Lecture Notes in Mathematics 2067

Frederik Herzberg

Stochastic Calculus with Infinitesimals

Springer

Lecture Notes in Mathematics 2067

For further volumes:
http://www.springer.com/series/304

Frederik S. Herzberg

Stochastic Calculus with Infinitesimals

Springer

Frederik S. Herzberg
Bielefeld University
Institute of Mathematical Economics
Bielefeld, Germany

and

Ludwig Maximilian University of Munich
Munich Center for Mathematical Philosophy
Munich, Germany

ISBN 978-3-642-33148-0 ISBN 978-3-642-33149-7 (eBook)
DOI 10.1007/978-3-642-33149-7
Springer Heidelberg New York Dordrecht London

Lecture Notes in Mathematics ISSN print edition: 0075-8434
ISSN electronic edition: 1617-9692

Library of Congress Control Number: 2012951626

Mathematics Subject Classification (2010): 03H05; 60G05; 91B25; 81Q30; 60G51; 60H05; 60H10

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Dedicated to Professor Edward Nelson on the occasion of his eightieth birthday in 2012

Abstract

This short monograph develops basic stochastic analysis—including Itô's formula, Girsanov's theorem, the Feynman–Kac formula, and results about Lévy processes with finite-variation jump part—and select applications in the framework of Edward Nelson's *Radically elementary probability theory* [Annals of Mathematics Studies, 117, Princeton, NJ: Princeton University Press, 1987]. This approach requires neither measure-theoretic probability theory nor functional analysis, but is based on a rigorous, yet elementary use of unlimited natural numbers and infinitesimals.

The underlying axiomatic framework, a modest subsystem of Nelson's Internal Set Theory (IST) [Bulletin of the American Mathematical Society, 83(6):1165–1198, 1977] and hence called Minimal Internal Set Theory, is truly elementary and can be easily motivated through the incompleteness of the Peano axioms or an ultrapower construction. (As a subsystem of IST, it is also conservative over—and hence consistent relative to—conventional mathematics, i.e. ZFC; moreover, a substantial fragment of it also admits an accessible relative consistency proof.)

In an excursion, the "radically elementary" approach to stochastic analysis will be employed to provide a "radically elementary" proof of the fundamental theorems of asset pricing. As an example for applications of Minimal Internal Set Theory in mathematical physics, a fully rigorous "radically elementary" definition of the Feynman path integral is proposed.

All these features recommend Minimal Internal Set Theory as a suitable framework for teaching stochastic analysis to finance or physics students without previous training in pure mathematics. The book is self-contained and written in expository style; in particular, it assumes no prior knowledge of nonstandard analysis.

Keywords Internal Set Theory; Infinitesimals; Nonstandard analysis; Itô's formula; Girsanov's theorem; Dynkin's formula; Feynman–Kac formula; Lévy processes; Fundamental theorems of asset pricing; Feynman path integral

Preface

This work continues Edward Nelson's programme of devising "radically elementary" approaches to analysis broadly conceived. This research agenda was initiated by Nelson in the mid-seventies through the invention of Internal Set Theory (**IST**) [59] and reached a first climax with the publication of *Radically Elementary Probability Theory*, which appeared in 1987 in the *Annals of Mathematics Studies* monograph series [60].

The objective of Nelson's 1987 monograph was to make the theory of stochastic processes (including continuous-time processes!) "readily available to anyone who can add, multiply, and reason" (from the preface [60, p. vii]) through an elementary, yet fully rigorous use of infinitesimals and unlimited numbers by invoking a very modest and easily accessible fragment of nonstandard analysis. The core concepts which make this possible are (a) the notion of a finite set with an unlimited number of elements and (b) the notion of a positive infinitesimal number; the point is that the employment of these concepts enables one to treat stochastic continuous-time phenomena as stochastic processes on finite probability spaces with discrete time lines of infinitesimal spacing.

This work extends Nelson's elementarization even to stochastic analysis, covering topics such as stochastic integration and differentiation (Itô's formula), change of measure (Girsanov's theorem), the link between diffusions and semi-elliptic partial differential equations (Dynkin's formula, Feynman–Kac formula), jump-diffusion processes (Lévy processes) as well as applications of stochastic analysis in financial economics (fundamental theorems of asset pricing), financial engineering (volatility invariance in the Black–Scholes model), and mathematical physics (rigorous definition of the Feynman path integral).

Viewed from an axiomatic perspective, we shall follow Nelson's example in assuming not just the axioms of conventional mathematics (say, Zermelo–Fraenkel set theory with Choice, **ZFC**) but also some elementary axioms that allow for basic nonstandard analysis; the resulting extension of **ZFC** will be called Minimal Internal Set Theory and is a subsystem of **IST**. Nelson [59] showed through an elaborate set-theoretic argument that **IST** is a conservative extension of **ZFC**; in Appendix A, we shall give a simple proof for the fact that at least a powerful subsystem of Minimal Internal Set Theory is a conservative extension of **ZFC** and hence consistent relative to **ZFC**. In Appendix B, the relation of Minimal Internal Set Theory to Robinsonian nonstandard analysis is briefly discussed. The remainder of the text, however, requires no acquaintance with model theory or any other part of mathematical logic whatsoever.

Munich, May 2012 Frederik S. Herzberg

Acknowledgments

Most of this short book was written while I was a Visiting Fellow, at Professor Edward Nelson's invitation, at the Mathematics Department of Princeton University, during the six months leading up to Easter 2010. My work was financially supported by a scholarship of the German Research Foundation (DFG), which I hereby acknowledge with thanks. The preparation of a revised version took place while I was a Visiting Fellow at the Munich Center for Mathematical Philosophy of Ludwig Maximilian University Munich, on a scholarship funded by the Alexander von Humboldt Foundation, which I also hereby gratefully acknowledge.

Words can hardly express how thankful I am to Edward and Sarah Jones Nelson for the warm and kind way in which they welcomed my wife and me to Princeton, for their manifold support throughout our sojourn and beyond, and not least for numerous extremely helpful discussions, both mathematically and otherwise. Furthermore, I also thank the Mathematics Department of Princeton University for their hospitality during my stay at Princeton. Moreover, I would like to thank five anonymous referees and the Editorial Board of Lecture Notes in Mathematics for their thorough perusal of this book and their many valuable suggestions. Thanks are also due to Ute McCrory and the staff at Springer for steering this manuscript through the reviewing and production process.

I take this opportunity to express my sincere gratitude to Professors Sergio Albeverio, Robert M. Anderson, Peter Koepke, Hannes Leitgeb, and Frank Riedel—both for discussions on the subject of this book and related topics as well as for their generous academic support in general over several years past.

Finally, I would like to include two more personal words of thanksgiving: to my family, above all my wonderful wife Angélique and my son Christian, for their love and faithfulness. And not least,

It is right and good
That with full hearts and minds and voices
We should praise You, Father Almighty, the unseen God,
Through Your only Son, Jesus Christ our Lord,

Who has saved us by His death, paid the price of Adam's sin,
And reconciled us once again to You.
Glory be to You forever.
Amen.

From the *Exsultet* of the Anglican Easter liturgy

Contents

Introduction

In a visionary monograph, Edward Nelson [60] has constructed the fundamental building blocks of a "radically elementary" theory of continuous-time stochastic processes, based on a simplified axiomatic version of nonstandard analysis, viz. a subsystem of Internal Set Theory (**IST**, also introduced by Nelson one decade earlier [59]). Nelson [60] extensively studied the Wiener process, including Donsker's invariance principle and Lévy's martingale characterization of the Wiener process (*nota bene:* in a single theorem [60, Theorem 18.1]), in such a "radically elementary" setting. However, he left the—significantly simpler—task of developing a radically elementary stochastic analysis from these building blocks to others.

The first and thus far only paper on radically elementary stochastic calculus was written by Benoît [10], who proved basic versions of both Itô's formula and Girsanov's theorem in a radically elementary setting. Benoît's [10] main concern, however, was the characterization of the measure induced by the Wiener walk. van den Berg [13] has authored a finance course based on radically elementary probability theory, but does not develop a fully fledged stochastic calculus therein. Moreover, after the first draft of this work had been written, the author came across the research by van den Berg and Amaro [16] who build upon Benoît's [10] work and link stochastic differential equations with partial differential equations—however, within the full framework of Internal Set Theory rather than within the framework of radically elementary probability theory, and without providing a systematic treatment of Itô diffusions.[1]

[1]There is, of course, also a significant body of research on stochastic integration and stochastic differential equations within the Robinsonian framework of nonstandard analysis (based on saturated enlargements of superstructures, cf. Robinson and Zakon [68]), starting from the seminal work of Loeb [51] and Anderson [4]. Major contributions to this area of research include those by Lindstrøm [45–48], Keisler [41], Hoover and Perkins [37, 38], Stroyan and Bayod [74], Capiński and Cutland [21–23], and Osswald [64]. A survey of some of the earlier results as well as a nonstandard approach to potential theory and the theory of Dirichlet forms can be found in the volume by Albeverio et al. [3]. The very first application of nonstandard analysis to (the foundations of) probability theory was given by Robinson's student Allen R. Bernstein and Frank

In this book, we develop basic stochastic analysis in the framework of radically elementary probability theory. First, we shall define (and briefly discuss) the axiomatic system of radically elementary probability theory. This axiom system will be a small subsystem of Nelson's Internal Set Theory [59] and thus a moderate extension of the conventional Zermelo–Fraenkel set theory including the Axiom of Choice (**ZFC**). This new axiomatic system, henceforth referred to as Minimal Internal Set Theory, comes in three variants of slightly different strength, viz. $\mathbf{minIST}^+$, **minIST** and $\mathbf{minIST}^-$, where $\mathbf{minIST}^+$ contains **minIST** and **minIST** contains $\mathbf{minIST}^-$. The results of this work will only depend on **minIST**, and much of radically elementary stochastic analysis can even be developed in $\mathbf{minIST}^-$, the weakest of these axiom systems.

A short review of radically elementary probability theory—which is nothing more than finite probability theory with the additional axioms of Minimal Internal Set Theory at hand—will follow. After defining Wiener walks, Wiener processes and recalling some important results such as the radically elementary equivalent of Lévy's characterization of Wiener processes (Nelson's "de Moivre–Laplace–Lindeberg–Feller–Wiener–Lévy–Doob–Erdős–Kac–Donsker–Prokhorov theorem" [60, Theorem 18.1]), we will present the original contributions of this work.

