a transition density. Denote by

$$\varphi(t, x) = \frac{1}{\sqrt{2\pi t}}\mathrm{e}^{-x^2/2t}, \quad x \in \mathbb{R},$$

the density of normal random variable $\xi \sim N(0, t)$. Since $B_s \perp\!\!\!\perp B_t - B_s$, the joint density of random variables B_s and $B_t - B_s$ is

$$p(x, y) = \varphi(s, x)\varphi(t - s, y), \quad x, y \in \mathbb{R}.$$

From this we easily find the joint density of B_s and B_t. Their joint distribution function is

$$\begin{aligned}F(x, y) &= \mathbf{P}\{B_s < x,\, B_t < y\} = \mathbf{P}\big\{B_s < x,\, B_s + (B_t - B_s) < y\big\}\\ &= \iint\limits_{\{u<x,u+v<y\}} p(u, v)\,\mathrm{d}u\,\mathrm{d}v = \int_{-\infty}^{x}\int_{-\infty}^{y-u} p(u, v)\,\mathrm{d}v\,\mathrm{d}u, \quad x, y \in \mathbb{R}.\end{aligned}$$

Differentiating, we get:

$$p_{B_s,B_t}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = p(x, y - x) = \varphi(s, x)\varphi(t - s, y - x).$$

Thus,

$$p(s, x, t, y) := p_{B_t|B_s}(y|x) = \frac{p_{B_s,B_t}(x, y)}{p_{B_s}(x)} = \varphi(t - s, y - x).$$

Similarly, since the increments

$$\begin{aligned}&B_{t_1} \sim N(0,t_1), \quad B_{t_2} - B_{t_1} \sim N(0,t_2-t_1), \quad \dots,\\ &\qquad B_s - B_{t_k} \sim N(0,s-t_k), \quad B_t - B_s \sim N(0,t-s)\end{aligned}$$

are independent random variables, the density of $(B_{t_1}, B_{t_2}, \dots, B_{t_k}, B_s, B_t)$ equals

$$\begin{aligned}&p(t_1,t_2,\dots,t_k,s,t;x_1,x_2,\dots,x_k,x,y)\\ &\quad = \varphi(t_1,x_1)\varphi(t_2-t_1,x_2-x_1)\cdots\\ &\qquad \times \varphi(t_k-t_{k-1},x_k-x_{k-1})\varphi(s-t_k,x-x_k)\varphi(t-s,y-x).\end{aligned}$$

From this we get that the conditional density on the right-hand side of equation [10.1] (when $X = B$) equals

$$\begin{aligned}&p_{B_t|B_{t_1},B_{t_2},\dots,B_{t_k},B_s}(y|x_1,x_2,\dots,x_k,x)\\ &\quad = \frac{p(t_1,t_2,\dots,t_k,s,t;x_1,x_2,\dots,x_k,x,y)}{p(t_1,t_2,\dots,t_k,s;x_1,x_2,\dots,x_k,x)}\\ &\quad = \tfrac{\varphi(t_1,x_1)\varphi(t_2-t_1,x_2-x_1)\cdots\varphi(t_k-t_{k-1},x_k-x_{k-1})\varphi(s-t_k,x-x_k)\varphi(t-s,y-x)}{\varphi(t_1,x_1)\varphi(t_2-t_1,x_2-x_1)\cdots\varphi(t_k-t_{k-1},x_k-x_{k-1})\varphi(s-t_k,x-x_k)}\\ &\quad = \varphi(t-s,y-x) = p(s,x,t,y).\end{aligned}$$

Unfortunately, in the general case, the possibility of explicitly writing a formula for transition density is a pleasant exception rather than a rule. It is worth mentioning two directions of the Markov process theory, analytic and probabilistic (or stochastic). The starting point of analytic theory is transition probabilities or densities. For their various classes, equations (for example, partial differential equations) are derived that often need to be solved by numerical methods; the existence of the corresponding Markov process is also proved, and any properties of transition densities or probabilities obtained usually can be interpreted as some properties of the process. Simplifying, this approach may be compared with the study of random variables on the basis of their distribution functions and densities. In the probabilistic approach, Markov processes are constructed directly as solutions of stochastic differential equations. Its main advantage is that the study of a Markov process as a solution of a concrete equation is often easier than that of a Markov process with implicitly defined transition density or probability. The reader probably has already understood that, in this book, we prefer the second, probabilistic approach.

DEFINITION 10.1.– *A solution X of the stochastic differential equation*

$$X_t = x + \int_0^t b(s,X_s)\,\mathrm{d}s + \int_0^t \sigma(s,X_s)\,\mathrm{d}B_s, \quad t \geqslant 0, \qquad [10.3]$$

with the coefficients b, σ satisfying the existence and uniqueness conditions is called a diffusion process. If the coefficients b and σ do not depend on t, X is called a (time)-homogeneous diffusion process. b and σ are called the drift and diffusion coefficients of process X, respectively. In order to emphasize the dependence of X on the initial point x, we denote it X^x, and by a diffusion process X we mean the whole family of solutions $\{X^x,\, x \in \mathbb{R}\}$. We also denote by $X^{s,x}$ the process in the time interval $[s,\infty)$ $(s \geqslant 0)$ that satisfies the equation

$$X_t^{s,x} = x + \int_s^t b\big(u, X_u^{s,x}\big)\,\mathrm{d}u + \int_s^t \sigma\big(u, X_u^{s,x}\big)\,\mathrm{d}B_u, \quad t \geqslant s. \qquad [10.4]$$

REMARK.– For every pair (x,s), the solution $X_t^{s,x}$, $t \geqslant 0$, is continuous almost surely. It is known that $X_t^{s,x}$ can be "corrected" so that it becomes continuous with respect to all variables (s,x,t).

PROPOSITION 10.2.– *For all $t \geqslant s$, we have:*

$$X_t^x = X_t^{s,X_s^x} \quad \textit{a.s.} \qquad [10.5]$$

REMARK.– This is, in fact, another formulation of the Markov property: at any time moment s, the process X_t^x starts anew from the value X_s^x at this moment (independently from the values of the process before the moment s). Later on, in Proposition 10.4, we shall show that this implies the Markov property [10.4] in terms of conditional expectations.

Proof. Substituting the random variable $X_s^x \in \mathcal{H}_s$ into the equation

$$X_t^{s,y} = y + \int_s^t b\big(u, X_u^{s,y}\big)\,\mathrm{d}u + \int_s^t \sigma\big(u, X_u^{s,y}\big)\,\mathrm{d}B_u, \quad t \geqslant s,\ y \in \mathbb{R},$$

we have:

$$X_t^{s,X_s^x} = X_s^x + \int_s^t b\big(u, X_u^{s,X_s^x}\big)\,\mathrm{d}u + \int_s^t \sigma\big(u, X_u^{s,X_s^x}\big)\,\mathrm{d}B_u, \quad t \geqslant s.$$

(Intuitively, such a substitution is rather clear, since the increments of Brownian motion in the interval $[s,t]$ are independent of $X_s^x \in \mathcal{H}_s$. A formal proof of this fact is similar to the proof of property $2'$ in Theorem 4.8.) On the other hand,

$$X_t^x = x + \int_0^t b\big(u, X_u^x\big)\,\mathrm{d}u + \int_0^t \sigma\big(u, X_u^x\big)\,\mathrm{d}B_u$$
$$= X_s^x + \int_s^t b\big(u, X_u^x\big)\,\mathrm{d}u + \int_s^t \sigma\big(u, X_u^x\big)\,\mathrm{d}B_u, \quad t \geqslant s.$$

By comparing the expressions obtained we see that, for $t \geqslant s$, the random processes X_t^x and X_t^{s,X_s^x} are solutions of the same SDE. Because of the uniqueness of a solution, they coincide in the interval $[s,\infty)$. △

DEFINITION 10.3.– *On the set of all bounded (measurable) functions* $f\colon \mathbb{R} \to \mathbb{R}$*, we define the family of linear operators* T_{st}*,* $0 \leqslant s \leqslant t$*, associated with the diffusion process* $\{X^x\}$ *defined by equation* [10.3]*:*

$$T_{st}f(x) := \mathbf{E}f\big(X_t^{s,x}\big), \quad x \in \mathbb{R},$$

where $X^{s,x}$ *is the solution of equation* [10.4]*.*

REMARK.– The boundedness is the simplest condition assuring the existence of expectations in the definition of T_{st}. For diffusion processes with Lipschitz coefficients, in fact, it suffices that f be a function with polynomial growth: $|f(x)| \leqslant C(1+|x|^k)$, $x \in \mathbb{R}$.

PROPOSITION 10.4 (Markov property of a diffusion process).– *For every* $x \in \mathbb{R}$*,* X^x *is a Markov process, and for all bounded (measurable) functions* $f\colon \mathbb{R} \to \mathbb{R}$*,*

$$\mathbf{E}\big(f\big(X_t^x\big)|X_{t_1}^x, X_{t_2}^x, \ldots, X_{t_k}^x, X_s^x\big) = \mathbf{E}\big(f\big(X_t^x\big)|X_s^x\big) = T_{st}f\big(X_s^x\big)$$

for all $0 \leqslant t_1 < \cdots < t_k < s < t$*. Moreover,*

$$T_{st}T_{tu} = T_{su}, \quad 0 \leqslant s \leqslant t \leqslant u,$$

i.e. $(T_{st}T_{tu})f := T_{st}(T_{tu}f) = T_{su}f$ *for all bounded* $($*measurable*$)$ *functions* $f\colon \mathbb{R} \to \mathbb{R}$*.*

Proof. Let us use the equality proved in Proposition 10.2,

$$X_t^x = X_t^{s,X_s^x} = X_t^{s,y}\Big|_{y=X_s^x}.$$

For every $y \in \mathbb{R}$, the random variable $X_t^{s,y}$ only depends on the increments of Brownian motion $B_u - B_v$, $u \geqslant v \geqslant s$, after time moment s. Therefore, we may substitute the random variable X_s^x, independent of all these increments, into the equality

$$\mathbf{E}f\big(X_t^{s,y}\big) = T_{st}f(y)$$

instead of y and obtain the equality for the conditional expectation:[1]

$$\mathbf{E}\big(f\big(X_t^x\big)|X_s^x\big) = \mathbf{E}\Big(f\big(X_t^{s,X_s^x}\big)|X_s^x\Big) = T_{st}f\big(X_s^x\big).$$

Similarly,

$$\mathbf{E}\big(f\big(X_t^{s,x}\big)|X_u^{s,x}\big) = T_{ut}f\big(X_u^{s,x}\big), \quad s \leqslant u \leqslant t.$$

Now take the expectations of both sides and apply the iteration rule:

$$\begin{aligned} T_{st}f(x) &= \mathbf{E}f\big(X_t^{s,x}\big) = \mathbf{E}\Big(\mathbf{E}\big(f\big(X_t^{s,x}\big)\big|X_u^{s,x}\big)\Big) \\ &= \mathbf{E}\big[(T_{ut}f)\big(X_u^{s,x}\big)\big] = T_{su}(T_{ut}f)(x) = (T_{su}T_{ut})f(x), \quad s \leqslant u \leqslant t. \end{aligned}$$

Since the random variables $X_{t_1}^x, X_{t_2}^x, \dots, X_{t_k}^x$ are independent of $B_u - B_v$, $u \geqslant v \geqslant s$, we obtain, as before:

$$\begin{aligned} \mathbf{E}\Big(f\big(X_t^x\big)\big|X_{t_1}^x, X_{t_2}^x, \dots, X_{t_k}^x, X_s^x\Big) &= \mathbf{E}\Big(f\big(X_t^{s,X_s^x}\big)\big|X_{t_1}^x, X_{t_2}^x, \dots, X_{t_k}^x, X_s^x\Big) \\ &= \mathbf{E}\Big(f\big(X_t^{s,X_s^x}\big)\big|X_s^x\Big) = T_{st}f\big(X_s^x\big). \end{aligned}$$

△

For simplicity, we further consider *time-homogeneous* diffusion processes $X = \{X^x\}$, i.e. we assume that the coefficients of equation [10.3] do not depend on time: $b = b(x)$, $\sigma = \sigma(x)$, $x \in \mathbb{R}$. Then the operators T_{st} (see Definition 10.3) depend only on the difference $t - s$. Therefore, it suffices to consider the operators $T_t := T_{0,t}$ defined by

$$T_tf(x) = \mathbf{E}f\big(X_t^x\big), \quad t \geqslant 0, \ x \in \mathbb{R}.$$

The property of operators proved in Proposition 10.4 becomes

$$T_sT_t = T_{s+t}, \quad t, s \geqslant 0.$$

Because of this property, the operator family $\{T_t, \, t \geqslant 0\}$ is called the operator semigroup corresponding to diffusion process X (the product of two operators from this family again is an operator from this family).

The transition density $p(s,x,t,y)$ of a homogeneous diffusion process $\{X^x\}$ also depends only on the difference $t-s$, and, therefore, it suffices to consider the function $p(t,x,y) := p(0,x,t,y) = p(s,x,s+t,y)$. The latter also is called the transition density of $\{X^x\}$.

1. This step is rather obvious intuitively but rather difficult to justify rigorously.

10.2. Backward and forward Kolmogorov equations

DEFINITION 10.5.– *The* generator (*or the infinitesimal operator*) *of a homogeneous diffusion process* $X = \{X^x\}$ *is the operator* A *defined on a set of functions* $f: \mathbb{R} \to \mathbb{R}$ *by*

$$Af(x) = \lim_{t \downarrow 0} \frac{T_t f(x) - f(x)}{t} = \lim_{t \downarrow 0} \frac{\mathbf{E} f(X_t^x) - f(x)}{t}, \quad x \in \mathbb{R}.$$

The domain of the definition of A *is the set* $\mathcal{D}_A$ *of all functions* $f: \mathbb{R} \to \mathbb{R}$ *for which such a* (*finite*) *limit exists for all* $x \in \mathbb{R}$.

PROPOSITION 10.6.– *Let* $X = \{X^x\}$ *be the homogeneous diffusion process defined by the SDE*

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t.$$

If a function $f \in C^2(\mathbb{R})$ *has bounded first and second derivatives, then* $f \in \mathcal{D}_A$, *and*

$$Af(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x), \quad x \in \mathbb{R},$$

or, in short, $Af = bf' + \frac{1}{2}\sigma^2 f''$. *Moreover, for every such function,*

$$\mathbf{E} f\big(X_t^x\big) = f(x) + \mathbf{E} \int_0^t Af\big(X_s^x\big)\,\mathrm{d}s, \quad t \geqslant 0,$$

or

$$T_t f(x) = f(x) + \int_0^t T_s A f(x)\,\mathrm{d}s, \quad t \geqslant 0$$

(*Dynkin*[2] *formula*).

---

2. Eugene B. Dynkin.

Proof. Denote $g(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x)$. By Itô's formula, we have:

$$\begin{aligned} f\big(X_t^x\big) &= f\big(X_0^x\big) + \int_0^t f'\big(X_s^x\big)\,\mathrm{d}X_s^x + \frac{1}{2}\int_0^t f''\big(X_s^x\big)\,\mathrm{d}\big\langle X^x\big\rangle_s \\ &= f(x) + \int_0^t \Big(b\big(X_s^x\big)f'\big(X_s^x\big) + \frac{1}{2}\sigma^2\big(X_s^x\big)f''\big(X_s^x\big)\Big)\,\mathrm{d}s \\ &\quad + \int_0^t \sigma\big(X_s^x\big)f'\big(X_s^x\big)\,\mathrm{d}B_s \\ &= f(x) + \int_0^t g\big(X_s^x\big)\,\mathrm{d}s + \int_0^t \sigma f'\big(X_s^x\big)\,\mathrm{d}B_s, \quad t \geqslant 0. \end{aligned}$$

Taking expectations gives

$$\mathbf{E}f\big(X_t^x\big) = f(x) + \mathbf{E}\int_0^t g\big(X_s^x\big)\,\mathrm{d}s = f(x) + \int_0^t \mathbf{E}g\big(X_s^x\big)\,\mathrm{d}s, \quad t \geqslant 0.$$

From this we get that there exists

$$Af(x) = \lim_{t\downarrow 0}\frac{\mathbf{E}f(X_t^x) - f(x)}{t} = \lim_{t\downarrow 0}\frac{\int_0^t \mathbf{E}g(X_s^x)\,\mathrm{d}s}{t} = \mathbf{E}g\big(X_0^x\big) = g(x).$$

Substituting $g(x) = Af(x)$ into the previous equality, we get the Dynkin formula. △

REMARKS.– 1. The generator of the diffusion process $\{X^x\}$ defined by the *Stratonovich* equation

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\circ\mathrm{d}B_t$$

is

$$Af = \Big(b + \frac{1}{2}\sigma\sigma'\Big)f' + \frac{1}{2}\sigma^2 f'',$$

which follows from the expression of the Stratonovich equation in the Itô form (Proposition 8.6),

$$\mathrm{d}X_t = \Big(b + \frac{1}{2}\sigma\sigma'\Big)(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t.$$

However, it is often convenient to use a more symmetric form for the generator of the solution of a Stratonovich equation,

$$Af = bf' + \frac{1}{2}\sigma(\sigma f')'$$

that can be directly checked.

2. Dynkin's formula is not only true for the fixed time moments t but also for Markov moments τ (that, in general, are random). Recall an important class of such moments, hitting moments of sets $B \subset \mathbb{R}$: $\tau = \inf\{t : X_t \in B\}$. If the expectation of such a moment is finite, that is, $\mathbf{E}\tau < +\infty$, then

$$\mathbf{E}f\big(X_\tau^x\big) = f(x) + \mathbf{E}\int_0^\tau Af\big(X_s^x\big)\,\mathrm{d}s.$$

3. From the proof of the proposition it follows that the process

$$M_f^x(t) := f\big(X_t^x\big) - \int_0^t Af\big(X_s^x\big)\,\mathrm{d}s, \quad t \geqslant 0,$$

is a martingale for all functions f as in the proposition and all $x \in \mathbb{R}$. In this case, the process X is called a solution of the *martingale problem* corresponding to the operator A. Construction of Markov processes in terms of the martingale problem currently is the most general way to construct Markov processes.

EXAMPLES 10.7.– 1. The generator of a Brownian motion ($X_t^x = x + B_t$) is

$$Af(x) = \frac{1}{2}f''(x).$$

2. The generator of the Ornstein–Uhlenbeck process (solution of the Langevin equation $\mathrm{d}X_t = -aX_t\,\mathrm{d}t + b\,\mathrm{d}B_t$) is

$$Af(x) = -axf'(x) + \frac{1}{2}b^2 f''(x).$$

3. The generator of the geometric Brownian motion, the solution of the growth equation $\mathrm{d}X_t = \mu X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t$, is

$$Af(x) = \mu x f'(x) + \frac{1}{2}\sigma^2 x^2 f''(x).$$

COROLLARY 10.8 (Backward Kolmogorov equation).– *Suppose that $X = \{X^x\}$ is a homogeneous diffusion process with generator A. For $f \in C_b^2(\mathbb{R})$, denote*

$$u(t,x) = T_t f(x) = \mathbf{E} f\big(X_t^x\big), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}.$$

Then the function u is a solution of the partial differential equation with initial condition

$$\begin{cases} \dfrac{\partial u}{\partial t} = Au, \\ u(0,x) = f(x), \end{cases} \qquad [10.6]$$

where the operator A is taken with respect to the argument x of a function $u = u(t,x)$.

Proof. By Dynkin's formula, we get that there exists the derivative

$$\frac{\partial u(t,x)}{\partial t} = \frac{\partial T_t f(x)}{\partial t} = T_t A f(x), \quad t \geqslant 0,\ x \in \mathbb{R}.$$

On the other hand,

$$\begin{aligned} \frac{\partial u(t,x)}{\partial t} &= \lim_{s \downarrow 0} \frac{T_{t+s} f(x) - T_t f(x)}{s} \\ &= \lim_{s \downarrow 0} \frac{T_s T_t f(x) - T_t f(x)}{s} = A T_t f(x) = Au(t,x). \end{aligned}$$

△

REMARKS.– 1. From the proof we see that, for $f \in C_b^2(\mathbb{R})$,

$$\frac{\partial}{\partial t} T_t f(x) = A T_t f(x) = T_t A f(x).$$

Therefore, an operator semi-group formally can be interpreted as the exponent of its generator A: $T_t = \mathrm{e}^{tA}$. It is proved that, for $t > 0$, the first equality holds for all $f \in C_b(\mathbb{R})$ and even for $f \in C(\mathbb{R})$ of polynomial growth ($|f(x)| \leqslant C(1 + |x|^k)$, $x \in \mathbb{R}$).

2. If a diffusion process $\{X^x\}$ has a transition density $p = p(t,x,y)$, then

$$T_t f(x) = \int_{\mathbb{R}} p(t,x,y) f(y)\,\mathrm{d}y, \quad t > 0.$$

Therefore the backward Kolmogorov equation for the function $u(t,x) := T_t f(x)$ can be written as

$$\begin{aligned} \frac{\partial}{\partial t} \int_{\mathbb{R}} p(t,x,y) f(y)\,\mathrm{d}y &= b(x) \frac{\partial}{\partial x} \int_{\mathbb{R}} p(t,x,y) f(y)\,\mathrm{d}y \\ &\quad + \frac{1}{2} \sigma^2(x) \frac{\partial^2}{\partial x^2} \int_{\mathbb{R}} p(t,x,y) f(y)\,\mathrm{d}y. \end{aligned}$$

Differentiating formally under the integral sign, we get

$$\int_{\mathbb{R}} \frac{\partial p(t,x,y)}{\partial t} f(y)\,\mathrm{d}y = b(x)\int_{\mathbb{R}} \frac{\partial p(t,x,y)}{\partial x} f(y)\,\mathrm{d}y + \frac{1}{2}\sigma^2(x)\int_{\mathbb{R}} \frac{\partial^2 p(t,x,y)}{\partial x^2} f(y)\,\mathrm{d}y,$$

$$\int_{\mathbb{R}} \Big[\frac{\partial p(t,x,y)}{\partial t} - b(x)\frac{\partial p(t,x,y)}{\partial x} - \frac{1}{2}\sigma^2(x)\frac{\partial^2 p(t,x,y)}{\partial x^2}\Big] f(y)\,\mathrm{d}y = 0,$$

or

$$\int_{\mathbb{R}} \Big[\frac{\partial p(t,x,y)}{\partial t} - A_x p(t,x,y)\Big] f(y)\,\mathrm{d}y = 0,$$

where the subscript x at the operator A shows that the latter is applied with respect to the argument x. Since the equality holds for all $f \in C_0^2(\mathbb{R})$, the expression in the square brackets is identically zero. Thus,

$$\frac{\partial p(t,x,y)}{\partial t} = A_x p(t,x,y), \quad t > 0,\ x \in \mathbb{R}. \qquad [10.7]$$

To justify the just made differentiation under the integral sign, let us require appropriate conditions on the transition density:

COROLLARY 10.9 (Backward Kolmogorov equation for transition density).– *Suppose that the transition density $p = p(t,x,y)$, $t > 0$, $x, y \in \mathbb{R}$, of a diffusion process with generator A is continuous in $(0,+\infty) \times \mathbb{R}^2$ together with partial derivatives $\partial p/\partial t$, $\partial p/\partial x$, and $\partial^2 p/\partial x^2$. Then p, in this domain, satisfies the backward Kolmogorov equation* [10.7].

EXAMPLE 10.10.– If the transition density of a process X is explicitly known, Corollary 10.9 allows us to explicitly write the solution of the corresponding partial differential equation. For example, in the case of Brownian motion, $Af(x) = \frac{1}{2}f''(x)$, and equation [10.6] coincides with the heat equation

$$\begin{cases} \frac{\partial u}{\partial t} = \frac{1}{2}\frac{\partial^2 u}{\partial x^2}, \\ u(0,x) = f(x). \end{cases}$$

Since $X_t^x = x + B_t \sim N(x,t)$, the transition density is

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}}\mathrm{e}^{-\frac{(x-y)^2}{2t}},$$

and thus the solution of the heat equation can be written as

$$u(t,x) = \frac{1}{\sqrt{2\pi t}} \int_{\mathbb{R}} f(y) \mathrm{e}^{-\frac{(x-y)^2}{2t}} \,\mathrm{d}y.$$

Corollary 10.8 has an important generalization:

THEOREM 10.11 (Feynman–Kac[3] formula).– *If $f \in C_b^2(\mathbb{R})$ and $g \in C_b(\mathbb{R})$, then the function*

$$v(t,x) = \mathbf{E}\left[\exp\left\{\int_0^t g\big(X_s^x\big)\,\mathrm{d}s\right\} f(X_t^x)\right], \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R},$$

is a solution of the partial differential equation

$$\frac{\partial v}{\partial t} = Av + gv$$

satisfying the initial condition $v(0,x) = f(x)$.

Proof. We first show that v is differentiable with respect to t. Denote $Y_t^x := f(X_t^x)$ and $Z_t^x := \exp\{\int_0^t g(X_s^x)\,\mathrm{d}s\}$. Then, as in the proof of Proposition 10.6, we have:

$$Y_t^x = f(x) + \int_0^t Af\big(X_s^x\big)\,\mathrm{d}s + \int_0^t \sigma f'(X_s^x)\,\mathrm{d}B_s, \quad t \geqslant 0.$$

By the main formula of integration, we get:

$$Z_t^x = Z_0^x + \int_0^t \exp\left\{\int_0^s g\big(X_u^x\big)\,\mathrm{d}u\right\} g\big(X_s^x\big)\,\mathrm{d}s = 1 + \int_0^t Z_s^x g\big(X_s^x\big)\,\mathrm{d}s, \quad t \geqslant 0.$$

3. Richard Feynman and Marc Kac.

Since Z^x is a regular process, $\langle Y^x, Z^x\rangle_t \equiv 0$. Therefore, by the integration-by-parts formula (Corollary 7.11) we get:

$$
\begin{aligned}
Y_t^x Z_t^x &= Y_0^x Z_0^x + \int_0^t Y_s^x \,\mathrm{d}Z_s^x + \int_0^t Z_s^x \,\mathrm{d}Y_s^x \\
&= f(x) + \int_0^t \big(Y_s^x Z_s^x g\big(X_s^x\big) + Z_s^x Af\big(X_s^x\big)\big)\,\mathrm{d}s \\
&\quad + \int_0^t Z_s^x (\sigma f')\big(X_s^x\big)\,\mathrm{d}B_s, \quad t \geqslant 0.
\end{aligned}
$$

Since the integrand processes are bounded and continuous, by taking expectations we get:

$$
\mathbf{E}\big(Y_t^x Z_t^x\big) = f(x) + \int_0^t \mathbf{E}\big(Y_s^x Z_s^x g\big(X_s^x\big) + Z_s^x Af\big(X_s^x\big)\big)\,\mathrm{d}s, \quad t \geqslant 0,
$$

and, thus, there exists

$$
\frac{\partial v(t,x)}{\partial t} = \mathbf{E}\big(Y_t^x Z_t^x g\big(X_t^x\big) + Z_t^x Af\big(X_t^x\big)\big).
$$

Now consider $Av(t,x) = \lim_{r\downarrow 0} r^{-1}(\mathbf{E}v(t,X_r^x) - v(t,x))$. Since the process is homogeneous, we have $X_s^y \overset{d}{=} X_{r+s}^{r,y}$. Applying the Markov property, we get:

$$
\begin{aligned}
v\big(t, X_r^x\big) &= \mathbf{E}\Bigg(\exp\bigg\{\int_0^t g\big(X_s^y\big)\,\mathrm{d}s\bigg\} f\big(X_t^y\big)\Bigg)\Bigg|_{y=X_r^x} \\
&= \mathbf{E}\Bigg(\exp\bigg\{\int_0^t g\big(X_{r+s}^{r,y}\big)\,\mathrm{d}s\bigg\} f\big(X_{r+t}^{r,y}\big)\Bigg)\Bigg|_{y=X_r^x} \\
&= \mathbf{E}\Bigg(\exp\bigg\{\int_0^t g\big(X_{r+s}^{r,X_r^x}\big)\,\mathrm{d}s\bigg\} f\big(X_{r+t}^{r,X_r^x}\big)\big|X_r^x\Bigg) \\
&= \mathbf{E}\Bigg(\exp\bigg\{\int_0^t g\big(X_{r+s}^x\big)\,\mathrm{d}s\bigg\} f\big(X_{r+t}^x\big)\big|X_r^x\Bigg).
\end{aligned}
$$

By the iteration rule for expectations,

$$\begin{aligned}
\mathbf{E}v\big(t, X_r^x\big) &= \mathbf{E}\left(\exp\left\{ \int_0^t g\big(X_{r+s}^x\big)\,\mathrm{d}s \right\} f\big(X_{r+t}^x\big) \right) \\
&= \mathbf{E}\left(\exp\left\{ \int_r^{r+t} g\big(X_s^x\big)\,\mathrm{d}s \right\} f\big(X_{r+t}^x\big) \right) \\
&= \mathbf{E}\left(Z_{r+t}^x f\big(X_{r+t}^x\big) \exp\left\{ -\int_0^r g\big(X_s^x\big)\,\mathrm{d}s \right\} \right),
\end{aligned}$$

and hence,

$$\begin{aligned}
Av(t,x) &= \lim_{r\downarrow 0} \frac{\mathbf{E}v(t, X_r^x) - v(t,x)}{r} \\
&= \lim_{r\downarrow 0} \frac{1}{r} \left[\mathbf{E}\left(Z_{r+t}^x f\big(X_{r+t}^x\big) \exp\left\{ -\int_0^r g\big(X_s^x\big)\,\mathrm{d}s \right\} \right) \right. \\
&\qquad\qquad \left. - \mathbf{E}\big(Z_t^x f\big(X_t^x\big)\big) \right] \\
&= \lim_{r\downarrow 0} \frac{1}{r} \Big[\mathbf{E}\big(Z_{r+t}^x f\big(X_{r+t}^x\big)\big) - \mathbf{E}\big(Z_t^x f\big(X_t^x\big)\big) \Big] \\
&\quad + \lim_{r\downarrow 0} \frac{1}{r} \mathbf{E}\left[Z_{r+t}^x f\big(X_{r+t}^x\big) \left(\exp\left\{ -\int_0^r g\big(X_s^x\big)\,\mathrm{d}s \right\} - 1 \right) \right] \\
&= \lim_{r\downarrow 0} \frac{1}{r} \big[v(t+r, x) - v(t,x) \big] \\
&\quad + \lim_{r\downarrow 0} \mathbf{E}\left[Z_{r+t}^x f\big(X_{r+t}^x\big) \frac{\exp\left\{ -\int_0^r g\big(X_s^x\big)\,\mathrm{d}s \right\} - 1}{r} \right] \\
&= \frac{\partial v(t,x)}{\partial t} + \mathbf{E}\Big(Z_t^x f\big(X_t^x\big) \big(-g\big(X_0^x\big)\big) \Big) = \frac{\partial v(t,x)}{\partial t} - v(t,x)g(x). \quad \triangle
\end{aligned}$$

We can say that the backward Kolmogorov equation for a transition density (Corollary 10.9) is the equality $\partial T_t f/\partial t = AT_t f$ of semi-group theory written in terms of the theory of differential equations. We can similarly write the equality $\partial T_t f/\partial t = T_t A f$. Let A^* be the linear operator which is formally adjoint to the generator A of a diffusion process $\{X^x\}$ and is defined by

$$A^* f(x) := -\big(b(x)f(x)\big)' + \frac{1}{2}\big(\sigma^2(x)f(x)\big)'',$$

where coefficients b and σ and function f are supposed to be sufficiently smooth, in order that the expression on the right-hand side is well defined. Let us check the following relation between the operators A^* and A:

$$\int_{-\infty}^{\infty} Af(x)g(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)A^*g(x)\,\mathrm{d}x, \tag{10.8}$$

which holds if, for example, $b, \sigma, g, f \in C^2(\mathbb{R})$ and at least one of the functions f, g has a compact support, i.e. belongs to class $C_0^2(\mathbb{R})$.

REMARK.– The last equation explains why A^* is called an operator *formally* adjoint to A. Two operators A and A^* in a linear space E with scalar product $(\cdot,\cdot)$ are called adjoint if for all $x, y \in E$, we have $(Ax, y) = (x, A^*y)$. If this is true not for all $x, y \in E$ but only for x and y from a "rather rich" subspace of E, then A^* and A are called formally adjoint. So, equation [10.8] can be interpreted as the equation $(Af, g) = (f, A^*g)$ in the Hilbert space $H = L^2(\mathbb{R})$ with scalar product $(f, g) := \int_{\mathbb{R}} f(x)g(x)\,\mathrm{d}x$.

Integrating by parts, we have

$$\begin{aligned}\int_{-\infty}^{\infty} b(x)f'(x)g(x)\,\mathrm{d}x &= \int_{-\infty}^{\infty} b(x)g(x)\,\mathrm{d}f(x)\\ &= b(x)g(x)f(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)\big(b(x)g(x)\big)'\,\mathrm{d}x\\ &= -\int_{-\infty}^{\infty} \big(b(x)g(x)\big)' f(x)\,\mathrm{d}x,\end{aligned}$$

where we used the fact that either f or g (and, thus, the product bgf) vanishes outside a finite interval and, thus,

$$b(x)f(x)g(x)\big|_{-\infty}^{\infty} = \lim_{R\to\infty}\big(b(x)f(x)g(x)\big|_{x=R} - b(x)f(x)g(x)\big|_{x=-R}\big) = 0.$$

Similarly, integrating by parts twice, we get

$$\frac{1}{2}\int_{-\infty}^{\infty} \sigma^2(x)f''(x)g(x)\,\mathrm{d}x = \frac{1}{2}\int_{-\infty}^{\infty} \big(\sigma^2(x)g(x)\big)'' f(x)\,\mathrm{d}x.$$

Summing the equalities obtained, we get equation [10.8].

PROPOSITION 10.12 (Forward Kolmogorov equation for transition density).– *Suppose that a diffusion process $\{X^x\}$ with coefficients $b, \sigma \in C^2(\mathbb{R})$ has a transition density $p = p(t, x, y)$, $t > 0$, $x, y \in \mathbb{R}$, that is continuous in $(0, +\infty) \times \mathbb{R}^2$ together with partial derivatives $\partial p/\partial t$, $\partial p/\partial y$, and $\partial^2 p/\partial y^2$. Then p, in this domain, satisfies the equation*

$$\frac{\partial p(t, x, y)}{\partial t} = A_y^* p(t, x, y), \qquad [10.9]$$

which is called the forward Kolmogorov, or Fokker–Planck,[4] *equation.*

Proof. The equality $\partial T_t f/\partial t = T_t A f$ or $\partial \mathbf{E} f(X_t^x)/\partial t = \mathbf{E} A f(X_t^x)$ can be rewritten, in terms of the transition density, as

$$\frac{\partial}{\partial t} \int_{-\infty}^{\infty} p(t, x, y) f(y)\, \mathrm{d}y = \int_{-\infty}^{\infty} p(t, x, y) A f(y)\, \mathrm{d}y.$$

The assumptions of the proposition allow the differentiation, under the integral signs, with respect to t on the left-hand side and to apply equation [10.8] on the right-hand side. Therefore,

$$\int_{-\infty}^{\infty} \frac{\partial p(t, x, y)}{\partial t} f(y)\, \mathrm{d}y = \int_{-\infty}^{\infty} A_y^* p(t, x, y) f(y)\, \mathrm{d}y,$$

or

$$\int_{-\infty}^{\infty} \Big[\frac{\partial p(t, x, y)}{\partial t} - A_y^* p(t, x, y)\Big] f(y)\, \mathrm{d}y = 0$$

for all $f \in C_0^2(\mathbb{R})$. Thus, the function in the brackets is identically zero. △

REMARK.– Why is equation [10.7] *backward* and [10.9] *forward*? This is simpler to explain in the case of inhomogeneous diffusion process. For its density $p(s, x, t, y)$, instead of equations [10.7] and [10.9], we should have:

$$-\frac{\partial p(s, x, t, y)}{\partial s} = A_x(s) p(s, x, t, y) \qquad [10.10]$$

and

$$\frac{\partial p(s, x, t, y)}{\partial t} = A_y^*(t) p(s, x, t, y), \qquad [10.11]$$

4. Adriaan Fokker and Max Planck.

respectively. Here, $A(t)$ and $A^*(t)$ are the analogs of the operators A and A^*, depending on the time parameter t (the subscript index x or y shows the argument with respect to which it is applied):

$$A(t)f(x) := b(t,x)f'(x) + \frac{1}{2}\sigma^2(t,x)f''(x), \quad x \in \mathbb{R},$$

$$A^*(t)f(x) := -\big(b(t,x)f(x)\big)'_x + \frac{1}{2}\big(\sigma^2(t,x)f(x)\big)''_{xx}, \quad x \in \mathbb{R}.$$

The first equation is interpreted as "backward" since it is related with the differentiation with respect to the starting point s of the time interval $[s,t]$, i.e. is "directed backward", while the second equation is directed "forward", since it is related to the differentiation with respect to the end t of the time interval $[s,t]$.

Note that the density $p(s,x,t,y)$ not only satisfies equation [10.10] for $(s,x) \in [0,t) \times \mathbb{R}$ and any fixed (t,y), but also satisfies the boundary (more precisely, the end) condition

$$\lim_{s\uparrow t} \int_{\mathbb{R}} p(s,x,t,y)f(y)\,\mathrm{d}y = f(x)$$

for all $f \in C_0(\mathbb{R})$ (or, in terms of generalized functions, $p(s,x,t,y) \to \delta(y-x)$ as $s \uparrow t$, where δ is the Dirac delta function). In terms of the theory of differential equations, this means that the transition density $p(s,x,t,y)$, $s < t$, $x,y \in \mathbb{R}$, is the *fundamental solution* of the equation

$$\frac{\partial u(s,x)}{\partial s} + A(s)u(s,x) = 0.$$

It is so called because the solution of the Cauchy problem with end condition

$$u(t-,x) := \lim_{s\uparrow t} u(s,x) = f(x)$$

can be expressed as

$$u(s,x) = \int_{\mathbb{R}} p(s,x,t,y)f(y)\,\mathrm{d}y.$$

Similarly, the transition density is also a fundamental solution of the equation $\frac{\partial v(t,y)}{\partial t} = A^*(t)v(t,y)$, that is, it satisfies equation [10.11] for all $(t,y) \in (s,+\infty)\times\mathbb{R}$ and any fixed (s,x), together with the initial condition $\int_{\mathbb{R}} p(s,x,t,y)f(y)\,\mathrm{d}y \to f(x)$ as $t \downarrow s$ for all $f \in C_0(\mathbb{R})$. The fact that the same function is a fundamental solution of two adjoint equations with respect to two different arguments (x and y) is well known in the theory of partial differential equations.

10.3. Stationary density of a diffusion process

DEFINITION 10.13.– *A stationary density* of a diffusion process X with transition density $p(t,x,y)$, $t>0$, $x,y\in\mathbb{R}$, is a density function $p_0(y)$, $y\in\mathbb{R}$, satisfying the equation

$$p_0(y) = \int_{\mathbb{R}} p(t,x,y)p_0(x)\,\mathrm{d}x, \quad t>0,\ x\in\mathbb{R}.$$

This relation can be interpreted as follows: if the initial value X_0 of a diffusion process X is a random variable with density p_0, then the values X_t of the process have the same distribution with density p_0 for all time moments $t\geqslant 0$. Moreover, the finite-dimensional distributions of X, i.e. those of all collections $(X_{t_1+t}, X_{t_2+t}, \ldots, X_{t_k+t})$ are the same for all $t\geqslant 0$. Such processes are called *stationary*. Thus, a diffusion process X with initial value X_0 distributed with the stationary density of X is a stationary process. In applications, stationary processes often play an important role, since many real-world processes are modeled by this kind of processes. Of course, in applications, we can hardly expect that the model process starts its history with stationary distribution. However, it appears that, under rather general conditions (as a rule, satisfied in practice) the stationary density of a diffusion process (provided that it does exist) has a certain attraction property: the distribution of a diffusion process X_t^x starting from any initial point x eventually stabilizes in the sense that the density of X_t^x becomes close to the stationary one. More precisely,

$$p(t,x,y)\to p_0(y), \quad t\to\infty,\ x,y\in\mathbb{R}.$$

Moreover, diffusion processes having a stationary density p_0 also possess the so-called *ergodicity* property: for every bounded (measurable) function $f\colon\mathbb{R}\to\mathbb{R}$,

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f\left(X_t^x\right)\mathrm{d}t = \int_{\mathbb{R}} f(y)p_0(y)\,\mathrm{d}y.$$

In particular,

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T 1\!\!1_{\{X_t^x\in A\}}\,\mathrm{d}t = \int_A p_0(y)\,\mathrm{d}y.$$

In other words, the average time the process X^x spends in a set $A\subset\mathbb{R}$ tends to the probability that a random variable with density p_0 takes a value in A.

REMARK.– It often happens that trajectories of a diffusion process X possess a certain closedness with respect to some interval $(a,b)\subset\mathbb{R}$: starting at any point $x\in(a,b)$, it stays in (a,b) forever. Then a and b are called *unattainable boundaries* of X. For example, for the solution of the Verhulst equation $\mathrm{d}X_t = (\lambda X_t - X_t^2)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t$, considered in section 9.3, such is the interval $(0, +\infty)$, while for Ginzburg–Landau equation, there are two such intervals, $(0, +\infty)$ and $(-\infty, 0)$. We can also give an example of a finite interval. The solution of the equation

$$\mathrm{d}X_t = -\sin X_t \cos^3 X_t\,\mathrm{d}t + \cos^2 X_t\,\mathrm{d}B_t, \qquad X_0 = x \in (-\pi/2, \pi/2),$$

(or, in Stratonovich form, $\mathrm{d}X_t = \cos^2 X_t \circ \mathrm{d}B_t$) is

$$X_t = \operatorname{arctg}(\operatorname{tg} x + B_t)$$

that stays in $(-\pi/2, \pi/2)$ for all t.

In such a situation, it makes sense to consider the stationary density in the interval (a, b):

DEFINITION 10.14.– *We say that a diffusion process X with transition density $p = p(t, x, y)$ has a stationary density p_0 in the interval (a, b) if*

$$\int_a^b p(t, x, y)\,\mathrm{d}y = 1, \quad t > 0,\ x \in (a, b),$$

and

$$p_0(y) = \int_a^b p(t, x, y) p_0(x)\,\mathrm{d}x, \quad t > 0,\ y \in (a, b).$$

There are two types of unattainable boundaries of a diffusion process. In the first case, $X_t^x \to a$ or $X_t^x \to b$ as $t \to \infty$; then a or b is called an *attracting* boundary of a diffusion process $\{X^x\}$. It is clear that, in this case, the diffusion process has no stationary density in the interval (a, b) since the limit distribution of the process is concentrated at point a or b. In the other case, the process $\{X^x\}$, when started at any point $x \in (a, b)$, infinitely often visits all points of the interval (a, b); then a and b are called the natural boundaries of the diffusion process.

PROPOSITION 10.15.– *Let X be a diffusion process with generator A and transition density $p = p(t, x, y)$ which has a stationary $p_0 = p_0(y)$ in the interval (a, b). Suppose that p and p_0 are continuous functions having continuous partial derivatives $\partial p/\partial t$, $\partial p/\partial y$, $\partial^2 p/\partial y^2$, $\partial p_0/\partial y$, and $\partial^2 p_0/\partial y^2$. Then, in the interval (a, b), p_0 satisfies the forward Kolmogorov (Fokker–Planck) equation*

$$A^* p_0 = 0.$$

Proof. It is similar to that of the forward Kolmogorov equation for transition density (Proposition 10.12). Again, we begin with the equality

$$\frac{\partial}{\partial t}\int_a^b p(t,x,y)f(y)\,\mathrm{d}y = \int_a^b p(t,x,y)Af(y)\,\mathrm{d}y, \quad x\in(a,b),$$

which holds for all $f \in C_0^2(a,b)$, i.e. for all f having two continuous derivatives in (a,b) and vanishing outside some closed interval $[a',b'] \subset (a,b)$. Here, the integration interval is changed since $\int_{\mathbb{R}\setminus(a,b)} p(t,x,y)\,\mathrm{d}y = 0$, $x\in(a,b)$. Multiplying both sides by $p_0(x)$ and integrating with respect to $x\in(a,b)$, we get:

$$\frac{\partial}{\partial t}\int_a^b \left(\int_a^b p(t,x,y)p_0(x)\,\mathrm{d}x\right) f(y)\,\mathrm{d}y$$
$$= \int_a^b \left(\int_a^b p(t,x,y)p_0(x)\,\mathrm{d}x\right) Af(y)\,\mathrm{d}y,$$

or

$$\frac{\partial}{\partial t}\int_a^b p_0(y)f(y)\,\mathrm{d}y = \int_a^b p_0(y)Af(y)\,\mathrm{d}y.$$

The left-hand side is zero, while the right-hand side equals $\int_a^b A^*p_0(y)f(y)\,\mathrm{d}y$. Thus,

$$\int_a^b A^*p_0(y)f(y)\,\mathrm{d}y = 0$$

for all $f\in C_0^2(a,b)$. Therefore, $A^*p_0 \equiv 0$ in (a,b). △

COROLLARY 10.16.– *Suppose that the conditions of Proposition* 10.15 *are satisfied and* $\sigma(x) > 0$, $x\in(a,b)$. *Then the stationary density* p_0 *of the diffusion process* X *is of the form*

$$p_0(y) = \frac{N}{\sigma^2(y)}\exp\left\{2\int_c^y \frac{b(u)}{\sigma^2(u)}\,\mathrm{d}u\right\}, \quad y\in(a,b),$$

where c is an arbitrary point from (a,b), *and* N *is the normalizing constant such that* $\int_a^b p_0(y)\,\mathrm{d}y = 1$.

Proof. Integrating the Kolmogorov equation for stationary density, we get:

$$-b(y)p_0(y) + \frac{1}{2}\big(\sigma^2(y)p_0(y)\big)' = \text{const.}$$

One can show that, in the case of natural boundaries a and b, the constant on the right-hand side is zero. Intuitively, this can be explained rather simply. Since a natural boundary, in a sense, repels the trajectory of a diffusion process, the latter spends less time near the boundary. Therefore the stationary density$p_0(y)$ decreases rather rapidly (together with the derivative) as y approaches the boundary. Denote $q(y) := \sigma^2(y)p_0(y)$. Then the equation becomes rather simple:

$$q'(y) = \frac{2b(y)}{\sigma^2(y)}q(y).$$

Its general solution is

$$q(y) = N\exp\left\{2\int_c^y \frac{b(u)}{\sigma^2(u)}\,\mathrm{d}u\right\}, \quad y \in (a, b),$$

where c is an arbitrary point from (a, b). From this the stated expression for p_0 follows. $\triangle$

REMARKS.– 1. If a diffusion process $X = \{X^x\}$ is given by the *Stratonovich* equation

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t) \circ \mathrm{d}B_t,$$

then a formula for its stationary density can be obtained by writing the equation in the Itô form, that is, instead of the drift coefficient b, taking the coefficient $\tilde{b} := b + \frac{1}{2}\sigma\sigma'$ (Proposition 8.6). However, we can write a simpler formula:

$$\begin{aligned}
p_0(y) &= \frac{N}{\sigma^2(y)}\exp\left\{2\int_c^y \frac{b(u) + \frac{1}{2}\sigma\sigma'(u)}{\sigma^2(u)}\,\mathrm{d}u\right\} \\
&= \frac{N}{\sigma^2(y)}\exp\left\{2\int_c^y \frac{b(u)}{\sigma^2(u)}\,\mathrm{d}u + \int_c^y \frac{\sigma'(u)}{\sigma(u)}\,\mathrm{d}u\right\} \\
&= \frac{N}{\sigma^2(y)}\exp\left\{2\int_c^y \frac{b(u)}{\sigma^2(u)}\,\mathrm{d}u\right\}\exp\big\{\ln\sigma(y) - \ln\sigma(c)\big\} \\
&= \frac{N}{\sigma(y)}\exp\left\{2\int_c^y \frac{b(u)}{\sigma^2(u)}\,\mathrm{d}u\right\}, \quad y \in (a, b),
\end{aligned}$$

where, in the last equality, the factor $\exp\{-\ln\sigma(c)\} = 1/\sigma(c)$, independent of y, is included in the normalizing constant N. Thus, we get a joint formula for the stationary density:

$$p_0(y) = \frac{N}{\sigma^\nu(y)} \exp\left\{2\int_c^y \frac{b(u)}{\sigma^2(u)}\,\mathrm{d}u\right\}, \quad y \in (a,b),$$

where $\nu = 2$ for Itô's equation, and $\nu = 1$ for the Stratonovich one.

2. The stationary density provides rather important information about the probabilistic behavior of the process. Also, it is very important that the stationary density can be written explicitly. Moreover, even in those (rather rare) cases where we can write explicitly the solution of the SDE (or its transition density), it is more difficult to conceive the features of the process from the latter.

10.4. Exercises

10.1. Write the generator of the Markov process $X_t = \operatorname{arctg}(1 + B_t)$, $t \geqslant 0$.

10.2. For any function $f \in C_b^2(\mathbb{R})$, denote $a(t,x) := \mathbf{E}f(xe^{B_t - t})$, $(t,x) \in [0,\infty) \times \mathbb{R}$. Show that the function a satisfies the partial differential equation

$$\frac{\partial}{\partial t}a(t,x) = -\frac{1}{2}x\frac{\partial}{\partial x}a(t,x) + \frac{1}{2}x^2\frac{\partial^2}{\partial x^2}a(t,x).$$

10.3. Consider the partial differential equation

$$\frac{\partial}{\partial t}u(t,x) = x\frac{\partial}{\partial x}u(t,x) + x^2\frac{\partial^2}{\partial x^2}u(t,x), \qquad (t,x) \in [0,\infty) \times \mathbb{R},$$

with the initial condition $u(0,x) = f(x)$ ($f \in C_b^2(\mathbb{R})$). Show that its solution can be written as

$$u(t,x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f\left(xe^{y\sqrt{2t}}\right)e^{-y^2/2}\,\mathrm{d}y, \qquad (t,x) \in (0,\infty) \times \mathbb{R}.$$

10.4. The parameter $\lambda \in \mathbb{R}$ is such that the stationary density of the diffusion defined by the Verhulst equation $dX_t = (\lambda - X_t)X_t\,dt + X_t\,dB_t$, $X_0 = x > 0$, coincides with that of the diffusion defined by the equation $dX_t = (1 - X_t)\,dt + \sqrt{X_t}\,dB_t$, $X_0 = x > 0$. Find λ.

10.5. The random process X_t, $t \geqslant 0$, satisfies the SDE

$$X_t = -\int_0^t X_s \, ds + B_t, \quad t \geqslant 0.$$

Let $p_t(x)$, $t \in \mathbb{R}$, be the density of X_t. Find the limit of $p_t(x)$ as $t \to \infty$.

10.6. A diffusion process $\{X^x\}$ is defined by the SDE $dX_t = \alpha X_t \, dt + \sqrt{1 + X_t^2} \, dB_t$. Find the values of the parameter α for which the process has a stationary density.

10.7. Can the diffusion defined by an SDE

$$dX_t = b(X_t) \, dt + \sigma(X_t) \, dB_t, \quad X_0 = x_0 \in (0,1),$$

($\sigma(0) = \sigma(1) = 0$, $\sigma(x) > 0$, $x \in (0,1)$) have the stationary density $p_0(x) = 1$, $x \in (0,1)$ (i.e. the density of the uniform distribution in the interval (0,1))?

Chapter 11

Examples

In this chapter, we give some examples illustrating the theory that has been presented thus far. The motivation of their choice is twofold. First, the equations considered are important and used in the physical sciences and finance. Second, they clearly show that outer perturbations ("noises") do not only quantitatively change the behavior of a macroscopic system by augmenting its deviations from the average behavior which is usually described by deterministic equations; we shall see that sufficiently strong multiplicative noises can induce *qualitative* changes in the behavior of the system that in principle cannot be described by ordinary (non-stochastic) equations that are oriented toward the *average* behavior.

In the figures, we present several typical trajectories of various stochastic differential equations. In the next chapter we shall see how such trajectories are obtained by a computer simulation. However, here we want to emphasize that though those trajectories are typical, they are *random*, since they depend on the random trajectories of a Brownian motion. Therefore, each time we run a computer program that solves or simulates an SDE, on a display (or on a paper sheet), we see another trajectory. However, the trajectories of a Brownian motion are usually simulated by means of sequences of the so-called pseudorandom numbers, that is, sequences of non-random numbers having properties of sequences of random numbers. Usually, a computer program allows us to set the mode where each time the same sequence of pseudorandom numbers is repeated anew. Then, each time we get the same trajectory of Brownian motion. We use this mode in the examples of this chapter. All solutions of SDEs are obtained with the same (and even generated in advance) trajectory of Brownian motion. The advantage of such an approach is that we can better visualize the dependence of the solution behavior on parameters (most often, on the noise intensity), since the influence of "randomness" is identical for all equations depending on the parameter. Therefore, in the examples, we can easily see both the similarities and differences of the solutions of SDEs. The similarities are stipulated by the same trajectory of the "driving" Brownian motion, while the differences are stipulated by different values of parameter(s).

11.1. Additive noise: Langevin equation

The stationary density of the Ornstein–Uhlenbeck diffusion process, described by the Langevin equation $\mathrm{d}X_t = -aX_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$, is

$$p_0(x) = N\exp\bigg\{-2\int_0^x \frac{au}{\sigma^2}\,\mathrm{d}u\bigg\} = N\exp\bigg\{-\frac{ax^2}{\sigma^2}\bigg\}.$$

We recognize that, for $a > 0$, up to a constant factor, p_0 is the density of normal random variable $\xi \sim N(0, \sigma^2/2a)$; therefore the normalizing constant N equals $1/\sqrt{\pi\sigma^2/a}$. Thus, the distribution of the Ornstein–Uhlenbeck process X_t^x eventually becomes close to the normal distribution $N(0, \sigma^2/2a)$. This fact complements those obtained in Chapter 9 about the behavior of its expectation and variance ($\mathbf{E}X_t^x = x\exp\{-at\} \to 0$ and $\mathbf{D}X_t^x = \sigma^2(2a)^{-1}(1-\exp\{-2at\}) \to \sigma^2/2a$).

When $a \leqslant 0$, the process has no stationary density. This is rather natural. When $a = 0$, up to constant factor, we have the Brownian motion $X_t^x = x + \sigma B_t \sim N(x, \sigma^2 t)$ the distribution of which has no limit as $t \to \infty$, since it has the constant expectation $\mathbf{E}X_t^x = x$ and infinitely increasing variance $\mathbf{D}X_t^x = \sigma^2 t$. For $a < 0$, it is even worse: not only the variance $\mathbf{D}X_t^x \to +\infty$, but also the expectation $\mathbf{E}X_t^x \to \operatorname{sgn} x \cdot \infty$.

11.2. Additive noise: general case

The latter example can be rather easily generalized to an arbitrary stochastic differential equation with *additive noise*:

$$\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t.$$

If the stationary density of the corresponding diffusion process exists, then it must be of the form

$$p_0(x) = N\exp\bigg\{\frac{2}{\sigma^2}\int_0^x f(u)\,\mathrm{d}u\bigg\}.$$

It is a density if the function p_0 is integrable. If so, then the normalizing constant N equals

$$N = \left(\int_{-\infty}^{\infty} \exp\left\{ \frac{2}{\sigma^2} \int_0^x f(u)\,\mathrm{d}u \right\} \mathrm{d}x \right)^{-1}.$$

We see that, for sufficiently large (in absolute value) x, the drift coefficient $f = f(x)$ must be the sign opposite to that of x, and the integral $\int_0^x f(u)\,\mathrm{d}u$ must tend to $-\infty$ (not "too slowly"). This can be interpreted as follows: for stabilization of the probabilistic behavior of process X^x, it must be sufficiently strongly pushed back in the direction opposite to its sign if it moves "too far away".

Let us first look what happens in the extreme case where there is no noise, that is, $\sigma = 0$. Since $f(x)x < 0$ for sufficiently large $|x|$, there is at least one point $x_0 \in \mathbb{R}$ such that $f(x_0) = 0$. In the theory of ordinary differential equations, such points are called *equilibrium points* (of the equation $\mathrm{d}X_t = f(X_t)\,\mathrm{d}t$): the process, starting from such a point x_0, stays there forever: $X_t^{x_0} = x_0$ for all $t \geqslant 0$. Equilibrium points are of two types, stable and unstable, depending on the behavior of X^x at the points that are close to the equilibrium point. If at the points x close to the equilibrium point x_0, the function f decreases (intersects the x-axis at x_0 downward, or $f'(x_0) < 0$), then, at these points, $f(x)(x - x_0) < 0$, and therefore the solution is pushed toward equilibrium point x_0. Such an equilibrium point is called *stable*. Otherwise, it is called *unstable*. It is convenient to characterize stable and unstable equilibrium points in terms of the notion of potential. Denote $U(x) = -\int_0^x f(y)\,\mathrm{d}y$, $x \in \mathbb{R}$. The function U is called a *potential* of the function f.[1] At an equilibrium point, potential U attains an extremum, minimum at a stable point and maximum at unstable one. It is typical that all solutions starting at non-equilibrium points converge to stable equilibrium points, or, in other words, "fall into a potential pit" (a minimum point of the potential). For illustration, it is rather convenient to use third-order polynomials $f(x) = a(x-x_1)(x-x_2)(x-x_3)$ with negative coefficient a at x^3 and three different real roots $x_1 < x_2 < x_3$. Then $f'(x_1) < 0$, $f'(x_2) > 0$, $f'(x_3) < 0$, and therefore x_2 is unstable, and x_1 and x_3 are stable equilibrium points. Let us consider a concrete example, $f(x) := -4x^3 - x^2 + 5x = -4x(x-1)(x+\frac{5}{4})$ with stable equilibrium points 1 and $-\frac{5}{4}$ and the unstable point 0 (the choice is motivated by a better graphical visualization). The potential of f is $U(x) = x^4 + \frac{1}{3}x^3 - \frac{5}{2}x^2$ (see Figure 11.1).

Several trajectories of the deterministic equation ($\sigma = 0$) are shown in the left part of Figure 11.2. We see that all trajectories, except the unstable equilibrium point $x = 0$, rather rapidly "fall into a potential pit", that is, tend to one of the stable

1. In the one-dimensional case, U is actually the antiderivative of f with opposite sign. In the multidimensional case, a function $U\colon \mathbb{R}^k \to \mathbb{R}$ is called a potential of a vector function $f\colon \mathbb{R}^k \to \mathbb{R}^k$ if $\operatorname{grad} U = -f$, i.e. $\partial U/\partial x_i = -f_i$ for all $i = 1, 2, \dots, k$.

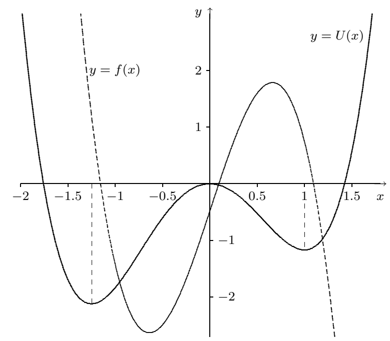


Figure 11.1. *The function* $f(x) = -4x^3 - x^2 + 5x$ *and its potential* U

equilibrium points, $x = 1$ or $x = -1.25$. In the right part of the figure, we see the typical trajectories of the stochastic equation with $\sigma = 0.8$ starting from the same points. They also tend to fall into one of the potential pits (stable equilibrium points), but, under the influence of random perturbations, deviate from them. Note also a certain regularity: trajectories that fall into the same potential pit eventually turn into a narrow beam of trajectories.

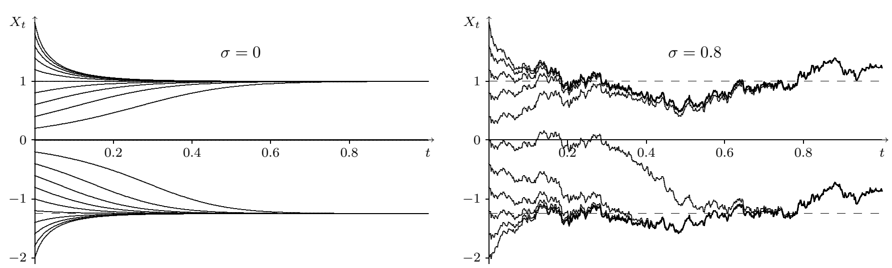


Figure 11.2. *Solutions of the equation* $\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$ *for* $\sigma = 0$ *and* $\sigma = 0.8$

Let us now return to the stationary density. For a process with additive noise, we can write it as

$$p_0(x) = N \exp\Big\{ -\frac{2}{\sigma^2} U(x) \Big\}.$$

We see that its maximum points coincide with the minimum points of the potential. The values of the maxima depend not only on the potential, but also on the coefficient σ showing the noise intensity. The lower σ, the greater the maxima of the stationary density. The stationary densities corresponding to several different values of σ are shown in Figure 11.3.

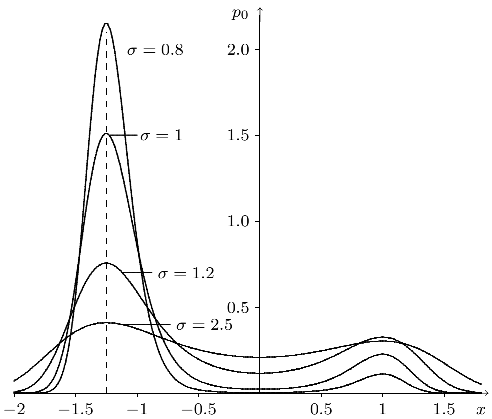


Figure 11.3. *The dependence of stationary density on σ*

The number and positions of extrema of the stationary density (or, equivalently, of the potential) are particularly important characteristics of the stationary behavior of a diffusion process. A diffusion process with additive noise, having a stationary density, spends most of the time near the maximum points of stationary density (or minimum points of the potential). Besides, the greater the maximum, the longer the time spent in a neighborhood of the maximum point. On the contrary, when getting near the maximum points of the potential (i.e. at the minimum points of the potential), the trajectories of a diffusion process tend to leave them relatively quickly. If the stationary density has several maximum points (the potential has several minima), then, after spending some time in one potential pit, the process will "roll" through a potential hill, into another potential pit. However, if this hill is rather high, or the noise (σ) is low, we would wait for such rolling a long time. For this reason, we did not observe such rolling through the hill in the time interval $[0, 1]$ (Figure 11.2, right), except for the trajectory starting from the hill itself, the point $x = 0$. Let us observe the behavior of the solution in a longer interval, say $[0, 100]$. It is shown in Figure 11.4 for several different values of σ. Note that, as σ increases, the rollings become more frequent, because of increasing values of the stationary densities near the point $x = 0$. The difference between times spent near equilibrium points $x = -1.25$ and $x = 1$ is perceptible for all σ, though it is less perceptible for higher σ. This is natural since, as σ decreases, so does the difference between the maximums of the stationary density (Figure 11.4).

Summarizing, we can state that an additive noise does not change *qualitatively* the stationary behavior of a diffusion process. By this we mean that the number and positions of maximum points of the stationary density do not depend on the noise intensity, which only influences the values of the maximums. The process mainly oscillates near one or another potential pit, changing pits from time to time, and the change in noise intensity has no qualitative influence on the form of the potential.

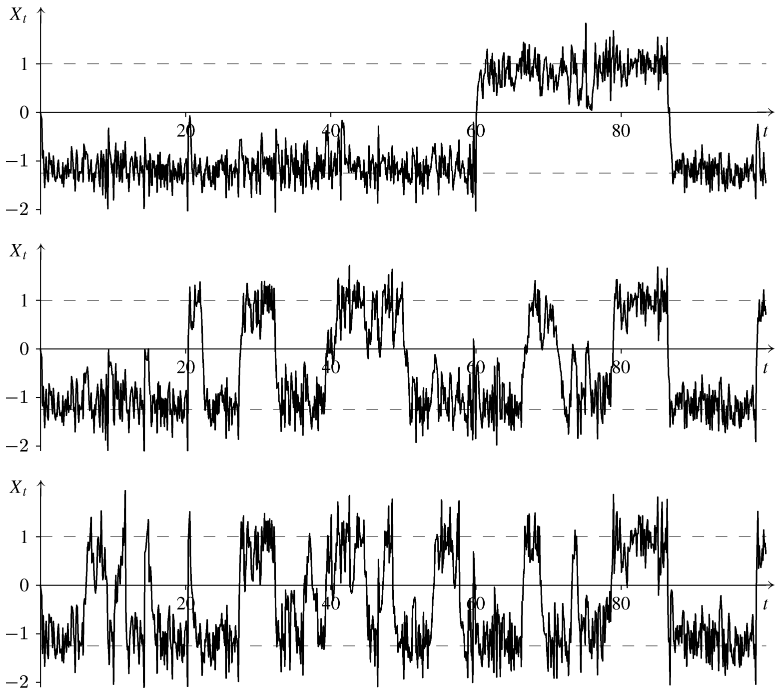


Figure 11.4. *Solutions of the equation* $\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t$, $X_0 = 0$, *for* $\sigma = 0.8;\ 1;\ 1.2$

11.3. Multiplicative noise: general remarks

Let us pass to examples with multiplicative noise. Suppose that a diffusion process $\{X^x\}$ described by the equation $\mathrm{d}X_t = f(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t$ has a stationary density p_0 in the interval (a, b). We already know that it is of the form

$$p_0(x) = \frac{N}{\sigma^2(x)} \exp\Big\{2\int_c^x \frac{f(u)}{\sigma^2(u)}\,\mathrm{d}u\Big\}, \qquad x \in (a, b).$$

To analyze how the stationary density varies under a proportional increase in noise, we assume that the diffusion coefficient is proportional to some function, i.e. $\sigma(x) = \sigma g(x)$, where $\sigma \geqslant 0$ is the proportionality coefficient. For analogy and comparison with the additive noise case, we write the stationary density in the form

$$p_0(x) = N \exp\Big\{2\int_c^x \frac{f(u)}{\sigma^2(u)}\,\mathrm{d}u - 2\ln\sigma(x)\Big\}$$

$$= N \exp\Big\{\frac{2}{\sigma^2}\int\limits_c^x \frac{f(u)}{g^2(u)}\,\mathrm{d}u - 2\ln g(x)\Big\}$$

$$= N \exp\{-\frac{2}{\sigma^2}\mathcal{U}(x)\}, \qquad x \in (a, b),$$

(the normalizing constants N vary from line to line) and call the function

$$\mathcal{U}(x) := -\Big[\int\limits_c^x \frac{f(u)}{g^2(u)}\,\mathrm{d}u - \sigma^2 \ln g(x)\Big], \qquad x \in (a, b),$$

the *stochastic potential* in the interval (a, b). In the additive case ($g(x) \equiv 1$), the stochastic potential coincides with the deterministic potential U. As already mentioned, the extremum points of the stationary density of a diffusion process and their positions are important characteristics of the behavior of the process. In the multiplicative case, these extremum points now coincide with those of the stochastic potential $\mathcal{U}$ and may essentially differ from the extremum points of the deterministic potential U. It is very important that a proportional increase in *multiplicative* noise can *qualitatively* change the behavior of the process. This means that not only the values of the extrema of stationary density depend on the noise intensity, but also their number and positions. Such qualitative changes in the stationary density are called *noise-induced transitions*. One can say that, in contrast to additive noise, multiplicative noise induces not only greater or smaller oscillations of a system described by an SDE in the pits of potential "landscape" but as the noise intensity varies, its "terrains" rise or descend and, at the same time, change the important characteristics of the stationary density, the number and positions of extremum points.

The extremum points of the stationary density coincide with those of the stochastic potential $\mathcal{U}$. Therefore, for such points x,

$$\mathcal{U}'(x) = \frac{f(x)}{g^2(x)} - \sigma^2 \frac{g'(x)}{g(x)} = 0,$$

i.e.

$$f(x) - \sigma^2 g(x) g'(x) = 0.$$

This simple equation is very important for analysis of noise-induced transitions. In the case of the Stratonovich equation, instead of the drift function $f(x)$, we have to take the function $f(x) + \sigma(x)\sigma'(x)/2 = f(x) + \frac{\sigma^2}{2} g(x) g'(x)$, and the latter equation becomes

$$f(x) - \frac{\sigma^2}{2} g(x) g'(x) = 0.$$

11.4. Multiplicative noise: Verhulst equation

Now consider a concrete example: the stochastic Verhulst equation

$$\mathrm{d}X_t = (\lambda X_t - X_t^2)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t.$$

As already mentioned, this equation was first used to model the biology of population growth. Later on, it appeared in various areas of natural sciences.[2] We have solved it in section 9.3:

$$X_t = \frac{X_0 \exp\left\{\left(\lambda - \frac{1}{2}\sigma^2\right)t + \sigma B_t\right\}}{1 + X_0 \int\limits_0^t \exp\left\{\left(\lambda - \dfrac{1}{2}\sigma^2\right)s + \sigma B_s\right\}\mathrm{d}s}.$$

From this expression we see that if $X_0 > 0$, then also $X_t > 0$ for all $t \geqslant 0$. If $\lambda < \sigma^2/2$, then $(\lambda - \sigma^2/2)t + \sigma B_t = t[(\lambda - \sigma^2/2) + \sigma B_t/t] \to -\infty$ as $t \to \infty$. Here we used $B_t/t \to 0$ as $t \to \infty$, which in turn follows from the fact that $B_t/\sqrt{t} \overset{d}{=} \xi \sim N(0,1)$ and thus $B_t/t = (B_t/\sqrt{t})/\sqrt{t} \to 0$ as $t \to \infty$. From this we find that $X_t \to 0$ as $t \to \infty$. Thus, for $\lambda < \sigma^2/2$, the point 0 is an attracting boundary, and therefore the process described by the Verhulst equation has no stationary density. So, we can expect the existence of stationary density p_0 in the interval $(0, \infty)$ only for $\lambda \geqslant \sigma^2/2$. It must be of the form (we take, e.g. $c = 1$)

$$\begin{aligned}
p_0(x) &= \frac{N}{\sigma^2 x^2} \exp\left\{2\int\limits_1^x \frac{\lambda u - u^2}{\sigma^2 u^2}\,\mathrm{d}u\right\} \\
&= \frac{N}{\sigma^2 x^2} \exp\left\{2\int\limits_1^x \left(\frac{\lambda}{\sigma^2 u} - \frac{1}{\sigma^2}\right)\mathrm{d}u\right\} \\
&= \frac{N}{\sigma^2 x^2} \exp\left\{\frac{2\lambda}{\sigma^2}\ln x - \frac{2}{\sigma^2}(x-1)\right\} \\
&= N\exp\{2/\sigma^2\} x^{2\lambda/\sigma^2 - 2} \exp\left\{-\frac{2}{\sigma^2}x\right\} \\
&= N x^{2\lambda/\sigma^2 - 2} \exp\left\{-\frac{2}{\sigma^2}x\right\}.
\end{aligned}$$

The function p_0 is integrable in $(0, \infty)$ if $2\lambda/\sigma^2 - 2 > -1$, that is, for $\sigma^2 < 2\lambda$. So, we can say that when $\sigma^2 = 2\lambda$ (or, equivalently, $\sigma = \sqrt{2\lambda}$), we have a noise-induced transition: the diffusion process described by the Verhulst equation has no

2. It is interesting to note that both Itô and Stratonovich interpretations are used—the choice depends on the concrete modeled process.

stationary density for $\sigma^2 \geqslant 2\lambda$ (in terms of generalized functions, we can say that, in this case, the stationary density $p_0(x) = \delta(x)$ is the Dirac delta function); and it has a density for $\sigma^2 < 2\lambda$. Moreover, note that

$$\lim_{x \downarrow 0} p_0(x) = \begin{cases} +\infty & \text{for } \sigma^2/2 < \lambda < \sigma^2, \\ N = \frac{2}{\lambda} & \text{for } \lambda = \sigma^2, \\ 0 & \text{for } \lambda > \sigma^2. \end{cases}$$

We see that the forms of stationary density are qualitatively different for $\lambda < \sigma^2 < 2\lambda$ and $\sigma^2 < \lambda$ (see Figure 11.5). Thus, we can say that, at $\sigma^2 = \lambda$ (or at

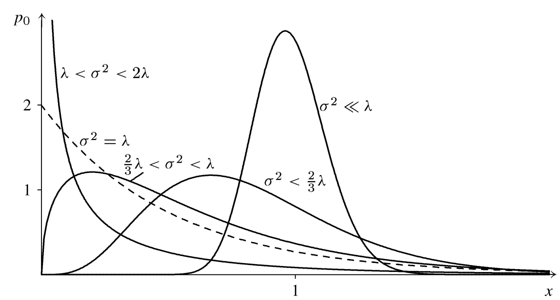


Figure 11.5. *Stationary densities of the Verhulst equation* ($\lambda = 1$, $\sigma^2 = 1/25;\ 1/3;\ 1;\ 4/5;\ 8/5$)

$\sigma = \sqrt{\lambda}$), we also have a noise-induced transition. A somewhat lesser qualitative difference (which can also be called a noise-induced transition) in the behavior of the stationary density p_0 in a neighborhood of $x = 0$ can be seen between the cases $\frac{2}{3}\lambda < \sigma^2 < \lambda$ and $\sigma^2 < \frac{2}{3}\lambda$. In the first case, $\lim_{x\downarrow 0} p_0'(x) = +\infty$, while in the second, $\lim_{x\downarrow 0} p_0'(x) = 0$.