These new results include radically elementary versions of the martingale representation theorem, Itô's formula, Girsanov's theorem, the diffusion invariance principle, the Markov property of Itô diffusions, Dynkin's formula, and the Feynman–Kac formula. Finally, we shall propose a radically elementary theory of Lévy processes. In addition, the book includes various excursions: a radically elementary discussion of certain "geometric" Itô processes (Sect. 3.4 of Chap. 3), a radically elementary approach to the fundamental theorems of asset pricing (Chap. 5), a rigorous radically elementary definition of the Feynman path integral (Chap. 8) as well as a proof of the conservativity of $\mathbf{minIST}^-$ as an extension of **ZFC** (Appendix A). One of the excursions in this book (Chap. 8) suggests another area of application of Minimal Internal Set Theory within mathematical physics. We shall provide a rigorous, yet radically elementary definition of the Feynman path integral.

Most challenging to prove among these results is the radically elementary version of Girsanov's theorem. Just as Lévy's [44] classical martingale characterization of the Wiener process is a pivotal ingredient in the classical proof of Girsanov's theorem [27], we shall use the aforementioned radically elementary analogue of Lévy's martingale characterization of the Wiener process established by Nelson [60, Theorem 18.1]) in order to prove our radically elementary version of Girsanov's theorem.

The logical interdependence of the various parts of the book is as follows. Chapter 1 (axiomatic framework), Sect. 2.1 of Chap. 2 (Random variables and stochastic processes), Sect. 2.3 of Chap. 2 (Wiener walks), and the definitions

---

Wattenberg [17], not long after the appearance of Robinson's groundbreaking monograph *Non-standard analysis* [67].

from Sect. 3.1 of Chap. 3 are basic and will be needed throughout this work. The discussion of Lévy processes (Chap. 9) assumes Sect. 3.5 of Chap. 3 (Lévy's characterization of Wiener processes). The proof of Girsanov's theorem (Chap. 4) assumes all of Chaps. 1–3, with the sole exception of Sect. 2.4 of Chap. 2 (which is optional). In particular, none of the results in the excursions will be used elsewhere in the text. The only exception to this rule is Sect. 3.4 of Chap. 3 (the excursion on certain "geometric" Itô processes), which will be used towards the end of the excursion on the fundamental theorems of asset pricing (Chap. 5). The brief informal introduction to Lévy finance in Sect. 9.5 of Chap. 9 assumes, of course, some familiarity with mathematical finance or financial economics, such as can be found in Chap. 5.

Thus the logical interdependencies within this book, excluding the contents of Chap. 4, may be visualized as follows:

$$
\begin{tikzcd}
 & & \text{Chap. 1} \arrow[d] \arrow[rrr] & & & \text{Chap. 8} \\
 & & \text{Sects. 2.1 and 2.3 of Chap. 2} \arrow[d] & & & \\
 & & \text{def.s from Sect. 3.1 of Chap. 3} \arrow[dll] \arrow[dl] \arrow[dr] \arrow[drr] \arrow[drrr] \arrow[rrr] & & & \text{Sect. 3.4 of Chap. 3} \arrow[d] \\
\begin{array}{c}\text{Chap. 5}\\ \text{except}\\ \text{Lemma 5.9}\end{array} & \text{Chap. 6} & & \text{Chap. 7} & \text{Chap. 9} & \begin{array}{c}\text{end of Chap. 5}\\ \text{(Lemma 5.9)}\end{array}
\end{tikzcd}
$$

For Chap. 4, we have the following, very simple, chart (which we only include for the sake of completeness):

$$
\begin{tikzcd}
\text{Chap. 1} \arrow[d] \\
\text{Chap. 2 w/o Sect. 2.4} \arrow[d] \\
\text{Chap. 3} \arrow[d] \\
\text{Chap. 4}
\end{tikzcd}
$$

This work is self-contained, except for occasional references to some important results from Nelson's monograph [60], the content of which is fully described in this book. These are:

- The underspill/overspill principle [60, Theorem 5.4] (see Remark 1.1)
- The radically elementary characterization of a.s. convergence [60, Theorem 7.1] (see Remark 2.4)
- The radically elementary Radon–Nikodym theorem [60, Theorem 8.1] (see Remark 2.2)

- The radically elementary Lebesgue theorem [60, Theorem 8.2] (see Remark 2.3)
- The (near) equivalence of a.s. infinitely close processes [60, Theorem 17.2] (see Remark 2.1)
- A radically elementary martingale inequality [60, paragraph following Theorems 11.1 and 11.2] (see Remark 2.12)
- The a.s. continuity of normalized martingales with infinitesimal increments [60, paragraph following Theorem 13.1] (see Remark 3.5)
- The unified "de Moivre–Laplace–Lindeberg–Feller–Wiener–Lévy–Doob–Erdős–Kac–Donsker–Prokhorov theorem" [60, Theorem 18.1] (see Remark 3.13)

For those readers who intend to study some or all of the above results in greater detail by consulting Nelson's original text [60], we briefly summarize the logical interdependencies:

(1) The radically elementary characterization of a.s. convergence [60, Theorem 13.1] follows from the underspill/overspill principle [60, Theorem 5.4].
(2) The radically elementary Lebesgue theorem [60, Theorem 8.2] is a consequence of the radically elementary versions of the Radon–Nikodym theorem [60, Theorem 8.1] and the characterization of a.s. convergence [60, Theorem 7.1].
(3) The (near) equivalence of a.s. infinitely close processes [60, Theorem 17.2] follows from the radically elementary Lebesgue theorem [60, Theorem 8.2].
(4) The proof of the unified "de Moivre–Laplace–Lindeberg–Feller–Wiener–Lévy–Doob–Erdős–Kac–Donsker–Prokhorov theorem" uses the following results:

   (a) The underspill/overspill principle [60, Theorem 5.4]
   (b) The radically elementary Lebesgue theorem [60, Theorem 8.2]
   (c) A radically elementary supermartingale inequality [60, Theorem 11.1]
   (d) A continuity result for martingales [60, Theorem 13.1], which in turn depends on the limited-fluctuation criterion [60, Theorem 12.3] and by that means on [60, Theorem 11.1] and some upcrossing inequalities [60, Theorem 12.1–12.2]
   (e) The fact that the Lindeberg condition makes (a.s.) increments infinitesimal [60, Theorem 14.1], which depends on the radically elementary characterization of a.s. convergence [60, Theorem 7.1]
   (f) A truncation lemma for martingales satisfying the Lindeberg condition [60, Theorem 14.3], which depends again on [60, Theorem 7.1], on [60, Theorem 14.1] and on the underspill/overspill principle [60, Theorem 5.4]
   (g) The fact that a small change of the probability measure transforms a process into a (nearly) equivalent one [60, Theorem 17.1]
   (h) The (near) equivalence of a.s. infinitely close processes [60, Theorem 17.2]
   (i) The fact that near equivalence respects continuity [60, Corollary 2 to Theorem 17.3], which depends on the underspill/overspill principle [60, Theorem 5.4]

Chapter 1
Infinitesimal Calculus, Consistently and Accessibly

The most important feature of Nelson's [60] radically elementary analysis is the discretization of the continuum. The crucial step herein is the consistent use of infinitely large ("nonstandard") numbers and infinitesimals, in a manner which was first proposed by Nelson through the axiom system of Internal Set Theory [59], motivated by the groundbreaking work of Abraham Robinson [66, 67]. One decade on, Nelson [60] introduced an even more elementary, yet still very powerful, formal system, which we shall review presently.

1.1 An Accessible Axiom System for Infinitesimal Calculus: Minimal Internal Set Theory

Mathematical analysis broadly conceived (including probability theory) can be made much more intuitive if one allows for the use of *infinitesimals*—as engineers, and partially also applied mathematicians, have done for centuries. A positive infinitesimal is a number which is greater than zero, yet in some sense arbitrarily small—viz. less than $1/2$, less than $1/3$, less than $1/4$, less than $1/5$ etc. In other words, it is a number which is positive, yet smaller than the reciprocal of any *standard* natural number—wherein, of course, the term "standard" still is in need of being defined.

So, on the one hand, the mathematical community has known infinitesimals since at least the days of Leibniz,[1] and practitioners successfully use them every day. On the other hand, it is not immediately obvious how to give a rigorous definition of the predicate "standard" or equivalently of the notion of an infinitesimal.

[1]For some fascinating insights into—and some polemical comments on—the history of infinitesimals, cf. e.g., Błasczcyk et al. [19].

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_1,

While there are several approaches to accomplish this, the first modern rigorous attempt to define infinitesimals—which will serve as our first motivation—is due to Robinson [66] with a precursor by Schmieden and Laugwitz [72]. Robinson extended the real line with a huge number of additional elements so that it became a real-ordered field which also contained infinitesimals and infinitely large numbers.

The technique that Robinson employed has some similarity to the construction of the reals out of Cauchy sequences of rational numbers: (a) The new numbers that he constructed are equivalence classes of real numbers (where the equivalence relation is such that two sequences are equivalent if and only if they agree on a set to which a given non-trivial $\{0, 1\}$-valued finitely-additive measure on the set of natural numbers assigns mass 1).[2] (b) The arithmetical operations and the order relation are defined element-wise (and can be verified to be well-defined).[3] (c) The original real numbers are embedded into the new number system as equivalence classes of constant sequences.

On this Robinsonian account, the *standard* natural numbers, are images of ordinary natural numbers under the canonical embedding. For an example of an infinitesimal, just consider the μ-equivalence class of any null sequence of real numbers. If one considers μ-equivalence classes of strictly increasing sequences of natural numbers, one obtains infinitely large numbers, which nevertheless have some relation to natural numbers and will therefore be called *nonstandard natural numbers*.

We will not go further into the details of Robinson's delicate construction of which we only sketched the very basic steps. Readers who are interested in learning more about Robinson's *nonstandard analysis* and its exciting applications are encouraged to have a look at Appendix B and the references therein. Instead

---

[2] Formally, let μ be a map which assigns each set of natural numbers either 0 or 1 and is such that whenever I, J are disjoint (i.e. $I \cap J = \varnothing$), $\mu(I \cup J) = \mu(I) + \mu(J)$ and such that there is no natural number k such that $\mu(I) = 1$ if and only if $k \in I$ for all I. Two infinite sequences of real numbers $(a_n)_n, (b_n)_n$ are called *μ-equivalent*, denoted $(a_n)_n \sim_\mu (b_n)_n$ if and only if $\mu(\{n : a_n = b_n\}) = 1$. It is not difficult to show that $\sim_\mu$ is indeed an equivalence relation. The new numbers are then just μ-equivalence classes of infinite sequences of real numbers.