Equating the derivative of the stationary density to zero or, equivalently, using the equation derived at the end of the previous section (which, in this case, is $\lambda x - x^2 - \sigma^2 x = 0$), we find the maximum point of the stationary density, $x_{\max} = \lambda - \sigma^2$ (when $0 < \sigma^2 < \lambda$). (We do not write the expression of the maximum $p_0(x_{\max})$ itself, since it is rather cumbersome.) When the noise is small (σ tends to zero), the stationary density becomes more acute, and the maximum point tends to $x = \lambda$. This is natural, since when the noise is small, the solutions of the SDE become similar to the solution of the deterministic equation ($\sigma = 0$), and for the latter, we have $\lim_{t\to\infty} X_t = \lambda$, provided that $X_0 > 0$.

In Figure 11.6, we see the solutions of the Verhulst equation with $\lambda = 1$ and various σ, beginning with zero (deterministic equation with the solutions tending to λ) and ending with large value $\sigma > \sqrt{2\lambda}$, when the equation has no stationary density and the solutions converge to zero.

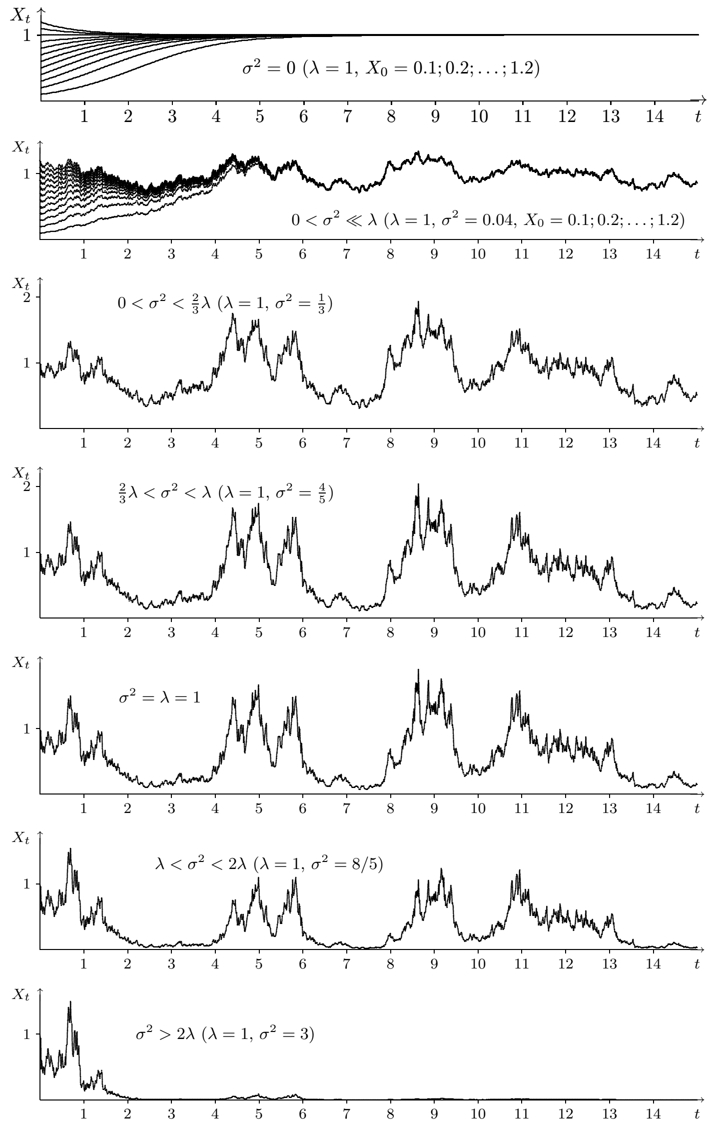


Figure 11.6. *Solutions of the Verhulst equation* $\mathrm{d}X_t = (\lambda X_t - X_t^2)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t$, $X_0 = 1$, *with various* σ ($\lambda = 1$; $\sigma^2 = 0;\ 1/25;\ 1/3;\ 4/5;\ 1;\ 8/5;\ 3$)

To end with this example, we talk about the Verhulst equation in the Stratonovich interpretation,

$$\mathrm{d}X_t = (\lambda X_t - X_t^2)\,\mathrm{d}t + \sigma X_t \circ \mathrm{d}B_t.$$

In the Itô form, it is written

$$\mathrm{d}X_t = (\tilde{\lambda}X_t - X_t^2)\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t$$

with $\tilde{\lambda} := \lambda + \frac{1}{2}\sigma^2$. Therefore, to pass to the Stratonovich interpretation, it suffices to replace λ by $\tilde{\lambda}$. For example, the solution of the equation is now

$$X_t = \frac{X_0 \exp\{\lambda t + \sigma B_t\}}{1 + X_0 \int_0^t \exp\{\lambda s + \sigma B_s\}\,\mathrm{d}s},$$

and the noise-induced transitions occur when $\lambda = 0$ ir $\lambda = \sigma^2/2$ (instead of $\lambda = \sigma^2/2$ and $\lambda = \sigma^2$, respectively).

11.5. Multiplicative noise: genetic model

The deterministic equation

$$\mathrm{d}X_t = \big(\alpha - X_t + \lambda X_t(1 - X_t)\big)\,\mathrm{d}t, \qquad X_0 \in (0,1),$$

with parameters $\alpha \in (0,1)$ and $\lambda \in \mathbb{R}$ is called the *genetic model*. Under certain conditions, its solution X_t models the relative quantity N_A/N of two genotypes, A and a, at time moment t, in a population; here $N = N_A + N_a$ is the size of the whole population. The equation also has applications in other areas. For example, it describes some chemical reactions with two reactants, where X_t means the relative quantity of one at time moment t.[3] Supposing that the parameter λ is perturbed by noise (random oscillations in the environment) close to white noise, we obtain the Stratonovich stochastic differential equation

$$\mathrm{d}X_t = \big(\alpha - X_t + \lambda X_t(1 - X_t)\big)\,\mathrm{d}t + \sigma X_t(1 - X_t) \circ \mathrm{d}B_t, \qquad X_0 \in (0,1).$$

The equivalent Itô equation is

$$\begin{aligned}\mathrm{d}X_t = \Big(\alpha - X_t + \lambda X_t(1 - X_t) + \frac{1}{2}\sigma^2 X_t(1 - X_t)(1 - 2X_t)\Big)\,\mathrm{d}t \\ + \sigma X_t(1 - X_t)\,\mathrm{d}B_t, \quad X_0 \in (0,1).\end{aligned}$$

The reason for choosing the Stratonovich interpretation is not mathematical—though "usual" integration and differentiation rules often allow us to explicitly solve

3. Chemists describe such a reaction using the scheme $A + X + Y \rightleftarrows 2Y + A^*$, $B + X + Y \rightleftarrows 2X + B^*$.

a Stratonovich equation, this time this does not work. It appears that the equation considered describes the above-mentioned biological and chemical SDEs better when they are understood in the Stratonovich sense.

We first check that if $X_0 \in (0,1)$, then $X_t \in (0,1)$ for all $t \geqslant 0$, that is, 0 and 1 are inaccessible boundaries of the process X starting from a point in $(0,1)$. Divide the equation by $X_t(1-X_t) > 0$, which is possible by the properties of the Stratonovich stochastic integral (Proposition 8.3):

$$\frac{1}{X_t(1-X_t)} \circ \mathrm{d}X_t = \Big[\frac{\alpha - X_t}{X_t(1-X_t)} + \lambda\Big]\,\mathrm{d}t + \sigma\,\mathrm{d}B_t.$$

Integrating and applying Itô's formula in the Stratonovich form (Theorem 8.4) to the left-hand side, we get:

$$\ln\Big(\frac{X_t}{1-X_t}\Big) - \ln\Big(\frac{X_0}{1-X_0}\Big) = \int_0^t \frac{\alpha - X_s}{X_s(1-X_s)}\,\mathrm{d}s + \lambda t + \sigma B_t.$$

Suppose that $0 < X_t \to 0$ as $t \to \tau$. Then the left-hand side of the equation tends to $-\infty$, while the integrand of the right-hand side $(\alpha - X_s)/(X_s(1-X_s)) \to +\infty$ as $s \to \tau$. Therefore the integral on the right-hand side cannot tend to $-\infty$ as $t \to \tau$, a contradiction. Similarly, we get a contradiction if we suppose that $1 > X_t \to 1$ as $t \to \tau$.

For simplicity, we further consider the case $\alpha = 1/2$. The stationary density is of the form (see Remark 1 on page 175)

$$\begin{aligned}
p_0(x) &= \frac{N}{x(1-x)} \exp\Big\{\frac{2}{\sigma^2}\int_{1/2}^{x} \frac{1/2 - u + \lambda u(1-u)}{u^2(1-u)^2}\,\mathrm{d}u\Big\}\\
&= \frac{N}{x(1-x)} \exp\Big\{\frac{2}{\sigma^2}\int_{1/2}^{x} \Big[\frac{1-2u}{2u^2(1-u)^2} + \frac{\lambda}{u(1-u)}\Big]\,\mathrm{d}u\Big\}\\
&= \frac{N}{x(1-x)} \exp\Big\{-\frac{2}{\sigma^2}\Big[\frac{1}{2x(1-x)} + \lambda \ln\frac{1-x}{x}\Big]\Big\}.
\end{aligned}$$

We easily see that this function is integrable in the interval $(0,1)$ for all values of parameters $\sigma > 0$ and $\lambda \in \mathbb{R}$; hence, the process considered always has a stationary density. The extremum points of the latter, x_m, satisfy the equation

$$q(x) := \frac{1}{2} - x + \lambda x(1-x) - \frac{\sigma^2}{2}x(1-x)(1-2x) = 0.$$

First, consider the *symmetric* case $\lambda = 0$. Then, for the extremum points of the stationary density, we get the equation

$$q(x_m) = \Big(\frac{1}{2} - x_m\Big)\Big(1 - \sigma^2 x_m(1 - x_m)\Big) = 0$$

or

$$\Big(x_m - \frac{1}{2}\Big)\Big(x_m^2 - x_m + \frac{1}{\sigma^2}\Big) = 0.$$

One its solution is $x_m = 1/2$. If $\sigma^2 \leqslant 4$, this equation has no more solutions, and the stationary density has a maximum at the point $x_{m1} = 1/2$. If $\sigma^2 > 4$, it has yet two solutions $x_{m\pm} = (1 \pm \sqrt{1 - 4/\sigma^2})/2$, and at these points, the stationary density has maximums, while at $x_{m1} = 1/2$, it already has a minimum (Figure 11.7).

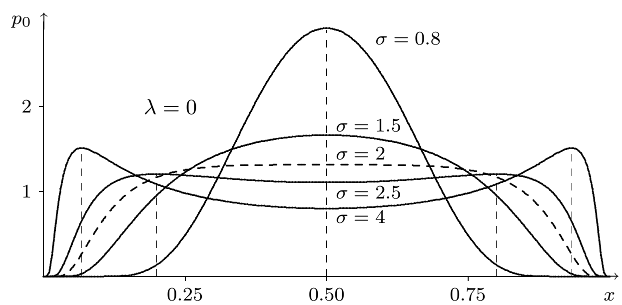


Figure 11.7. *Symmetric stationary densities of genetic model* $(\lambda = 0)$

Thus, we can say that, at $\sigma^2 = 4$, we have a noise-induced transition. Returning to the chemical-reaction interpretation, imagine that one reactant in the reaction is blue, and the other is yellow. While there is no "noise" or its intensity is small, we see a mixture of green color. When the noise intensity increases up to the value that is needed for noise-induced transition, the relative quantity of the first reactant, X_t, will be close to one or the other maximum point of the stationary density. So, we will see that the mixture of reactants is, in succession, of blue and yellow color. It is interesting that similar macroscopic phenomena are observable in real laboratory experiments, and they cannot be explained in the framework of deterministic equations.

In the *asymmetric* case $\lambda > 0$ (or $\lambda < 0$), the situation is qualitatively similar, though there are some differences (Figure 11.8).

First, the form of the stationary density becomes asymmetric. The deterministic equation ($\sigma = 0$) has a stable equilibrium point $x_0 = 1/2 + (\sqrt{1 + \lambda^2} - 1)/(2\lambda)$ (the solution of the equation $1/2 - x + \lambda x(1 - x) = 0$ in the interval $(0, 1)$). When $\sigma > 0$

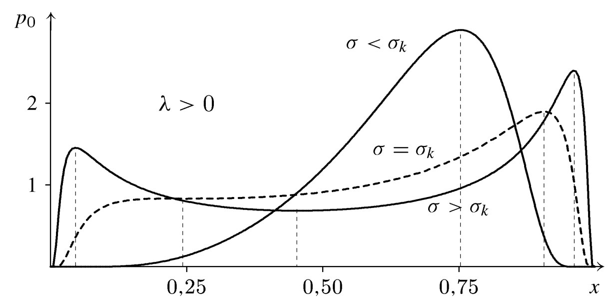


Figure 11.8. *Asymmetric stationary densities of genetic model* ($\lambda = 1$, $\sigma = \sigma_k \approx 3.054$, $\sigma = 1.2 < \sigma_k$, $\sigma = 5 > \sigma_k$)

starts to increase from zero, the stationary density has a unique maximum point x_{m1}, the solution of the equation $q(x) = 0$, which approaches 1 (or 0 if $\lambda < 0$). When σ reaches a critical value σ_k, the equation $q(x) = 0$ has yet one (double) solution x_{m2} (this solution, together with σ_k, can be obtained by solving the equation system $q(x) = 0$, $q'(x) = 0$). For $\sigma = \sigma_k$, the stationary density has, at this point, an inflection. When $\sigma > \sigma_k$, the double solution of the equation $q(x) = 0$ separates into two solutions $x_{m2} < x_{m3}$. The stationary density has a maximum at the point x_{m2}, and it has a minimum at the point x_{m3}. Thus, we can say that, at the point $\sigma = \sigma_k$, we have a noise-induced transition that is characterized by a change in the number of extrema of the stationary density.

In comparison to the symmetric case, we can notice another interesting difference. In the symmetric case, the transition from one to three extrema is "gentle": as σ passes through the critical value, the maximum point x_{m1} continuously separates into two maximum points, x_{m+} and x_{m-} ($x_{m+} - x_{m-} \to 0$ as $\sigma \downarrow \sigma_k = 2$). However, in the asymmetric case, one new maximum "sprouts" near the other end of the interval $(0, 1)$. In Figure 11.9, we see typical trajectories of solutions of the genetic model in symmetric ($\lambda = 0$) and asymmetric ($\lambda > 0$) cases with various noise intensities—smaller than, equal to, and greater than critical values. The behavior of the trajectories is adequate to the particularities of the stationary densities.

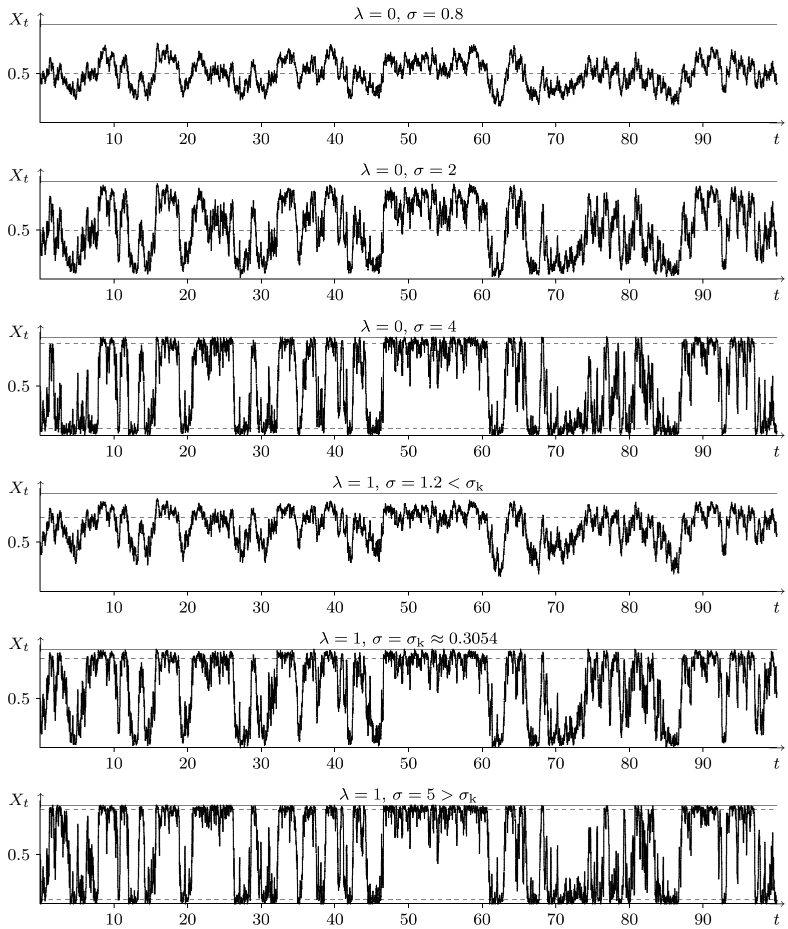


Figure 11.9. *Solutions of genetic model in symmetric* $(\lambda = 0)$ *and asymmetric* $(\lambda = 1 > 0)$ *cases. Dashed lines show the maximum points of stationary densities*

Chapter 12

Example in Finance: Black–Scholes Model

12.1. Introduction: what is an option?

Option is a contract between two parties, buyer and seller, which gives one party the right, but not the obligation, to buy or to sell some asset until some agreed date, while the second party has the obligation to fulfill the contract if requested.

Let us illustrate this notion by using a real-life example. Suppose that you have decided to buy a house and have found somewhere you like. At the moment, you do not have enough money, but you hope to get it in the near future (say, by getting credit or by selling another property). Therefore you make a contract (called an option) with the seller of a house that he will wait for a half of year and will sell the house for an agreed price of, say, \$100,000. However, he agrees to wait only under the condition that you pay for this, say, \$1,000. Further, we can imagine the following two scenarios:

1. Within half year, the prices of real estate have increased, and the market price of the house has increased to \$120,000. Since the owner has signed a contract with you and is paid for this (by selling the option), he is obliged to sell you the house for \$100,000. Thus, in this case, you have a significant profit of

$$120{,}000 - 100{,}000 - 1{,}000 = \$19{,}000.$$

2. Checking the details on that house and talking to the neighbors, you learn that there are many dysfunctional neighbors, and that the house has a very expensive and time-consuming solid-fuel heating boiler. So, although at the beginning, you thought you had found a dream house, now you clearly see that it is not worth the agreed price. Happily, according to the contract, you are not obliged to buy the house. Of course, you loose the sum of \$1,000 paid for the contract.

This example illustrates some important properties of an option. First, buying an option, you gain the right, but not the obligation, to buy something. If you do not exercise the option after the expiration time of the contract, it becomes worthless, and you loose only the sum paid for the option. Second, the option is a contract giving the right to a certain asset. Therefore, an option is a *derivative* financial instrument, whose price depends on the price of some other asset. In our example, this asset is a house. In the options used by investors, the role of such an asset is usually played by stocks or market indices.

The main two types of options are the following:

– *Call option* gives its owner the right to buy some asset (commonly a stock, a bond, currency, etc.) for a price called the *exercise price* or *strike price* until a specified time called the *maturity* or *expiration time* of the option. If, at the expiration time, the option is not exercised (that is, the owner decides not to buy the underlying asset), it becomes void and worthless. Call options are similar to long-term stock position as the buyer hopes that the price of the underlying stock will significantly increase up to the expiration time.

– *Put option* gives its owner the right to sell some asset for an exercise price until the expiration time of the option. Put options are similar to short-term stock position as the buyer hopes that the price of the underlying stock will significantly decrease until the expiration time.

The option price of an option (not to be mixed with the price of underlying stock) is called the *premium*. The premium depends on factors such as the price of the underlying stock, exercise price, the time remaining until the expiration time, and on the *volatility* of the stock. Determining the theoretical fair value (premium) for a call or put option, the so-called *option pricing* is rather complicated and one of the main problems in financial mathematics.

There are two main types of options:

– *European options*: they can be only exercised at the expiration time.

– *American options*: in contrast to European options, they can be exercised at any time before the expiration time.

Let S_t be the stock price at time moment t. The exercise price of the European call option with exercise price K and expiration time T, that is, the profit of the option buyer is $S_T - K$ if $S_T > K$ and zero if $S_T \leqslant K$. Thus, the profit of the call option buyer is

$$f(S_T) = (S_T - K)^+ = \begin{cases} S_T - K & \text{if } S_T > K; \\ 0 & \text{if } S_T \leqslant K. \end{cases}$$

Similarly, the profit of a buyer of the European put option with exercise price K and expiration time T is

$$g(S_T) = (K - S_T)^+ = \begin{cases} K - S_T & \text{if } S_T < K; \\ 0, & \text{if } S_T \geqslant K. \end{cases}$$

The following two main questions arise at the time of option purchase:

1) What is the option premium to be paid by the buyer to the seller of, say, a call option? Or, in other words, how can we evaluate, at the time of signing the contract ($t = 0$), the future profit of the buyer $(S_T - K)^+$ at time moment T? This is the so-called problem of option pricing.

2) How can the seller of an option, after getting the premium at the moment $t = 0$, to guarantee the profit $(S_T - K)^+$ at time moment T? This is the problem of option risk management or hedging.

An answer to these closely related questions is based on modeling stock prices (usually by, stochastic differential equations) and working on the assumption of no-arbitrage in the financial market which essentially means that there is no guaranteed profit without risk.

12.2. Self-financing strategies

12.2.1. *Portfolio and its trading strategy*

In financial mathematics, a financial market is modeled as a collection of investor's (or agent's) assets which is called a portfolio. It may contain stocks, bonds, and other securities, bank accounts, investment funds, or derivatives (financial instruments that are based on the expected future price movements of an asset). Let S_t^i be the price of the ith asset at time t. All prices S^i are random processes in some probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The collection of all prices $X_t = (S_t^0, S_t^1, \dots, S_t^N)$, $t \geqslant 0$, is a random vectorial process. Usually, the asset with index 0 is assumed to be riskless (e.g. a bond or banc account) with constant short-time (or spot) interest rate $r \geqslant 0$, that is, its price is a non-random process $S_t^0 = S_0^0 e^{rt}$.[1] For short, we call it a bond. The remaining assets ($i = 1, 2, \dots, N$) are assumed to be risky, and we call them stocks, denoting their collection by $S_t = (S_t^1, \dots, S_t^N)$. The investors may change the contents of the portfolio by buying or selling stocks and bonds or their parts. This is modeled by a trading strategy, which is a random vectorial process $\varphi_t = (\theta_t^0, \theta_t) = (\theta_t^0, \theta_t^1, \dots, \theta_t^N)$, $t \geqslant 0$, where θ_t^0 is the number of shares in a bond at time t, and θ_t^i

1. Without loss of generality, we can assume that $S_0^0 = 1$ (it is not important whether we have one banknote of \$100 or 100 banknotes of \$1).

($i = 1, 2, \dots, N$) is the number of shares of the ith stock at time t. The wealth of the portfolio at time t equals

$$V_t = V_t(\varphi) = \varphi_t \cdot X_t = \theta_t^0 S_t^0 + \sum_{i=1}^{N} \theta_t^i S_t^i. \qquad [12.1]$$

Usually, we shall consider trading strategies in a finite time interval $[0, T]$.

EXAMPLE 12.1.– The most simple and well-known example is the Black–Scholes model, where the portfolio consists of two assets, a bond with price $S_t^0 = S_0^0 e^{rt}$ and stock with price S_t satisfying the linear stochastic differential equation

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t$$

or, in the integral form,

$$S_t = S_0 + \mu \int_0^t S_u \;\mathrm{d}u + \sigma \int_0^t S_u \;\mathrm{d}B_u, \quad t \in [0, T]; \qquad [12.2]$$

the constant μ is called the mean rate of return, and $\sigma > 0$ is called the volatility and shows, in some sense, the degree of risk of that stock. As we know, the solution of equation [12.2] is the so-called geometric Brownian motion

$$S_t = S_0 \exp\big\{(\mu - \sigma^2/2)t + \sigma B_t\big\}, \quad t \in [0, T].$$

In the multidimensional Black–Scholes model with N stocks, it is assumed that the stock prices $S^1, \dots, S^N$ satisfy the SDEs

$$\mathrm{d}S_t^i = S_t^i \Big(\mu^i \,\mathrm{d}t + \sum_{j=1}^{d} \sigma^{ij} \;\mathrm{d}B_t^j\Big), \quad i = 1, \dots, N,$$

where $B^1, \dots, B^d$ are independent Brownian motions, μ^i are the corresponding mean rates of return, and $\sigma = (\sigma^{ij})$ is the matrix of volatilities. In more general models, b^i and σ may be functions of time t or even random processes.

12.2.2. *Self-financing strategies*

In financial mathematics, the main interest is in strategies that are self-financing, which means that the wealth of a portfolio varies without exogenous infusion or withdrawal of money, i.e. the purchase of a new asset must be financed by the sale of an old one. Formally, a self-financing strategy must satisfy the condition

$$dV_t(\varphi) = \varphi_t \cdot \,\mathrm{d}X_t = \sum_{i=0}^{N} \theta_t^i \;\mathrm{d}S_t^i. \qquad [12.3]$$

Mathematically, it makes sense if we assume that, say, S_i are Itô processes, and θ^i are adapted processes (with respect to some Brownian motion or, more generally, with respect to some filtration $\mathbb{F} = \{\mathcal{F}_t,\, t \in [0,T]\}$). Then condition [12.3] can be rewritten in the integral form

$$V_t(\varphi) - V_0(\varphi) = \int_0^t \varphi_u \cdot \mathrm{d}X_u = \sum_{i=0}^{N} \int_0^t \theta_u^i \,\mathrm{d}S_u^i.$$

Of course, the processes θ^i must satisfy some integrability conditions that guarantee the existence of the integrals on the right-hand, for example,

$$\int_0^T |\theta_t^0| \,\mathrm{d}t < \infty, \qquad \int_0^T (\theta_t^i)^2 \,\mathrm{d}\langle S^i \rangle_t < \infty, \qquad i = 1, \ldots, N.$$

This can be motivated as follows. In discrete-time financial models, changes of the contents of a portfolio are allowed only at discrete time moments $t_0, t_1, \ldots, t_k$. In this case, the self-financing condition is written as

$$\varphi_{t_k} \cdot X_{t_{k+1}} = \varphi_{t_{k+1}} \cdot X_{t_{k+1}}.$$

It can be interpreted as follows. At each time moment t_k, the investor, taking into account known stock prices $S_{t_k}^i$, may redistribute the shares φ_{t_k} of assets of the portfolio without any receipt from, or deduction to, an outside source of money. In other words, the wealth of the portfolio at time moment t_{k+1} must remain the same as if the stock prices do not change. Relation with the continuous-time self-financing condition becomes clearer if we rewrite the last equality in the form

$$\begin{aligned} V_{t_k+1}(\varphi) - V_{t_k}(\varphi) &= \varphi_{t_{k+1}} \cdot X_{t_{k+1}} - \varphi_{t_k} \cdot X_{t_k} = \varphi_{t_k} \cdot X_{t_{k+1}} - \varphi_{t_k} \cdot X_{t_k} \\ &= \varphi_{t_k} \cdot (X_{t_{k+1}} - X_{t_k}). \end{aligned}$$

In view of equation [12.1], the self-financing condition [12.3] can be rewritten as

$$\begin{aligned} \mathrm{d}V_t(\varphi) &= \theta_t^0 \,\mathrm{d}S_t^0 + \theta_t \cdot \mathrm{d}S_t = \theta_t^0 r S_t^0 \,\mathrm{d}t + \theta_t \cdot \mathrm{d}S_t \\ &= \big(V_t(\varphi) - \theta_t \cdot S_t\big) r \,\mathrm{d}t + \theta_t \cdot \mathrm{d}S_t \\ &= rV_t(\varphi) \,\mathrm{d}t + \theta_t \cdot (-rS_t \,\mathrm{d}t + \mathrm{d}S_t). \end{aligned} \qquad [12.4]$$

Note that in the equation obtained, there is no “riskless” part $\theta^0 S^0$.

12.2.3. *Stock discount*

Denote $\widetilde{S}_t^i = S_t^i/S_t^0 = \mathrm{e}^{-rt}S_t^i$, $i = 1, \dots, N$, and $\widetilde{V}_t(\varphi) = V_t(\varphi)/S_t^0 = \mathrm{e}^{-rt}V_t(\varphi)$, where the stock and portfolio prices are discounted with respect to the bond price. Using Itô's formula, by equation [12.4] we have:

$$\begin{aligned}
\mathrm{d}\widetilde{V}_t(\varphi) &= -r\mathrm{e}^{-rt}V_t(\varphi)\,\mathrm{d}t + \mathrm{e}^{-rt}\,\mathrm{d}V_t(\varphi)\\
&= \mathrm{e}^{-rt}\big(-rV_t(\varphi)\,\mathrm{d}t + \mathrm{d}V_t(\varphi)\big)\\
&= \mathrm{e}^{-rt}\theta_t\cdot(-rS_t\,\mathrm{d}t + \mathrm{d}S_t)\\
&= \theta_t\cdot\big(S_t\,\mathrm{d}(\mathrm{e}^{-rt}) + \mathrm{e}^{-rt}\,\mathrm{d}S_t\big) = \theta_t\cdot\mathrm{d}\widetilde{S}_t
\end{aligned}$$

or, in the integral form,

$$\widetilde{V}_t(\varphi) = \widetilde{V}_0(\varphi) + \int_0^t \theta_u\cdot\mathrm{d}\widetilde{S}_u. \qquad [12.5]$$

The relation obtained shows that every self-financing strategy is completely defined by the initial portfolio wealth $v = V_0(\varphi)$ and the risky part of the portfolio, the stock shares $\theta = (\theta^1, \dots, \theta^N)$, since the bond share θ_t^0 at time t can be calculated by using the equalities

$$V_t = V_t(\varphi) = S_t^0\widetilde{V}_t(\varphi) = S_t^0\Big(v + \int_0^t \theta_u\cdot\mathrm{d}\widetilde{S}_u\Big), \qquad [12.6]$$

$$\theta_t^0 = \frac{V_t - \theta_t\cdot S_t}{S_t^0}. \qquad [12.7]$$

In other words, every self-financing strategy is defined by a pair of random processes (V, θ) taking values in $\mathbb{R}\times\mathbb{R}^N$ and satisfying the stochastic differential equation

$$\mathrm{d}V_t = rV_t\,\mathrm{d}t + \theta_t\cdot(-rS_t\,\mathrm{d}t + \mathrm{d}S_t),$$

where V is the portfolio wealth, and θ is the vector of stock shares. Then the bond share is

$$\theta^0 = (V - \theta\cdot S)/S^0.$$

Therefore, it is sometimes convenient, instead of the self-financing strategy $\varphi = (\theta^0, \theta)$, to speak of the self-financing strategy θ and to denote the wealth of the portfolio by $V_t(\theta)$, instead of $V_t(\varphi) = V_t(\theta^0, \theta)$.

EXAMPLE 12.2.– In the Black–Scholes model, the discounted stock price $\widetilde{S}$ satisfies the equation

$$\begin{aligned}
\mathrm{d}\widetilde{S}_t &= \mathrm{d}(\mathrm{e}^{-rt} S_t) = -r\,\mathrm{e}^{-rt} S_t\,\mathrm{d}t + \mathrm{e}^{-rt}\,\mathrm{d}S_t \\
&= -r\,\widetilde{S}_t\,\mathrm{d}t + \mathrm{e}^{-rt} S_t(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}B_t) \\
&= \widetilde{S}_t\big((\mu - r)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t\big).
\end{aligned}$$

Therefore, the discounted portfolio wealth under the strategy θ for stock price satisfies the equation

$$\mathrm{d}\widetilde{V}_t = \theta_t\,\mathrm{d}\widetilde{S}_t = \theta_t \widetilde{S}_t\big((\mu - r)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t\big).$$

DEFINITION 12.3.– *A self-financing strategy φ is called an arbitrage strategy if*

$$V_0(\varphi) = 0, \quad V_T(\varphi) \geqslant 0 \quad \textit{and} \quad \mathbf{P}\{V_T(\varphi) > 0\} > 0.$$

A financial market is called arbitrage-free or viable[2] *if there are no arbitrage strategies.*

In an arbitrage-free market, there is no self-financing strategy that allows getting a positive profit without risk.

EXAMPLES 12.4.– 1) For simplicity, we first give an example of arbitrage strategy on an infinite time interval $[0, T] = [0, +\infty]$. Suppose that the financial market consists of a bond without dividends, that is, $S_t^0 = 1$, $r = 0$, and the stock is $S_t = B_t$. For any $x > 0$, denote $\tau_x = \min\{t \geqslant 0 : S_t = x\}$. Consider the strategy $\theta_t := 1\!\!1_{(0,\tau_x]}(t)$. Then, starting at zero wealth $V_0 = 0$, the wealth of the portfolio at time t will be $V_t = \int_0^t 1\!\!1_{(0,\tau_x]}(u)\,\mathrm{d}B_u = B_{t\wedge\tau_x} \to x$, $t \to \infty$. Thus, $\mathbf{P}\{V_\infty = x > 0\} = 1$, and we have an arbitrage opportunity.

2) It is known that, for every positive random variable $Y \in \mathcal{F}_T$, there exists an adapted random process θ such that $\int_0^T \theta_t\,\mathrm{d}B_t = Y$.[3] Consider the financial market

2. Or we say that there are no arbitrage opportunities.

3. Do not confuse this with the representation $Y = \mathbf{E}Y + \int_0^T H_t\,\mathrm{d}Bt$ for $Y \in L^2(\Omega)$!

$(S_t^0, S_t) = (1, B_t)$ and self-financing strategy $\varphi_t = (\theta_t^0, \theta_t) = (\int_0^t \theta_s \, \mathrm{d}B_s - \theta_t B_t, \theta_t)$ (see equations [12.6]–[12.7]). Then

$$V_0(\varphi) = \theta_0^0 \cdot 1 + \theta_0 B_0 = 0,$$

but

$$V_T(\varphi) = \theta_T^0 \cdot 1 + \theta_T B_T = \left(\int_0^T \theta_s \, \mathrm{d}B_s - \theta_T B_T \right) + \theta_T B_T = Y > 0.$$

Thus, we again have an opportunity of arbitrage.

In financial mathematics, as a rule, markets with no opportunity of arbitrage are considered. This is not too restrictive, since in real markets, arbitrage opportunities are small and do not last a long time, because the market immediately reacts and moves to the equilibrium state. To assure the no-arbitrage, usually, strategies satisfying some integrability or boundedness restrictions (to be formulated below) are considered. Such strategies are called *admissible*. Their set must be *sufficiently* wide to have the possibility of calculating and realizing various derivative financial instruments but, at the same time, *not too* wide so that arbitrage strategies could not be admissible.