[3] It is also not difficult to verify that the following relation and operations are well-defined. For all sequences of real numbers $(a_n)_n, (b_n)_n$,

$$\begin{aligned}
[(a_n)_n]_{\sim_\mu} > [(b_n)_n]_{\sim_\mu} &:\Leftrightarrow \mu(\{n : a_n > b_n\}) = 1 \\
[(a_n)_n]_{\sim_\mu} + [(b_n)_n]_{\sim_\mu} &:= [(a_n + b_n)_n]_{\sim_\mu} \\
[(a_n)_n]_{\sim_\mu} - [(b_n)_n]_{\sim_\mu} &:= [(a_n - b_n)_n]_{\sim_\mu} \\
[(a_n)_n]_{\sim_\mu} [(b_n)_n]_{\sim_\mu} &:= [(a_n b_n)_n]_{\sim_\mu} \\
[(a_n)_n]_{\sim_\mu}^{-1} &:= [(1/a_n)_n]_{\sim_\mu} \text{ if } a_n \neq 0 \text{ for all } n
\end{aligned}$$

we will present a simple axiom system which captures a minimal fragment of nonstandard analysis, but is just powerful enough for our purposes of developing a stochastic calculus with infinitesimals.

In order to simplify both the presentation of the axiom system and the later material, we take an important, at first sight radical step: Henceforth, **when we refer to "real numbers" or to the set R**, **we mean (elements of) the extended number system**—which, of course, not only contains ordinary real numbers, but also other, "nonstandard" real numbers such as infinitesimals and infinitely large numbers. If we want to refer to the ordinary natural numbers, we will refer to them as **standard** natural numbers. Otherwise, **the term "natural number" can refer to a standard or nonstandard natural number, and N will be used to denote the set of all (standard and nonstandard) natural numbers in this sense.**[4]

With these conventions, we now introduce the following collection of axioms and axiom schemes, which we shall henceforth refer to—for historical reasons (see Sect. A.1 of Appendix A)—as *Minimal Internal Set Theory*, abbreviated **minIST**[5]:

- All theorems[6] of conventional mathematics are axioms of Minimal Internal Set Theory, even with the new interpretation of **N** and **R**.[7]
- 0 is standard.
- For every $n \in \mathbf{N}$, if n is standard, then $n + 1$ is standard, too.
- There exists a *nonstandard natural number* n, i.e. some $n \in \mathbf{N}$ which is not standard.

---

[4]We will hardly ever have the need to refer to standard real numbers; we will, however, often refer to limited (standard or nonstandard) real numbers (see below).

[5]The name is derived from Nelson's [59] Internal Set Theory (**IST**), of which even Minimal Internal Set Theory combined with the Sequence Principle (see footnote 5 on p. 3) is only a small subsystem, see Sect. A.1 of Appendix A. Although Nelson [62] did not state this explicitly, an axiom system such as **minIST** is most probably what he had in mind when suggesting the use of "minimal nonstandard analysis" [62, p. 30].

In his 1987 monograph on *Radically elementary probability theory* [60], Nelson proposes an axiom system which enlarges **minIST** by the following axiom scheme:

- (**Sequence Principle**) If $A(v_0, v_1)$ is any formula (which may involve the predicate "standard") with the property that for all standard natural numbers n there exists some x with $A(n, x)$, then there exists a sequence $(x_n)_{n \in \mathbf{N}}$ such that $A(n, x_n)$ holds for all standard n.

However, Nelson [60] only uses the Sequence Principle occasionally and conveniently marks those results which are proved through the Sequence Principle by an asterisk; the greater part of radically elementary probability theory—and in particular, *all results from radically elementary probability theory which we use in this book*—can be developed in **minIST**. Again, none of the results of the present work depend on the Sequence Principle.

[6]Equivalently one could write: "All axioms . . . ".

[7]This principle is known as the *Transfer Principle* of nonstandard analysis. It is beyond the scope of this book to give a rigorous justification. We only point out that the extended, nonstandard real number system was devised to preserve very simple mathematical propositions (such as "$x^2 \geq 0$ for all real numbers x") and that by a beautiful theorem due to Łoś [52] this preservation property can be shown to hold for complex mathematical propositions as well.

- (**External Induction**) If $A(v)$ is any formula of the new, extended language[8] such that $A(0)$ holds and such that $A(n)$ entails $A\,(n+1)$ for all standard n, then $A(n)$ readily holds for all standard n.

Unless explicitly stated otherwise, we will in this book always assume the axioms of Minimal Internal Set Theory (minIST).[9]

Formulae which do not involve the predicate "standard" will be called *internal*, because they can already be expressed in conventional mathematics. All other formulae (i.e. precisely those that involve the predicate "standard") are called *external*.

Note that we have not allowed that external formulae may be used to define new sets; violation of this rule is called *illegal set formation*. (Robinsonian nonstandard analysis has methods to treat external sets, too; in that framework, these sets are no longer "illegal".) For example, the usual principle of mathematical induction only pertains to internal formulae; if one wishes to prove an external formula by means of induction, one has to apply the above axiom scheme of External Induction.

1.2 Finer Classification of the Reals: Finite vs. Limited

By the first axiom of **minIST**, we just inherit all results and concepts from conventional mathematics. For instance, a set v is *finite* if and only if v is bijective to $\{0, \cdots, n-1\}$ for some $n \in \mathbf{N}$, which is then called the *cardinality* of v.

Now the number of elements of a finite set may be nonstandard. But any nonstandard natural number is greater than every standard natural number,[10] which—combined with the fact that $0, 1, 2, 3, 4, \ldots, 1000000000000, \ldots$ are all standard—shows that nonstandard natural numbers are very large indeed, yea, in

[8]Thus, the formula $A(v)$ may involve the predicate "standard"!

[9]In Appendix A, we shall consider an even weaker system than **minIST**, denoted by **minIST**$^-$, which still allows for much of radically elementary mathematics to be developed and also admits a simple relative consistency proof.

[10]This could be an instructive exercise for students interested in the foundational aspects of **minIST**. By External Induction in k, one can prove for all standard $k \in \mathbf{N}$ that if $n \in \mathbf{N}$ and $n \leq k$, then n is standard:

- For the base step of the External Induction, note that the only $n \in \mathbf{N}$ with $n \leq 0$ is 0, hence standard by an axiom of **minIST**.
- For the induction step of the External Induction note that whenever $n \in \mathbf{N}$ with $n \leq k+1$, one has
  (1) either $n \leq k$, in which case n is standard by induction hypothesis of the External Induction,
  (2) or $n = k+1$, whence n again is standard (as k is standard and successors of standard natural numbers are standard by another axiom of **minIST**).

Thus, there can be no pair of a nonstandard $n \in \mathbf{N}$ and a standard k such that one would have $n \leq k$. Hence nonstandard natural numbers are always greater than all standard natural numbers.

some sense *unlimited*. In particular, finite probability spaces can have an unlimited number of elements and thus be very rich.

Any real number x which satisfies $|x| \leq k$ for some standard k is called *limited* (denoted $|x| \ll \infty$), and any real number which is not limited is called *unlimited* (denoted $|x| \simeq \infty$). Any real number x which satisfies $|x| \leq \frac{1}{k}$ for all standard $k \neq 0$ is called *infinitesimal* (denoted $x \simeq 0$). In particular, for every nonstandard n, the reciprocal $\frac{1}{n}$ is a strictly positive infinitesimal. Given $x, y \in \mathbf{R}$, we write:

- $x \simeq y$ if and only if $x - y$ is infinitesimal,
- $x \ll y$ if and only if both $x < y$ and $x \not\simeq y$,
- $x \lessapprox y$ if and only if $x < y$ or $x \simeq y$.

Remark 1.1 (Underspill and Overspill Principles). In **minIST**, one can prove (cf. Nelson [60, Theorem 5.4, p. 18]) that there are no sets which would consist of either

- all the standard natural numbers, or
- all the nonstandard natural numbers, or
- all the limited reals, or
- all the unlimited reals, or
- all the infinitesimal reals.

This allows, for example, for the following proof principles. Let $A(x)$ be an internal formula.

- *Underspill in* **N**. If $A(n)$ holds for all nonstandard $n \in \mathbf{N}$, then also for some standard $n \in \mathbf{N}$.
- *Overspill in* **R**. If $A(x)$ holds for all infinitesimal $x \in \mathbf{R}$, then also for some non-infinitesimal $x \in \mathbf{R}$.

Chapter 2
Radically Elementary Probability Theory

2.1 Random Variables and Stochastic Processes

The expressive power of **minIST** comes from the fact that it allows for the notions of

- finite sets with unlimited cardinality, and
- finite subsets of the reals whose distance is at most an infinitesimal from every point in some non-empty open interval.

In particular, there exist—in **minIST** and even some of its subsystems[1]—finite probability spaces with a sample space Ω of unlimited cardinality, and every compact interval $[t_0, t_1]$ allows for a discrete, finite subset $\mathbf{T}'$ of infinitesimal spacing. The Cartesian product $\Omega \times \mathbf{T}'$ of such sets will still be a finite set. Radically elementary probability theory approaches continuous-time random phenomena using discrete, finite methods by studying stochastic processes $\xi : \Omega \times \mathbf{T}' \to \mathbf{R}$.

For the rest of this book, we fix a finite set Ω. Unless stated otherwise, P will denote a probability measure on the power-set of Ω. A *(real-valued) random variable* is a map $x : \Omega \to \mathbf{R}$. The expectation operator with respect to P will be denoted by E^P, or just E, if no ambiguity can arise. Similarly, the variance operator with respect to P will be denoted by Var^P, or just Var, if no ambiguity can arise. Since Ω is finite, $E[x]$ and $Var[x]$ are well-defined for all random variables x.

Let $A(\omega)$ be a formula (internal or external). We shall say that *A holds almost surely with respect to P* (abbreviated *P-a.s. A* or *a.s. A* where no ambiguity can arise) or *A holds for P-almost every $\omega \in \Omega$* (abbreviated *$A(\omega)$ for P-a.e. $\omega \in \Omega$* or just *$A(\omega)$ for a.e. $\omega \in \Omega$*) if and only if for all $\varepsilon \gg 0$ there exists some $N \subseteq \Omega$ such that $P(N) \leq \varepsilon$ whilst $A(\omega)$ holds for all $\omega \notin N$. Moreover, if A is internal, then we define the event $\{A\}$ by

$$\{A\} = \{\omega \in \Omega \ : \ A(\omega)\} ,$$

[1]Such as the system **minIST**$^-$ discussed in Appendix A.

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_2,

and one can show that $P\{A\} \simeq 1$ if and only if a.s. A.[2] Again for any internal formula $A(\omega)$, we shall say that A *with P-probability* 1 or just A if and only if $P\{A\} = 1$.