Thus, we further simply assume the following (no arbitrage opportunity) condition is satisfied:

(NAO) *There are no arbitrage strategies in the class of admissible strategies.*

THEOREM 12.5.– *Let condition* **(NAO)** *be satisfied. Then any two admissible strategies, having the same portfolio wealth V_T at time moment T, have the same portfolio wealth V_t at every time $t \leqslant T$.*

Proof. For simplicity, let $t = 0$. Suppose that there are two admissible strategies $\varphi = (\theta^0, \theta)$ and $\tilde{\varphi} = (\tilde{\theta}^0, \tilde{\theta})$ in a market $X = (S^0, S)$ such that $V_T(\varphi) = V_T(\tilde{\varphi})$ but $V_0(\varphi) > V_0(\tilde{\varphi})$. Consider the new self-financing strategy $\bar{\varphi}$ which consists of the strategy $\tilde{\varphi} - \varphi$ with initial wealth $V_0(\tilde{\varphi}) - V_0(\varphi) < 0$ for portfolio $X = (S^0, S)$, together with the investment of the profit $V_0(\varphi) - V_0(\tilde{\varphi})$ into bond; that is,

$$\bar{\varphi}_t := \tilde{\varphi}_t - \varphi_t + \big(V_0(\varphi) - V_0(\tilde{\varphi}), 0\big),$$

and

$$V_t(\bar{\varphi}) = (\tilde{\varphi}_t - \varphi_t) \cdot X_t + \big(V_0(\varphi) - V_0(\tilde{\varphi})\big) S_t^0.$$

Then, at time moment $t = 0$,

$$\begin{aligned} V_0(\bar{\varphi}) &= (\tilde{\varphi}_0 - \varphi_0) \cdot X_0 + \big(V_0(\varphi) - V_0(\tilde{\varphi})\big) S_0^0 \\ &= V_0(\tilde{\varphi}) - V_0(\varphi) + V_0(\varphi) - V_0(\tilde{\varphi}) = 0, \end{aligned}$$

while, at time moment $t = T$,

$$\begin{aligned} V_T(\bar{\varphi}) &= (\tilde{\varphi}_T - \varphi_T) \cdot X_T + \big(V_0(\varphi) - V_0(\tilde{\varphi})\big) S_T^0 \\ &= V_T(\tilde{\varphi}) - V_T(\varphi) + \big(V_0(\varphi) - V_0(\tilde{\varphi})\big) S_T^0 \\ &= \big(V_0(\varphi) - V_0(\tilde{\varphi})\big) S_T^0 > 0. \end{aligned}$$

Thus, we obtained the arbitrage strategy $\bar{\varphi}$, a contradiction to **(NAO)**. △

PROPOSITION 12.6.– *Suppose that the stocks $S = (S^1, S^2)$ in a two-dimensional Black–Scholes model satisfy two stochastic differential equations with the same volatility σ and the same driving Brownian motion B (i.e. affected by the same perturbations),*

$$\mathrm{d}S_t^1 = S_t^1(\mu_1\,\mathrm{d}t + \sigma\,\mathrm{d}B_t) \quad \textit{and} \quad \mathrm{d}S_t^2 = S_t^2(\mu_2\,\mathrm{d}t + \sigma\,\mathrm{d}B_t).$$

Then, in the no-arbitrage market, $\mu_1 = \mu_2$.

Proof. Suppose that, say, $\mu_1 > \mu_2$. Consider the strategy where we buy one stock S^1 and sell S_0^1/S_0^2 stocks S^2, that is, the portfolio wealth is

$$V_t = S_t^1 - \frac{S_0^1}{S_0^2} S_t^2, \quad t \in [0, T].$$

Then the initial portfolio wealth is

$$V_0 = S_0^1 - \frac{S_0^1}{S_0^2} S_0^2 = 0,$$

while the final one is

$$\begin{aligned} V_T &= S_T^1 - \frac{S_0^1}{S_0^2} S_T^2 \\ &= S_0^1 \exp\big\{\big(\mu_1 - \frac{\sigma^2}{2}\big)T + \sigma B_T\big\} - \frac{S_0^1}{S_0^2} S_0^2 \exp\big\{\big(\mu_2 - \frac{\sigma^2}{2}\big)T + \sigma B_T\big\} \\ &= S_0^1 \exp\big\{\sigma B_T - \frac{\sigma^2}{2}T\big\}\Big(\mathrm{e}^{\mu_1 T} - \mathrm{e}^{\mu_2 T}\Big) > 0, \end{aligned}$$

and so we have the arbitrage strategy, a contradiction. △

REMARK.– The proposition can be interpreted as follows: In the no-arbitrage market, there cannot be stocks affected by the same perturbations (B and σ) but with different growth tendencies (μ). Also, note that if $\sigma = 0$, i.e. S^1 and S^2 are riskless instruments, then their investment return must be the same as that of S^0, i.e. $\mu_1 = \mu_2 = r$; otherwise, we could easily construct an arbitrage as in the proposition just proved.

12.3. Option pricing problem: the Black–Scholes model

Recall that, in the Black–Scholes model, the portfolio consists of two assets, riskless asset (bond) S_t^0 and risky asset (stock) S_t. The bond price is

$$S_t^0 = S_0^0 e^{rt}, \quad t \in [0, T],$$

where $r > 0$ is the constant interest rate. It satisfies the simple ordinary differential equation $\mathrm{d}S_t^0 = rS_t^0 \,\mathrm{d}t$. The stock price S_t is a random process satisfying the stochastic differential equation $\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t$ or, in the integral form,

$$S_t = S_0 + \mu \int_0^t S_u \,\mathrm{d}u + \sigma \int_0^t S_u \,\mathrm{d}B_u, \quad t \in [0, T],$$

where the constant μ is called the mean rate of return, and $\sigma > 0$ is the volatility. The solution of the equation is the geometric Brownian motion

$$S_t = S_0 \exp\left\{(\mu - \sigma^2/2)t + \sigma B_t\right\}, \quad t \in [0, T].$$

Suppose that we know the initial stock price S_0 and that we want to buy a European call option with maturity T and exercise price K. Since we do not know the stock price S_T at time T, a natural question arises: what price we would like or would we agree to pay for this option, i.e. w hat is its true price (premium) at time moment $t = 0$?

Black and Scholes defined this price according to the following principles:

1) The investor after investment, at the initial time moment $t = 0$, has in his portfolio (S_t^0, S_t) this true price $V_0 = \theta_0^0 S_0^0 + \theta_0 S_0$. He can control this portfolio using a self-financing strategy $\theta = (\theta_t^0, \theta_t)$ so that, at maturity T, he would get the same profit $V_T = (S_T - K)^+$, as he would get by buying the option.

2) If the option were offered for any other price, we would have an opportunity of arbitrage.

Note that by Theorem 12.5 such a true option price in an arbitrage-free market must be unique.

Thus, let us try to find a self-financing strategy $\tilde{\theta} = (\theta_t^0, \theta_t)$ such that the portfolio wealth $V_t = \theta_t^0 S_t^0 + \theta_t S_t$ satisfies the *end* condition $V_T = h(S_T) := (S_T - K)^+$. Assume that the portfolio wealth can be written in the form $V_t = F(t, S_t)$, $t \in [0, T]$, with a sufficiently smooth deterministic function $F(t, x)$, $t \in [0, T]$, $x \in \mathbb{R}$.[4] By the self-financing condition, we have:

$$\mathrm{d}F(t, S_t) = \mathrm{d}V(t) = \theta_t^0 \, \mathrm{d}S_t^0 + \theta_t \, \mathrm{d}S_t.$$

On the other hand, by Itô's formula, we have:

$$\begin{aligned} \mathrm{d}F(t, S_t) &= F_t'(t, S_t)\,\mathrm{d}t + F_x'(t, S_t)\,\mathrm{d}S_t + \frac{1}{2} F_{xx}''(t, S_t)\,\mathrm{d}\langle S\rangle_t \\ &= \big(F_t'(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 F_{xx}''(t, S_t)\big)\,\mathrm{d}t + F_x'(t, S_t)\,\mathrm{d}S_t. \end{aligned}$$

From the last two equations we get that $\theta_t = F_x'(t, S_t)$ and

$$\theta_t^0 \, \mathrm{d}S_t^0 = \big(F_t'(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 F_{xx}''(t, S_t)\big)\,\mathrm{d}t.$$

Since $\mathrm{d}S_t^0 = rS_t^0\,\mathrm{d}t$, from this we get

$$\theta_t^0 S_t^0 r\,\mathrm{d}t = \big(F_t'(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 F_{xx}''(t, S_t)\big)\,\mathrm{d}t,$$

i.e.

$$\theta_t^0 S_t^0 r = F_t'(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 F_{xx}''(t, S_t). \qquad [12.8]$$

Since $F(t, S_t) = \theta_t^0 S_t^0 + \theta_t S_t = \theta_t^0 S_t^0 + F_x'(t, S_t) S_t$, substituting $\theta_t^0 S_t^0 = F(t, S_t) - F_x'(t, S_t) S_t$ into equation [12.8], we get:

$$\big(F(t, S_t) - F_x'(t, S_t) S_t\big) r = F_t'(t, S_t) + \frac{1}{2}\sigma^2 S_t^2 F_{xx}''(t, S_t).$$

Obviously, the latter equality is satisfied, provided that the function F satisfies the partial differential equation

$$r\big(xF_x'(t, x) - F(t, x)\big) + F_t'(t, x) + \frac{1}{2}\sigma^2 x^2 F_{xx}''(t, x) = 0. \qquad [12.9]$$

This equation is called the Black–Scholes partial differential equation. Note that, in this equation, the end condition $V_T = F(T, S_T) = h(S_T)$ (or $F(T, x) = h(x)$)

4. Of course, this is a serious restriction. Happily, it is not too strong.

is not reflected in any way. If a function F is its solution, then the random process $V_t = F(t, S_t)$, $t \in [0, T]$, shows the portfolio wealth with some self-financing strategy (θ^0, θ). The latter can be expressed in terms of F:

$$\theta_t^0 = \frac{F(t, S_t) - F_x'(t, S_t)S_t}{S_t^0}, \quad \theta_t = F_x'(t, S_t).$$

Another important property of equation [12.9] is the absence of the coefficient μ. This is important in practice since, in a real financial market, it is difficult to effectively estimate such parameters. Note also that to determine the initial wealth $V_0 = \theta_0^0 S_0^0 + \theta_0 S_0$ of the portfolio, we only need know the initial values of the portfolio, S_0^0 and S_0.

12.4. Black–Scholes formula

In the Black–Scholes model, to find the European call option price $V_0 = F(0, S_0)$, we have to find the solution $F = F(t, x)$ of equation [12.9] satisfying the end condition $F(t, x) = h(x) = (x - K)^+$. Happily, the equation is solvable in an explicit form; without this, it would be difficult to convince practitioners to recognize any model. We shall solve it by using stochastic analysis methods. To this end, we first reverse the time in equation [12.9] by considering the function $u(t, x) := F(T-t, x)$, $t \in [0, T]$, $x > 0$. Then

$$\begin{aligned}
u_t'(t,x) &= -F_t'(T-t,x),\\
u_x'(t,x) &= F_x'(T-t,x),\\
u_{xx}''(t,x) &= F_{xx}''(T-t,x),
\end{aligned}$$

and equation [12.9] becomes

$$u_t'(t,x) = rxu_x'(t,x) + \frac{1}{2}\sigma^2 x^2 u_{xx}''(t,x) - ru(t,x) \qquad [12.10]$$

with the *initial* condition $u(0, x) = h(x)$. For any function $f \in C^2(\mathbb{R})$, denote

$$Lf(x) := rxf'(x) + \frac{1}{2}\sigma^2 x^2 f''(x).$$

L is the generator of the solution of SDE $\mathrm{d}X_t = rX_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t$. Denote by X_t^x, $t \in [0, T]$, the solution of the latter starting at $X_0^x = x$. Then, by Corollary 10.8, the function $v(t, x) := \mathbf{E}h(X_t^x)$ solves the equation $v_t'(t, x) = Lv(t, x)$ with the initial condition $v(0, x) = h(x)$. Then we directly check that the function $u(t, x) :=$

$e^{-rt}v(t,x)$ solves the equation $u'_t(t,x) = Lu(t,x) - ru(t,x)$ (i.e. equation [12.10]) with the same initial condition $u(0,x) = h(x)$.[5]

Now we find the function $v(t,x) = \mathbf{E}h(X_t^x)$ by applying the explicit formula

$X_t^x = x\exp\{(r-\sigma^2/2)t + \sigma B_t\}$.

Let $\varphi(x) := \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$, $x \in \mathbb{R}$, denote the density of the standard normal distribution, and $\Phi(x) := \int_{-\infty}^{x}\varphi(y)\,\mathrm{d}y$, $x \in \mathbb{R}$, its distribution function. Since $B_t \sim \sqrt{t}N(0,1)$, we have:

$$\begin{aligned}
v(t,x) &= \int_{\mathbb{R}} h(x\exp\{(r-\sigma^2/2)t + \sigma\sqrt{t}y\})\varphi(y)\,\mathrm{d}y \\
&= \int_{\mathbb{R}} (x\exp\{(r-\sigma^2/2)t + \sigma\sqrt{t}y\} - K)^+\varphi(y)\,\mathrm{d}y \\
&= \int_{D_K} (x\exp\{(r-\sigma^2/2)t + \sigma\sqrt{t}y\} - K)\varphi(y)\,\mathrm{d}y,
\end{aligned}$$

where

$$\begin{aligned}
D_K : &= \{y \in \mathbb{R} : x\exp\{(r-\sigma^2/2)t + \sigma\sqrt{t}y\} - K \geqslant 0\} \\
&= \{y \in \mathbb{R} : (r-\sigma^2/2)t + \sigma\sqrt{t}y \geqslant \ln(K/x)\} \\
&= \Big\{y \in \mathbb{R} : y \geqslant \frac{\ln(K/x) - (r-\sigma^2/2)t}{\sigma\sqrt{t}}\Big\},
\end{aligned}$$

that is, $D_K = [\widetilde{K}, +\infty)$ with $\widetilde{K} = \frac{\ln(K/x)-(r-\sigma^2/2)t}{\sigma\sqrt{t}}$. Thus, continuing, we have:

$$\begin{aligned}
v(t,x) &= \int_{\widetilde{K}}^{\infty} (x\exp\{(r-\sigma^2/2)t + \sigma\sqrt{t}y\} - K)\varphi(y)\,\mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}}\int_{\widetilde{K}}^{\infty} x\exp\{(r-\sigma^2/2)t + \sigma\sqrt{t}y\}e^{-y^2/2}\,\mathrm{d}y - K\int_{\widetilde{K}}^{\infty}\varphi(y)\,\mathrm{d}y \\
&= \frac{1}{\sqrt{2\pi}}\int_{\widetilde{K}}^{\infty} x\exp\{-y^2/2 + \sigma\sqrt{t}y + (r-\sigma^2/2)t\}\,\mathrm{d}y - K\big(1-\Phi(\widetilde{K})\big) \\
&= \frac{x}{\sqrt{2\pi}}\int_{\widetilde{K}}^{\infty} \exp\Big\{-\frac{y^2 - 2\sigma\sqrt{t}y + \sigma^2 t}{2} + rt\Big\}\,\mathrm{d}y - K\Phi(-\widetilde{K})
\end{aligned}$$

5. The same result can be obtained immediately by applying the Feynman–Kac formula (Theorem 10.11).

$$
\begin{aligned}
&= \frac{x\mathrm{e}^{rt}}{\sqrt{2\pi}} \int_{\widetilde{K}}^{\infty} \mathrm{e}^{-(y-\sigma\sqrt{t})^2)/2}\,\mathrm{d}y - K\Phi(-\widetilde{K}) \\
&= \frac{x\mathrm{e}^{rt}}{\sqrt{2\pi}} \int_{\widetilde{K}-\sigma\sqrt{t}}^{\infty} \mathrm{e}^{-y^2/2}\,\mathrm{d}y - K\Phi(-\widetilde{K}) \\
&= x\mathrm{e}^{rt}\big(1-\Phi(\widetilde{K}-\sigma\sqrt{t})\big) - K\Phi(-\widetilde{K}) \\
&= x\mathrm{e}^{rt}\Phi(\sigma\sqrt{t}-\widetilde{K}) - K\Phi(-\widetilde{K}) = x\mathrm{e}^{rt}\Phi(d_1) - K\Phi(d_2),
\end{aligned}
$$

where we denoted

$$d_1 = d_1(t,x) = \sigma\sqrt{t} - \widetilde{K} = \frac{\ln(x/K) + (r+\sigma^2/2)t}{\sigma\sqrt{t}}$$

and

$$d_2 = d_2(t,x) = -\widetilde{K} = \frac{\ln(x/K) + (r-\sigma^2/2)t}{\sigma\sqrt{t}} = d_1 - \sigma\sqrt{t}.$$

Substituting the expression obtained into the equality

$$F(t,x) = u(T-t,x) = \mathrm{e}^{-r(T-t)}v(T-t,x),$$

we finally obtain:

$$F(t,x) = x\Phi\big(d_1(T-t,x)\big) - K\mathrm{e}^{-r(T-t)}\Phi\big(d_2(T-t,x)\big), \qquad [12.11]$$

$$d_1(t,x) = \frac{\ln(x/K) + (r+\sigma^2/2)t}{\sigma\sqrt{t}}, \quad d_2(t,x) = d_1 - \sigma\sqrt{t}. \qquad [12.12]$$

To emphasize that the formula obtained gives the *call* option price, the function $F(t,x)$ in equation [12.11] is commonly denoted by $C(t,x)$; moreover, when we to show its dependence on all the parameters, it is typically denoted $C(t,x,K,T,r,\sigma)$.

So, let us summarize the results just obtained:

THEOREM 12.7.– *The price of the European call option in the Black–Scholes model is*

$$V_0 = C(0,S_0) = S_0\Phi\big(d_1(T,S_0)\big) - K\mathrm{e}^{-rT}\Phi\big(d_2(T,S_0)\big), \qquad [12.13]$$

where d_1 and d_2 are the functions defined in equation [12.12]*.*

The corresponding wealth process is

$$V_t = C(t,S_t) = S_t\Phi\big(d_1(T-t,S_t)\big) - K\mathrm{e}^{-r(T-t)}\Phi\big(d_2(T-t,S_t)\big), \qquad t \in [0,T],$$

realized by the self-financing strategy (θ_t^0, θ_t), $t \in [0, T]$, *with*

$$\theta_t = C'_x(t, S_t) \quad \textit{and} \quad \theta_t^0 = \frac{C(t, S_t) - \theta_t S_t}{S_t^0}. \qquad [12.14]$$

Formula [12.13] is the famous Black–Scholes formula for call option price. Note that the option price does not depend on the mean rate of return μ (usually unknown) but does depend on the volatility σ.

The European put option price $P(t, x)$ in the Black–Scholes model is calculated very similarly, replacing the end condition $h(x) = (x - K)^+$ by the end condition $h(x) = (K - x)^+$:

$$P(t, x) = K\mathrm{e}^{-r(T-t)}\Phi\big(-d_2(T - t, x)\big) - x\Phi\big(-d_1(T - t, x)\big). \qquad [12.15]$$

Typical graphs of prices of call and put options are shown in Figures 12.1 and 12.2.

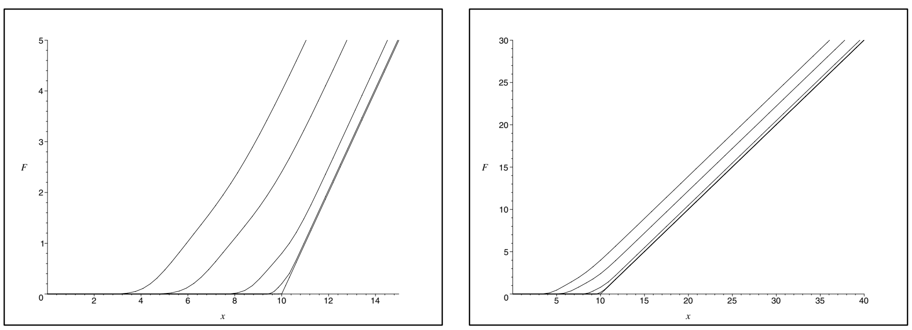


Figure 12.1. *Call option prices* $C(t, x)$ *as functions of* x *at different scales;* $K = 10$, $T = 10$, $r = 0.05$, $\sigma = 0.1$, $t = 0, 5, 9, 9.9, 10$

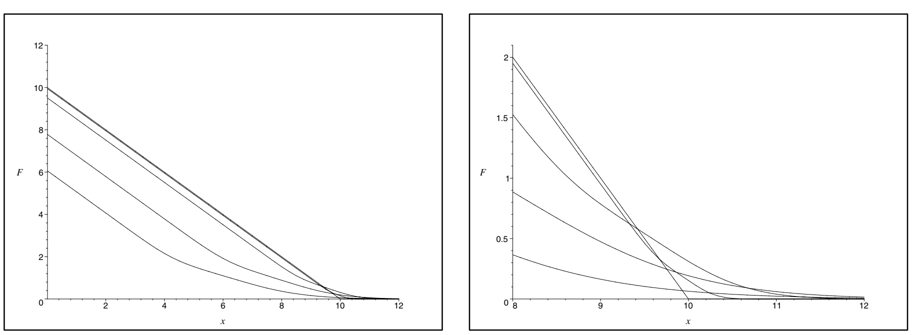


Figure 12.2. *Put option prices* $P(t, x)$ *as functions of* x *at different scales;* $K = 10$, $T = 10$, $r = 0.05$, $\sigma = 0.1$, $t = 0, 5, 9, 9.9, 10$

12.5. Risk-neutral probabilities: alternative derivation of Black–Scholes formula

Recall that, in the Black–Scholes model, the stock price satisfies the equation

$$\mathrm{d}S_t = \mu S_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t,$$

and the discounted stock price $\widetilde{S}_t = e^{-rt}S_t$ satisfies the equation

$$\mathrm{d}\widetilde{S}_t = \widetilde{S}_t((\mu - r)\,\mathrm{d}t + \sigma \,\mathrm{d}B_t).$$

Denote $\lambda = (\mu - r)/\sigma$. The number λ is called *market price of risk*.[6] Then the last equation writes

$$\mathrm{d}\widetilde{S}_t = \sigma \widetilde{S}_t(\lambda \,\mathrm{d}t + \mathrm{d}B_t)$$

Denoting by $\widetilde{B}_t = B_t + \lambda t$, $t \in [0,T]$ (the Brownian motion with drift), we get the equation

$$\mathrm{d}\widetilde{S}_t = \sigma \widetilde{S}_t \,\mathrm{d}\widetilde{B}_t.$$

At this point, we need the new notion of equivalent probabilities, which is extremely important in financial mathematics. Until now, we have "lived" in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with some fixed probability $\mathbf{P}$. Now we need to consider other probabilities on the same σ-algebra $\mathcal{F}$. A probability $\widetilde{\mathbf{P}}$ on $\mathcal{F}$ is said to be equivalent to $\mathbf{P}$ if, for all $A \in \mathcal{F}$,

$$\widetilde{\mathbf{P}}(A) > 0 \Longleftrightarrow \mathbf{P}(A) > 0,$$

that is, the events of positive probability (or, equivalently, of zero probability) are the same for both probabilities. For our purposes in this book, we only need to know a simple version of the so-called Girsanov theorem, which states that:

If B_t, $t \in [0,T]$, is a Brownian motion in the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and $\lambda \in \mathbb{R}$ is a constant, then a probability $\widetilde{\mathbf{P}}$ exists on the σ-algebra $\mathcal{F}$, which is equivalent to $\mathbf{P}$ and such that the Brownian motion with drift, $\widetilde{B}_t := B_t + \lambda t$, $t \in [0,T]$, is a Brownian motion with respect to $\widetilde{\mathbf{P}}$ (i.e. in the probability space $(\Omega, \mathcal{F}, \widetilde{\mathbf{P}})$).

So, by Girsanov's theorem, there is a probability $\widetilde{\mathbf{P}}$, equivalent to $\mathbf{P}$, such that $\widetilde{B}_t = B_t + \lambda t$, $t \in [0,T]$, is a Brownian motion with respect to $\widetilde{\mathbf{P}}$. Thus, from the last equation we see that the discounted stock price $\widetilde{S}$ is a martingale with respect to $\widetilde{\mathbf{P}}$ (or a $\widetilde{\mathbf{P}}$-martingale for short). Such a probability $\widetilde{\mathbf{P}}$ is called a risk-neutral probability or a martingale probability. In the general case, we have the following definition.

---

6. Interpretation: in a financial market, the return of a risky investment must be, on average, higher than that of a riskless one. The market price of risk is the relative (with respect to risk, σ) rate of extra return above the risk-free rate, r.

DEFINITION 12.8.– *A probability $\widetilde{\mathbf{P}}$ is called a risk-neutral probability or a martingale probability if it is equivalent to the probability $\mathbf{P}$ and all discounted stock prices $\widetilde{S}_t^i = \mathrm{e}^{-rt} S_t^i$, $t \in [0, T]$, are $\widetilde{\mathbf{P}}$-martingales.*

From the definition it is clear why $\widetilde{\mathbf{P}}$ is called a *martingale* probability. The term *risk-neutral* probability is used because the average growth rates of all (risky) stocks $S_t^i = \widetilde{S}_t^i \mathrm{e}^{rt}$ with respect to $\widetilde{\mathbf{P}}$ are equal to the interest rate r of the (riskless) bond, that is, $\widetilde{\mathbf{P}}$ is neutral with respect to risk.

So, we have shown the existence of a risk-neutral probability in the Black–Scholes model. The existence of such a probability in a discrete-time model is equivalent to no-arbitrage opportunity condition **(NAO)**. Moreover, in the discrete-time case, all self-financing strategies appear to be admissible. An analogous general fact in continuous-time models is known as *the first fundamental theorem of financial mathematics*; however, its formulation is much more subtle. To avoid this, we therefore, replace condition **(NAO)** by the close condition of the existence of risk-neutral probability:

(RNP) *In the market, there exists a risk-neutral probability.*

So, every arbitrage-free financial market likely exists in two parallel spaces, the initial "risky" probability space $(\Omega, \mathcal{F}, \mathbf{P})$, where the "true" portfolio wealth process is defined, and, at the same time, "riskless" probability space $(\Omega, \mathcal{F}, \widetilde{\mathbf{P}})$, where the average growth rates of stocks are equal to the riskless interest rate r. The first space serves for modeling price dynamics, while the second, as we shall see, serves as a tool for option pricing.

Since the wealth of a self-financing portfolio satisfies equation [12.5], i.e.

$$\widetilde{V}_t(\varphi) = \widetilde{V}_0(\varphi) + \int_0^t \theta_u \cdot \mathrm{d}\widetilde{S}_u,$$

we see that the discounted wealth $\widetilde{V}_t(\varphi)$ is an integral with respect to the $\widetilde{\mathbf{P}}$-martingale $\widetilde{S}$. Therefore, in "good" cases, it is also a $\widetilde{\mathbf{P}}$-martingale.

DEFINITION 12.9.– *A self-financing strategy φ is called $\widetilde{\mathbf{P}}$-admissible if the discounted portfolio wealth $\widetilde{V}_t(\varphi)$ is a $\widetilde{\mathbf{P}}$-martingale.*

REMARK.– In the Black–Scholes model, the discounted portfolio wealth $\widetilde{V}_t = \widetilde{V}_t(\varphi)$ satisfies the equation

$$\mathrm{d}\widetilde{V}_t = \sigma \theta_t \widetilde{S}_t \, \mathrm{d}\widetilde{B}_t.$$

A sufficient condition for $\widetilde{V}$ to be a $\widetilde{\mathbf{P}}$-martingale is $\theta\widetilde{S} \in H^2[0,T]$, i.e.

$$\widetilde{\mathbf{E}}\left[\int_0^T (\theta_t \widetilde{S}_t)^2 \,\mathrm{d}t\right] < +\infty,$$

where $\widetilde{\mathbf{E}}$ denotes the expectation with respect to $\widetilde{\mathbf{P}}$.

PROPOSITION 12.10.– *There are no-arbitrage strategies in the class of $\widetilde{\mathbf{P}}$-admissible strategies .*

Proof. Suppose, on the contrary, that a $\widetilde{\mathbf{P}}$-admissible strategy φ is an arbitrage strategy. Then $\widetilde{V}_0(\varphi) = V_0(\varphi) = 0$, $\widetilde{V}_T(\varphi) = V_T(\varphi)\mathrm{e}^{-rT} \geqslant 0$, and $\mathbf{P}\{\widetilde{V}_T(\varphi) > 0\} > 0$. Since $\widetilde{\mathbf{P}} \sim \mathbf{P}$, we also have $\widetilde{\mathbf{P}}\{\widetilde{V}_T(\varphi) > 0\} > 0$. Therefore, $\widetilde{\mathbf{E}}[\widetilde{V}_T(\varphi)] > 0$. On the other hand, since $\widetilde{V}(\varphi)$ is a $\widetilde{\mathbf{P}}$-martingale, we have $\widetilde{\mathbf{E}}[\widetilde{V}_T(\varphi)] = \widetilde{\mathbf{E}}[\widetilde{V}_0(\varphi)] = 0$, a contradiction! △

REMARK.– Another common way to avoid the opportunity of arbitrage when condition **(RNP)** is satisfied is limitation to the strategies φ for which the process $M_t = \int_0^t \varphi_s \,\mathrm{d}X_s$ is bounded from below (i.e. admissible strategies cannot lead to bankruptcy).

DEFINITION 12.11.– *A contingent claim is any non-negative ($\mathcal{F}_T$-measurable) random variable H. A self-financing strategy φ is said to replicate a contingent claim H if $V_T(\varphi) = H$. The initial portfolio wealth $V_0 = V_0(\varphi)$ under such a strategy is called a fair price of H.*

A financial market is said to be complete if every contingent claim is replicatable by an admissible self-financing strategy.

THEOREM 12.12 (Second fundamental theorem of FM).– *Arbitrage-free market is complete if and only if a risk-neutral probability is unique.*

We further consider a complete arbitrage-free market with risk-neutral probability $\widetilde{\mathbf{P}}$.

THEOREM 12.13.– *If a contingent claim H has a finite expectation, $\widetilde{\mathbf{E}}H < +\infty$, then the portfolio wealth of replicating strategy equals*

$$V_t = V_t(\varphi) = \widetilde{\mathbf{E}}\left[\mathrm{e}^{-r(T-t)} H \middle| \mathcal{F}_t\right]. \qquad [12.16]$$

The fair price of the claim is $V_0 = \widetilde{\mathbf{E}}[\mathrm{e}^{-rT} H]$.

Proof. Every discounted portfolio wealth $\widetilde{V}_t(\varphi) = \mathrm{e}^{-rt} V_t(\varphi)$ with $\widetilde{\mathbf{P}}$-admissible strategy φ is a $\widetilde{\mathbf{P}}$-martingale. If a self-financing strategy φ replicates the contingent claim H, then $V_T(\varphi) = H$. From this we get:

$$\mathrm{e}^{-rt} V_t(\varphi) = \widetilde{\mathbf{E}}\left[\mathrm{e}^{-rT} V_T(\varphi) \middle| \mathcal{F}_t\right] = \widetilde{\mathbf{E}}\left[\mathrm{e}^{-rT} H \middle| \mathcal{F}_t\right],$$

and, therefore,

$$V_t(\varphi) = \widetilde{\mathbf{E}}\left[\mathrm{e}^{-r(T-t)}H\big|\mathcal{F}_t\right]. \qquad \triangle$$

So, we know the fair price of a contingent claim. How can we realize it and find the corresponding replicating self-financing strategy $\varphi = (\theta^0, \theta)$ for which $V_T(\varphi) = H$? By Theorem 12.13 and self-financing condition [12.5], we have

$$\widetilde{\mathbf{E}}[\mathrm{e}^{-rT}H|\mathcal{F}_t] = V_t(\varphi)\mathrm{e}^{-rt} = \widetilde{V}_t(\varphi) = V_0 + \int_0^t \theta_u \,\mathrm{d}\widetilde{S}_u.$$

This is an integral equation with a stochastic integral of the unknown process θ on the right-hand side. For example, in the case of Black–Scholes model, we have the equation

$$\widetilde{\mathbf{E}}[\mathrm{e}^{-rT}H|\mathcal{F}_t] = V_0 + \sigma\int_0^t \theta_u\widetilde{S}_u \,\mathrm{d}\widetilde{B}_u.$$

How can we solve such an equation? If the integral were ordinary, we could find θ by simple differentiation with respect to t. This does not work in the case of a stochastic integral. In the case of diffusion models[7] with European contingent claims $H = h(S_T)$, the equation can be solved by means of Itô's formula. We illustrate this for the Black–Scholes model. In this case,

$$V_t = V_t(\varphi) = \widetilde{\mathbf{E}}\left[\mathrm{e}^{-r(T-t)}h(S_T)\big|\mathcal{F}_t\right]$$

with stock price S satisfying the equation

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}\widetilde{B}_t,$$

where $\widetilde{B}$ is a Brownian motion with respect to a risk-neutral probability $\widetilde{\mathbf{P}}$. We have

$$\begin{aligned}
S_T &= S_0\exp\left\{\sigma\widetilde{B}_T + (r-\sigma^2/2)T\right\}\\
&= S_0\exp\left\{\sigma\widetilde{B}_t + (r-\sigma^2/2)t\right\}\cdot\exp\left\{\sigma(\widetilde{B}_T-\widetilde{B}_t) + (r-\sigma^2/2)(T-t)\right\}\\
&= S_t\exp\left\{\sigma(\widetilde{B}_T-\widetilde{B}_t) + (r-\sigma^2/2)(T-t)\right\}.
\end{aligned}$$

Since $\widetilde{B}_T - \widetilde{B}_t \perp\!\!\!\perp S_t \in \mathcal{F}_t$ and $\widetilde{B}_T - \widetilde{B}_t \stackrel{d}{=} \xi\sqrt{T-t}$ with $\xi \sim N(0,1)$, taking the conditional expectation with respect to $\mathcal{F}_t$, we get:

$$\begin{aligned}
\widetilde{\mathbf{E}}\left[h(S_T)\big|\mathcal{F}_t\right] &= \mathbf{E}\,h\left(x\exp\left\{\sigma\xi\sqrt{T-t} + (r-\sigma^2/2)(T-t)\right\}\right)\Big|_{x=S_t}\\
&= v(T-t, S_t)
\end{aligned}$$

7. When the stock prices are diffusion processes.

with

$$\begin{aligned} v(t,x) &= \mathbf{E}\, h\left(x \exp\left\{\sigma\xi\sqrt{t} + (r-\sigma^2/2)t\right\}\right) \\ &= \int_{\mathbb{R}} h\left(x \exp\left\{\sigma y\sqrt{t} + (r-\sigma^2/2)t\right\}\right) \varphi(y)\,\mathrm{d}y; \end{aligned}$$

where, as before, $\varphi(y) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{y^2/2}$, $y \in \mathbb{R}$, is the density of the standard normal distribution.

Finally, we obtain:

$$V_t = V_t(\varphi) = \widetilde{\mathbf{E}}\left[\mathrm{e}^{-r(T-t)} h(S_T) \big| \mathcal{F}_t\right] = \mathrm{e}^{-r(T-t)} v(T-t, S_t).$$

Denoting $F(t,x) = \mathrm{e}^{-r(T-t)} v(T-t,x)$, we get the portfolio wealth under the strategy φ replicating the contingent claim $h(S_T)$:

$$V_t(\varphi) = F(t, S_t),$$

which coincides with that obtained in sections 12.3 and 12.4. Therefore, in the same way, we derive the Black–Scholes formula [12.13] for call option (and [12.15] for put option), and applying the equalities

$$V_t(\varphi) = F(t,S_t) = \theta_t^0 S_t^0 + \theta_t S_t \quad \text{ir} \quad \mathrm{d}V_t(\varphi) = \theta_t^0\,\mathrm{d}S_t^0 + \theta_t\,\mathrm{d}S_t,$$

we find the replicating strategy $\varphi = (\theta^0, \theta)$ defined by formulas [12.14].