We fix a nonstandard natural number N, and we put

$$\mathbf{T} := \left\{0, \frac{1}{N}, \dots, \frac{N-1}{N}, 1\right\}, \qquad \mathrm{d}t := \frac{1}{N}.$$

The normalized counting measure on the power-set of the time line will be denoted ν:

$$\forall I \subseteq \mathbf{T} \qquad \nu(I) = \frac{\mathrm{card}(I)}{\mathrm{card}(\mathbf{T})} = \frac{\mathrm{card}(I)}{N+1}.$$

Given an arbitrary $p > 0$, the term $\mathcal{O}\left((\mathrm{d}t)^p\right)$ denotes a random variable x such that $\frac{x}{(\mathrm{d}t)^p}$ is limited, and the term $\mathfrak{o}\left((\mathrm{d}t)^p\right)$ denotes a random variable x such that $\frac{x}{(\mathrm{d}t)^p} \simeq 0$. (This is the radically elementary analogue of Landau's $\mathcal{O}$ and $\mathfrak{o}$.)

In this setting, a (real-valued) *stochastic process* is a map $\xi : \mathbf{T}' \to \mathbf{R}^\Omega$ for some $\mathbf{T}'$ such that $\mathbf{T}' = \mathbf{T} \cap [t_0, t_1]$ for some $t_0, t_1 \in \mathbf{T}$. For any such stochastic process $\xi : \mathbf{T}' \to \mathbf{R}^\Omega$, we put

$$\forall t \in \mathbf{T}' \setminus \{t_1\} \qquad \mathrm{d}\xi(t) := \xi(t + \mathrm{d}t) - \xi(t).$$

Through a slight abuse of notation, one can also view ξ as a random, real-valued *trajectory* (or *sample path*), i.e. as a map $\Omega \to \mathbf{R}^{\mathbf{T}'}$, $\omega \mapsto (\xi(t)(\omega))_{t \in \mathbf{T}'}$ so that $\xi(\omega)(t) = \xi(t)(\omega)$. The set

$$\Lambda_\xi := \{\xi(\omega) \ : \ \omega \in \Omega\} \subseteq \mathbf{R}^{\mathbf{T}'}$$

is a finite subset of $\mathbf{R}^{\mathbf{T}'}$, the *set of trajectories* of ξ.

Given a stochastic process ξ, a trajectory $\lambda \in \Lambda_\xi$ is said to be *(nearly) continuous* if and only if for all $s, t \in \mathbf{T}'$, one has $\lambda(s)(\omega) \simeq \lambda(t)(\omega)$ whenever $s \simeq t$. A stochastic process ξ is called *P-a.s. continuous* if and only if for P-a.e. $\omega \in \Omega$, the trajectory $\xi(\omega)$ is continuous.

We will often exploit the fact that every stochastic process ξ is uniquely determined by $\xi(t_0)$ and $\mathrm{d}\xi$ (and trivially vice versa), as is evident from the telescoping sum identity

[2] Proving this could be a useful exercise for students. If $P\{A\} \simeq 1$, then clearly a.s. A (simply choose $N = \Omega \setminus \{A\}$). Conversely, if a.s. A, then the set

$$M = \{n \in \mathbf{N} \ : \ \exists N \subseteq \Omega \left(P(N) \leq 1/n \ \& \ \Omega \setminus \{A\} \subseteq N\right)\}$$

contains all standard elements of $\mathbf{N}$. Since there is no set which consists of all standard natural numbers (see Remark 1.1), M must contain some nonstandard $n_0 \in \mathbf{N}$, too. But then, $P(\Omega \setminus \{A\}) \leq 1/n_0 \simeq 0$, so $P\{A\} \simeq 1$.

$$\forall s \in \mathbf{T}' \qquad \xi(s) - \xi(t_0) = \sum_{t_0 < t < s} \mathrm{d}\xi(t).$$

Let now $\Lambda \subseteq \mathbf{R}^{\mathbf{T}}$ be finite. A *functional* on Λ is a map $F : \Lambda \to \mathbf{R}$. If $\Lambda_\xi \subseteq \Lambda$ for a stochastic process $\xi : \mathbf{T} \to \mathbf{R}^\Omega$ and F is a functional on Λ, then $F(\xi)$ is defined as the random variable

$$\omega \mapsto F(\xi)(\omega) := F(\xi(\omega)) = F\left((\xi(t)(\omega))_{t\in\mathbf{T}}\right).$$

A functional F on Λ is called

- *continuous* if and only if
$$F(\lambda) \simeq F(\mu)$$
for all $\lambda, \mu \in \Lambda$ which satisfy $\lambda(t) \simeq \mu(t)$ for all $t \in \mathbf{T}$,
- *limited* if and only if $F(\lambda)$ is limited for all $\lambda \in \Lambda$.

Two stochastic processes $\xi, \eta : \mathbf{T} \to \mathbf{R}^\Omega$ are called *nearly equivalent* if and only if $E[F(\xi)] \simeq E[F(\eta)]$ for all limited continuous functionals F on $\Lambda_\xi \cup \Lambda_\eta$.

Remark 2.1. Nelson [60, Theorem 17.2, p. 73] has shown that two processes ξ and η are already nearly equivalent if a.s. $\forall t \in \mathbf{T} \quad \xi(t) \simeq \eta(t)$.

2.2 Integrability and Limitedness

Recall that given any event $A \subseteq \Omega$, the *indicator function* of A is defined as

$$\chi_A : \Omega \to \{0,1\}, \qquad \omega \mapsto \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A \end{cases}.$$

Likewise, whenever $I \subseteq \mathbf{T}$, the indicator function of $\mathbf{T}$ is the function

$$\chi_I : \mathbf{T} \to \{0,1\}, \qquad t \mapsto \begin{cases} 1, & t \in I, \\ 0, & t \notin I \end{cases}.$$

With this definition, a random variable x is $L^1(P)$ or *integrable* if and only if $E\left[|x|\chi_{\{|x|>a\}}\right] \simeq 0$ for all positive unlimited a, and x is $L^p(P)$ or *integrable of p-th order* (for any $p > 0$) if and only if $|x|^p$ is $L^1(P)$. The real number $E[|x|^p]$ is called the *p-th moment* of x.

Remark 2.2. The radically elementary Radon–Nikodym theorem (cf. Nelson [60, Theorem 8.1, p. 30]; [62, Theorem 4]) says that a random variable x is $L^1(P)$ if and only if $E[|x|]$ is limited and $E[|x|\chi_M] \simeq 0$ holds for all events $M \subseteq \Omega$ with $P(M) \simeq 0$.

Remark 2.3. If x and y are two $L^1(P)$ random variables such that $x \simeq y$ a.s., then by the radically elementary Lebesgue theorem (cf. Nelson [60, Theorem 8.2, p. 31]) $E[x] \simeq E[y]$.

The proof uses the radically elementary Radon–Nikodym theorem (Remark 2.2) and the following fact, which is interesting in its own right and whose proof uses underspill/overspill:

Remark 2.4. For any random variable z, the following are equivalent:

- $z \simeq 0$ a.s.
- For all $\lambda \gg 0$, $P\{|z| \geq \lambda\} \simeq 0$.
- There exists some $\varepsilon \simeq 0$ such that $P\{|z| \geq \varepsilon\} \simeq 0$.

(cf. Nelson [60, Theorem 7.1]).

It is easy to prove a converse of the radically elementary Lebesgue theorem:

Theorem 2.5. *Let $x : \Omega \to \mathbf{R}$. If $E[|x|] \simeq 0$, then $x \simeq 0$ P-almost surely.*

Proof. For all standard $n \in \mathbf{N}$, one has

$$\frac{1}{n} P\left\{|x| \geq \frac{1}{n}\right\} \leq E[|x|] \leq \frac{1}{n^2},$$

hence by overspill/underspill, there exists some nonstandard $n \in \mathbf{N}$ such that

$$P\left\{|x| \geq \frac{1}{n}\right\} \leq \frac{1}{n}.$$

Obviously $x \simeq 0$ on $\Omega \setminus \left\{|x| \geq \frac{1}{n}\right\}$ and $P\left\{|x| \geq \frac{1}{n}\right\} < \varepsilon$ for all $\varepsilon \gg 0$. $\square$

For the special case $\Omega = \mathbf{T} \setminus \{T\}$ and $P = \nu$, this reads:

Theorem 2.6. *Let $f : \mathbf{T} \setminus \{T\} \to \mathbf{R}$. If $\int_0^T |f(t)|\,\mathrm{d}t \simeq 0$, then $f(t) \simeq 0$ for ν-almost every t.*

As a corollary to these theorems, one arrives at:

Theorem 2.7. *Let $\xi : \Omega \times \mathbf{T} \setminus \{T\} \to \mathbf{R}$. If $E\left[\int_0^T |\xi(t)|\,\mathrm{d}t\right] \simeq 0$, then for P-a.e. $\omega \in \Omega$ and ν-a.e. $t \in \mathbf{T} \setminus \{T\}$, $\xi(t)(\omega) \simeq 0$.*

If a random variable is $L^1(P)$, then also a.s. limited. This can easily be shown through an application of the *underspill* proof principle (see Remark 1.1):

Remark 2.8. If x is $L^1(P)$, then a.s. x is limited.

Proof. Let x is $L^1(P)$ and fix $\varepsilon \gg 0$. Then for every nonstandard $n \in \mathbf{N}$,

$$n\, P\{|x| > n\} \leq E\left[|x| \chi_{\{|x|>n\}}\right] \simeq 0,$$

whence the formula

$$n\,P\,\{|x| > n\} < \varepsilon$$

holds for all nonstandard natural numbers n. However, the set of such n cannot equal $\mathbf{N}$ (see Remark 1.1), therefore it must also contain a standard n_ε. Hence the event $N_\varepsilon = \{|x| > n_\varepsilon\}$ has probability $< \varepsilon$ while x is limited on $\Omega \setminus N_\varepsilon$. Since $\varepsilon \gg 0$ was arbitrary, x is a.s. limited. □

Finally, we have the following simple sufficient condition for integrability, due to Nelson (personal communication):

Remark 2.9. For all $p > q > 0$ and every random variable x, if $E\,[|x|^p] \ll \infty$ and $\frac{q}{p} \ll 1$, then x is $L^q(P)$.