Finally, it is worth noting the advantages of this new derivation of the Black–Scholes formula:

– We did not derive and need not solve the Black–Scholes PDE equation for the function F.

– We have found the wealth of replicating strategy by using the existence of a risk-neutral probability and formula [12.16] for the portfolio wealth under the replicating self-financing strategy.

– This approach allows us to expect success in finding the prices of options with more complicate contingent claims H.

12.6. Exercises

12.1. In the Black–Scholes model, consider the options with contingent claims $H_1 = S_T \mathbb{1}_{\{S_T>K\}}$ (asset-or-nothing option) and $H_2 = \mathbb{1}_{\{S_T>K\}}$ (cash-or-nothing option). Calculate the prices of these options.

12.2. Show the so-called put-call parity formula

$$C(t,x) - P(t,x) = x - K\mathrm{e}^{-r(T-t)}.$$

12.3. Show the following limit properties of the call option price:

$$\lim_{\sigma \downarrow 0} C(t,x,K,T,r,\sigma) = \left(x - K\mathrm{e}^{-r(T-t)}\right)^{+},$$
$$\lim_{\sigma \to +\infty} C(t,x,K,T,r,\sigma) = x.$$

Give a financial interpretation of these properties.

12.4. Consider the Black–Scholes model with varying deterministic volatility where stock price follows, under the risk-neutral probability $\widetilde{\mathbf{P}}$, the equation

$$\mathrm{d}S_t = rSt\,\mathrm{d}t + \sigma(t)S_t\,\mathrm{d}B_t, \quad S_0 = x,$$

where σ is a continuous function. Show (without calculations) that the call option price can be written in the form

$$C(0,x) = \widetilde{\mathbf{E}}(S_T - K)^{+} = x\Phi(D_1) - K\mathrm{e}^{-rT}\Phi(D_2)$$

and find D_1 and D_2.

Chapter 13

Numerical Solution of Stochastic Differential Equations

Consider the stochastic differential equation

$$X_t = x + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s, \quad t \geqslant 0. \tag{13.1}$$

As we know, in an explicit form such an equation is rarely solvable. It is natural that, as in the case of ordinary differential equations, we need numerical methods enabling us to use computers to solve SDEs. However, what is meant by an approximate solution of an SDE the solution of which is a *random* process?

We typically proceed as follows. Consider a discretization of a fixed time interval $[0, T]$:

$$0 = t_0 < t_1 < t_2 < \cdots < t_N = T,$$

with fixed time step $h = T/N$. For all such h, we construct discrete-time random processes X^h_{kh}, $k = 0, 1, 2, \ldots, N$, that depend on the values of Brownian motion B at discretization moments, i.e. on B_{kh}, $k = 0, 1, 2, \ldots, N$, and that "as well as possible" approximate the exact solution X of the equation as $h = T/N \to 0$. It is convenient to extend the processes X^h_t to the whole time interval $t \in [0, T]$. For example, we can set

$$X^h_t := X^h_{[t/h]h}, \quad t \in [kh, (k+1)h)$$

(step process), or

$$X^h_t := X^h_{kh} + \bigl(X_{(k+1)h} - X_{kh}\bigr)\frac{t - kh}{h}, \quad t \in [kh, (k+1)h)$$

(polygonal line).

How can we estimate the closeness of X^h and X? There are two different ways that this can be done.

DEFINITION 13.1.– *$\{X^h\}$ is said to be an nth-order strong approximation of X if, for all $t \in [0, T]$,*

$$\mathbf{E}\big|X_t^h - X_t\big| = \mathrm{O}(h^n), \quad h \to 0.$$[1]

$\{X^h\}$ is said to be an nth-order weak approximation of X if, for all $t \in [0, T]$,

$$\mathbf{E}f\big(X_t^h\big) - \mathbf{E}f(X_t) = \mathrm{O}(h^n), \quad h \to 0,$$

for a sufficiently wide class of functions $f : \mathbb{R} \to \mathbb{R}$ (for example, $C_b^\infty(\mathbb{R})$).

Having a theoretical approximation X^h, we can make (or generate) and observe its (random) trajectories on a computer screen or printout. Since X^h, in contrast to the true solution X, depends only on the values of Brownian motion at time moments kh, $k = 1, 2, \ldots$, we need to know how to computer generate the random variables B_{kh}. This is not a difficult task, since $B_{kh} = \sum_{i=0}^{k-1}(B_{(i+1)h} - B_{ih})$, and the increments $\Delta B_i = B_{(i+1)h} - B_{ih} \sim N(0, h)$ are independent random variables. Therefore, we only have to generate a sequence of independent random variables $\xi_i \sim N(0, 1)$, $i \in \mathbb{N}$, and then take $\Delta B_i := \xi_i\sqrt{h}$. Mathematical software often offers the possibility of generating such a sequence $\{\xi_i\}$. However, some programming languages only have functions or procedures that generate uniformly distributed random variables, for example, the RANDOM function in the programming language PASCAL. In such a case, we can use methods of generating normal variables from those that are uniformly distributed. One of the simplest is Box–Muller[2] method. It is based on the fact that if U_1 and U_2 are two independent random variables uniformly distributed in the interval $(0, 1)$, then

$$\xi_1 := \sqrt{-2\ln U_1}\cos(2\pi U_2) \quad \text{and} \quad \xi_2 := \sqrt{-2\ln U_1}\sin(2\pi U_2)$$

are two independent standard normal random variables.

13.1. Memories of approximations of ordinary differential equations

First consider the ordinary differential equation

$$X_t' = b(t, X_t), \qquad X_0 = x, \quad \text{or} \quad X_t = x + \int_0^t b(s, X_s)\,\mathrm{d}s, \quad t \in [0, T].$$

---

1. Recall that $g(h) = \mathrm{O}(h^n)$, $h \to 0$, means that there are $h_0 > 0$ and $C > 0$ such that $|g(h)| \leqslant Ch^n$ for $0 < |h| \leqslant h_0$ (in our case, $0 < h \leqslant h_0$ since $h > 0$).

2. George Edward Pelham Box and Mervin Edgar Muller.

Denoting $h = T/N$, $t_k = kh$, we have:

$$X_0 = x, \qquad X_{t_{k+1}} = X_{t_k} + \int_{t_k}^{t_{k+1}} b(s, X_s)\,\mathrm{d}s, \quad k = 0, 1, \ldots, N-1.$$

Let us approximate the integral by using the rectangle formula:

$$X_{t_{k+1}} \approx X_{t_k} + \int_{t_k}^{t_{k+1}} b(t_k, X_{t_k})\,\mathrm{d}s = X_{t_k} + b(t_k, X_{t_k})\,h, \quad k = 0, 1, \ldots, N-1.$$

This approximate equality prompts us to use the simplest, Euler[3] approximation:

$$X_0^h = x, \qquad X_{t_{k+1}}^h = X_{t_k}^h + b\big(t_k, X_{t_k}^h\big)h, \quad k = 0, 1, \ldots, N-1.$$

It is known that the Euler approximation is a first-order one:

$$\sup_{t \leqslant T} \big|X_t^h - X_t\big| = \mathrm{O}(h), \quad h \to 0.$$

Replacing the rectangle formula by the trapezoidal one, we get:

$$X_{t_{k+1}} \approx X_{t_k} + \frac{1}{2}\big[b(t_k, X_{t_k}) + b(t_{k+1}, X_{t_{k+1}})\big]h, \quad k = 0, 1, \ldots, N-1,$$

which prompts the trapezoidal approximation

$$X_{t_{k+1}}^h = X_{t_k}^h + \frac{1}{2}\big[b\big(t_k, X_{t_k}^h\big) + b\big(t_{k+1}, X_{t_{k+1}}^h\big)\big]h, \quad k = 0, 1, \ldots, N-1.$$

However, this scheme has an inconvenience, since at every step, we have to solve an equation with respect to $X_{t_{k+1}}^h$.[4] This inconvenience can be avoided by replacing the value of $X_{t_{k+1}}^h$ by its approximate value borrowed from the Euler method. So, we obtain the *modified trapezoidal* approximation

$$\begin{aligned}
X_{t_{k+1}}^h &= X_{t_k}^h + \frac{1}{2}\big[b\big(t_k, X_{t_k}^h\big) + b\big(t_{k+1}, \overline{X}_{t_{k+1}}^h\big)\big]h, \\
\overline{X}_{t_{k+1}}^h &= X_{t_k}^h + b\big(t_k, X_{t_k}^h\big)h, \quad k = 0, 1, \ldots, N-1.
\end{aligned}$$

This approximation is also called the improved Euler or Heun[5] approximation. It has second-order accuracy:

$$\sup_{t \leqslant T} \big|X_t^h - X_t\big| = \mathrm{O}(h^2), \quad h \to 0.$$

3. Johann Albrecht Euler.
4. Such schemes are called implicit.
5. Karl Ludwig Wilhelm Heun.

Another way to improve the accuracy order is using Taylor's formula. Applying it to the solution X in the interval $[t_k, t_{k+1}]$, we get:

$$\begin{aligned} X_{t_{k+1}} &= X_{t_k} + X'_{t_k} h + \frac{1}{2} X''_{t_k} h^2 + \cdots + \frac{1}{n!} X^{(n)}_{t_k} h^n + R_{nk}, \\ R_{nk} &= \frac{1}{(n+1)!} X^{(n+1)}_{\theta_{nk}} h^{n+1}, \quad t_k \leqslant \theta_{nk} \leqslant t_{k+1}. \end{aligned}$$

Iteratively differentiating the equation $X'_t = b(t, X_t)$, we get:

$$\begin{aligned} X''_t &= b'_t(t, X_t) + b'_x(t, X_t) X'_t = b'_t(t, X_t) + b'_x(t, X_t) b(t, X_t) \\ &= (b'_t + b b'_x)(t, X_t), \\ X'''_t &= \big(b''_{tt} + 2b''_{tx} b + b''_{xx} b^2 + b'_t b'_x + b {b'_x}^2\big)(t, X_t), \end{aligned}$$

and so on. Substituting these expressions into Taylor's formula and discarding the remainder term R_n, we get the so-called nth-order Taylor's approximation, and it is, indeed, the nth-order approximation.[6] For example, the first-order Taylor approximation coincides with the Euler approximation, the second-order Taylor's approximation is defined by

$$X^h_{t_{k+1}} = X^h_{t_k} + b\big(t_k, X^h_{t_k}\big) h + \frac{1}{2!} \big[b'_t + b b'_x\big] \big(t_k, X^h_{t_k}\big) h^2,$$

the third-order one is defined by

$$\begin{aligned} X^h_{t_{k+1}} = X^h_{t_k} &+ b\big(t_k, X^h_{t_k}\big) h + \frac{1}{2!} \big[b'_t + b b'_x\big] \big(t_k, X^h_{t_k}\big) h^2 \\ &+ \frac{1}{3!} \big[b''_{tt} + 2b''_{tx} b + b''_{xx} b^2 + b'_t b'_x + b {b'_x}^2\big] \big(t_k, X^h_{t_k}\big) h^3, \end{aligned}$$

and so on. Taylor approximations are particularly simple if the coefficient b does not depend on t. For example, the fourth-order approximation is defined by

$$X^h_{t_{k+1}} = X^h_{t_k} + bh + \frac{1}{2!} b b' h^2 + \frac{1}{3!} b (b b')' h^3 + \frac{1}{4!} b \big(b (b b')'\big)' h^4,$$

where the function b and all its derivatives are taken at the point $(t_k, X^h_{t_k})$.

6. For deterministic equations, the notions of strong and weak approximations are equivalent.

13.2. Euler approximation

The analog of the Euler approximation for the SDE [13.1] is the Euler (or Euler–Maruyama[7]) approximation defined by

$$X_0^h = x, \qquad X_{t_{k+1}}^h = X_{t_k}^h + b\big(t_k, X_{t_k}^h\big)h + \sigma\big(t_k, X_{t_k}^h\big)\Delta B_k,$$
$$\Delta B_k = \Delta B_k^h = B_{t_{k+1}} - B_{t_k}, \quad t_k = kh.$$

It is convenient to extend it to the whole interval $[0,T]$ as follows:

$$X_0^h = x,$$
$$X_t^h = X_{t_k}^h + b\big(t_k, X_{t_k}^h\big)(t - t_k) + \sigma\big(t_k, X_{t_k}^h\big)(B_t - B_{t_k}), \quad t \in [t_k, t_{k+1}].$$

THEOREM 13.2.– *Suppose that the coefficients of equation* [13.1] *satisfy the Lipschitz condition*

$$\big|b(t,x) - b(s,y)\big|^2 + \big|\sigma(t,x) - \sigma(s,y)\big|^2 \leqslant C\big(|x-y|^2 + |t-s|\big).$$

Then

$$\sup_{t\leqslant T} \mathbf{E}\big|X_t^h - X_t\big| = \mathrm{O}(h^{1/2}), \quad h \to 0,$$

that is, the Euler approximation (*as a strong approximation*) *has the order* $\frac{1}{2}$.

Proof. The order of convergence essentially depends on the "worse" diffusion part of the equation. Therefore, for simplicity, we suppose that the "better" drift part equals zero, that is, $b \equiv 0$. The Euler approximation satisfies the equation with "delaying" coefficient,

$$X_t^h = x + \int_0^t \sigma^h\big(s, X^h\big)\,\mathrm{d}B_s,$$

where $\sigma^h(s,X) := \sigma(t_k, X_{t_k})$, $s \in [t_k, t_{k+1})$. First, let us estimate the difference of the coefficients. For $s \in [t_k, t_{k+1})$ and all $h > 0$, we have:

$$\begin{aligned}\big|\sigma^h\big(s,X^h\big) - \sigma\big(s,X_s\big)\big|^2 &= \big|\sigma\big(t_k, X_{t_k}^h\big) - \sigma(s,X_s)\big|^2\\ &\leqslant C\big(\big|X_{t_k}^h - X_s\big|^2 + |t_k - s|\big)\\ &\leqslant C\big(2\big|X_{t_k}^h - X_{t_k}\big|^2 + 2|X_{t_k} - X_s|^2 + h\big).\end{aligned}$$

7. Gishiro Maruyama.

By taking the expectations we have:

$$\mathbf{E}|\sigma^h(s,X^h)-\sigma(s,X_s)|^2$$
$$\leqslant C\big(2\mathbf{E}|X^h_{t_k}-X_{t_k}|^2+2\mathbf{E}|X_{t_k}-X_s|^2+h\big),\quad s\in[t_k,t_{k+1}]:$$

Denoting

$$\varphi^h(t):=\sup_{s\leqslant t}\mathbf{E}\big|X^h_s-X_s\big|^2,\quad t\in[0,T],$$
$$\psi(h):=\sup_{|s-u|\leqslant h}\mathbf{E}|X_s-X_u|^2,\quad h>0,$$

from the last inequality we have

$$\mathbf{E}\big|\sigma^h\big(s,X^h\big)-\sigma(s,X_s)\big|^2\leqslant 2C\big(\varphi^h(s)+\psi(h)+h\big),\quad s\in[0,T],\ h>0.$$

Therefore,

$$\varphi^h(t)=\sup_{s\leqslant t}\mathbf{E}\big|X^h_s-X_s\big|^2=\sup_{s\leqslant t}\mathbf{E}\left(\int_0^s\big[\sigma^h\big(u,X^h\big)-\sigma(u,X_u)\big]\,\mathrm{d}B_u\right)^2$$
$$=\sup_{s\leqslant t}\mathbf{E}\int_0^s\big(\sigma^h\big(u,X^h\big)-\sigma(u,X_u)\big)^2\,\mathrm{d}u$$
$$\leqslant 2C\int_0^t\big(\varphi^h(u)+\psi(h)+h\big)\,\mathrm{d}u$$
$$\leqslant 2C\int_0^t\varphi^h(u)\,\mathrm{d}u+2CT\big(\psi(h)+h\big),\quad t\in[0,T].$$

Using the Gronwall lemma (Lemma 6.3), we get:

$$\varphi^h(T)\leqslant 2CT\big(\psi(h)+h\big)\mathrm{e}^{2CT},\quad h>0.$$

Noting that

$$\mathbf{E}|X_s-X_u|^2=\mathbf{E}\left(\int_u^s\sigma(v,X_v)\,\mathrm{d}B_v\right)^2$$
$$=\int_u^s\mathbf{E}\sigma^2(v,X_v)\,\mathrm{d}v\leqslant\mathrm{const}\cdot(s-u),\quad 0\leqslant u\leqslant s,$$

we therefore have:

$$\psi(h) = \sup_{|s-u|\leqslant h} \mathbf{E}|X_s - X_u|^2 = \mathrm{O}(h), \quad h \to 0.$$

Hence,

$$\varphi^h(T) = \mathrm{O}(h), \quad h \to 0$$
$$\implies \sup_{t\leqslant T} \mathbf{E}\big|X_t^h - X_t\big| \leqslant \big(\varphi^h(T)\big)^{1/2} = \mathrm{O}\big(h^{1/2}\big), \quad h \to 0. \qquad \triangle$$

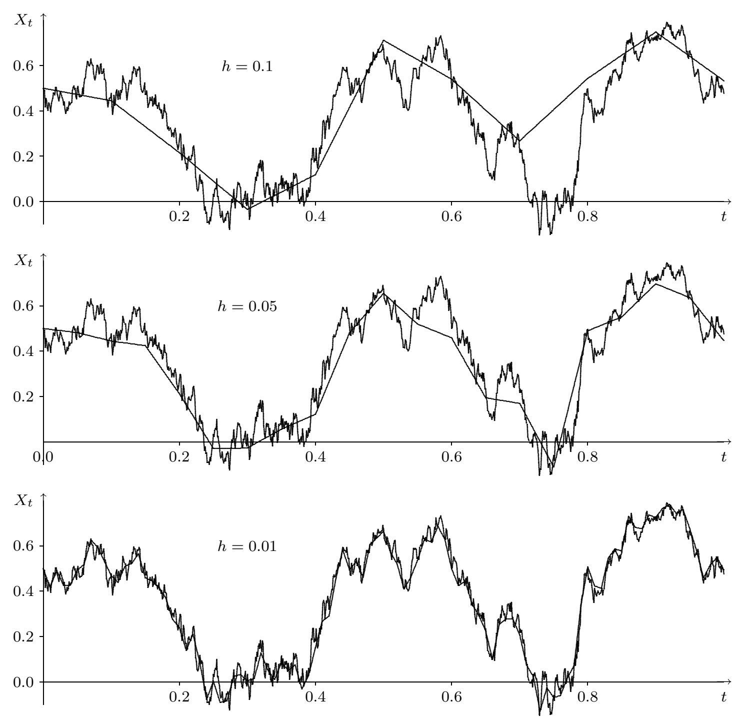


Figure 13.1. *Euler approximations with various steps*

In Figure 13.1, we see a trajectory of the true solution of the stochastic differential equation

$$\mathrm{d}X_t = -\sin X_t \cos^3 X_t \,\mathrm{d}t + \cos^2 X_t \,\mathrm{d}B_t, \qquad X_0 = 1/2,$$

and its Euler approximations for various steps h. The equation is chosen for two reasons. First, we know the true solution $X_t = \operatorname{arctg}(\operatorname{tg} X_0 + B_t)$, which, in particular, is bounded; second, its coefficients are rather complicated. In the figure, we clearly see that the smaller step h, the closer the approximation to the true solution; for sufficiently small $h = 0.01$, the Euler approximation is visually hard to distinguish from the true solution.

13.3. Higher-order strong approximations

We now try to improve the approximation order by using an analog of the Taylor formula, the Itô–Taylor formula for stochastic differential equations. For simplicity, we consider the time-homogeneous equation. We have already proved (see the proof of Proposition 10.6) the formula

$$f(X_t) = f(X_{t_k}) + \int_{t_k}^{t} Af(X_s)\,\mathrm{d}s + \int_{t_k}^{t} Sf(X_s)\,\mathrm{d}B_s, \quad t \geqslant t_k, \qquad [13.2]$$

where $Af = bf' + \frac{1}{2}\sigma^2 f''$, $Sf = \sigma f'$. In particular, taking $f(x) = x$, we have $Af = b$, $Sf = \sigma$, and equation [13.2] becomes the initial Itô equation

$$X_t = X_{t_k} + \int_{t_k}^{t} b(X_s)\,\mathrm{d}s + \int_{t_k}^{t} \sigma(X_s)\,\mathrm{d}B_s, \qquad t \geqslant t_k.$$

Similarly to the case of deterministic equation, applying formula [13.2] to the functions $f = b$ and $f = \sigma$, we get:

$$\begin{aligned} X_t &= X_{t_k} + \int_{t_k}^{t} \left(b(X_{t_k}) + \int_{t_k}^{s} Ab(X_u)\,\mathrm{d}u + \int_{t_k}^{s} Sb(X_u)\,\mathrm{d}B_u \right) \mathrm{d}s \\ &\quad + \int_{t_k}^{t} \left(\sigma(X_{t_k}) + \int_{t_k}^{s} A\sigma(X_u)\,\mathrm{d}u + \int_{t_k}^{s} S\sigma(X_u)\,\mathrm{d}B_u \right) \mathrm{d}B_s \\ &= X_{t_k} + b(X_{t_k}) \int_{t_k}^{t} \mathrm{d}s + \sigma(X_{t_k}) \int_{t_k}^{t} \mathrm{d}B_s + R \end{aligned}$$

with the remainder term

$$\begin{aligned} R &= \int_{t_k}^{t}\int_{t_k}^{s} Ab(X_u)\,\mathrm{d}u\,\mathrm{d}s + \int_{t_k}^{t}\int_{t_k}^{s} Sb(X_u)\,\mathrm{d}B_u\,\mathrm{d}s \\ &\quad + \int_{t_k}^{t}\int_{t_k}^{s} A\sigma(X_u)\,\mathrm{d}u\,\mathrm{d}B_s + \int_{t_k}^{t}\int_{t_k}^{s} S\sigma(X_u)\,\mathrm{d}B_u\,\mathrm{d}B_s. \end{aligned}$$

This is the simplest Itô–Taylor (IT) formula. It can be extended by using formula [13.2] to the integrand functions of the remainder term. For example, applying formula [13.2] to the integrand $f = S\sigma$ of the last ("worst") summand, we get the Itô–Taylor formula

$$X_t = X_{t_k} + b(X_{t_k}) \int_{t_k}^{t} \mathrm{d}s + \sigma(X_{t_k}) \int_{t_k}^{t} \mathrm{d}B_s + S\sigma(X_{t_k}) \int_{t_k}^{t}\int_{t_k}^{s} \mathrm{d}B_u \,\mathrm{d}B_s + R_1$$

with the remainder term

$$\begin{aligned} R_1 &= \int_{t_k}^{t}\int_{t_k}^{s} Ab(X_u)\,\mathrm{d}u\,\mathrm{d}s + \int_{t_k}^{t}\int_{t_k}^{s} Sb(X_u)\,\mathrm{d}B_u\,\mathrm{d}s \\ &\quad + \int_{t_k}^{t}\int_{t_k}^{s} A\sigma(X_u)\,\mathrm{d}u\,\mathrm{d}B_s + \int_{t_k}^{t}\int_{t_k}^{s}\int_{t_k}^{u} AS\sigma(X_v)\,\mathrm{d}v\,\mathrm{d}B_u\,\mathrm{d}B_s \\ &\quad + \int_{t_k}^{t}\int_{t_k}^{s}\int_{t_k}^{u} S^2\sigma(X_v)\,\mathrm{d}B_v\,\mathrm{d}B_u\,\mathrm{d}B_s =: R_{11} + R_{12} + R_{13} + R_{14} + R_{15}. \end{aligned}$$

The remainder can be estimated as follows:

$$\mathbf{E}R_1^2 \leqslant C(t - t_k)^3.$$

For example,

$$\begin{aligned} \mathbf{E}R_{15}^2 &= \mathbf{E}\left(\int_{t_k}^{t}\int_{t_k}^{s}\int_{t_k}^{u} S^2\sigma(X_v)\,\mathrm{d}B_v\,\mathrm{d}B_u\,\mathrm{d}B_s \right)^2 \\ &= \int_{t_k}^{t} \mathbf{E}\left(\int_{t_k}^{s}\int_{t_k}^{u} S^2\sigma(X_v)\,\mathrm{d}B_v\,\mathrm{d}B_u \right)^2 \mathrm{d}s \\ &= \int_{t_k}^{t}\int_{t_k}^{s} \left(\mathbf{E} \int_{t_k}^{u} S^2\sigma(X_v)\,\mathrm{d}B_v \right)^2 \mathrm{d}u\,\mathrm{d}s = \int_{t_k}^{t}\int_{t_k}^{s}\int_{t_k}^{u} \mathbf{E}\big(S^2\sigma(X_v)\big)^2\,\mathrm{d}v\,\mathrm{d}u\,\mathrm{d}s \\ &\leqslant \sup_{v\in[0,T]} \mathbf{E}\big(\sigma(\sigma\sigma')'(X_v)\big)^2 \int_{t_k}^{t}\int_{t_k}^{s}\int_{t_k}^{u} \mathrm{d}v\,\mathrm{d}u\,\mathrm{d}s = C\frac{(t - t_k)^3}{6}. \end{aligned}$$

Taking $t = t_{k+1}$, we have:

$$X_{t_{k+1}} = X_{t_k} + b(X_{t_k})h + \sigma(X_{t_k})\Delta B_k + \frac{1}{2}\sigma\sigma'(X_{t_k})(\Delta B_k^2 - h) + R_1,$$

$$\Delta B_k = B_{t_{k+1}} - B_{t_k}, \qquad \mathbf{E}R_1^2 \leqslant Ch^3.$$

Discarding the remainder, we get the approximation

$$\begin{aligned} X_{t_{k+1}}^h &= X_{t_k}^h + b\big(X_{t_k}^h\big)h + \sigma\big(X_{t_k}^h\big)\Delta B_k + \frac{1}{2}\sigma\sigma'\big(X_{t_k}^h\big)\big(\Delta B_k^2 - h\big) \\ &= X_{t_k}^h + \Big(b - \frac{1}{2}\sigma\sigma'\Big)\big(X_{t_k}^h\big)h + \sigma\big(X_{t_k}^h\big)\Delta B_k + \frac{1}{2}\sigma\sigma'\big(X_{t_k}^h\big)\Delta B_k^2. \end{aligned}$$

It is called the Milstein[8] approximation, and its order is 1. In Figure 13.2, we compare the Euler and approximations of the same equation, with two different steps, $h = 0.1$ and 0.05 (for smaller values of h, it would be visually difficult to distinguish them). The equation is the same as in Figure 13.1. From several simulated trajectories of the solution we have chosen those for which the differences of approximations are better seen. We can notice that the larger the error of the Euler approximation, the clearer the accuracy of the Milstein approximation.

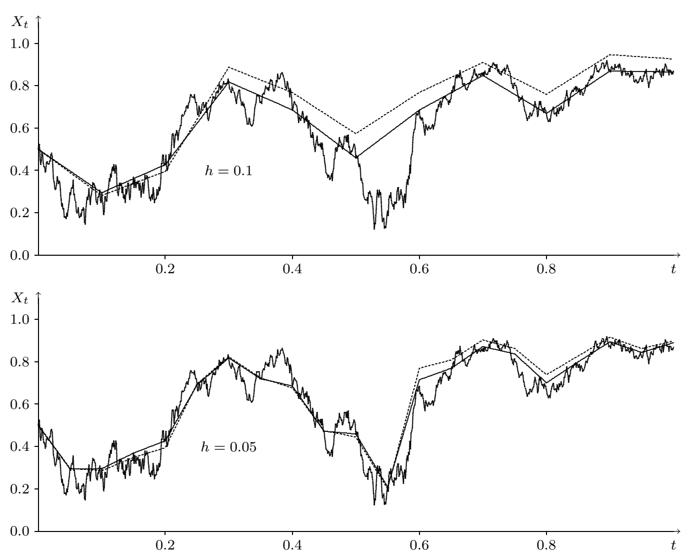


Figure 13.2. *Comparison of Euler (dashed line) and Milstein (solid line) approximations*

8. Grigori N. Milstein.

Let us try to further improve the approximation order by specifying the "worst" summands of the remainder $R_1 = \sum_{i=1}^{5} R_{1i}$. For $t = t_{k+1}$, $\mathbf{E}R_{11}^2 = \mathrm{O}(h^4)$ and $\mathbf{E}R_{14}^2 = \mathrm{O}(h^4)$; therefore, R_{12}, R_{13}, and R_{15} with the second moments of order $\mathrm{O}(h^3)$, are the "worse" ones. Again, applying Itô's formula to the integrands of these summands, we get:

$$R_{12} = \int_{t_k}^{t_{k+1}}\int_{t_k}^{s} Sb(X_u)\,\mathrm{d}B_u\,\mathrm{d}s = Sb(X_{t_k})\int_{t_k}^{t_{k+1}}\int_{t_k}^{s}\mathrm{d}B_u\,\mathrm{d}s + R_{22},$$

$$R_{13} = \int_{t_k}^{t_{k+1}}\int_{t_k}^{s} A\sigma(X_u)\,\mathrm{d}u\,\mathrm{d}B_s = A\sigma(X_{t_k})\int_{t_k}^{t_{k+1}}\int_{t_k}^{s}\mathrm{d}u\,\mathrm{d}B_s + R_{23},$$

$$\begin{aligned} R_{15} &= \int_{t_k}^{t_{k+1}}\int_{t_k}^{s}\int_{t_k}^{u} S^2\sigma(X_v)\,\mathrm{d}B_v\,\mathrm{d}B_u\,\mathrm{d}B_s \\ &= S^2\sigma(X_{t_k})\int_{t_k}^{t_{k+1}}\int_{t_k}^{s}\int_{t_k}^{u}\mathrm{d}B_v\,\mathrm{d}B_u\,\mathrm{d}B_s + R_{25} \end{aligned}$$

with estimates $\mathbf{E}R_{2i}^2 = \mathrm{O}(h^4)$, $i = 2, 3, 5$. Denoting the integral

$$\Delta Z_k = \int_{t_k}^{t_{k+1}}\int_{t_k}^{s} dB_u\,\mathrm{d}s = \int_{t_k}^{t_{k+1}}(B_s - B_{t_k})\,\mathrm{d}s,$$

the other two integrals can be expressed in terms of ΔZ_k and ΔB_k:

$$\begin{aligned} \int_{t_k}^{t_{k+1}}\int_{t_k}^{s} du\,\mathrm{d}B_s &= \int_{t_k}^{t_{k+1}}(s - t_k)\,\mathrm{d}B_s \\ &= (s - t_k)(B_s - B_{t_k})\Big|_{s=t_k}^{t_{k+1}} - \int_{t_k}^{t_{k+1}}(B_s - B_{t_k})\,\mathrm{d}s \\ &= \Delta B_k h - \Delta Z_k; \end{aligned}$$

$$\int_{t_k}^{t_{k+1}}\int_{t_k}^{s}\int_{t_k}^{u}\mathrm{d}B_v\,\mathrm{d}B_u\,\mathrm{d}B_s = \int_{t_k}^{t_{k+1}}\int_{t_k}^{s}(B_u - B_{t_k})\,\mathrm{d}B_u\,\mathrm{d}B_s$$

$$\stackrel{4.10}{=\!=} \int_{t_k}^{t_{k+1}} \Big[\frac{(B_s - B_{t_k})^2}{2} - \frac{s-t_k}{2}\Big]\,\mathrm{d}B_s$$

$$\stackrel{5.3}{=\!=} \frac{\Delta B_k^3}{6} - \frac{1}{2}\int_{t_k}^{t_{k+1}} (B_s - B_{t_k})\,\mathrm{d}s - \frac{1}{2}\int_{t_k}^{t_{k+1}} (s-t_k)\,\mathrm{d}B_s$$

$$= \frac{\Delta B_k^3}{6} - \frac{1}{2}\Delta B_k h.$$

Thus,

$$\begin{aligned} X_{t_{k+1}} &= X_{t_k} + b(X_{t_k})h + \sigma(X_{t_k})\Delta B_k + \frac{1}{2}\sigma\sigma'(X_{t_k})\big(\Delta B_k^2 - h\big) \\ &\quad + Sb(X_{t_k})\Delta Z_k + A\sigma(X_{t_k})(\Delta B_k h - \Delta Z_k) \\ &\quad + S^2\sigma(X_{t_k})\Big(\frac{1}{6}\Delta B_k^3 - \frac{1}{2}\Delta B_k h\Big) + R_2 \end{aligned}$$

with estimate $\mathbf{E}R_2^2 = \mathrm{O}(h^4)$. Discarding the remainder term, we get the approximation

$$\begin{aligned} X_{t_{k+1}}^h &= X_{t_k}^h + b\big(X_{t_k}^h\big)h + \sigma\big(X_{t_k}^h\big)\Delta B_k + \frac{1}{2}\sigma\sigma'\big(X_{t_k}^h\big)\big(\Delta B_k^2 - h\big) \\ &\quad + Sb\big(X_{t_k}^h\big)\Delta Z_k + A\sigma\big(X_{t_k}^h\big)(\Delta B_k h - \Delta Z_k) \\ &\quad + S^2\sigma\big(X_{t_k}^h\big)\Big(\frac{1}{6}\Delta B_k^3 - \frac{1}{2}\Delta B_k h\Big) \\ &= X_{t_k}^h + \Big(b - \frac{1}{2}\sigma\sigma'\Big)\big(X_{t_k}^h\big)h + \sigma\big(X_{t_k}^h\big)\Delta B_k + \frac{1}{2}\sigma\sigma'\big(X_{t_k}^h\big)\Delta B_k^2 \\ &\quad + \frac{1}{6}S^2\sigma\big(X_{t_k}^h\big)\Delta B_k^3 + \Big(A\sigma - \frac{1}{2}S^2\sigma\Big)\big(X_{t_k}^h\big)\Delta B_k h \\ &\quad + (Sb - A\sigma)\big(X_{t_k}^h\big)\Delta Z_k. \end{aligned}$$

We could expect the order of this approximation to be better and equal $\frac{3}{2}$. However, for this, we have to complement it by an additional term obtained by specifying yet one summand of the remainder term:

$$R_{11} = \int_{t_k}^{t_{k+1}}\int_{t_k}^{s} Ab(X_u)\,\mathrm{d}u\,\mathrm{d}s = \frac{1}{2}Ab(X_{t_k})h^2 + R_{22}$$

with estimate $R_{22} = \mathrm{O}(h^5)$. So, we get the following strong approximation of order $\frac{3}{2}$:

$$\begin{aligned} X^h_{t_{k+1}} &= X^h_{t_k} + \Big(b - \frac{1}{2}\sigma\sigma'\Big)\big(X^h_{t_k}\big)h + \frac{1}{2}Ab(X_{t_k})h^2 \\ &\quad + \sigma\big(X^h_{t_k}\big)\Delta B_k + \frac{1}{2}\sigma\sigma'\big(X^h_{t_k}\big)\Delta B_k^2 + \frac{1}{6}S^2\sigma\big(X^h_{t_k}\big)\Delta B_k^3 \\ &\quad + \Big(A\sigma - \frac{1}{2}S^2\sigma\Big)\big(X^h_{t_k}\big)\Delta B_k h + (Sb - A\sigma)\big(X^h_{t_k}\big)\Delta Z_k. \end{aligned}$$

Note that, in this approximation, the random variables ΔZ_k "participate" that cannot be expressed by the values of Brownian motion B at discretization points t_k. Therefore, at each step, we have to generate a pair of random variables, $(\Delta B_k, \Delta Z_k)$. Though not being independent, they can easily be generated. ΔZ_k is a normal random variable with mean $\mathbf{E}\Delta Z_k = 0$ and variance

$$\begin{aligned} \mathbf{E}\Delta Z_k^2 &= \mathbf{E}\Delta Z_1^2 = \mathbf{E}\bigg(\int_0^h B_s\,\mathrm{d}s\bigg)^2 = \mathbf{E}\bigg(\int_0^h B_s\,\mathrm{d}s \cdot \int_0^h B_u\,\mathrm{d}u\bigg) \\ &= \int_0^h\int_0^h \mathbf{E}(B_sB_u)\,\mathrm{d}s\,\mathrm{d}u = \int_0^h\int_0^h (s\wedge u)\,\mathrm{d}s\,\mathrm{d}u \\ &= \int_0^h\bigg(\int_0^u s\,\mathrm{d}s + \int_u^h u\,\mathrm{d}s\bigg)\mathrm{d}u = \int_0^h\Big(\frac{u^2}{2} + u(h-u)\Big)\mathrm{d}u \\ &= \int_0^h\Big(hu - \frac{u^2}{2}\Big)\mathrm{d}u = \frac{h^3}{2} - \frac{h^3}{6} = \frac{1}{3}h^3. \end{aligned}$$

The covariance of ΔB_k and ΔZ_k is

$$\begin{aligned} \mathbf{E}(\Delta B_k\Delta Z_k) &= \mathbf{E}(\Delta B_1\Delta Z_1) = \mathbf{E}\int_0^h B_hB_s\,\mathrm{d}s \\ &= \int_0^h \mathbf{E}(B_hB_s)\,\mathrm{d}s = \int_0^h s\,\mathrm{d}s = \frac{1}{2}h^2. \end{aligned}$$

The pair of correlated normal random variables $(\Delta B_k, \Delta Z_k)$ can be obtained from a pair of independent normal random variables $U_1, U_2 \sim N(0,1)$ as follows:

$$\Delta B_k = U_1\sqrt{h}, \qquad \Delta Z_k = \frac{1}{2}\Big(U_1 + \frac{1}{\sqrt{3}}U_2\Big)h^{3/2}.$$

For further improvements of approximation order, we would need new random variables $\int_{t_k}^{t_{k+1}}(B_s - B_{t_k})^2\,\mathrm{d}s$ that are more difficult to generate. And what about the Heun method? Let us write its analog for a stochastic differential equation, with the main attention to its diffusion part:

$$\begin{aligned}
X_{t_{k+1}}^h &= X_{t_k}^h + b\big(X_{t_k}^h\big)h + \frac{1}{2}\big[\sigma\big(X_{t_k}^h\big) + \sigma\big(\overline{X}_{t_{k+1}}^h\big)\big]\Delta B_k,\\
\overline{X}_{t_{k+1}}^h &= X_{t_k}^h + b\big(X_{t_k}^h\big)h + \sigma\big(X_{t_k}^h\big)\Delta B_k.
\end{aligned}$$

By the Taylor formula, we have:

$$\begin{aligned}
\sigma\big(\overline{X}_{t_{k+1}}^h\big) &= \sigma\big(X_{t_k}^h\big) + \sigma'\big(X_{t_k}^h\big)\big(\overline{X}_{t_{k+1}}^h - X_{t_k}^h\big) + \mathrm{O}\big((\overline{X}_{t_{k+1}}^h - X_{t_k}^h)^2\big)\\
&= \sigma\big(X_{t_k}^h\big) + \sigma\sigma'\big(X_{t_k}^h\big)\Delta B_k + b\sigma'\big(X_{t_k}^h\big)h + \mathrm{O}\big(\big(X_{t_{k+1}}^h - X_{t_k}^h\big)^2\big).
\end{aligned}$$

Therefore, for the Heun approximation, we can write:

$$\begin{aligned}
X_{t_{k+1}}^h = X_{t_k}^h + b\big(X_{t_k}^h\big)h + \sigma\big(X_{t_k}^h\big)\Delta B_k + \frac{1}{2}\sigma\sigma'\big(X_{t_k}^h\big)\Delta B_k^2&\\
+ \text{ higher-order terms.}&
\end{aligned}$$

Thus, the Heun approximation is “equivalent” to the approximation

$$X_{t_{k+1}}^h = X_{t_k}^h + b\big(X_{t_k}^h\big)h + \sigma\big(X_{t_k}^h\big)\Delta B_k + \frac{1}{2}\sigma\sigma'\big(X_{t_k}^h\big)\Delta B_k^2.$$

Note that this is the Milstein approximation for the equation

$$\mathrm{d}X_t = \Big(b + \frac{1}{2}\sigma\sigma'\Big)(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t,$$

that is, for the Stratonovich equation

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sigma(X_t)\circ\mathrm{d}B_t.$$

There is nothing surprising here. Moreover, comparing the Heun approximation with the definition of the Stratonovich integral, we can see that this is exactly what we should expect. The advantage of the Heun method is that it does not use the derivatives of the coefficients. Therefore, it can be considered the simplest Runge–Kutta[9] method for numerically solving stochastic differential equations. Unfortunately, there are no first-order methods for Itô equations.

9. Karl David Tolmé Runge and Martin Wilhelm Kutta.

13.4. First-order weak approximations

For the homogeneous stochastic differential equation

$$X_t = x + \int_0^t b(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}B_s, \quad t \geqslant 0,$$

we shall consider weak approximations of the form

$$X_0 = x, \qquad X^h_{t_{k+1}} = a\big(X^h_{t_k}, h, \Delta B_k\big),$$

where $a = a(x,s,y)$, $(x,s,y) \in \mathbb{R} \times [0,T] \times \mathbb{R}$, is a "sufficiently good" real function satisfying the condition $a(x,0,0) \equiv x$. We shall call such a function an increment function of an approximation or simply, an approximation. By formula [13.2] we have:

$$f(X_h) = f(X_0) + \int_0^h Af(X_t)\,\mathrm{d}t + \int_0^h Sf(X_t)\,\mathrm{d}B_t,$$

and taking the expectations yields

$$\mathbf{E}f(X_h) = f(x) + \int_0^h \mathbf{E}Af(X_t)\,\mathrm{d}t, \quad h \geqslant 0.$$

Iterating this formula for the integrand function $\mathbf{E}Af(X_t)$, $t \in [0,h]$, we get:

$$\begin{aligned}\mathbf{E}f(X_h) &= f(x) + \int_0^h \left[Af(x) + \int_0^t \mathbf{E}A^2 f(X_s)\,\mathrm{d}s\right]\mathrm{d}t \\ &= f(x) + Af(x)h + \int_0^h\!\!\int_0^t \mathbf{E}A^2 f(X_s)\,\mathrm{d}s\,\mathrm{d}t. \qquad [13.3]\end{aligned}$$

A similar formula can be written for the approximation X^h. Denote

$$\tilde{f}(x,s,y) = f(a(x,s,y))$$

and

$$\Delta\tilde{f} = \Big(\frac{\mathrm{d}}{\mathrm{d}s} + \frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2}\Big)\tilde{f}.$$

Applying Itô's formula to the process $\tilde{f}(x,t,B_t)$, $t \in [0,h]$, we have:

$$\begin{aligned} f(X_h^h) &= f\big(a(x,h,B_h)\big) = \tilde{f}(x,h,B_h) \\ &= \tilde{f}(x,0,0) + \int_0^h \frac{\mathrm{d}}{\mathrm{d}t}\tilde{f}(x,t,B_t)\,\mathrm{d}t + \int_0^h \frac{\mathrm{d}}{\mathrm{d}y}\tilde{f}(x,t,B_t)\,\mathrm{d}B_t \\ &\quad + \frac{1}{2}\int_0^h \frac{\mathrm{d}^2}{\mathrm{d}y^2}\tilde{f}(x,t,B_t)\,\mathrm{d}t \\ &= \tilde{f}(x,0,0) + \int_0^h \Delta\tilde{f}(x,t,B_t)\,\mathrm{d}t + \int_0^h \frac{\mathrm{d}}{\mathrm{d}y}\tilde{f}(x,t,B_t)\,\mathrm{d}B_t. \end{aligned}$$

Taking the expectations, we get

$$\mathbf{E}f(X_h^h) = \tilde{f}(x,0,0) + \int_0^h \mathbf{E}\Delta\tilde{f}(x,t,B_t)\,\mathrm{d}t :$$

Iterating this formula for the integrand function $\mathbf{E}\Delta\tilde{f}(x,t,B_t)$, $t \in [0,h]$, we get:

$$\begin{aligned} \mathbf{E}f(X_h^h) &= f(x) + \int_0^h \left[\Delta\tilde{f}(x,0,0) + \int_0^t \mathbf{E}\Delta^2\tilde{f}(x,s,B_s)\,\mathrm{d}s\right]\mathrm{d}t \\ &= f(x) + \Delta\tilde{f}(x,0,0)h + \int_0^h\int_0^t \mathbf{E}\Delta^2\tilde{f}(x,s,B_s)\,\mathrm{d}s\,\mathrm{d}t. \end{aligned} \qquad [13.4]$$

If, for example, $A^2 f$ and $\Delta^2\tilde{f}$ (or, more generally, $\mathbf{E}A^2 f(X_s)$ and $\mathbf{E}\Delta^2\tilde{f}(x,s,B_s)$, $s \in [0,T]$) are bounded functions, we can write:

$$\begin{aligned} \mathbf{E}f(X_h) &= f(x) + Af(x)h + \mathrm{O}(h^2), \\ \mathbf{E}f(X_h^h) &= f(x) + \Delta\tilde{f}(x,0,0)h + \mathrm{O}(h^2). \end{aligned}$$

Suppose that the approximation a is chosen so that

$$\Delta\tilde{f}(x,0,0) = Af(x), \quad x \in \mathbb{R}. \qquad [13.5]$$

Then, with any initial value $x \in \mathbb{R}$,

$$\mathbf{E}f\big(X_h^h\big) - \mathbf{E}f(X_h) = \mathrm{O}(h^2).$$

Using the Markov property of X and X^h, we can also write that, for conditional expectations,

$$\mathbf{E}\big(f\big(X^h_{t_{k+1}}\big)\big|X^h_{t_k}=x\big)-\mathbf{E}\big(f(X_{t_{k+1}})\big|X_{t_k}=x\big)=\mathrm{O}(h^2)$$

for all $t_k = kh$ and $x \in \mathbb{R}$. In this case, we say that the weak approximation X^h has second-order *one-step* accuracy (or that it is *locally* of second order). Under certain regularity conditions, it is proved that, as in the deterministic case, one-step second-order accuracy implies first-order ("global") accuracy. In order to apply condition [13.5], let us express $\Delta\tilde{f}$ in terms of the function a:

$$\frac{\mathrm{d}}{\mathrm{d}s}\tilde{f}=f'(a)a'_s, \qquad \frac{\mathrm{d}}{\mathrm{d}y}\tilde{f}=f'(a)a'_y,$$

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2}\tilde{f}=f''(a){a'_y}^2+f'(a)a''_{yy},$$

$$\Delta\tilde{f}=\Big(a'_s+\frac{1}{2}a''_{yy}\Big)f'(a)+\frac{1}{2}{a'_y}^2f''(a).$$

At the point $\bar{x} := (x, 0, 0)$, we have

$$\Delta\tilde{f}(\bar{x})=\Big(a'_s+\frac{1}{2}a''_{yy}\Big)(\bar{x})f'(x)+\frac{1}{2}{a'_y}^2(\bar{x})f''(x).$$

Condition [13.5] is satisfied if, in the obtained expressions of $\Delta\tilde{f}(\bar{x})$ and $Af(x)$, the coefficients at the derivatives $f'(x)$ and $f''(x)$ coincide, that is,

$$\begin{cases}\big(a'_s+\frac{1}{2}a''_{yy}\big)(\bar{x}) & = b(x),\\ {a'_y}^2(\bar{x}) & = \sigma^2(x).\end{cases}$$

It is rather natural to choose $a'_y(\bar{x}) = \sigma(x)$ (though we can also take, e.g. $a'_y(\bar{x}) = -\sigma(x)$). Thus, we finally get the following sufficient conditions for the first-order accuracy of approximation X^h in terms of function a:

$$\begin{cases}a'_y(\bar{x}) & = \sigma(x),\\ \big(a'_s+\frac{1}{2}a''_{yy}\big)(\bar{x}) & = b(x).\end{cases} \qquad [13.6a]$$

In the case of Stratonovich equation

$$X_t=x+\int_0^t b(X_s)\,\mathrm{d}s+\int_0^t\sigma(X_s)\circ\mathrm{d}B_s, \quad t\geqslant 0,$$

in conditions [13.6a], the coefficient b has to be replaced by $\tilde{b} = b + \frac{1}{2}\sigma\sigma'$, since this equation is equivalent to the equation

$$X_t=x+\int_0^t\tilde{b}(X_s)\,\mathrm{d}s+\int_0^t\sigma(X_s)\,\mathrm{d}B_s, \quad t\geqslant 0.$$

So, we get the following sufficient conditions for the first-order accuracy of the Stratonovich equation:

$$\begin{cases} a'_y(\bar{x}) & = \sigma(x), \\ \left(a'_s + \frac{1}{2}a''_{yy}\right)(\bar{x}) & = \left(b + \frac{1}{2}\sigma\sigma'\right)(x). \end{cases} \qquad [13.6b]$$

REMARK.– To simplify the calculations and especially to speed them up, we can replace the increments $\Delta B_k \sim N(0,h)$ of Brownian motion with random variables $\xi_k = \eta_k\sqrt{h}$ such that $\mathbf{E}(\xi_k)^i = \mathbf{E}(\Delta B_k)^i$, $i = 1,2,3$. Indeed, instead of formula [13.4], let us write the Taylor formula for $\tilde{f}$ up to the third-order term:

$$\tilde{f}(x,h,\Delta B_k) = \sum_{i=0}^{3} \frac{1}{i!}\frac{\partial^i \tilde{f}(x,h,0)}{\partial y^i}(\Delta B_k)^i + \frac{1}{4!}\frac{\partial^4 \tilde{f}(x,h,\theta)}{\partial y^4}(\Delta B_k)^4,$$

where $|\theta| \leqslant |\Delta B_k|$. Similarly,

$$\tilde{f}(x,h,\xi_k) = \sum_{i=0}^{3} \frac{1}{i!}\frac{\partial^i \tilde{f}(x,h,0)}{\partial y^i}(\xi_k)^i + \frac{1}{4!}\frac{\partial^4 \tilde{f}(x,h,\bar{\theta})}{\partial y^4}(\xi_k)^4,$$

where $|\bar{\theta}| \leqslant |\xi_k|$. Subtracting one equality from the other and taking the expectations, we get:

$$\begin{aligned} &\left|\mathbf{E}\tilde{f}(x,h,\xi_k) - \mathbf{E}\tilde{f}(x,h,\Delta B_k)\right| \\ &\quad \leqslant \frac{1}{4!}\left(\mathbf{E}\left|\frac{\partial^4 \tilde{f}(x,h,\bar{\theta})}{\partial y^4}(\xi_k)^4\right| + \mathbf{E}\left|\frac{\partial^4 \tilde{f}(x,h,\theta)}{\partial y^4}(\Delta B_k)^4\right|\right) \\ &\quad \leqslant \mathbf{E}\big(r(x,h,\xi_k)\,\xi_k^4\big) + \mathbf{E}\big(r(x,h,\Delta B_k)\,\Delta B_k^4\big), \end{aligned}$$

where

$$r(x,h,y) := \frac{1}{4!}\sup_{\theta\in[-y,y]}\left|\frac{\partial^4 \tilde{f}(x,h,\theta)}{\partial y^4}\right|.$$

If, for example, the function r is bounded by some constant C, then

$$\begin{aligned} &\mathbf{E}\big(r(x,h,\xi_k)\,\xi_k^4\big) + \mathbf{E}\big(r(x,h,\Delta B_k)\,\Delta B_k^4\big) \\ &\qquad\qquad \leqslant C\big(\mathbf{E}(\eta_k\sqrt{h})^4 + \mathbf{E}(\Delta B_k)^4\big) \\ &\qquad\qquad = C\big(C_1 h^2 + 3h^2\big) = \mathrm{O}\big(h^2\big). \end{aligned}$$

One can get the same estimate by applying the Hölder inequality in a more general case where r increases "not too rapidly" with y. Thus, replacing the increments ΔB_k

with random variables ξ_k does not decrease the one-step accuracy order of the weak approximation X^h. We can take, for example, $\xi_k = \pm\sqrt{h}$ (i.e. $\eta_k = \pm 1$) with probability $\frac{1}{2}$.

EXAMPLES 13.3.– Consider (increment functions of) approximations of the form

$$a(x, s, y) = x + c_1 s + c_2 y + c_3 y^2$$

(with $c_i = c_i(x)$). At points $\bar{x}$, we have $a'_s = c_1$, $a'_y = c_2$, $a''_{yy} = 2c_3$; therefore, conditions [13.6a] become

$$\begin{cases} c_1 + c_3 & = b(x), \\ c_2 & = \sigma(x). \end{cases}$$

For c_2, we have no choice: $c_2 = \sigma(x)$. Note that both the Euler approximation ($a = x + b(x)s + \sigma(x)y$) and Milstein approximation ($a = x + \big(b - \frac{1}{2}\sigma\sigma'\big)(x)s + \sigma(x)y + \frac{1}{2}\sigma\sigma'(x)y^2$) satisfy these conditions. Thus, they are both first-order *weak* approximations. Since the Euler approximation is evidently simpler than the Milstein one, in practice it is usually preferred as a weak first-order approximation, unless, for some reason, it is important that the approximation used also be used as a first-order *strong* one.

In the case of Stratonovich equation, from conditions [13.6b] we get the following conditions for the coefficients of function a:

$$\begin{cases} c_2 & = \sigma(x), \\ c_1 + c_3 & = b(x) + \frac{1}{2}\sigma\sigma'(x). \end{cases}$$

Here, it is natural to take

$$c_1 = b(x), \qquad c_2 = \sigma(x), \qquad c_3 = \frac{1}{2}\sigma\sigma'(x)$$

or

$$c_1 = b(x) + \frac{1}{2}\sigma\sigma'(x), \qquad c_2 = \sigma(x), \qquad c_3 = 0.$$

Now let us consider the Heun approximation

$$a = x + b(x)s + \frac{1}{2}\Big[\sigma(x) + \sigma\big(x + b(x)s + \sigma(x)y\big)\Big]y.$$

Denoting $\bar{a} = x + b(x)s + \sigma(x)y$, we have:

$$\begin{aligned}
a'_s &= b(x) + \frac{1}{2}\sigma'(\bar{a})b(x)y \implies a'_s(\bar{x}) = b(x),\\
a'_y &= \frac{1}{2}\big(\sigma(x) + \sigma(\bar{a})\big) + \frac{1}{2}\sigma'(\bar{a})\sigma(x)y \implies a'_y(\bar{x}) = \sigma(x),\\
a''_{yy} &= \frac{1}{2}\sigma'(\bar{a})\sigma(x) + \frac{1}{2}\sigma''(\bar{a})\sigma^2(x)y + \frac{1}{2}\sigma'(\bar{a})\sigma(x)\\
&= \sigma'(\bar{a})\sigma(x) + \frac{1}{2}\sigma''(\bar{a})\sigma^2(x)y \implies a''_{yy}(\bar{x}) = \sigma\sigma'(x).
\end{aligned}$$

Thus, the Heun approximation satisfies conditions [13.6b] and therefore is a first-order weak approximation for the Stratonovich equation (as we can expect for any approximation which is a *strong* first-order approximation).

How are such approximations realized in practice? Suppose that we have already chosen some (say, Euler) approximation $\{X^h, h > 0\}$ of a solution of equation [13.1] and would like to calculate approximate expectation $\mathbf{E}f(X_t)$, knowing that, theoretically, $\mathbf{E}f(X_t^h) \approx \mathbf{E}f(X_t)$ with some accuracy when h is "small". Then we generate by computer "sufficiently many" independent realizations $X^{h,i}$, $i = 1, 2, \ldots, N$, of the process X^h and calculate the arithmetic average

$$\frac{1}{N}\sum_{i=1}^{N} f\big(X_t^{h,i}\big).$$

By the law of large numbers, this sum converges to $\mathbf{E}f(X_t^h)$ as $N \to \infty$ with probability one. Therefore, this average can be considered an estimate of the (theoretical) expectation $\mathbf{E}f(X_t)$ with error

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N} f\big(X_t^{h,i}\big) - \mathbf{E}f(X_t)\right|\\
&\quad \leqslant \left|\frac{1}{N}\sum_{i=1}^{N} f\big(X_t^{h,i}\big) - \mathbf{E}f\big(X_t^h\big)\right| + \left|\mathbf{E}f\big(X_t^h\big) - \mathbf{E}f(X_t)\right| =: \varepsilon_1 + \varepsilon_2.
\end{aligned}$$

The quantity $\varepsilon_2 = \varepsilon_2(h)$ is called the *system error*. It depends only on the weak approximation ("system") $\{X^h\}$ and tends more or less slowly (depending on the chosen approximation) to 0 as $h \to 0$. The (random) quantity $\varepsilon_1 = e_1(h, N, \omega)$ is called the *statistical error*. To minimize it, we have to take a rather large number of realizations N. On the other hand, to reduce the computation time, it is desirable to reduce N as much as possible. These, rather complicated questions of choosing optimal N are considered in the so-called theory of Monte Carlo methods.

We shall look at how the Euler approximation behaves in a concrete example. For a rather illustrative example, we have to choose an SDE and a test function f such that: a) the equation must be "non-trivial" but explicitly solvable; b) the function f also has to be "non-trivial," though such that theoretical expectation $\mathbf{E}f(X_t)$ is known. As an example, we take the equation

$$\mathrm{d}X_t = \frac{1}{2}X_t\,\mathrm{d}t + \sqrt{X_t^2+1}\,\mathrm{d}B_t, \qquad X_0 = x,$$

and the test function $f(x) := \operatorname{arcsinh}^4(x)$, $x \in \mathbb{R}$ (recall that $\operatorname{arcsinh} = \ln\big(x + \sqrt{x^2+1}\big)$ is the inverse of the hyperbolic sine $\sinh(x) = (\mathrm{e}^x - \mathrm{e}^{-x})/2$). The solution of the equation is $X_t = \sinh(B_t + \operatorname{arcsinh} x)$. Denoting $y := \operatorname{arcsinh} x$, we get:

$$\begin{aligned}\mathbf{E}f(X_t) &= \mathbf{E}(B_t+y)^4 = \mathbf{E}B_t^4 + 4\mathbf{E}B_t^3 y + 6\mathbf{E}B_t^2 y^2 + 4\mathbf{E}B_t y^3 + y^4\\ &= 3t^2 + 6y^2 t + y^4.\end{aligned}$$

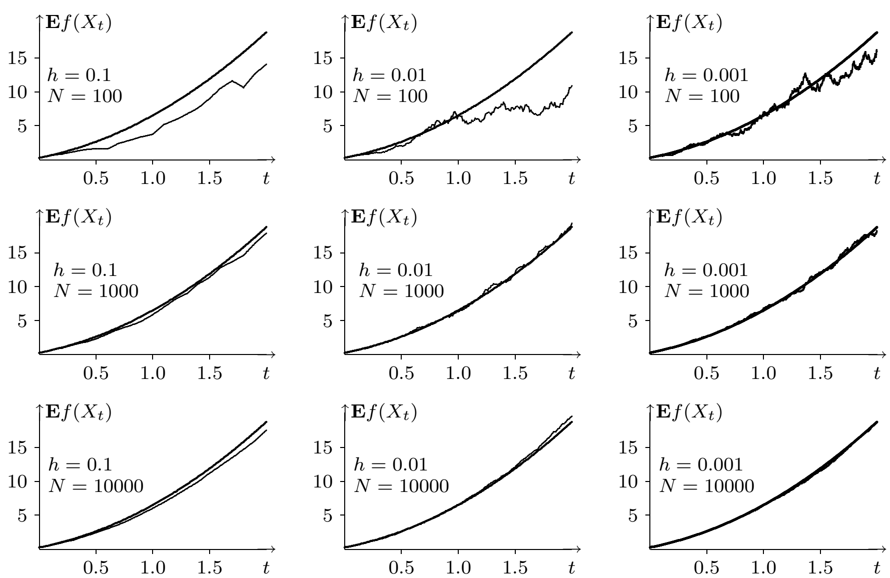
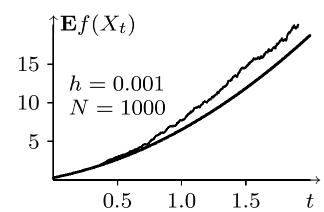


Figure 13.3. *Weak Euler approximation with various steps h and numbers N of simulated trajectories*

Examples of application of the Euler approximation for this equation with various values of h and N are shown in Figure 13.3. Having chosen $x = 0.8$ ($y \approx 0.28816$), the thick (and smooth) lines are the graphs of the expectation $\mathbf{E}f(X_t) = 3t^2 + 6y^2 t + y^4$. The thin and less regular lines show the values of the generated averages $N^{-1}\sum_{i=1}^{N} f(X_t^{h,i})$ at the points $t = kh$, $k = 1, 2, \ldots$, joined by a polygonal line. We see that the accuracy of approximation is more perceptible by increasing N than by decreasing h. This can be explained rather simply: if h is decreased, say, 10 times, then the *relative* system error decreases by approximately 10 times. If h is already chosen and is "reasonably small", a subsequent decrease would not significantly decrease the *absolute* system error. Therefore, in the total error, the main contribution is due to statistical error, which more perceptibly decreases with increasing N. Thus, for further diminishing the error, it is more reasonable to decrease N.

In practical calculations, one has to keep in mind that approximations, as the solutions of SDEs, are *random* processes. Therefore, though in rare cases (with small probability), approximations may differ from the theoretically expected results, we see one "accidental" image which is rather far from the one theoretically expected as observed in Figure 13.3 ($h =$ 0.001, $N = 1{,}000$). Thus, we have to take some care when interpreting simulation results.



13.5. Higher-order weak approximations

Let us try to improve the order of weak approximations. Extending equations [13.3] and [13.4], we get:

$$\mathbf{E}f(X_h) = f(x) + Af(x)h + A^2 f(x)\frac{h^2}{2} + \int_0^h\int_0^t\int_0^s \mathbf{E}A^3 f(X_u)\,\mathrm{d}u\,\mathrm{d}s\,\mathrm{d}t,$$

$$\mathbf{E}f(X_h^h) = f(x) + \Delta\tilde{f}(x,0,0)h + \Delta^2\tilde{f}(x,0,0)\frac{h^2}{2} + \int_0^h\int_0^t\int_0^s \mathbf{E}\Delta^3\tilde{f}(x,u,B_u)\,\mathrm{d}u\,\mathrm{d}s\,\mathrm{d}t.$$

If, for example, $\mathbf{E}A^3 f(X_s)$ and $\mathbf{E}\Delta^3\tilde{f}(x,s,B_s)$, $s \in [0,T]$, are bounded functions, then we have:

$$\mathbf{E}f(X_h) = f(x) + Af(x)h + A^2 f(x)\frac{h^2}{2} + \mathrm{O}(h^3),$$

$$\mathbf{E}f(X_h^h) = f(x) + \Delta\tilde{f}(x,0,0)h + \Delta^2\tilde{f}(x,0,0)\frac{h^2}{2} + \mathrm{O}(h^3). \qquad [13.7]$$

Therefore, if, in addition to condition [13.5], we require that

$$\Delta^2 \tilde{f}(x,0,0) = A^2 f(x), \quad x \in \mathbb{R}, \tag{13.8}$$

we get sufficient conditions for the third-order one-step accuracy or second-order "global" accuracy.

By elementary but tedious calculations, we get the following expressions for $A^2 f$ and $\Delta^2 \tilde{f}$:

$$\begin{aligned}
A^2 f(x) &= A(Af)(x) = A(bf' + \frac{1}{2}\sigma^2 f'')(x) \\
&= b\Big(bf' + \frac{1}{2}\sigma^2 f''\Big)'(x) + \frac{1}{2}\sigma^2\Big(bf' + \frac{1}{2}\sigma^2 f''\Big)''(x) \\
ex &= \Big[b\Big(bf'' + b'f' + \frac{1}{2}\sigma^2 f''' + \sigma\sigma' f''\Big) \\
&\quad + \frac{1}{2}\sigma^2\Big(bf'' + b'f' + \frac{1}{2}\sigma^2 f''' + \sigma\sigma' f''\Big)'\Big](x) \\
&= \Big[b^2 f'' + bb'f' + \frac{1}{2}b\sigma^2 f''' + b\sigma\sigma' f'' \\
&\quad + \frac{1}{2}\sigma^2\Big(b''f' + b'f'' + b'f'' + bf''' + \sigma\sigma' f''' + \frac{1}{2}\sigma^2 f'''' \\
&\quad + \sigma'^2 f'' + \sigma\sigma'' f'' + \sigma\sigma' f'''\Big)\Big](x) \\
&= \Big(bb' + \frac{1}{2}\sigma^2 b''\Big)(x) f'(x) \\
&\quad + \Big(b^2 + \sigma^2 b' + b\sigma\sigma' + \frac{1}{2}\sigma^2\sigma'^2 + \frac{1}{2}\sigma^3\sigma''\Big)(x) f''(x) \\
&\quad + \big(b\sigma^2 + \sigma^3\sigma'\big)(x) f'''(x) + \frac{1}{4}\sigma^4(x) f''''(x)
\end{aligned}$$

and

$$\begin{aligned}
\Delta^2 \tilde{f}(\bar{x}) &= \Big(a''_{ss} + a'''_{yys} + \frac{1}{4}a''''_{yyyy}\Big)(\bar{x}) f'(x) \\
&\quad + \Big[\Big(a'_s + \frac{1}{2}a''_{yy}\Big)^2 + 2a'_y a''_{ys} + a'_y a'''_{yyy} + \frac{1}{2}\big(a''_{yy}\big)^2\Big](\bar{x}) f''(x) \\
&\quad + \Big[\Big(a'_s + \frac{1}{2}a''_{yy}\Big)a'^2_y + a'^2_y a''_{yy}\Big](\bar{x}) f'''(x) + \frac{1}{4}a'^4_y(\bar{x}) f''''(x).
\end{aligned}$$

Equating the coefficients of $\Delta^2 \tilde{f}(\bar{x})$ and $A^2 f(x)$ at the equal-order derivatives, we get the equations

$$\begin{aligned}
\left(a''_{ss} + a'''_{yys} + \frac{1}{4}a''''_{yyyy}\right)(\bar{x}) &= \Big(bb' + \frac{1}{2}\sigma^2 b''\Big)(x), \\
\Big[\left(a'_s + \frac{1}{2}a''_{yy}\right)^2 + 2a'_y a''_{ys} + a'_y a'''_{yyy} + \frac{1}{2}\left(a''_{yy}\right)^2\Big](\bar{x}) & \\
&= \Big(b^2 + \sigma^2 b' + b\sigma\sigma' + \frac{1}{2}\sigma^2\sigma'^2 + \frac{1}{2}\sigma^3\sigma''\Big)(x), \\
\Big[\left(a'_s + \frac{1}{2}a''_{yy}\right)a'^2_y + a'^2_y a''_{yy}\Big](\bar{x}) &= \left(b\sigma^2 + \sigma^3\sigma'\right)(x), \\
a'^4_y(\bar{x}) &= \sigma^4(x).
\end{aligned}$$

Recalling conditions [13.6a], the fourth equation is automatically satisfied. Reducing the equal terms

$$\Big(a'_s + \frac{1}{2}a''_{yy}\Big)^2(\bar{x}) = b^2(x) \qquad \text{and} \qquad \Big(a'_s + \frac{1}{2}a''_{yy}\Big)a'^2_y(\bar{x}) = b\sigma^2(x),$$

in the second and third equations and then dividing the third equation by $a'^2_y(\bar{x}) = \sigma^2(x)$, we get:

$$\begin{gathered}
\Big[2a'_y a''_{ys} + a'_y a'''_{yyy} + \frac{1}{2}\left(a''_{yy}\right)^2\Big](\bar{x}) = \Big(\sigma^2 b' + b\sigma\sigma' + \frac{1}{2}\sigma^2\sigma'^2 + \frac{1}{2}\sigma^3\sigma''\Big)(x), \\
a''_{yy}(\bar{x}) = \sigma\sigma'(x).
\end{gathered}$$

Now, in the former equation, subtract the equal terms

$$\frac{1}{2}(a''_{yy})^2(\bar{x}) = \frac{1}{2}\sigma^2\sigma'^2(x)$$

(the latter equation) and then divide it by $a'_y(\bar{x}) = \sigma(x)$. We get:

$$\left(2a''_{ys} + a'''_{yyy}\right)(\bar{x}) = \Big(\sigma b' + b\sigma' + \frac{1}{2}\sigma^2\sigma''\Big)(x).$$

Recalling once more the equality $a''_{yy}(\bar{x}) = \sigma\sigma'(x)$, this time in the second equation in [13.6a], we finally get the following conditions for the second-order accuracy for weak approximations of Itô equations in terms of the function a:

$$\begin{cases}
a'_y(\bar{x}) & = \sigma(x), \\
a'_s(\bar{x}) & = b(x) - \frac{1}{2}\sigma\sigma'(x), \\
a'_{yy}(\bar{x}) & = \sigma\sigma'(x), \\
\left(2a''_{ys} + a'''_{yyy}\right)(\bar{x}) & = (b\sigma)'(x) + \frac{1}{2}\sigma^2\sigma''(x), \\
\left(a''_{ss} + a'''_{yys} + \frac{1}{4}a''''_{yyyy}\right)(\bar{x}) & = bb'(x) + \frac{1}{2}\sigma^2 b''(x).
\end{cases} \tag{13.9a}$$

For the Stratonovich equations, we obtain such conditions by replacing in conditions [13.9a] the coefficient b by the coefficient $b + \frac{1}{2}\sigma\sigma'$:

$$\begin{cases} a'_y(\bar{x}) & = \sigma(x), \\ a'_s(\bar{x}) & = b(x), \\ a'_{yy}(\bar{x}) & = \sigma\sigma'(x), \\ \big(2a''_{ys} + a'''_{yyy}\big)(\bar{x}) & = (b\sigma' + b'\sigma)(x) + (\sigma^2\sigma'' + \sigma\sigma'^2)(x), \\ \big(a''_{ss} + a'''_{yys} + \frac{1}{4}a''''_{yyyy}\big)(\bar{x}) & = bb'(x) \\ & \quad + \frac{1}{2}(b\sigma'^2 + b\sigma\sigma'' + b'\sigma\sigma' + b''\sigma^2)(x) \\ & \quad + (\sigma^3\sigma''' + 4\sigma^2\sigma'\sigma'' + \sigma\sigma'^3)(x). \end{cases} \qquad [13.9b]$$

After some rearrangements in the last two equations, we get a more symmetric version of conditions [13.9b]:

$$\begin{cases} a'_y(\bar{x}) & = \sigma(x), \\ a'_s(\bar{x}) & = b(x), \\ a''_{yy}(\bar{x}) & = \sigma\sigma'(x), \\ \big(2a''_{ys} + a'''_{yyy}\big)(\bar{x}) & = (b\sigma)'(x) + \sigma(\sigma\sigma')'(x), \\ \big(a''_{ss} + a'''_{yys} + \frac{1}{4}a''''_{yyyy}\big)(\bar{x}) & = bb'(x) + \frac{1}{2}b(\sigma\sigma')'(x) + \frac{1}{2}\sigma(\sigma b')'(x) \\ & \quad + \frac{1}{4}\sigma\big(\sigma(\sigma\sigma')'\big)'(x). \end{cases} \qquad [13.9b']$$

Similarly to the remark on page 234, preserving the second-order accuracy, we can replace the increments ΔB_k‘ $\sim N(0,h)$ by random variables $\xi_k = \eta_k\sqrt{h}$ such that $\mathbf{E}(\xi_k)^i = \mathbf{E}(\Delta B_k)^i$, $i = 1, 2, 3, 4, 5$. For example, we can use random variables η_k taking values $\pm\sqrt{3}$ with probability $\frac{1}{6}$ and the value 0 with probability $\frac{2}{3}$.