Proof. Let $p > q > 0$ with $\frac{q}{p} \ll 1$ and suppose $E\,[|x|^p] \ll \infty$. The function $z \mapsto z^{p/q}$ is convex on $\mathbf{R}_{>0}$ (its second derivative being $z \mapsto \frac{p}{q}\left(\frac{p}{q} - 1\right) z^{p/q-2} > 0$), hence by Jensen's inequality

$$E\,[|x|^q]^{p/q} \le E\,[|x|^p] \ll \infty.$$

Moreover, for all $M \subseteq \Omega$ with $P(M) \simeq 0$, the Hölder inequality yields

$$\begin{aligned} E\,[|x|^q \chi_M] &\le E\,[|x|^p]^{q/p}\, E\left[\chi_M^{\frac{1}{1-q/p}}\right]^{1-q/p} \\ &= E\,[|x|^p]^{q/p}\, P(M)^{1-q/p} \simeq 0. \end{aligned}$$

Hence, by Nelson's Radon–Nikodym theorem (see Remark 2.2), $|x|^q$ is $L^1(P)$ and thus x is $L^1(P)$. □

For stochastic processes, there are several notions of limitedness. A stochastic process η is said to be *limited* if and only if $\eta(t)$ is limited for all $t \in \mathbf{T}$ (with P-probability 1).

Remark 2.10. A stochastic process η is limited if and only if there exists a limited real number C such that $\max_{t\in\mathbf{T}} |\eta(t)| \le C$ (with P-probability 1).

Proof. Since Ω and $\mathbf{T}$ are finite (though possibly of unlimited cardinality), the maximum

$$C := \max_{\omega\in\Omega} \max_{t\in\mathbf{T}} |\eta(t)(\omega)|\, \chi_{\{P\{\cdot\}>0\}}(\omega)$$

is well-defined. It will be a limited real if and only if η is limited. Moreover, $\max_{t\in\mathbf{T}} |\eta(t)| \le C$ with P-probability 1 by definition. □

Definition 2.11. A process $\xi = (\xi(t))_{t\in\mathbf{T}'}$ is called *a.s. limited* if and only if a.s. $\max_{t\in\mathbf{T}'} |\xi(t)|$ is limited. (Because $\mathbf{T}'$ is finite, this is equivalent to asserting that a.s. for all $t \in \mathbf{T}'$, $\xi(t)$ is limited.)

The notions of independence, filtrations, conditional expectations, and martingales are inherited from conventional mathematics. Note, however, that (Ω, P) is a finite probability space and that therefore, one only needs the concepts of finite probability theory; radically elementary stochastic calculus does not require any measure theory.

For example:

- An *algebra of random variables* is a subset $\mathcal{A}$ of $\mathbf{R}^{\Omega}$ containing all the constant maps from Ω to $\mathbf{R}$ and such that for all $x, y \in \mathcal{A}$, also $x + y \in \mathcal{A}$ and $xy \in \mathcal{A}$. Let $\mathcal{A}$ be an algebra of random variables. An *atom* of $\mathcal{A}$ is a maximal subset A of Ω such that all elements of $\mathcal{A}$ are constant on A. One can show that $\mathcal{A}$ equals the set of all random variables which are constant on each of the atoms of $\mathcal{A}$ (cf. Nelson [60, p. 6]). A random variable is called $\mathcal{A}$-*measurable* if and only if it is an element of $\mathcal{A}$. A subset of Ω is called $\mathcal{A}$-*measurable* if and only if its characteristic function is $\mathcal{A}$-measurable. The orthogonal projection operator $\mathbf{R}^{\Omega} \to \mathcal{A}$ is denoted $E[\cdot|\mathcal{A}]$ and called *conditional expectation* with respect to $\mathcal{A}$.
- Let $x_1, \ldots, x_m$ be random variables, and let $\mathcal{A}$ be the smallest algebra containing $x_1, \ldots, x_m$ (the *algebra generated by* $x_1, \ldots, x_m$). Then, for every $x \in \mathcal{A}$, there exists some function $f : \mathbf{R}^m \to \mathbf{R}$ such that
$$x(\omega) = f\left(x_1(\omega), \ldots, x_m(\omega)\right)$$
for all $\omega \in \Omega$.
- A *filtration* is a family $(\mathcal{G}_t)_{t\in\mathbf{T}'}$ such that $\mathcal{G}_t \subseteq \mathcal{G}_s$ whenever $t \leq s$ and each $\mathcal{G}_t$ is an algebra of random variables.
- Let $(\mathcal{G}_t)_{t\in\mathbf{T}}$ be a filtration. We say that a stochastic process ξ is $\mathcal{G}$-*adapted* if and only if $\xi(t)$ is $\mathcal{G}_t$-measurable for all $t \in \mathbf{T}$. A stochastic process ξ is called a $(\mathcal{G}, P)$-*supermartingale* if and only if ξ is $\mathcal{G}$-adapted and
$$E\left[\xi(t)|\,\mathcal{G}_s\right] \leq \xi(s)$$
(with P-probability 1) for all $s, t \in \mathbf{T}$ with $s \leq t$. A $(\mathcal{G}, P)$-supermartingale is simply called a *supermartingale* where no ambiguity as to $\mathcal{G}$ or P can arise. A process ξ is a *submartingale* if and only if $-\xi$ is a supermartingale, and it is a *martingale* if and only if it is both a supermartingale and a submartingale.
- A $\mathcal{G}$-adapted stochastic process ξ is a martingale if and only if $E\left[\xi(t + \mathrm{d}t)|\,\mathcal{G}_t\right] = \xi(t)$ for all $t \in \mathbf{T} \setminus \{1\}$, in other words, if $E\left[\mathrm{d}\xi(t)|\,\mathcal{G}_t\right] = 0$ for all $t \in \mathbf{T} \setminus \{1\}$.
- The random variables $\xi(t_0), \ldots$ of a process $(\xi(t))_{t\in\mathbf{T}'}$ are *independent* if and only if
$$P\left\{\xi = \lambda\right\} = \prod_{t\in\mathbf{T}'} P\left\{\xi(t) = \lambda(t)\right\}$$
for all trajectories $\lambda : \mathbf{T}' \to \mathbf{R}$ of ξ.

All results from finite probability theory are directly inherited, since our axiom system is a simple extension of conventional mathematics.

A criterion for the a.s. limitedness of martingales can be deduced from the following martingale inequality (again due to Nelson [60]):

Remark 2.12. If $\xi = (\xi(t))_{t\in\mathbf{T}}$ is a super- or submartingale, then for all $\lambda \in \mathbf{R}_{>0}$,

$$P\left\{\max_{t\in\mathbf{T}} |\xi(t) - \xi(0)| \geq \lambda\right\} \leq \frac{2}{\lambda} E\left[|\xi(1) - \xi(0)|\right].$$

(Cf. Nelson [60, Theorem 11.1].) If $\xi = (\xi(t))_{t\in\mathbf{T}}$ is a martingale, then for all $\lambda \in \mathbf{R}_{>0}$,

$$P\left\{\max_{t\in\mathbf{T}} |\xi(t) - \xi(0)| \geq \lambda\right\} \leq \frac{1}{\lambda} E\left[|\xi(1) - \xi(0)|\right].$$

(Cf. Nelson [60, remark after Theorem 11.2].)

Corollary 2.13. *If $\xi = (\xi(t))_{t\in\mathbf{T}}$ is a supermartingale or submartingale with limited $\xi(0)$ and limited $E\left[|\xi(1) - \xi(0)|\right]$, then ξ is a.s. limited.*

Proof. Let $\varepsilon \gg 0$, let k be the limited real

$$k = \frac{1}{\varepsilon} E\left[|\xi(1) - \xi(0)|\right],$$

and consider the event $N = \{\max_{t\in\mathbf{T}} |\xi(t) - \xi(0)| \geq k\}$. Then, by Remark 2.12 applied to $\lambda = k$, one has $P(N) \leq \varepsilon$, whilst

$$\max_{t\in\mathbf{T}} |\xi(t)| \leq |\xi(0)| + \max_{t\in\mathbf{T}} |\xi(t) - \xi(0)| \leq |\xi(0)| + k$$

on $\Omega \setminus N$. □

2.3 Wiener Walks and Wiener Processes

Now we introduce a fundamental object of radically elementary probability theory and stochastic calculus: the radically elementary analogue of N. Wiener's process. A *Wiener walk* on (Ω, P) is a process $W = (W(t))_{t\in\mathbf{T}}$ such that

- $W(0) = 0$,
- $\mathrm{d}W(0), \ldots, \mathrm{d}W(1 - \mathrm{d}t)$ are independent, and
- for all $t \in \mathbf{T} \setminus \{1\}$,

$$P\left\{\mathrm{d}W(t) = \sqrt{\mathrm{d}t}\right\} = P\left\{\mathrm{d}W(t) = -\sqrt{\mathrm{d}t}\right\} = \frac{1}{2}.$$

A stochastic process is called a (near) *Wiener process* on (Ω, P) if and only if it is nearly equivalent to some Wiener walk on (Ω, P). It is worthwhile to note that any Wiener process allows for a strong approximation by a Wiener walk through a coupling construction (cf. Lawler [43, Sects. 7.5, 7.6]). Note that a Wiener process ξ does not necessarily have to be exactly a martingale; if it is a martingale, ξ will be called a *Wiener martingale*.

Remark 2.14. If ξ is a Wiener process, then $\xi(1)$ is $L^2(P)$.

Proof. In the first part of his proof of the radically elementary analogue of Wiener's characterization, Nelson [60, Proof of Theorem 18.1, part (i)⇒(ii), p. 76] shows exactly this. □

Remark 2.15. Clearly, one can choose a finite probability space (Ω, P) in such a way that there exists a Wiener walk (and hence a Wiener process) on (Ω, P): Simply let $\Omega = \{\pm\sqrt{\mathrm{d}t}\}^{\mathbf{T}\setminus\{1\}}$, let P be the uniform distribution on Ω, and let

$$\forall s \in \mathbf{T} \qquad W(s) = \sum_{t<s} \pi(t),$$

wherein $\pi(t) : \{\pm\sqrt{\mathrm{d}t}\}^{\mathbf{T}\setminus\{1\}} \to \{\pm\sqrt{\mathrm{d}t}\}$, for any $t \in \mathbf{T} \setminus \{1\}$, is the projection onto the t-th Cartesian factor in $\{\pm\sqrt{\mathrm{d}t}\}^{\mathbf{T}\setminus\{1\}}$. Since $P = \otimes_{t\in\mathbf{T}\setminus\{1\}} P_0$ if P_0 denotes the uniform distribution on $\{\pm\sqrt{\mathrm{d}t}\}$, it is obvious that the W thus defined is a Wiener walk on (Ω, P).[3]

In all that follows, we assume that W is a Wiener walk on (Ω, P).

In a similar spirit, one can define a radically elementary analogue of Poisson's process. A *Poisson walk* on (Ω, P) is a process $\zeta = (\zeta(t))_{t\in\mathbf{T}}$ such that

- $\zeta(0) = 0$,
- $\mathrm{d}\zeta(0), \ldots, \mathrm{d}\zeta(1 - \mathrm{d}t)$ are independent, and
- for all $t \in \mathbf{T} \setminus \{1\}$,

$$P\{\mathrm{d}\zeta(t) = 0\} = 1 - \mathrm{d}t, \quad P\{\mathrm{d}\zeta(t) = 1\} = \frac{1}{2}\mathrm{d}t, \quad P\{\mathrm{d}\zeta(t) = -1\} = \frac{1}{2}\mathrm{d}t.$$

Remark 2.16. Again, it is easy to construct a finite probability space (Ω, P) in such a way that there exists a Poisson walk on (Ω, P): Let $\Omega = \{-1, 0, 1\}^{\mathbf{T}\setminus\{1\}}$, let P_0 be a probability measure on $\{-1, 0, 1\}$ defined by

$$P_0\{-1\} = \frac{1}{2}\mathrm{d}t, \quad P_0\{0\} = 1 - \mathrm{d}t, \quad P_0\{1\} = \frac{1}{2}\mathrm{d}t,$$

let P be the product measure $P = \otimes_{t\in\mathbf{T}\setminus\{1\}} P_0$, and let

$$\forall s \in \mathbf{T} \qquad \zeta(s) = \sum_{t<s} \pi(t),$$

[3] In Robinsonian nonstandard analysis, this Wiener walk is known as Anderson's [4] construction of the Wiener process.

wherein $\pi(t) : \{1, 0, 1\}^{\mathbf{T}\setminus\{1\}} \to \{-1, 0, 1\}$, for any $t \in \mathbf{T} \setminus \{1\}$, is the projection onto the t-th Cartesian factor in $\{-1, 0, 1\}^{\mathbf{T}\setminus\{1\}}$. Clearly then, ζ will be a Poisson walk on (Ω, P).[4]

2.4 Distribution of the Wiener Walk

Benoît [10, Proposition 4.2.1] has given an elementary proof which shows that a Wiener walk at time t has essentially a Gaussian normal distribution with mean zero and variance t:

Lemma 2.17 (Distribution of the Wiener walk). *For any $n \in \mathbf{N} \cap [0, 1/\mathrm{d}t]$ and every $k \in \mathbf{Z} \cap [-n, n]$,*

$$P\left\{W(n\,\mathrm{d}t) = k\sqrt{\mathrm{d}t}\right\} = \frac{2^{-n} n!}{\left(\frac{n+k}{2}\right)!\left(\frac{n-k}{2}\right)!},$$

hence for all limited $x \in \mathbf{R}$ and all $t \in \mathbf{T}$ with $t \gg 0$,

$$\frac{P\left\{x - \sqrt{\mathrm{d}t} \le W(t) < x + \sqrt{\mathrm{d}t}\right\}}{2\sqrt{\mathrm{d}t}} \simeq \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right).$$

The first, exact equation in this lemma is just elementary combinatorics. In order to obtain the approximate formula, Benoît [10, pp. 73–74] approximates—for $x = k\sqrt{\mathrm{d}t}$ and $t = n\,\mathrm{d}t$—the binomial coefficients by means of the infinitesimal Stirling formula

$$\forall \ell \simeq \infty \quad \exists \varepsilon \simeq 0 \qquad \ell! = (1+\varepsilon)\ell^{\ell} \mathrm{e}^{-\ell}\sqrt{\ell}\sqrt{2\pi}$$

(cf. van den Berg [11, Sect. 8.4.2, p. 180]),[5] which is tantamount to

$$\forall \ell \simeq \infty \quad \exists \varepsilon \simeq 0 \qquad \ln(\ell!) = \left(\ell + \frac{1}{2}\right)\ln \ell - \ell + \frac{\ln(2\pi)}{2} + \varepsilon$$

and can be applied to $\ell = n$, $\ell = n + k$ and $\ell = n - k$ since $n = t/\mathrm{d}t \simeq \infty$ and $k/n = \sqrt{\mathrm{d}t}\, x/t \simeq 0$ (as $t \gg 0$ and x is limited); from there, the second-order Taylor expansion of the logarithm function yields the result.

---

[4]In Robinsonian nonstandard analysis, this Poisson walk is known as Loeb's [51] construction of the Poisson process.

[5]One should note that this infinitesimal version of Stirling's formula can also be proved in radically elementary probability theory, cf. van den Berg [11, last paragraph on p. 172].

2.5 Integrability and Limited Paths of the Wiener Walk

In Chap. 4, we shall present radically elementary versions of both Girsanov's theorem and the diffusion invariance principle. Because the density in Girsanov's theorem involves the exponential of W, it will be helpful to know that $\exp(W)$ is integrable of any limited order.

Remark 2.18. $\exp(W(s))$ is $L^p(P)$ for all $s \in \mathbf{T}$ and every limited $p > 0$.

Proof. Let $s \in \mathbf{T}$. Since $W(s) = W(s) - W(0) = \sum_{t<s} \mathrm{d}W(t)$, we have

$$E\left[\exp(W(s))^{2p}\right] = E\left[\exp\left(2p\sum_{t<s}\mathrm{d}W(t)\right)\right] = E\left[\prod_{t<s}\exp(2p\mathrm{d}W(t))\right],$$

and since the increments $\mathrm{d}W(0), \ldots, \mathrm{d}W(s-\mathrm{d}t)$ of W are independent and identically distributed, we deduce

$$E\left[\exp(W(s))^{2p}\right] = \prod_{t<s} E\left[\exp(2p\mathrm{d}W(t))\right] = E\left[\exp(2p\mathrm{d}W(0))\right]^{s/\mathrm{d}t}. \tag{2.1}$$

Since p was assumed to be limited, a first-order Taylor expansion of exp yields the existence of limited numbers a, b such that

$$\exp(2p\sqrt{\mathrm{d}t}) = 1 + 2p\sqrt{\mathrm{d}t} + a\mathrm{d}t, \qquad \exp(-2p\sqrt{\mathrm{d}t}) = 1 - 2p\sqrt{\mathrm{d}t} + b\mathrm{d}t.$$

Hence, exploiting that $\exp(x) = \sum_{k=0}^{\infty} x^k/k! \geq 1 + x$ for all $x \in \mathbf{R}_{\geq 0}$,

$$\begin{aligned} E\left[\exp(2p\mathrm{d}W(0))\right] &= \frac{\exp(2p\sqrt{\mathrm{d}t}) + \exp(-2p\sqrt{\mathrm{d}t})}{2} = 1 + \frac{a+b}{2}\mathrm{d}t \\ &\leq 1 + \frac{|a+b|}{2}\mathrm{d}t \leq \exp\left(\frac{|a+b|}{2}\mathrm{d}t\right) \end{aligned}$$

and therefore by Eq. (2.1),

$$E\left[\exp(W(s))^{2p}\right] \leq \exp\left(\frac{|a+b|}{2}\mathrm{d}t\right)^{s/\mathrm{d}t} = \exp\left(\frac{s|a+b|}{2}\right).$$

Since a, b are limited, we arrive at

$$E\left[\exp(W(s))^{2p}\right] \ll \infty. \tag{2.2}$$

Combining estimate (2.2) with Remark 2.9, we obtain that $\exp(W(s))^p = \exp(pW(s))$ is $L^1(P)$, hence $\exp(W(s))$ is $L^p(P)$. □

To know that the exponential of the Wiener walk is L^1 at any time, also means that W is always a.s. limited at all times.

Corollary 2.19. *W is a.s. limited.*

Proof. Due to the Nelson's Radon–Nikodym theorem (Remark 2.2) and the previous Remark 2.18, we see that $E\left[\exp\left(W(1)\right)\right]$ is limited. Since $\mathrm{e}^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ and thus $2\mathrm{e}^x \geq x^2$ for all $x \in \mathbf{R}$, we conclude that $E\left[|W(1)|^2\right]$ is limited. By the Cauchy–Schwarz inequality, this means that $E\left[|W(1)|\right]$ is limited. Therefore, Corollary 2.13 shows that W is a.s. limited. □

Chapter 3
Radically Elementary Stochastic Integrals

3.1 Martingales and Itô Integrals

For any two processes ξ, η, the *stochastic integral* of η with respect to ξ is the process $\int \eta \mathrm{d}\xi$ defined by

$$\int_0^s \eta \mathrm{d}\xi = \int_0^s \eta(t)\mathrm{d}\xi(t) = \sum_{t<s} \eta(t)\mathrm{d}\xi(t)$$

for all $s \in \mathbf{T}$. Note that $\mathrm{d}(t) = t + \mathrm{d}t - t = \mathrm{d}t$ for all $t \in \mathbf{T}\setminus\{1\}$, whence for the process $\mathrm{id} = (t)_{t\in\mathbf{T}}$ we have $\int_0^s \eta \mathrm{d}\,\mathrm{id} = \int_0^s \eta(t)\mathrm{d}(t) = \int_0^s \eta(t)\mathrm{d}t$. (Since the radically elementary approach to stochastic processes does not use *conventional* Riemann integrals, there is no danger of confusion attached to the notation $\int_0^s \eta(t)\mathrm{d}t$.)

Note that since Ω and $\mathbf{T}$ are finite, the expectation operator E and the finite integral $\int \cdot \mathrm{d}t$ always commute.

Theorem 3.1. *Let $(\mathcal{G}_t)_{t\in\mathbf{T}}$ be a filtration. A process m is a $(\mathcal{G}, P)$-martingale if and only if $\int \eta \mathrm{d}m$ is a $(\mathcal{G}, P)$-martingale for all $\mathcal{G}$-adapted m.*

Proof. The constant deterministic process $(1)_{t\in\mathbf{T}}$ is clearly adapted and m can be written as $m = \int 1\mathrm{d}m$.