13.6. Example: Milstein-type approximations

We say that $a = a(x, s, y)$ is a Milstein-type approximation if it is a polynomial with respect to s and y. Let

$$a(x,s,y) = x + c_0 y + c_1 s + c_2 y^2 + c_3 ys + c_4 y^3 + c_5 s^2 + c_6 y^2 s + c_7 y^4$$

with $c_i = c_i(x)$, $i = 0, 1, \ldots, 7$. The partial derivatives of a the points $\bar{x}$ are:

$$\begin{aligned} &a'_s = c_1, && a''_{ss} = 2c_5, && && \\ &a'_y = c_0, && a''_{yy} = 2c_2, && a'''_{yyy} = 6c_4, && a''''_{yyyy} = 24c_7, \\ &a''_{ys} = c_3, && a'''_{yys} = 2c_6. && && \end{aligned}$$

From conditions [13.9a] we have

$$\begin{aligned}
c_0 &= \sigma(x),\\
c_1 &= b(x) - \frac{1}{2}\sigma\sigma'(x),\\
c_2 &= \frac{1}{2}\sigma\sigma'(x),\\
c_3 + 3c_4 &= \frac{1}{2}(b\sigma)'(x) + \frac{1}{4}\sigma^2\sigma''(x),\\
c_5 + c_6 + 3c_7 &= \frac{1}{2}bb'(x) + \frac{1}{4}\sigma^2 b''(x).
\end{aligned}$$

We have a great deal of freedom in choosing coefficients c_3–c_7. Let us simply take $c_4 = c_6 = c_7 = 0$. So, we get the following second-order weak approximation (called the Milstein second-order weak approximation):

$$\begin{aligned}
a(x,s,y) &= x + \sigma(x)y + \Big(b - \frac{1}{2}\sigma\sigma'\Big)(x)s + \frac{1}{2}\sigma\sigma'(x)y^2\\
&\quad + \Big(\frac{1}{2}(b\sigma)' + \frac{1}{4}\sigma^2\sigma''\Big)(x)ys + \Big(\frac{1}{2}bb' + \frac{1}{4}\sigma^2 b''\Big)(x)s^2.
\end{aligned}$$

Similarly, we can construct the Milstein approximation for Stratonovich equation. From conditions [13.9b′] we have:

$$\begin{aligned}
c_0 &= \sigma(x),\\
c_1 &= b(x),\\
c_2 &= \frac{1}{2}\sigma\sigma'(x),\\
c_3 + 3c_4 &= \frac{1}{2}(b\sigma)'(x) + \frac{1}{2}\sigma(\sigma\sigma')'(x),\\
c_5 + c_6 + 3c_7 &= \frac{1}{2}bb'(x) + \frac{1}{4}\sigma(\sigma b')'(x) + \frac{1}{4}b(\sigma\sigma')'(x)\\
&\quad + \frac{1}{8}\sigma\big(\sigma(\sigma\sigma')'\big)'(x).
\end{aligned}$$

Again, there is much freedom in choosing coefficients c_3–c_7. The following is an especially symmetric version of the Milstein-type approximation:

$$\begin{aligned}
a(x,s,y) &= x + b(x)s + \frac{1}{2}bb'(x)s^2\\
&\quad + \sigma(x)y + \frac{1}{2}\sigma\sigma'(x)y^2 + \frac{1}{6}\sigma(\sigma\sigma')'(x)y^3 + \frac{1}{24}\sigma\big(\sigma(\sigma\sigma')'\big)'(x)y^4\\
&\quad + \frac{1}{2}(\sigma b)'(x)ys + \frac{1}{4}\big(\sigma(\sigma b')' + b(\sigma\sigma')'\big)(x)y^2 s. \qquad [13.10]
\end{aligned}$$

REMARK. When $\sigma = 0$, we have the deterministic equation $\mathrm{d}X_t = b(X_t)\,\mathrm{d}t$. Therefore, there is no surprise that in this case the latter approximation coincides with the second-order Taylor approximation defined by

$$X^h_{t_{k+1}} = a\big(X^h_{t_k}, h, 0\big) = X^h_{t_k} + b\big(X^h_{t_k}\big)h + \frac{1}{2!}bb'\big(X^h_{t_k}\big)h^2.$$

The more surprising fact is that, in the case of a "purely" stochastic Stratonovich equation where $b = 0$, the coefficients of the weak second-order approximation

$$X^h_{t_{k+1}} = a\big(X^h_{t_k}, 0, \Delta B_{t_k}\big) = X^h_{t_k} + \sigma\big(X^h_{t_k}\big)\Delta B_{t_k} + \frac{1}{2}\sigma\sigma'\big(X^h_{t_k}\big)\Delta B^2_{t_k}$$
$$+ \frac{1}{6}\sigma(\sigma\sigma')'(X^h_{t_k})\Delta B^3_{t_k} + \frac{1}{24}\sigma\big(\sigma(\sigma\sigma')'\big)'\big(X^h_{t_k}\big)\Delta B^4_{t_k},$$

where function a is defined by equation [13.10], coincide with those of the fourth-order Taylor approximation

$$X^h_{t_{k+1}} = a\big(X^h_{t_k}, 0, h\big) = X^h_{t_k} + \sigma\big(X^h_{t_k}\big)h + \frac{1}{2}\sigma\sigma'\big(X^h_{t_k}\big)h^2$$
$$+ \frac{1}{6}\sigma(\sigma\sigma')'(X^h_{t_k})h^3 + \frac{1}{24}\sigma\big(\sigma(\sigma\sigma')'\big)'\big(X^h_{t_k}\big)h^4$$

of the deterministic equation $\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}t$. This can be roughly explained by the fact that, for Stratonovich integrals and equations, the "usual" differential rules hold. In the deterministic equation, taking, instead of the "deterministic" differential $\mathrm{d}t$, taking the stochastic differential $\mathrm{d}B_t$, we arrive at the "purely" stochastic Stratonovich equation. Therefore, in the approximation of the deterministic equation, replacing increments $t_{k+1} - t_k = h$ by increments $\Delta B_k = B_{t_{k+1}} - B_{t_k}$, we get the approximation of the stochastic equation the order of which is, as usual (unfortunately), two times lower.

To compare the Milstein approximation with the Euler one, we consider the equation of the previous example, $\mathrm{d}X_t = 0.5X_t\,\mathrm{d}t + \sqrt{X_t^2 + 1}\,\mathrm{d}B_t$, $X_0 = 0.8$, as well as the test function $f(x) := \operatorname{arcsinh}^4(x)$, $x \in \mathbb{R}$. After elementary calculations, in this case, the Milstein approximation is

$$a(x, s, y) = x + \sqrt{x^2 + 1}\,y(1 + s/2) + x(y^2/2 + s^2/8).$$

To diminish the role of statistical error, we took a rather large number of simulations, $N = 10{,}000$. On the other hand, we took an h larger than before, since with small h it would be difficult to visually distinguish the difference between the Euler and Milstein approximation (e.g. on the computer display, it would difficult to distinguish the errors 10^{-1} and 10^{-2}, though the latter is 10 times smaller). The simulation results in two different time intervals are shown in Figure 13.4. Its right part corresponds to the shaded rectangle in the left part.

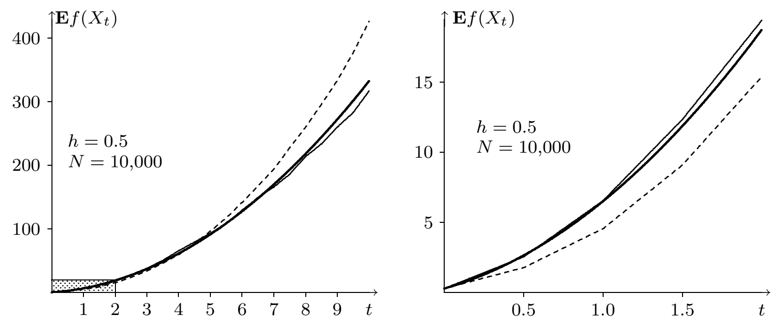


Figure 13.4. *Comparison of weak Euler and Milstein approximations. Thick line: exact values of the expectation $\mathbf{E}f(X_t)$; thin line: Milstein approximation; dashed line: Euler approximation*

13.7. Example: Runge–Kutta approximations

In a wide sense, Runge–Kutta (RK) approximations are derivative-free approximations, i.e. such that, in their approximations, there are no derivatives of the coefficients. This is particularly convenient when programming the simulations since, for each new equation, we have to reprogram only the coefficients; moreover, in the case of complicate coefficients, we do have to waste time on "manual" calculations of derivatives. We shall consider RK approximations of the form similar to popular four-step RK approximations for solving deterministic differential equations:

$$
\begin{aligned}
a(x,s,y) &= x + \sum_{i=0}^{3} q_i F_i s + \sum_{i=0}^{3} r_i G_i y, \quad \text{where}\\
F_0 &= b(x), \qquad G_0 = \sigma(x + \alpha_{00} F_0 s),\\
H_1 &= x + \alpha_{10} F_0 s + \beta_{10} G_0 y,\\
F_1 &= b(H_1), \qquad G_1 = \sigma(H_1),\\
H_2 &= x + (\alpha_{20} F_0 + \alpha_{21} F_1) s + (\beta_{20} G_0 + \beta_{21} G_1) y,\\
F_2 &= b(H_2), \qquad G_2 = \sigma(H_2),\\
H_3 &= x + (\alpha_{30} F_0 + \alpha_{31} F_1 + \alpha_{32} F_2) s + (\beta_{30} G_0 + \beta_{31} G_1 + \beta_{32} G_2) y,\\
F_3 &= b(H_3), \qquad G_3 = \sigma(H_3).
\end{aligned}
$$

Here, the additional somewhat unusual parameter α_{00} is introduced for convenience. It will allow us to get simpler collections of the parameters $\{\alpha_{ij}, \beta_{ij}, r_i, q_i\}$.

It is convenient to write the parameters of a Runge–Kutta method in the so-called Butcher[10] array. We shall define the above approximation by using a similar (double-width) array in which $\alpha_i = \sum_j \alpha_{ij}$, $\beta_i = \sum_j \beta_{ij}$:

$$
\begin{array}{c|cccc|cccc}
\alpha_0 & \alpha_{00} & & & 0 & & & & \\
\alpha_1 & \alpha_{10} & & & \beta_1 & \beta_{10} & & & \\
\alpha_2 & \alpha_{20} & \alpha_{21} & & \beta_2 & \beta_{20} & \beta_{21} & & \\
\alpha_3 & \alpha_{30} & \alpha_{31} & \alpha_{32} & \beta_3 & \beta_{30} & \beta_{31} & \beta_{32} & \\
\hline
 & q_0 & q_1 & q_2 & q_3 & r_0 & r_1 & r_2 & r_3
\end{array}
$$

Using conditions [13.9b], let us try to derive equations for 21 unknown parameters α_{ij}, β_{ij}, q_i, r_i. Why do we consider the conditions for Stratonovich, and not for Itô equations? As we shall see, for an Itô equation, there is no collection of parameters satisfying conditions [13.9a]. However, we can apply an RK method for an Itô equation by solving the equivalent Stratonovich equation with the coefficient $\tilde{b} = b - \frac{1}{2}\sigma\sigma'(x)$ instead of b. Using one derivative of the coefficient σ is better than using two derivatives of both coefficients σ and b, as in just considered Milstein approximation.

We first calculate the partial derivatives a'_y, a'_s, a''_{yy}, a''_{ys}, a'''_{yyy}, a''_{ss}, a'''_{yys}, and a''''_{yyyy} of a involved in conditions [13.9b]. "Manual" calculation would be rather tedious; happily, the MAPLE package does this job perfectly. Here, we do not write the derivatives themselves, since their expressions span multiple pages! We only need the values of the derivatives at the points $\bar{x} = (x, 0, 0)$, where a major part of the derivatives vanish. We also gather the expressions obtained into three groups. First, we write the partial derivatives with respect to y, then with respect to s, and, finally, the mixed partial derivatives:

$$
\begin{aligned}
a'_y(\bar{x}) &= (r_0 + r_1 + r_2 + r_3)\sigma(x), \\
a''_{yy}(\bar{x}) &= 2(r_1\beta_1 + r_2\beta_2 + r_3\beta_3)\sigma\sigma'(x), \\
a'''_{yyy}(\bar{x}) &= 3\big(r_1\beta_1^2 + r_2\beta_2^2 + r_3\beta_3^2\big)\sigma^2\sigma''(x) \\
&\quad + 6(r_2\beta_{21}\beta_1 + r_3\beta_{31}\beta_1 + r_3\beta_{32}\beta_2)\sigma\sigma'^2(x), \\
a''''_{yyyy}(\bar{x}) &= 4\big(r_1\beta_1^3 + r_2\beta_2^3 + r_3\beta_3^3\big)\sigma^3\sigma'''(x) \\
&\quad + \big[12\big(r_2\beta_{21}\beta_1^2 + r_3\beta_{31}\beta_1^2 + r_3\beta_{32}\beta_2^2\big) \\
&\quad + 24(r_2\beta_{21}\beta_1\beta_2 + r_3\beta_{31}\beta_1\beta_3 + r_3\beta_{32}\beta_2\beta_3)\big]\sigma^2\sigma'\sigma''(x)x \\
&\quad + 24(r_3\beta_{32}\beta_{21}\beta_1)\sigma\sigma'^3(x);
\end{aligned}
$$

10. John Butcher.

$$
\begin{aligned}
a'_s(\bar{x}) &= (q_0 + q_1 + q_2 + q_3)b(x),\\
a''_{ss}(\bar{x}) &= 2(q_1\alpha_1 + q_2\alpha_2 + q_3\alpha_3)bb'(x);\\
a''_{ys}(\bar{x}) &= 2(r_0\alpha_0 + r_1\alpha_1 + r_2\alpha_2 + r_3\alpha_3)b\sigma'(x)\\
&\quad + 2(q_1\beta_1 + q_2\beta_2 + q_3\beta_3)b'\sigma(x),\\
a'''_{yys}(\bar{x}) &= 2\big[r_1\beta_1\alpha_0 + r_2(\beta_{20}\alpha_0 + \beta_{21}\alpha_1)\\
&\quad + r_3(\beta_{30}\alpha_0 + \beta_{31}\alpha_1 + \beta_{32}\alpha_2)\big]b\sigma'^2(x)\\
&\quad + 2(r_1\alpha_1\beta_1 + r_2\alpha_2\beta_2 + r_3\alpha_3\beta_3)b\sigma\sigma''(x)\\
&\quad + 2(q_2\beta_{21}\beta_1 + q_3\beta_{31}\beta_1 + q_3\beta_{32}\beta_2\\
&\quad + r_2\alpha_{21}\beta_1 + r_3\alpha_{31}\beta_1 + r_3\alpha_{32}\beta_2)b'\sigma\sigma'(x)\\
&\quad + 2\big(q_1\beta_1^2 + q_2\beta_2^2 + q_3\beta_3^2\big)b''\sigma^2(x).
\end{aligned}
$$

Note that in the derivatives with respect to y, only the coefficient σ is involved, with respect to s, only b, while in mixed derivatives, we see both coefficients. Inserting all values of partial derivatives into equations [13.9b], we get the following 15 equations for the parameters of an RK approximation:

$$r_0 + r_1 + r_2 + r_3 = 1, \qquad [13.11a]$$

$$r_1\beta_1 + r_2\beta_2 + r_3\beta_3 = \frac{1}{2}, \qquad [13.11b]$$

$$1\beta_1^2 + r_2\beta_2^2 + r_3\beta_3^2 = \frac{1}{3}, \qquad [13.11c]$$

$$r_2\beta_{21}\beta_1 + r_3\beta_{31}\beta_1 + r_3\beta_{32}\beta_2 = \frac{1}{6}, \qquad [13.11d]$$

$$r_1\beta_1^3 + r_2\beta_2^3 + r_3\beta_3^3 = \frac{1}{4}, \qquad [13.11e]$$

$$
\begin{aligned}
&\big(r_2\beta_{21}\beta_1^2 + r_3\beta_{31}\beta_1^2 + r_3\beta_{32}\beta_2^2\big)\\
&\quad + 2(r_2\beta_{21}\beta_1\beta_2 + r_3\beta_{31}\beta_1\beta_3 + r_3\beta_{32}\beta_2\beta_3) = \frac{1}{3}, \qquad [13.11f]
\end{aligned}
$$

$$r_3\beta_{32}\beta_{21}\beta_1 = \frac{1}{24}; \qquad [13.11g]$$

$$q_0 + q_1 + q_2 + q_3 = 1, \qquad [13.11h]$$

$$q_1\alpha_1 + q_2\alpha_2 + q_3\alpha_3 = \frac{1}{2}; \qquad [13.11i]$$

$$r_0\alpha_0 + r_1\alpha_1 + r_2\alpha_2 + r_3\alpha_3 = \frac{1}{2}, \qquad [13.11j]$$

$$q_1\beta_1 + q_2\beta_2 + q_3\beta_3 = \frac{1}{2}, \qquad [13.11k]$$

$$r_1\beta_1\alpha_0 + r_2(\beta_{20}\alpha_0 + \beta_{21}\alpha_1) + r_3(\beta_{30}\alpha_0 + \beta_{31}\alpha_1 + \beta_{32}\alpha_2) = \frac{1}{4}, \qquad [13.11l]$$

$$q_2\beta_{21}\beta_1 + q_3\beta_{31}\beta_1 + q_3\beta_{32}\beta_2 + r_2\alpha_{21}\beta_1 + r_3\alpha_{31}\beta_1 + r_3\alpha_{32}\beta_2 = \frac{1}{4}, \qquad [13.11m]$$

$$r_1\alpha_1\beta_1 + r_2\alpha_2\beta_2 + r_3\alpha_3\beta_3 = \frac{1}{4}, \qquad [13.11n]$$

$$q_1\beta_1^2 + q_2\beta_2^2 + q_3\beta_3^2 = \frac{1}{2}. \qquad [13.11o]$$

Here, as before, the equations are grouped according the "participation" of the coefficients σ and b. If, for example, the coefficient b were zero, the first seven equations [13.11a–g] would suffice, since the remaining equations would be satisfied with arbitrary values of parameters.

REMARK. Here, it becomes clear why, for Itô equations, there is no weak second-order Runge–Kutta approximation. From the second of equations [13.9a] we would have $(q_0+q_1+q_2+q_3)b(x) = b(x) - \frac{1}{2}\sigma\sigma'(x)$. Obviously, this is impossible, unless the product $\sigma\sigma'$ is proportional to the coefficient b.

How can we find a solution to this system of 15 equations? At first sight, the problem seems to be complicated. Happily, the first seven are well known in numerical methods for solving ordinary differential equations. They coincide with those that are to be satisfied by the parameters of a *fourth-order* RK method for an *ordinary* differential equation $\mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}t$! Possibly, we have already got used to similar and even coinciding results for stochastic Stratonovich and deterministic differential equations. Unfortunately, *strong* approximations cannot "brag" about such nice coincidences. Actually, instead of seven, usually, eight equations are taken that can also be applied in the time-inhomogeneous case where the coefficients depend on time ($\mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}t$), or the equation is multidimensional. More precisely, instead of equation [13.11f], the following equations are considered:

$$r_2\beta_{21}\beta_1^2 + r_3\beta_{31}\beta_1^2 + r_3\beta_{32}\beta_2^2 = \frac{1}{12}, \qquad [13.11f']$$

$$r_2\beta_{21}\beta_1\beta_2 + r_3\beta_{31}\beta_1\beta_3 + r_3\beta_{32}\beta_2\beta_3 = \frac{1}{8}. \qquad [13.11f'']$$

Note that equation [13.11f] is a linear combination of the latter two ($\frac{1}{12} + 2\cdot\frac{1}{8} = \frac{1}{3}$).

Thus, we can use what is known about the coefficients of RK methods for deterministic equations. In the literature, we can find all of the solutions of the system of eight equations [13.11a–e], [13.11f′], [13.11f″], and [13.11g]. Among them, the most popular are the *classical Runge–Kutta* method and *Kutta's* $3/8$-*rule*. Their parameters β_{ij} and r_i are given by the following Butcher arrays:

$$
\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{2} & \frac{1}{2} & & & \\
\frac{1}{2} & 0 & \frac{1}{2} & & \\
1 & 0 & 0 & 1 & \\
\hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
\end{array}
\qquad\qquad
\begin{array}{c|cccc}
0 & & & & \\
\frac{1}{3} & \frac{1}{3} & & & \\
\frac{2}{3} & -\frac{1}{3} & 1 & & \\
1 & 1 & -1 & 1 & \\
\hline
 & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array}
$$

Classical RK method　　　Kutta's $3/8$ rule

Thus, we have a good start: 10 coefficients are already known. We leave choosing the remaining 11 parameters (α_{ij} and q_i) to the computer. It makes sense to look for systems of parameters with the maximum number of zeros (the more zeros, the less computation time required). Here are the best two obtained on the basis of the classical RK method and Kutta's $3/8$-rule:

$$
\begin{array}{c|cccc|c|cccc}
1 & 1 & & & & 0 & & & & \\
\frac{1}{2} & \frac{1}{2} & & & & \frac{1}{2} & \frac{1}{2} & & & \\
0 & 0 & 0 & & & \frac{1}{2} & 0 & \frac{1}{2} & & \\
1 & 1 & 0 & 0 & & 1 & 0 & 0 & 1 & \\
\hline
 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
\end{array}
\qquad
\begin{array}{c|cccc|c|cccc}
0 & 0 & & & & 0 & & & & \\
1 & 1 & & & & \frac{1}{3} & \frac{1}{3} & & & \\
0 & 0 & 0 & & & \frac{2}{3} & -\frac{1}{3} & 1 & & \\
1 & 0 & 0 & 1 & & 1 & 1 & -1 & 1 & \\
\hline
 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & & \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & \frac{1}{8}
\end{array}
$$

Finally, we give a simulation example with the first (left) RK approximation. As before, we consider the equation $\mathrm{d}X_t = 0.5X_t\,\mathrm{d}t + \sqrt{X_t^2+1}\,\mathrm{d}B_t$, $X_0 = 0.8$, and the test function $f(x) := \operatorname{arcsinh}^4(x)$, $x \in \mathbb{R}$.; the Stratonovich form of the equation is $\mathrm{d}X_t = \sqrt{X_t^2+1} \circ \mathrm{d}B_t$, $X_0 = 0.8$. Since the drift coefficient of the latter is $b \equiv 0$, its RK approximation is particularly simple:

$$a(x, s, y) = x + \frac{1}{6}(G_0 + 2G_1 + 2G_2 + G_3)\,y,$$

where

$$G_0 = \sigma(x), \quad G_1 = \sigma\Big(x + \frac{1}{2}G_0 y\Big),$$

$$G_2 = \sigma\Big(x + \frac{1}{2}G_1 y\Big), \quad G_3 = \sigma(x + G_2 y).$$

The simulation results are shown in Figure 13.5. For comparison, the calculations are also performed with the first-order Heun approximation (see examples 13.3).

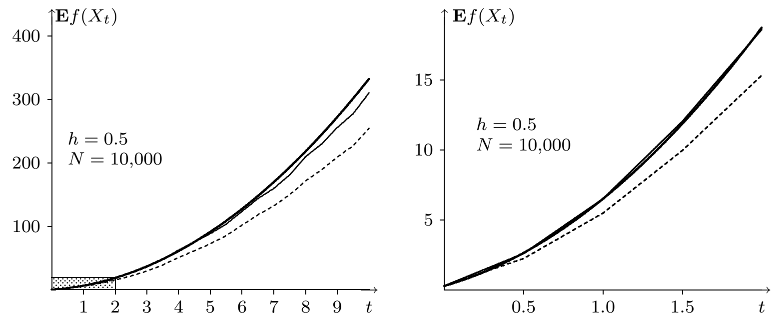


Figure 13.5. *Runge–Kutta approximations. Thick line: exact values of the expectation $\mathbf{E}f(X_t)$; thin line: second-order Runge–Kutta approximation (in the right-hand figure, the approximate expectation is almost merged with the true one); dashed line: Heun approximation*

13.8. Exercises

13.1. The solution $X_t = x\exp\{(\mu - \sigma^2/2)t + \sigma B_t\}$, $t \geqslant 0$, of the SDE $dX_t = \mu X_t\,dt + \sigma X_t\,dB_t$, $X_0 = x$, is positive for any initial condition $x > 0$.

1. Show that if $0 < \sigma^2 \leq 2\mu$, then the Milstein approximation (but not Euler) is also positive for any initial condition $x > 0$.

2. Let $2\mu < \sigma^2 < 4\mu$. Construct a Milstein-type first-order weak approximation which is positive for or any initial condition $x > 0$.

13.2. A weak approximation $\{X^h,\ h > 0\}$ of an SDE $dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t$, $X_0 = x$, is defined by

$$X_0^h = x, \quad X_{(k+1)h}^h = X_{kh}^h + b(X_{kh}^h)h + \sigma(X_{kh}^h)\xi_k\sqrt{h}, \quad k = 0, 1, 2, \dots,$$

where $\{\xi_k\}$ is a sequence of independent random variables uniformly distributed in the interval $[-a, a]$. Find the value of parameter $a > 0$ such that the approximation is a first-order one.

13.3. Solve the previous problem with $\{\xi_k\}$ such that

$$\mathbf{P}\{\xi_k = 0\} = \frac{1}{2} \quad \text{and} \quad \mathbf{P}\{\xi_k = a\} = \mathbf{P}\{\xi_k = -a\} = \frac{1}{4}.$$

13.4. Give an example of an SDE with non-zero coefficients such that its Euler approximation (as a strong approximation) is of order 1 (though, in general, it is only of order $\frac{1}{2}$).

13.5. For the SDE $dX_t = \mu X_t\,dt + \sigma X_t\,dB_t$, $X_0 = x$, consider weak approximations $\{X^h,\ h > 0\}$ of the form

$$X_0^h = x, \quad X_{(k+1)h}^h = X_{kh}^h - \mu X_{kh}^h h - \sigma X_{kh}^h \Delta B_k + a(X_{kh}^h)\Delta B_k^2, \quad k = 0, 1, 2, \ldots.$$

Choose a function $a(x)$, $x \in \mathbb{R}$, so that the approximation is a first-order one.

13.6. For an SDE $dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t$, $X_0 = x$, consider weak approximations $\{X^h,\ h > 0\}$ of the form

$$X_0^h = x, \quad X_{(k+1)h}^h = X_{kh}^h + c_k b(X_{kh}^h)h + d_k \sigma(X_{kh}^h)\Delta B_k, \quad k = 0, 1, 2, \ldots.$$

In which of the following cases it is a first-order one?

a) $c_k = 1$, $d_k = 1$;

b) $c_k = 1$, $d_k = -1$;

c) $c_k = -1$, $d_k = 1$;

d) $c_k = -1$, $d_k = -1$;

e) $c_k = 1$, $d_k = (-1)^k$.

Chapter 14

Elements of Multidimensional Stochastic Analysis

In this book, we have essentially restricted ourselves to one-dimensional processes. For the reader's convenience, in this chapter we present a short overview of the main definitions and facts in the multidimensional case. Some of them are rather simple and natural generalizations of the facts already stated, other facts can be checked without any principal difficulties, while for a deeper understanding of some facts, additional, more exhaustive sources will be necessary.

14.1. Multidimensional Brownian motion

A random process

$$B_t = (B_t^1, B_t^2, \ldots, B_t^k), \quad t \geqslant 0,$$

with values in the k-dimensional real space $\mathbb{R}^k$, such that all coordinates B^i, $i = 1, 2, \ldots, k$, are independent one-dimensional Brownian motions is called a k-dimensional (standard) Brownian motion.

The set $\mathcal{H}_t = \mathcal{H}_t^B$ of all random variables that depend only on the values of a k-dimensional Brownian B until moment t is called its history (or past) until moment t. If a random variable $X \in \mathcal{H}_t$, then $X \perp\!\!\!\perp B_u - B_s$ for all $u \geqslant s \geqslant t$. In view of remark 2 in Chapter 2, it is convenient to suppose that the history $\mathcal{H}_t^{B^i}$ of each coordinate $\mathcal{H}_t^{B^i}$ coincides with the history $\mathcal{H}_t^B$ of the whole Brownian motion B.

$\mathcal{M}$ Let $\mathcal{F}_t$ denote the σ-algebra generated by all (k-dimensional) random variables B_s, $s \leqslant t$, and zero-probability events. Then the history $\mathcal{H}_t$ of k-dimensional Brownian motion is the set of all $\mathcal{F}_t$-measurable random variables.

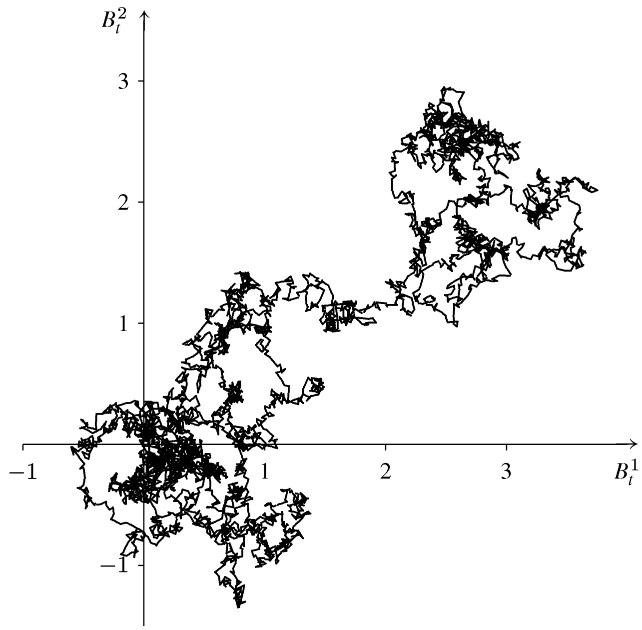


Figure 14.1. *Typical trajectory of a two-dimensional Brownian motion* $B_t = (B_t^1, B_t^2)$*,* $t \in [0, 5]$*; its values are generated with step* $\Delta t = 0.001$

As before, a (real) random process X is called adapted (to Brownian motion B) in the interval $I = [0, T]$ (or $I = [0, \infty)$) if $X_t \in \mathcal{H}_t^B$ for all $t \in I$. As already mentioned, we may suppose that $\mathcal{H}_t^{B^i} = \mathcal{H}_t^B$, and thus, if X is adapted to a k-dimensional Brownian motion B, it is adapted to every coordinate B^i. Therefore, for all stochastic integrals $\int_0^t X_s \, \mathrm{d}B_s^i$, all definitions and statements of Chapter 3 hold without any changes.

14.2. Itô's formula for a multidimensional Brownian motion

If F is of class $C^2([0, T] \times \mathbb{R}^k)$, then

$$F(T, B_T) - F(0, B_0)$$
$$= \sum_{i=1}^{k} \int_0^T \frac{\partial F}{\partial x_i}(t, B_t) \, \mathrm{d}B_t^i + \int_0^T \frac{\partial F}{\partial t}(t, B_t) \, \mathrm{d}t + \frac{1}{2} \sum_{i=1}^{k} \int_0^T \frac{\partial^2 F}{\partial x_i^2}(t, B_t) \, \mathrm{d}t.$$

14.3. Stochastic differential equations

Suppose that we are given the functions $b_i\colon\ \mathbb{R}^m \times I \to \mathbb{R}$ and $\sigma_{ij}\colon\ \mathbb{R}^m \times I \to \mathbb{R}$, $i = 1,\dots,m$, $j = 1,\dots,k$ ($I = [0,T]$ or $I = [0,\infty)$). Consider the system of stochastic differential equations

$$X_t^i = x_0^i + \int_0^t b_i(X_s,s)\,\mathrm{d}s + \sum_{j=1}^k \int_0^t \sigma_{ij}(X_s,s)\,\mathrm{d}B_s^j, \quad t \in I,\ i = 1,\dots,m,$$

or, in the differential form,

$$\mathrm{d}X_t^i = b_i(X_t,t)\,\mathrm{d}t + \sum_{j=1}^k \sigma_{ij}(X_t,t)\,\mathrm{d}B_t^j, \quad i = 1,\dots,m.$$

It is often convenient to write it in the vector form

$$\mathrm{d}\begin{pmatrix} X_t^1 \\ X_t^2 \\ \vdots \\ X_t^m \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}(X_t,t)\,\mathrm{d}t + \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \dots & \sigma_{mk} \end{pmatrix}(X_t,t)\,\mathrm{d}\begin{pmatrix} B_t^1 \\ B_t^2 \\ \vdots \\ B_t^k \end{pmatrix},$$

or abbreviate it to

$$\mathrm{d}X_t = b(X_t,t)\,\mathrm{d}t + \sigma(X_t,t)\,\mathrm{d}B_t$$

with vector and matrix coefficients $b = (b_i, i = 1,\dots,m)\colon\ \mathbb{R}^m \times I \to \mathbb{R}^m$ and $\sigma = (\sigma_{ij}, i = 1,\dots,m, j = 1,\dots,k)\colon\ \mathbb{R}^m \times I \to \mathbb{R}^{m\times k}$. Therefore, a system of stochastic differential equations may be interpreted as a single multidimensional stochastic differential equation.

If all the coordinates b_i of the functions b and all elements σ_{ij} of the matrix σ are Lipschitz functions, this stochastic differential equation has a unique solution.

Multidimensional stochastic differential equations not only generalize one-dimensional equations, but sometimes allow us to write in the vectorial form some scalar stochastic equations that cannot be written as one-dimensional equations, as was done in Chapter 6. For example, in some models the coefficients of a one-dimensional equation also depend on the "driving" (say, one-dimensional) Brownian motion B:

$$dY_t = b(Y_t,B_t,t)\,\mathrm{d}t + \sigma(Y_t,B_t,t)\,\mathrm{d}B_t.$$

Denoting $X_t = (X_t^1, X_t^2) := (Y_t, B_t)$, such an equation can be written as the usual two-dimensional stochastic differential equation

$$\mathrm{d}X_t = \begin{pmatrix} b(X_t,t) \\ 0 \end{pmatrix}\,\mathrm{d}t + \begin{pmatrix} \sigma(X_t,t) \\ 1 \end{pmatrix}\,\mathrm{d}B_t.$$

Deterministic higher-order normal differential equations are commonly reduced to first-order differential equation systems. We can similarly proceed in the case of stochastic equations. For example, consider the second-order equation with white noise

$$Y_t'' = b(Y_t, Y_t', t) + \sigma(Y_t, Y_t', t)\,\eta_t.$$

Denoting $X_t = (Y_t, Y_t')$, we can rewrite this equation as the first-order differential equation system

$$\begin{aligned}
\big(X^1\big)_t' &= X_t^2,\\
\big(X^2\big)_t' &= b\big(X_t^1, X_t^2, t\big) + \sigma\big(X_t^1, X_t^2, t\big)\,\eta_t,
\end{aligned}$$

which is modeled by the two-dimensional stochastic differential equation with one-dimensional driving Brownian motion

$$\mathrm{d}X_t = \begin{pmatrix} X_t^2 \\ b(X_t, t) \end{pmatrix} \mathrm{d}t + \begin{pmatrix} 0 \\ \sigma(X_t, t) \end{pmatrix} \mathrm{d}B_t.$$

14.4. Itô processes

An adapted random process X_t, $t \in [0, T]$, is called an *Itô process* (or *diffusion-type process*) if it can be expressed in the form

$$X_t = X_0 + \int_0^t K_s\,\mathrm{d}s + \sum_{i=1}^k \int_0^t H_s^i\,\mathrm{d}B_s^i, \quad t \in [0, T], \qquad [14.1]$$

where K and H^i, $i = 1, \dots, k$, are adapted random processes such that the above integrals exist (a.s.), i.e.

$$\int_0^T |K_s|\,\mathrm{d}s < +\infty \quad \text{and} \quad \int_0^T \big(H^i\big)_s^2\,\mathrm{d}s < +\infty \quad \text{a.s.}$$

In this case, we say that the process X has the stochastic differential

$$\mathrm{d}X_t = K_t\,\mathrm{d}t + \sum_{i=1}^k H_t^i\,\mathrm{d}B_t^i$$

abbreviated to

$$\mathrm{d}X = K\,\mathrm{d}t + \sum_{i=1}^k H^i\,\mathrm{d}B^i.$$

In particular, when all $H^i = 0$, we say that the Itô process X is *regular*.

The stochastic (Itô) integral of an adapted stochastic process Z_t, $t \in [0,T]$, with respect to an Itô process X (expressed in the form [14.1]) is the random process

$$Z \bullet X_t = \int_0^t Z_s \, \mathrm{d}X_s := \int_0^t Z_s K_s \, \mathrm{d}s + \sum_{i=1}^k \int_0^t Z_s H_s^i \, \mathrm{d}B_s^i, \quad t \in [0,T],$$

provided that the integrals on the right-hand side exist for all $t \in [0,T]$. Note that, in this case, the stochastic integral $Z \bullet X$ is also an Itô process.