Conversely, suppose m is a martingale and let η be $\mathcal{G}$-adapted. Then for all $t \in \mathbf{T} \setminus \{1\}$,

$$E\left[\eta(t)\mathrm{d}m(t)\right| \mathcal{G}_t] = \eta(t) \underbrace{E\left[\mathrm{d}m(t)\right| \mathcal{G}_t]}_{=0} = 0,$$

so $\int \eta \mathrm{d}m$ is indeed a $(\mathcal{G}, P)$-martingale. □

Stochastic integrals with respect to W are also called *Itô integrals*. A martingale with respect to $\mathcal{F} = (\mathcal{F}_t)_{t\in\mathbf{T}}$ (the filtration generated by W) is just an Itô integral of some adapted process, and vice versa:

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_3,

Theorem 3.2 (Martingale representation theorem and converse). *A stochastic process $(m_t)_{t \in \mathbf{T}}$ is an $(\mathcal{F}, P)$-martingale if and only if there exists a unique $\mathcal{F}$-adapted process $\phi = (\phi_t)_{t \in \mathbf{T} \setminus \{1\}}$ such that for all $s \in \mathbf{T}$,*

$$m(s) = m(0) + \int_0^s \phi(t) \mathrm{d}W(t).$$

$E\left[|m(s)|^2\right]$ *is limited for all* $s \in \mathbf{T}$ *if and only if* $E\left[\int_0^1 \left|\phi(t)^2\right| \mathrm{d}t\right]$ *is limited.*

Proof. First, let m be a martingale. Let $t \in \mathbf{T} \setminus \{1\}$. Since m is $\mathcal{F}$-adapted, $\mathrm{d}m(t)$ is $\mathcal{F}_{t+\mathrm{d}t}$-measurable and therefore, there is some $f : \mathbf{R}^{t/dt+1} \to \mathbf{R}$ such that

$$\mathrm{d}m(t)(\omega) = f\left(\mathrm{d}W(0)(\omega), \ldots, \mathrm{d}W(t)(\omega)\right)$$

for all $\omega \in \Omega$. Therefore, exploiting that W has independent increments, each with distribution $\frac{\delta_{\sqrt{\mathrm{d}t}} + \delta_{-\sqrt{\mathrm{d}t}}}{2}$, we obtain

$$\begin{aligned}
&E\left[\mathrm{d}m(t) | \mathcal{F}_t\right] \\
&\quad = E\left[f\left(\mathrm{d}W(0), \ldots, \mathrm{d}W(t - \mathrm{d}t), \sqrt{\mathrm{d}t}\right) \chi_{\left\{\mathrm{d}W(t) = \sqrt{\mathrm{d}t}\right\}} \middle| \mathcal{F}_t \right] \\
&\qquad + E\left[f\left(\mathrm{d}W(0), \ldots, \mathrm{d}W(t - \mathrm{d}t), -\sqrt{\mathrm{d}t}\right) \chi_{\left\{\mathrm{d}W(t) = -\sqrt{\mathrm{d}t}\right\}} \middle| \mathcal{F}_t \right] \\
&\quad = f\left(\mathrm{d}W(0), \ldots, \mathrm{d}W(t - \mathrm{d}t), \sqrt{\mathrm{d}t}\right) P\left\{\mathrm{d}W(t) = \sqrt{\mathrm{d}t}\right\} \\
&\qquad + f\left(\mathrm{d}W(0), \ldots, \mathrm{d}W(t - \mathrm{d}t), -\sqrt{\mathrm{d}t}\right) P\left\{\mathrm{d}W(t) = -\sqrt{\mathrm{d}t}\right\} \\
&\quad = \frac{1}{2} f\left(\mathrm{d}W(0), \ldots, \mathrm{d}W(t - \mathrm{d}t), \sqrt{\mathrm{d}t}\right) \\
&\qquad + \frac{1}{2} f\left(\mathrm{d}W(0), \ldots, \mathrm{d}W(t - \mathrm{d}t), -\sqrt{\mathrm{d}t}\right)
\end{aligned}$$

Since m is a martingale, $E\left[\mathrm{d}m(t) | \mathcal{F}_t\right] = 0$, hence

$$\begin{aligned}
&f\left(\mathrm{d}W(0), \ldots, \mathrm{d}W(t - \mathrm{d}t), \sqrt{\mathrm{d}t}\right) \\
&\quad = -f\left(\mathrm{d}W(0), \ldots, \mathrm{d}W(t - \mathrm{d}t), -\sqrt{\mathrm{d}t}\right).
\end{aligned}$$

Defining

$$\vartheta(t) = f\left(\mathrm{d}W(0), \ldots, \mathrm{d}W(t - \mathrm{d}t), \sqrt{\mathrm{d}t}\right),$$

we get

$$\begin{aligned}
\mathrm{d}m(t) &= f\left(\mathrm{d}W(0),\ldots,\mathrm{d}W\left(t-\mathrm{d}t\right),\sqrt{\mathrm{d}t}\right)\chi_{\{\mathrm{d}W(t)=\sqrt{\mathrm{d}t}\}}\\
&\quad + f\left(\mathrm{d}W(0),\ldots,\mathrm{d}W\left(t-\mathrm{d}t\right),-\sqrt{\mathrm{d}t}\right)\chi_{\{\mathrm{d}W(t)=-\sqrt{\mathrm{d}t}\}}\\
&= f\left(\mathrm{d}W(0),\ldots,\mathrm{d}W\left(t-\mathrm{d}t\right),\sqrt{\mathrm{d}t}\right)\chi_{\{\mathrm{d}W(t)=\sqrt{\mathrm{d}t}\}}\\
&\quad - f\left(\mathrm{d}W(0),\ldots,\mathrm{d}W\left(t-\mathrm{d}t\right),\sqrt{\mathrm{d}t}\right)\chi_{\{\mathrm{d}W(t)=-\sqrt{\mathrm{d}t}\}}\\
&= \vartheta(t)\chi_{\{\mathrm{d}W(t)=\sqrt{\mathrm{d}t}\}} - \vartheta(t)\chi_{\{\mathrm{d}W(t)=-\sqrt{\mathrm{d}t}\}}\\
&= \vartheta(t)\mathrm{d}W(t)/\sqrt{\mathrm{d}t}.
\end{aligned}$$

By definition, $\vartheta(t)$ is $\mathcal{F}_t$-measurable. Hence, if we define $\phi(t) = \vartheta(t)/\sqrt{\mathrm{d}t}$, it is also $\mathcal{F}_t$-measurable and

$$\mathrm{d}m(t) = \phi(t)\mathrm{d}W(t).$$

Since t was arbitrary, this holds for any t, and yields, after writing $m(s)-m(0)$ as a telescoping sum,

$$m(s) = m(0) + \sum_{t<s}\mathrm{d}m(t) = m(0) + \sum_{t<s}\phi(t)\mathrm{d}W(t).$$

If there were another process $\tilde{\phi} = \left(\tilde{\phi}_t\right)_{t\in\mathbf{T}\setminus\{1\}}$ such that $\int\tilde{\phi}\mathrm{d}W = m = \int\phi\mathrm{d}W$, then for all $t\in\mathbf{T}\setminus\{1\}$,

$$\tilde{\phi}(t)\mathrm{d}W(t) = \mathrm{d}m(t) = \phi(t)\mathrm{d}W(t),$$

whence $\tilde{\phi}(t) = \phi(t)$ since $\mathrm{d}W(t) = \pm\sqrt{\mathrm{d}t} \neq 0$, therefore $\tilde{\phi} = \phi$, proving the uniqueness of ϕ.

Conversely, suppose $m(s) = m(0) + \int_0^s \phi(t)\mathrm{d}W(t)$ for all $s\in\mathbf{T}$ for some $\mathcal{F}$-adapted ϕ. The definition of the stochastic integral and the $\mathcal{F}$-adaptedness of W imply that m is $\mathcal{F}$-adapted. It remains to be shown that $E\left[\mathrm{d}m(t)|\,\mathcal{F}_t\right] = 0$ for all $t\in\mathbf{T}\setminus\{1\}$. This is straightforward:

$$E\left[\mathrm{d}m(t)|\,\mathcal{F}_t\right] = E\left[\phi(t)\mathrm{d}W(t)|\,\mathcal{F}_t\right] = \phi(t)E\left[\mathrm{d}W(t)|\,\mathcal{F}_t\right] = \phi(t)\underbrace{E\left[\mathrm{d}W(t)\right]}_{=0}.$$

By the Itô isometry (Lemma 3.4), one has

$$E\left[|m(s)|^2\right] = E\left[\int_0^s \left|\phi(t)^2\right|\mathrm{d}t\right],$$

and the right-hand side is monotonely increasing in s. Hence, $E\left[|m(s)|^2\right]$ is limited for all $s \in \mathbf{T}$ if and only if $E\left[\int_0^s \left|\phi(t)^2\right| \mathrm{d}t\right]$ is. □

Definition 3.3. A stochastic process $\xi = (\xi(t))_{t\in\mathbf{T}}$ is called a *normalized martingale* (or just *normalized*) if and only if

$$\forall t \in \mathbf{T} \setminus \{1\} \qquad E\left[\mathrm{d}\xi(t) | \mathcal{F}_t\right] = 0, \qquad E\left[(\mathrm{d}\xi(t))^2 \middle| \mathcal{F}_t\right] = \mathrm{d}t.$$

The Wiener walk W, for example, is normalized.

Lemma 3.4 (Radically elementary Itô isometry). *Let m be a normalized martingale and η be an $\mathcal{F}$-adapted stochastic process. Then for all $s, v \in \mathbf{T}$ with $s \geq v$,*

$$E\left[\left|\int_v^s \eta(t)\mathrm{d}m(t)\right|^2 \middle| \mathcal{F}_v\right] = E\left[\int_v^s \eta(t)^2 \mathrm{d}t \middle| \mathcal{F}_v\right].$$

Proof.

$$\begin{aligned}
E\left[\left|\int_v^s \eta(t)\mathrm{d}m(t)\right|^2 \middle| \mathcal{F}_v\right] &= E\left[\left(\sum_{t=v}^{s} \eta(t)\mathrm{d}m(t)\right)^2 \middle| \mathcal{F}_v\right] \\
&= 2\sum_{v\leq u<t<s} E\left[\eta(t)\mathrm{d}m(t)\eta(u)\mathrm{d}m(u) | \mathcal{F}_v\right] + \sum_{v\leq t<s} E\left[\eta(t)^2\mathrm{d}m(t)^2 \middle| \mathcal{F}_v\right] \\
&= 2\sum_{v\leq u<t<s} E\left[E\left[\eta(t)\mathrm{d}m(t)\eta(u)\mathrm{d}m(u) | \mathcal{F}_t\right] | \mathcal{F}_v\right] \\
&\quad + \sum_{v\leq t<s} E\left[E\left[\eta(t)^2\mathrm{d}m(t)^2 \middle| \mathcal{F}_t\right] \middle| \mathcal{F}_v\right] \\
&= 2\sum_{v\leq u<t<s} E\left[\eta(t)E\left[\mathrm{d}m(t) | \mathcal{F}_t\right]\eta(u)\mathrm{d}m(u) | \mathcal{F}_v\right] \\
&\quad + \sum_{v\leq t<s} E\left[\eta(t)^2 E\left[\mathrm{d}m(t)^2 \middle| \mathcal{F}_t\right] \middle| \mathcal{F}_v\right] \\
&= 0 + \sum_{v\leq t<s} E\left[\eta(t)^2\mathrm{d}t \middle| \mathcal{F}_v\right] = E\left[\sum_{v\leq t<s} \eta(t)^2\mathrm{d}t \middle| \mathcal{F}_v\right] \\
&= E\left[\int_v^s \eta(t)^2\mathrm{d}t \middle| \mathcal{F}_v\right].
\end{aligned}$$

□

Remark 3.5. Nelson [60, deliberations following Theorem 13.1, p. 55] has shown that if m is a normalized martingale such that $\mathrm{d}m(t)$ is infinitesimal for all $t \in \mathbf{T} \setminus \{1\}$, then m is P-a.s. continuous.