The *covariation* of two Itô processes

$$X_t = X_0 + \int_0^t K_s \, \mathrm{d}s + \sum_{i=1}^k \int_0^t H_s^i \, \mathrm{d}B_s^i \quad \text{and} \quad Y_t = Y_0 + \int_0^t \widetilde{K}_s \, \mathrm{d}s + \sum_{i=1}^k \int_0^t \widetilde{H}_s^i \, \mathrm{d}B_s^i$$

is the random process $\langle X, Y \rangle_t := \sum_{i=1}^k \int_0^t H_s^i \widetilde{H}_s^i \, ds$, $t \in [0,T]$. The process $\langle X \rangle_t := \langle X, X \rangle_t$, $t \in [0,T]$, is called the *quadratic variation* of an Itô process X. Note that the covariation $\langle X, Y \rangle = 0$ if at least one of Itô processes X and Y is regular.

With these definitions at hand, all the properties of the stochastic integral with respect to an Itô process (see Chapter 7) remain valid. The definition and properties of the Stratonovich integral (Chapter 8) also remain unchanged.

As in the one-dimensional case, we can consider multidimensional stochastic equations (equation systems), where the stochastic integrals are understood in the Stratonovich case:

$$X_t^i = x_0^i + \int_0^t b_i(X_s, s) \, \mathrm{d}s + \sum_{j=1}^k \int_0^t \sigma_{ij}(X_s, s) \circ \mathrm{d}B_s^j, \quad t \in I, \ i = 1, \ldots, m,$$

or, in the matrix form,

$$\mathrm{d}X_t = b(X_t, t) \, \mathrm{d}t + \sigma(X_t, t) \circ \mathrm{d}B_t.$$

If the coefficients $\sigma_{ij} \in C^2$, this equation is equivalent to the multidimensional Itô equation

$$\mathrm{d}X_t = \tilde{b}(X_t, t) \, \mathrm{d}t + \sigma(X_t, t) \, \mathrm{d}B_t,$$

where the drift coefficient $\tilde{b} = (\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_m)$ is defined by

$$\tilde{b}_i = b_i + \frac{1}{2} \sum_{l=1}^m \sum_{j=1}^k \sigma_{lj} \frac{\partial \sigma_{ij}}{\partial x_l}, \quad i = 1, \ldots, m.$$

14.5. Itô's formula for multidimensional Itô processes

If

$$X_t^i = X_0^i + \int_0^t K_s^i \, \mathrm{d}s + \sum_{l=1}^{k} \int_0^t H_s^{il} \, \mathrm{d}B_s^l, \quad t \in [0,T],$$

$i = 1, \ldots, m$, are Itô processes and $F \in C^2([0,T] \times \mathbb{R}^m)$, then

$$\begin{aligned}
&F(T, X_T) - F(0, X_0) \\
&= \sum_{i=1}^{m} \int_0^T \frac{\partial F}{\partial x_i}(t, X_t) \, \mathrm{d}X_t^i + \int_0^T \frac{\partial F}{\partial t}(t, X_t) \, \mathrm{d}t \\
&\quad + \frac{1}{2} \sum_{i,j=1}^{m} \int_0^T \frac{\partial^2 F}{\partial x_i \partial x_j}(t, X_t) \, \mathrm{d}\langle X^i, X^j \rangle_t \\
&= \sum_{i=1}^{m} \sum_{l=1}^{k} \int_0^T \frac{\partial F}{\partial x_i}(t, X_t) H_t^{il} \, \mathrm{d}B_t^l + \sum_{i=1}^{m} \int_0^T \frac{\partial F}{\partial x_i}(t, X_t) K_t^i \, \mathrm{d}t \\
&\quad + \int_0^T \frac{\partial F}{\partial t}(t, X_t) \, \mathrm{d}t + \frac{1}{2} \sum_{i,j=1}^{m} \sum_{l=1}^{k} \int_0^T \frac{\partial^2 F}{\partial x_i x_j}(t, X_t) H_t^{il} H_t^{jl} \, \mathrm{d}t.
\end{aligned}$$

If, moreover, $F \in C^{2,3}([0,T] \times \mathbb{R}^m)$, then this formula can be written in terms of the Stratonovich integral:

$$\begin{aligned}
&F(T, X_T) - F(0, X_0) \\
&= \sum_{i=1}^{m} \int_0^T \frac{\partial F}{\partial x_i}(t, X_t) \circ \mathrm{d}X_t^i + \int_0^T \frac{\partial F}{\partial t}(t, X_t) \, \mathrm{d}t \\
&= \sum_{i=1}^{m} \sum_{l=1}^{k} \int_0^T \frac{\partial F}{\partial x_i}(t, X_t) H_t^{il} \circ \mathrm{d}B_t^l + \sum_{i=1}^{m} \int_0^T \frac{\partial F}{\partial x_i}(t, X_t) K_t^i \, \mathrm{d}t + \int_0^T \frac{\partial F}{\partial t}(t, X_t) \, \mathrm{d}t.
\end{aligned}$$

14.6. Linear stochastic differential equations

The general linear stochastic differential equation is

$$\mathrm{d}X_t = \big(a_1(t) X_t + a_2(t)\big) \, \mathrm{d}t + \sum_{j=1}^{k} \big(b_{1j}(t) X_t + b_{2j}(t)\big) \, \mathrm{d}B_t^j,$$

or, in a coordinate-wise form,

$$\mathrm{d}X_t^i = \left(\sum_{p=1}^{m} a_1^{ip}(t) X_t^p + a_2^i(t) \right) \mathrm{d}t + \sum_{j=1}^{k} \left(\sum_{p=1}^{m} b_{1j}^{ip}(t) X_t^p + b_{2j}^i(t) \right) \mathrm{d}B_t^j,$$

$i = 1, \dots, m$, where a_1 and b_{1j}, $j = 1, \dots, k$, are (non-random) $m \times m$ matrix functions, and a_2 and b_{2j}, $j = 1, \dots, k$, are m-dimensional vector functions. Its solution is the m-dimensional random process

$$X_t = \Phi_t \left\{ X_0 + \int_0^t \Phi_s^{-1} \left(a_2(s) - \sum_{j=1}^{k} b_{1j}(s) b_{2j}(s) \right) \mathrm{d}s + \sum_{j=1}^{k} \int_0^t \Phi_s^{-1} b_{2j}(s)\, \mathrm{d}B_s^j \right\},$$

$t \geqslant 0$, where Φ is the $m \times m$ matrix random process which is the solution of the stochastic differential equation

$$\mathrm{d}\Phi_t = a_1(t)\Phi_t\, \mathrm{d}t + \sum_{j=1}^{k} b_{1j}(t)\Phi_t\, \mathrm{d}B_t^j$$

with initial condition $\Phi_0 = E$ (unit $m \times m$ matrix). In contrast to the one-dimensional case, this has no explicit solution. However, if a_1 and b_{1j} are constants (do not depend on t) and commute, i.e. $a_1 b_{1j} = b_{1j} a_1$ and $b_{1i} b_{1j} = b_{1j} b_{1i}$ for all i, j, then[1]

$$\Phi_t = \exp\left\{ \left(a_1 - \frac{1}{2} \sum_{j=1}^{k} (b_{1j})^2 \right) t + \sum_{j=1}^{k} b_{1j} B_t^j \right\}, \quad t \geqslant 0.$$

For example, if we have an equation with several *additive* noises, that is, all b_{1j} are zero matrices, then $\Phi_t = \exp\{a_1 t\}$, and therefore the solution of such a equation is

$$X_t = \mathrm{e}^{a_1 t} \left\{ X_0 + \int_0^t \mathrm{e}^{-a_1 s} a_2(s)\, \mathrm{d}s + \sum_{j=1}^{k} \int_0^t \mathrm{e}^{-a_1 s} b_{2j}(s)\, \mathrm{d}B_s^j \right\}, \quad t \geqslant 0.$$

14.7. Diffusion processes

The solutions X^x of the stochastic differential equation

$$\mathrm{d}X_t = b(X_t)\, \mathrm{d}t + \sigma(X_t)\, \mathrm{d}B_t$$

1. The exponential of a matrix A is defined as the sum of series $\exp A = \mathrm{e}^A := \sum_{n=0}^{\infty} A^n/n!$.

with vector and matrix coefficients $b = (b_i, i = 1, \dots, m)\colon \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma = (\sigma_{ij}, i = 1, \dots, m, j = 1, \dots, k)\colon \mathbb{R}^m \to \mathbb{R}^{m\times k}$ satisfying Lipschitz condition and with initial condition $X_0^x = x$, $x \in \mathbb{R}^m$, define the time-homogeneous Markov diffusion process with values in $\mathbb{R}^m$. Its generator (or infinitesimal operator) A, defined, as before, by

$$Af(x) = \lim_{t\downarrow 0} \frac{T_t f(x) - f(x)}{t} = \lim_{t\downarrow 0} \frac{\mathbf{E} f(X_t^x) - f(x)}{t}, \quad x \in \mathbb{R}^m,$$

is

$$Af(x) = \sum_{i=1}^{m} b_i(x) \frac{\partial f}{\partial x_i}(x) + \frac{1}{2} \sum_{i,j=1}^{m} (\sigma\sigma^*)_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x), \quad x \in \mathbb{R},$$

where σ^* is the transpose matrix of σ, and $(\sigma\sigma^*)_{ij} = \sum_{l=1}^{k} \sigma_{il}\sigma^*_{lj} = \sum_{l=1}^{k} \sigma_{il}\sigma_{jl}$.

For multidimensional diffusion processes $\{X^x, x \in \mathbb{R}^m\}$, all statements characterizing their relations with partial differential equations also hold. For example, if $f \in C_b^2(\mathbb{R}^m)$, then the function

$$u(t,x) = T_t f(x) = \mathbf{E} f(X_t^x), \quad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^m,$$

satisfies the *backward Kolmogorov equation* with initial condition

$$\begin{cases} \dfrac{\partial u}{\partial t} = Au, \\ u(0,x) = f(x). \end{cases}$$

The same equation (with respect to x) is also satisfied by the transition density $p = p(t,x,y)$, $t > 0$, $x, y \in \mathbb{R}^m$, of the diffusion process (provided that it exists and is sufficiently smooth):

$$\frac{\partial p(t,x,y)}{\partial t} = A_x p(t,x,y), \quad t > 0,\ x, y \in \mathbb{R}^m.$$

Moreover, it satisfies the forward Kolmogorov (Fokker–Planck) equation

$$\frac{\partial p(t,x,y)}{\partial t} = A_y^* p(t,x,y), \quad t > 0,\ x, y \in \mathbb{R}^m,$$

where A^* is the formal adjoint of the operator A:

$$A^* f(x) = -\sum_{i=1}^{m} \frac{\partial}{\partial x_i}\big(b_i(x) f(x)\big) + \frac{1}{2} \sum_{i,j=1}^{m} \frac{\partial^2}{\partial x_i \partial x_j}\big((\sigma\sigma^*)_{ij}(x) f(x)\big), \quad x \in \mathbb{R}.$$

A stationary density $p_0 = p_0(y)$, $y \in \mathbb{R}^m$, of a multidimensional diffusion process is defined as in the one-dimensional case and also satisfies the forward Kolmogorov (Fokker–Planck) equation $A^* p_0 = 0$. The Feynman–Kac formula (Theorem 10.11) also holds with obvious replacement of $\mathbb{R}$ by $\mathbb{R}^m$.

14.8. Approximations of stochastic differential equations

The *Euler approximation* of an m-dimensional stochastic differential equation (with k-dimensional driving Brownian motion B)

$$X_t = x + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s, \quad t \in [0, T],$$

is defined by

$$X_0^h = x, \qquad X_{t_{p+1}}^h = X_{t_p}^h + b\big(t_p, X_{t_p}^h\big)h + \sigma\big(t_p, X_{t_p}^h\big)\Delta B_p,$$

$$\Delta B_p = \Delta B_p^{(h)} = B_{t_{p+1}} - B_{t_p}, \qquad t_p = ph.$$

As a strong approximation, it is of order $\frac{1}{2}$, while as a weak approximation it is of order 1. In other words,

$$\sup_{t\leqslant T} \mathbf{E}\big|X_t^h - X_t\big| = \mathrm{O}(h^{1/2}) \qquad \text{and} \qquad \mathbf{E}f(X_t^h) - \mathbf{E}f(X_t) = \mathrm{O}(h), \quad h \to 0,$$

for all $f \in C_b^4(\mathbb{R}^m)$.

The *Milstein approximation* $X^h = (X^{1h}, X^{2h}, \dots, X^{mh})$ is defined by

$$X_0^{ih} = x^i, \qquad X_{t_{p+1}}^{ih} = X_{t_p}^{ih} + b_i\big(t_p, X_{t_p}^{ih}\big)h + \sum_j \sigma_{ij}\big(t_p, X_{t_p}^{ih}\big)\Delta B_p^j$$

$$+ \sum_{j,l,q} \frac{\partial \sigma_{iq}}{\partial x_j}\sigma_{jl}\Delta C_p^{lq}, \qquad t_p = ph;$$

$$\Delta C_p^{lq} := \int_{t_p}^{t_{p+1}} \big(B_s^l - B_{t_p}^l\big)\,\mathrm{d}B_s^q.$$

It is of order 1 both as a strong and weak approximation. Because of its complexity, it does not make sense to use it as a weak approximation, since its (weak) order coincides with that of Euler. Besides, generating random variables C_p^{lq} is rather complicated for $l \neq q$. Therefore, this approximation in the multidimensional case in practise is reserved for some "symmetric" cases where the coefficients at C_p^{lq} and C_p^{ql} coincide, i.e. for all i, l, q, satisfy the so-called commutation condition

$$\sum_j \frac{\partial \sigma_{iq}}{\partial x_j}\sigma_{jl} = \sum_j \frac{\partial \sigma_{il}}{\partial x_j}\sigma_{jq}.$$

Then the generation of the random variables ΔC_p^{lq} and ΔC_p^{ql} can be avoided by using the equalities $\Delta C_k^{lq} + \Delta C_k^{ql} = \Delta B_p^l \Delta B_p^q$ for $l \neq q$ and $\Delta C_p^{lq} + \Delta C_p^{ql} = (\Delta B_p^l)^2 - h$ for $l = q$.

Solutions, Hints, and Answers

Chapter 2 exercises

2.1. For $s \geqslant t$, use the independence of B_t and $B_s - B_t$:

$$\mathbf{E}B_tB_s = \mathbf{E}B_t[B_s - B_t) + B_t] = \mathbf{E}B_t\,\mathbf{E}(B_s - B_t) + \mathbf{E}B_t^2 = 0 \cdot 0 + t = t,$$

$$\mathbf{E}B_t^2B_s^2 = \mathbf{E}B_t^2[(B_s - B_t) + B_t]^2 = \mathbf{E}B_t^2\,\mathbf{E}(B_s - B_t)^2 + 2\mathbf{E}B_t^3\,\mathbf{E}(B_s - B_t) + \mathbf{E}B_t^4,$$

and so on.

2.2. The first and second expectations are zeros because of the symmetry of distributions of the random variables. For the second expectation, use the equality

$$\mathbf{E}\Big(\int_0^t B_s\,\mathrm{d}s\Big)^2 = \mathbf{E}\Big(\int_0^t B_s\,\mathrm{d}s \int_0^t B_u\,\mathrm{d}u\Big) = \int_0^t\int_0^t \mathbf{E}B_sB_u\,\mathrm{d}s\,\mathrm{d}u$$
$$= \int_0^t\int_0^t s \wedge u\,\mathrm{d}s\,\mathrm{d}u.$$

2.3. By part 2 of Theorem 2.4,

$$\mathbf{P}\{B_t \leqslant 0\;\forall\, t \in [0,T]\} = \mathbf{P}\{\sup_{t\leqslant T} B_t = 0\} = \mathbf{P}\{|B_T| = 0\} = 0.$$

2.4.

$$\mathbf{E}\sup_{t\leqslant T} B_t^2 = \mathbf{E}\max\{(\sup_{t\leqslant T} B_t)^2, (\inf_{t\leqslant T} B_t)^2\}$$
$$\leqslant \mathbf{E}(\sup_{t\leqslant T} B_t)^2 + \mathbf{E}(\inf_{t\leqslant T} B_t)^2 = \mathbf{E}|B_T|^2 + \mathbf{E}|-B_T|^2 = 2T.$$

2.5. $X_t = \xi\sqrt{t}$, where $\xi \sim N(0,1)$.

2.6. A careful reading of the proof of Theorem 2.4 shows that it also works in this case.

2.7. *Hint*: Check that W satisfies all the properties of a Brownian motion (stated in Theorem 2.3).

2.8. *Answer*: $\operatorname{cov}(W_t, B_t) = \rho t$.

2.9. *Answer*: $\operatorname{cov}(Z_t, Z_s) = t \wedge s - ts$.

2.10. Using the independence of X and Y, we have:

$$\begin{aligned}\mathbf{E}\left(\sum_i \Delta X_i \Delta Y_i\right)^2 &= \sum_{i,j} \mathbf{E}\Delta X_i^2 \mathbf{E}\Delta Y_i^2 + 2\sum_{i<j} \mathbf{E}\Delta X_i \mathbf{E}\Delta Y_i \mathbf{E}\Delta X_j \mathbf{E}\Delta Y_j \\ &= \sum_{i,j} \Delta t_i^2 + 0 \to 0.\end{aligned}$$

2.11. *Hint*: Use Exercises 2.8 and 2.9 to show that the possible limits fill the interval $[0,1]$.

Chapter 4 exercises

4.1. *Answer*: $C = 3$.

4.2. First, note that, for $n \in \mathbb{N}$ such that $n \geqslant t$,

$$\begin{aligned}\mathbf{E}\left(B_t \int_0^n f(s)\,dB_s\right) &= \mathbf{E}\left(B_t \int_0^t f(s)\,dB_s\right) + \mathbf{E}\left(B_t \int_t^n f(s)\,dB_s\right) \\ &= \mathbf{E}\left(\int_0^t dB_s \int_0^t f(s)\,dB_s\right) + 0 \\ &= \int_0^t f(s)\,ds\end{aligned}$$

and then (correctly) pass to the limit, as $n \to \infty$, on the left-hand side.

4.3. *Hint*: First, show the "partial summation" formula

$$\sum_i f_i \Delta B_i = f(t)B_t - \sum_i B_i \Delta f_i$$

(with the sums taken over a partition of the interval $[0,t]$) and then pass to the limit under the refinement of partitions.

4.4. No contradiction, since the process $\mathbb{1}_{[0,\tau]} \notin H^2[0,\infty)$, and thus, the stochastic integral $\int_0^\infty \mathbb{1}_{[0,\tau]}(t)\,\mathrm{d}B_t$ must not have zero expectation.

4.5. *Hint*: Consider the process $X = \mathbb{1}_{[0,\tau]}$ from exercise 4.5.

4.6. The integrand in the last integral, B_1, is not an adapted process, and thus, even if we define this integral as $B_1 \int_0^1 \mathrm{d}B_s$, it must not have zero expectation.

4.7. By the properties of the stochastic integral, $\mathbf{E}X = 0$ and

$$\begin{aligned}\mathbf{E}X &= \mathbf{E}\int_0^T (B_t + B_t^2)^2\,\mathrm{d}t = \int_0^T \mathbf{E}(B_t^2 + 2B_t^3 + B_t^4)\,\mathrm{d}t \\ &= \int_0^T (t + 3t^2)\,\mathrm{d}t = T^2/2 + T^3.\end{aligned}$$

4.8.

$$\mathbf{E}Y_t = \int_0^t \mathbf{E}X_s^2\,\mathrm{d}s = \int_0^t \mathbf{E}\int_0^s B_u^4\,\mathrm{d}u = \int_0^t\int_0^s 3u^2\,\mathrm{d}u\,\mathrm{d}s = t^4/4.$$

For $s \geqslant t$,

$$\mathbf{E}Y_tY_s = \mathbf{E}Y_t^2 + \mathbf{E}[Y_t(Y_s - Y_t)] = \mathbf{E}Y_t^2,$$

and so, $\mathbf{E}Y_tY_s = (t \wedge s)^4/4$.

4.9.

$$\begin{aligned}I &= \sum_k B_{k+1}\Delta B_k = \sum_k (\Delta B_k + B_k)\,\Delta B_k \\ &= \sum_k \Delta B_k^2 + \sum_k B_k\,\Delta B_k \xrightarrow{L^2} 1 + \int_0^1 B_t\,\mathrm{d}B_t = B_1^2/2 + 1/2.\end{aligned}$$

4.10. *Hint*: Apply the inequality $\mathbf{E}|X| \leqslant \sqrt{\mathbf{E}X^2}$.

Chapter 5 exercises

5.1. *Hint*: To find a stochastic integral $\int_0^t f(B_s,s)\,\mathrm{d}s$, apply Itô's formula to the antiderivative $F(x,t)$ of $f(x,t)$ with respect to x (i.e. for F such that $F'_x(x,t) = f(x,t)$).

5.2. *Hint*: Applying Itô's formula to the $\cos(2B_t)$ and taking the expectation, derive and solve a differential equation for the function $f(t) := \mathbf{E}\cos(2B_t)$.

5.3. *Hint*: Apply Itô's formula to $F(B_t)$ with $F \in C_b^2(\mathbb{R})$ and then take $X_s = F'(B_s)$.

5.4. *Hint*: Notice that, by Itô's formula, $X_t = \operatorname{arctg}(B_t)$.

Chapter 6 exercises

6.1. *Hint*: Apply Itô's formula for Brownian motion and the function $f(x) = \operatorname{tg}(\pi/4 + x)$.

6.2. *Answer*: $X_t = 1 + \frac{1}{2}\int_0^t X_s\,\mathrm{d}s + \int_0^t X_s\,\mathrm{d}B_s$.

6.3. *Hint*: Take the expectation of the equation written in the integral form and then derive and solve a differential equation for $f(t) = \mathbf{E}X_t$.

6.4. Similarly to the proof of existence in Theorem 6.4, we have:

$$\begin{aligned}
\mathbf{E}|X_t^x - X_t^{x_0}|^2 &= \mathbf{E}\Bigg[x - x_0 + \int_0^t \big(b(X_s^x,s) - b(X_s^{x_0},s)\big)\,\mathrm{d}s \\
&\qquad + \int_0^t \big(\sigma(X_s^x,s) - \sigma(X_s^{x_0},s)\big)\,\mathrm{d}B_s\Bigg]^2 \\
&\leqslant 3(x-x_0)^2 + 3\mathbf{E}\Bigg[\int_0^t \big(b(X_s^x,s) - b(X_s^{x_0},s)\big)\,\mathrm{d}s\Bigg]^2 \\
&\qquad + 3\mathbf{E}\Bigg[\int_0^t \big(\sigma(X_s^x,s) - \sigma(X_s^{x_0},s)\big)\,\mathrm{d}B_s\Bigg]^2 \\
&\leqslant \cdots \leqslant 3(x-x_0)^2 + M_t\int_0^t \mathbf{E}|X_s^x - X_s^{x_0}|^2\,\mathrm{d}s, \quad t \geqslant 0.
\end{aligned}$$

Now, by Gronwall's lemma, for $t \in [0, T]$, we have:

$$\mathbf{E}|X_t^x - X_t^{x_0}|^2 \leqslant 3(x - x_0)^2 \mathrm{e}^{M_T t} \to 0 \quad \text{as } x \to x_0.$$

6.5. Applying Itô's formula, we have:

$$\begin{aligned} X_t &= X_0 + 3\int_0^t B_s^2 \, \mathrm{d}B_s + 3\int_0^t B_s \, \mathrm{d}s \\ &= 3\int_0^t X_s^{2/3} \, \mathrm{d}B_s + 3\int_0^t X_s^{1/3} \, \mathrm{d}s. \end{aligned}$$

The solution $X_t = B_t^3$ of the equation obtained is not unique since we have at least one more solution, the trivial one $X_t = 0$, $t \geqslant 0$.

Chapter 7 exercises

7.1. By Itô's formula we have:

$$\begin{aligned} \mathrm{d}Z_t &= \mathrm{d}(X_t^2 + Y_t^2) = 2X_t \, \mathrm{d}X_t + \langle X \rangle_t + Y_t \, \mathrm{d}Y_t + \langle Y \rangle_t \\ &= 2X_t Y_t \, \mathrm{d}B_t + Y_t^2 \, \mathrm{d}t - 2Y_t X_t \, \mathrm{d}t + X_t^2 \, \mathrm{d}t = Z_t \, \mathrm{d}t, \end{aligned}$$

and thus $Z_t = Z_0 \, \mathrm{e}^t = (x^2 + y^2) \, \mathrm{e}^t$.

7.2. (1) Take X a Brownian motion B, and $X^n = B^n$ its polygonal approximation. (2) $X^n = (B^n + B)/2$.

7.3. *Answers*: (1) $A_t = \int_0^t (B_s^2 + s) \, ds$; (2) $f(s) = -2s$; (3) $b(x) = -x/2$.

7.4. (1) *Hint*: The limit is $\langle X \rangle_T$. To find it, apply Itô's formula to the process X. (2) *Answer*: 0.

7.5. *Hint*: Apply Proposition 7.8 and Itô's formula.

7.6. *Hint*: Notice that, by Itô's formula, $Y_t = \operatorname{arctg}(X_t) - \operatorname{arctg}(X_0)$.

7.7. By Itô's formula,

$$B_t^4 = 4\int_0^t B_s^3 \, \mathrm{d}B_s + 6\int_0^t B_s^2 \, \mathrm{d}s \quad \text{and} \quad B_t^2 = 2\int_0^t B_s \, \mathrm{d}B_s + \int_0^t \mathrm{d}s.$$

Therefore, $\mathbf{E}\langle B^4, B^2 \rangle_t = \mathbf{E}\int_0^t B_s^3 B_s \, \mathrm{d}s = \int_0^t \mathbf{E}B_s^4 \, \mathrm{d}s = \int_0^t 3s^2 \, \mathrm{d}s = t^3$.

7.8. $\mathbf{E}\langle X,Y\rangle_t = \mathbf{E}\langle B^3,1\rangle_t = \mathbf{E}\int_0^t 3B_s^2\,\mathrm{d}s = t^3$.

7.9. *Hint*: Apply Itô's formula to the process $Y_t = f(X_t)$ and the definition of the quadratic variation of an Itô process.

7.10. *Hint*: Apply Itô's formula to the process $Y_t = e^{-X_t^2}$.

7.11. By Itô's formula,

$$X_t^2 = x^2 + 2\int_0^t X_s\,\mathrm{d}X_s + \langle X\rangle_t = x^2 + 2\int_0^t X_s\sqrt{X_s}\,\mathrm{d}B_s + \int_0^t X_s\,\mathrm{d}s$$

$$\implies \mathbf{E}X_t^2 = x^2 + \int_0^t \mathbf{E}X_s\,\mathrm{d}s = x^2 + xt;$$

$$X_t^3 = x^3 + 3\int_0^t X_s^2\,\mathrm{d}X_s + 3\int_0^t X_s\,\mathrm{d}\langle X\rangle_s$$

$$= x^3 + 2\int_0^t X_s^2\sqrt{X_s}\,\mathrm{d}B_s + 3\int_0^t X_s^2\,\mathrm{d}s$$

$$\implies \mathbf{E}X_t^3 = x^3 + \int_0^t \mathbf{E}X_s^2\,\mathrm{d}s = x^3 + 3x^2t + \frac{3}{2}xt^2.$$

7.12. *Answer:* $\mathbf{E}X_t^2 = \frac{3}{2} + 3(x-1)e^{-t} + (x^2 - 3x + \frac{3}{2})e^{-2t}$.

Chapter 8 exercises

8.1.

$$B\circ(B\circ B)_t = B^2\circ B_t = \frac{B_t^3}{3};$$

$$\mathbf{E}\int_0^t B_s^3\circ dY_s = \mathbf{E}\int_0^t B_s^3\circ\mathrm{d}(B^2\circ B)_t = \mathbf{E}\int_0^t B_s^5\circ\mathrm{d}B_t = \mathbf{E}\left(\frac{B_t^6}{6}\right) = \frac{5t^3}{2};$$

$$\mathbf{E}\int_0^t X_s\circ\mathrm{d}Y_s = \mathbf{E}\int_0^t X_s\,\mathrm{d}Y_s = \mathbf{E}\int_0^t X_s\,s\,\mathrm{d}B_s = 0.$$

8.2. $\mathrm{d}X_t = -X_t\,\mathrm{d}t + \sigma(X_t)\circ\mathrm{d}B_t = \big(-X_t + \frac{1}{2}\sigma\sigma'(X_t)\big)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}B_t$. Therefore, the solution is a martingale if $\frac{1}{2}\sigma\sigma'(x) = x$, i.e. $(\sigma^2)'(x) = x \implies \sigma^2(x) = \frac{x^2}{2} + C$. So, for example, we can take $\sigma(x) = \sqrt{\frac{x^2}{2} + a}$ with arbitrary $a > 0$.

8.3. Rewrite the equation in the Stratonovich form

$$\mathrm{d}X_t^n = \frac{1}{2}X_t^n\,\mathrm{d}t + X_t^n\circ\mathrm{d}(B^n - B)_t, \qquad X_0^n = 1.$$

By Corollary 8.8 we find that the limit X satisfies $\mathrm{d}X_t = \frac{1}{2} X_t \, \mathrm{d}t$ with $X_0 = 1$, and thus $X_t = \mathrm{e}^{t/2}$.

8.4. *Answer*: All the cases are possible.

8.5. *Hint*: Write the equation in the Itô form and use Example 6.2.

Chapter 9 exercises

9.1. *Answer*: $\mathbf{E}X_t^3 = x^3$, $t \geqslant 0$.

9.2. *Hint*: Apply Itô's formula.

9.3. *Hint*: Consider $Y_t = X_t/(1 - X_t)$.

9.4. *Hint*: Apply equation [9.4].

9.5. *Hint*: By the change of variable $Y_t = 1/X_t^3$ reduce the equation to a linear one.

9.6. *Answer*: a) $X_t = \operatorname{arctg} B_t$; b) $X_t = \frac{\pi}{2}$; c) $X_t = \frac{3\pi}{2} + \operatorname{arctg} B_t$.

9.7. Use Proposition 9.9 to get $W = \mathcal{E}(-\langle X \rangle)$; thus, the equation is $\mathrm{d}W_t = -W_t \, \mathrm{d}\langle X \rangle_t$ (with $W_0 = 1$).

9.8. *Hint*: Apply Theorem 9.13.

Chapter 10 exercises

10.1. *Hint*: Using Itô's formula, write an SDE for X and then use the drift and diffusion coefficients to write the generator.

10.2. *Hint*: Apply Corollary 10.8 (see also example 10.10).

10.3. *Hint*: Apply Corollary 10.8.

10.4. *Answer*: $\lambda = 3/2$.

10.5. *Hint*: The limit is the stationary density of the diffusion.
Answer: $\lim_{t \to \infty} p_t(x) = \frac{1}{\sqrt{\pi}} \mathrm{e}^{-x^2}$.

10.6. *Answer*: $\alpha < \frac{1}{2}$.

10.7. The equation

$$\frac{N}{\sigma^2(x)}\exp\left\{2\int\frac{b(x)}{\sigma^2(x)}\,\mathrm{d}x\right\}=1$$

with, say, $N=1$ reduces to the relation $2b(x)=(\sigma^2)'(x)$. So, we can take, for example, $\sigma(x)=x(1-x)$ and $b(x)=\frac{1}{2}x(1-x)(1-2x)$.

Chapter 12 exercises

12.1. *Answer*: $S_0\Phi\big(d_1(T,S_0)\big)$ and $\mathrm{e}^{-rT}\Phi\big(d_2(T,S_0)\big)$, respectively, with the same notation as in the Black–Scholes formula [12.13].

12.2. *Hint*: Apply the simple relation $(S_T-K)^+-(K-S_T)^+=S_T-K$.

12.3. *Hint*: Show first that

$$\lim_{\sigma\downarrow 0}d_i=+\infty\cdot\operatorname{sgn}(\ln(x/K)+rT),\quad i=1,2,$$

and

$$\lim_{\sigma\to+\infty}d_1=+\infty,\quad \lim_{\sigma\to+\infty}d_2=-\infty.$$

12.4. *Hint*: In this case, as before, we have to calculate $\mathrm{e}^{-rT}\widetilde{\mathbf{E}}(S_T-K)^+$, where, instead of $S_T=S_0\exp\left\{\sigma\widetilde{B}_T+(r-\sigma^2/2)T\right\}$, we now have

$$S_T=S_0\exp\left\{\int_0^T\sigma(u)\,\mathrm{d}B_u-\frac{1}{2}\int_0^T\sigma^2(u)\,\mathrm{d}u+rT\right\};$$

that is, the normal random variable $\xi=\sigma\widetilde{B}_T+(r-\sigma^2/2)T\sim N(rT,\sigma^2T)$ has to be replaced by the normal random variable

$$\tilde{\xi}=\int_0^T\sigma(u)\,\mathrm{d}B_u-\frac{1}{2}\int_0^T\sigma^2(u)\,\mathrm{d}u+rT\sim N(rT,\int_0^T\sigma^2(u)\,\mathrm{d}u).$$

Answer: $D_1=\frac{\ln(x/K)+rT+\Sigma^2/2}{\Sigma}$, $\quad D_2=d_1-\Sigma$, where $\Sigma=\int_0^T\sigma^2(u)\,\mathrm{d}u$.

Chapter 13 exercises

13.1. 1. The increment function for the Milstein approximation is

$$
\begin{aligned}
a(x, s, y) &= x + \mu x s + \sigma x y + \frac{1}{2}\sigma^2 x(y^2 - s) \\
&= x\left((\mu - \frac{1}{2}\sigma^2)s + 1 + \sigma y + \frac{1}{2}\sigma^2 y^2\right) > 0,
\end{aligned}
$$

provided that $0 < \sigma^2 < 2\mu$.

2. *Answer*: $a(x, s, y) = x + \sigma x y + \mu y^2$.

13.2 and 13.3. *Hint*: The second moment of ξ_k must equal 1, the same as that of a standard random variable (the second and third moments are zeros for all values of a).

13.4. *Hint*: For an SDE with constant (non-zero) diffusion coefficient, the Milstein and Euler approximations coincide.

13.5. *Answer*: $a(x) = 2\mu x$.

13.6. *Answer*: a), b), e).

Bibliography

[DAN 07] R.-A. DANA, M. JEANBLANC, *Financial Markets in Continuous Time*, Springer, Dordrecht Heidelberg, 2007.

[GAR 04] C.W. GARDINER, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer Series in Synergetics), 3rd edition, Springer, 2004.

[HID 80] T. HIDA, *Brownian Motion*, Springer, 1980.

[HOR 06] W. HORSTHEMKE, R. LEFEVER, *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology* (Springer Series in Synergetics), Springer, 2006.

[KLO 80] P. E. KLOEDEN, E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Springer, 1980.

[LAM 96] D. LAMBERTON, B. LAPEYERE, *Introduction to Stochastic Calculus Applied to Finance*, Chapman & Hall, London, 1996.

[MAC 05] V. MACKEVIČIUS, *Stochastic Analysis: Stochastic Integrals and Stochastic Differential Equations*, Vilnius University Press, 2005 (in Lithuanian).

[MIK 99] T. MIKOSCH, *Elementary Stochastic Calculus with Finance in View*, World Scientific Publishing, Singapore, 1999.

[ØKS 80] B. ØKSENDAL, *Stochastic Differential Equations*, Springer, 1980.

[PHA 07] H. PHAM, *Introduction aux Mathématiques et Modèles Stochastiques des Marchés Financiers*, Lecture Notes, University of Paris 7, Version: 2006–2007.

[PRO 90] P. PROTTER, *Stochastic Integration and Differential Equations: A New Approach*, Springer, 1990.

[SCH 80] Z. SCHUSS, *Theory and Applications of Stochastic Differential Equations*, John Wiley & Sons, 1980.

[SKO 96] A.V. SKOROKHOD, *Lectures on the Theory of Stochastic Processes*, VSP, Utrecht, 1996.

Index