3.2 Radically Elementary Itô Processes

An Itô process is essentially an Itô integral plus an absolutely continuous process.

Definition 3.6. Let $\bar{W}$ be a Wiener process, let $\xi(0) \in \mathbf{R}$, and let $\mu = (\mu(t))_{t \in \mathbf{T}\setminus\{1\}}$ and $\sigma = (\sigma(t))_{t \in \mathbf{T}\setminus\{1\}}$ be two $\mathcal{F}$-adapted processes. A stochastic process ξ is called an *Itô process on* (Ω, P) *with respect to* $\bar{W}$ *and with drift coefficient* μ, *diffusion coefficient* σ *and initial value* $\xi(0)$ if and only if

$$\xi(s) = \xi(0) + \int_0^s \mu(t)\mathrm{d}t + \int_0^s \sigma(t)\mathrm{d}\bar{W}(t)$$

for all $s \in \mathbf{T}$. The equation

$$\forall t \in \mathbf{T} \setminus \{1\} \qquad \mathrm{d}\xi(t) = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}\bar{W}(t)$$

is called the *stochastic differential equation* solved by ξ.

The representation of an Itô process in the form $\xi = \xi(0) + \int \mu(t)\mathrm{d}t + \int \sigma(t)\mathrm{d}W(t)$ is called *Itô decomposition*. Under certain assumptions, the Itô decomposition is essentially unique. We give a proof under fairly restrictive assumptions (recall that ν denotes the normalized counting measure on $\mathbf{T} \setminus \{T\}$):

Theorem 3.7 (Uniqueness of the Itô decomposition). *Let $\mu_1, \mu_2, \sigma_1, \sigma_2$ be $\mathcal{F}$-adapted processes. Suppose for all $t \in \mathbf{T} \setminus \{1\}$, we have*

$$\mu_1(t)\mathrm{d}t + \sigma_1(t)\mathrm{d}W(t) = \mu_2(t)\mathrm{d}t + \sigma_2(t)\mathrm{d}W(t) + R\,(t + \mathrm{d}t)\,(\mathrm{d}t)^{3/2} \tag{3.1}$$

for some $R(t+\mathrm{d}t)$ such that $E\left[\int_0^1 R(t+\mathrm{d}t)^2\mathrm{d}t\right]$ is limited.[1] *Assume $E\left[\int_0^1 |\mu_1(t) - \mu_2(t)|^2\mathrm{d}t\right]$ is limited. Then for P-a.e. $\omega \in \Omega$ and ν-a.e. $t \in \mathbf{T} \setminus \{1\}$,*

$$\sigma_1(t)(\omega) \simeq \sigma_2(t)(\omega), \qquad \mu_1(t)(\omega) \simeq \mu_2(t)(\omega).$$

Proof. Put $\mu = \mu_1 - \mu_2$ and $\sigma = \sigma_1 - \sigma_2$. We need to verify that for P-a.e. $\omega \in \Omega$ and ν-a.e. $t \in \mathbf{T} \setminus \{1\}$, $\sigma(t)(\omega) \simeq 0 \simeq \mu(t)(\omega)$.

[1] We denote this random variable by $R\,(t + \mathrm{d}t)$ rather than $R\,(t)$ because it is $\mathcal{F}_{t+\mathrm{d}t}$-measurable, but in general not $\mathcal{F}_t$-measurable.

For this purpose, first note that by definition of μ and σ and by the assumption (3.1) in the Theorem, we have for all $t \in \mathbf{T} \setminus \{1\}$,

$$\mu(t)\mathrm{d}t = \sigma(t)\mathrm{d}W(t) + R\,(t + \mathrm{d}t)\,(\mathrm{d}t)^{3/2}. \tag{3.2}$$

Squaring both sides of this equality and afterwards rearranging terms yields $\mu(t)^2(\mathrm{d}t)^2 - R\,(t+\mathrm{d}t)^2\,(\mathrm{d}t)^3 - 2R\,(t+\mathrm{d}t)\,(\mathrm{d}t)^{3/2}\mathrm{d}W(t) = \sigma(t)^2\mathrm{d}t$, hence (dropping nonnegative terms and using the triangle inequality)

$$\begin{aligned}
E\left[\int_0^s \sigma(t)^2\mathrm{d}t\right] &= \int_0^s E\left[\sigma(t)^2\right]\mathrm{d}t \\
&= \sum_{t<s} E\left[\mu(t)^2\right](\mathrm{d}t)^2 - \sum_{t<s} E\left[R(t+\mathrm{d}t)^2\right](\mathrm{d}t)^3 \\
&\quad -2\sum_{t<s} E\left[R\,(t+\mathrm{d}t)\,\mathrm{d}W(t)\right](\mathrm{d}t)^{3/2} \\
&\leq \sum_{t<s} E\left[\mu(t)^2\right](\mathrm{d}t)^2 - 2\sum_{t<s} E\left[R\,(t+\mathrm{d}t)\,\mathrm{d}W(t)\right](\mathrm{d}t)^{3/2} \\
&\leq \sum_{t<s} E\left[\mu(t)^2\right](\mathrm{d}t)^2 \\
&\quad +2\sum_{t<s} \left|E\left[R\,(t+\mathrm{d}t)\,\mathrm{d}W(t)\right]\right|(\mathrm{d}t)^{3/2}.
\end{aligned}$$

However, the last expression can be estimated, due to Jensen's inequality, as follows:

$$\left|E\left[R\,(t+\mathrm{d}t)\,\mathrm{d}W(t)\right]\right| \leq E\left[R\,(t+\mathrm{d}t)^2\,\mathrm{d}t\right]^{1/2},$$

so we actually have shown that

$$\begin{aligned}
E\left[\int_0^s \sigma(t)^2\mathrm{d}t\right] &\leq E\left[\int_0^s \mu(t)^2\mathrm{d}t\right]\mathrm{d}t \\
&\quad +2\sum_{t<s}\left(E\left[R\,(t+\mathrm{d}t)^2\right]\mathrm{d}t\right)^{1/2}(\mathrm{d}t)^{3/2}.
\end{aligned} \tag{3.3}$$

In order to further simplify the right-hand side, we apply Jensen's inequality again (this time for the average on $\mathbf{T} \cap [0, s)$ as expectation operator):

$$\begin{aligned}
&\sum_{t<s}\left(E\left[R\,(t+\mathrm{d}t)^2\right]\mathrm{d}t\right)^{1/2} \\
&= \operatorname{card}\left(\mathbf{T}\cap[0,s)\right)\frac{1}{\operatorname{card}\left(\mathbf{T}\cap[0,s)\right)}\sum_{t<s}\left(E\left[R\,(t+\mathrm{d}t)^2\right]\mathrm{d}t\right)^{1/2}
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathrm{card}\,(\mathbf{T}\cap[0,s))\left(\frac{1}{\mathrm{card}\,(\mathbf{T}\cap[0,s))}\sum_{t<s}E\left[R\,(t+\mathrm{d}t)^2\right]\mathrm{d}t\right)^{1/2}\\
&=\mathrm{card}\,(\mathbf{T}\cap[0,s))^{1/2}\,E\left[\int_0^s R\,(t+\mathrm{d}t)^2\,\mathrm{d}t\right]^{1/2}.
\end{aligned}
$$

Inserting this into Eq. (3.3) and exploiting that $\mathrm{card}\,(\mathbf{T}\cap[0,s)) = s/\mathrm{d}t \leq 1/\mathrm{d}t$, hence $\mathrm{card}\,(\mathbf{T}\cap[0,s))^{1/2} = (\mathrm{d}t)^{-1/2}$, we conclude that

$$
E\left[\int_0^s \sigma(t)^2\mathrm{d}t\right] \leq E\left[\int_0^1 \mu(t)^2\mathrm{d}t\right]\mathrm{d}t + 2E\left[\int_0^1 R\,(t+\mathrm{d}t)^2\,\mathrm{d}t\right]^{1/2}\mathrm{d}t.
$$

However, by assumption, both $E\left[\int_0^1 R\,(t+\mathrm{d}t)^2\,\mathrm{d}t\right]$ and $E\left[\int_0^1 \mu(t)^2\mathrm{d}t\right]$ are limited, whence

$$
E\left[\int_0^s \sigma(t)^2\mathrm{d}t\right] = \mathcal{O}(\mathrm{d}t) \simeq 0.
$$

This entails that for P-a.e. $\omega \in \Omega$ and ν-a.e. $t \in \mathbf{T}\setminus\{1\}$, $\sigma(t)(\omega) \simeq 0$ (by Theorem 2.7).

In order to complete the proof, we also need to verify that $\mu(t)(\omega) \simeq 0$ for P-a.e. $\omega \in \Omega$ and ν-a.e. t. To achieve this, we first compute (the conditional expectation of) $\mu(t)\mathrm{d}t$. Now, according to Eq. (3.2), the latter term is the same as $\sigma(t)\mathrm{d}W(t) + R\,(t+\mathrm{d}t)\,(\mathrm{d}t)^{3/2}$, hence, using the $\mathcal{F}_t$-linearity of the operator $E\,[\cdot|\,\mathcal{F}_t]$, we get

$$
\begin{aligned}
\mu(t)\mathrm{d}t = E\,[\mu(t)\mathrm{d}t|\,\mathcal{F}_t] = \sigma(t)\,&\underbrace{E\,[\mathrm{d}W(t)|\,\mathcal{F}_t]}_{=0}\\
&+E\,[R\,(t+\mathrm{d}t)|\,\mathcal{F}_t]\,(\mathrm{d}t)^{3/2}.
\end{aligned}
$$

Therefore, $\mu(t) = E\,[R\,(t+\mathrm{d}t)|\,\mathcal{F}_t]\,(\mathrm{d}t)^{1/2}$, hence (applying the conditional Jensen inequality)

$$
\begin{aligned}
\mu(t)^2 &= E\,[R\,(t+\mathrm{d}t)|\,\mathcal{F}_t]^2\,\mathrm{d}t\\
&\leq E\left[R\,(t+\mathrm{d}t)^2\middle|\,\mathcal{F}_t\right]\mathrm{d}t.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
E\left[\int_0^1 \mu(t)^2\mathrm{d}t\right] &\leq E\left[\int_0^1 E\left[R\,(t+\mathrm{d}t)^2\middle|\,\mathcal{F}_t\right]\mathrm{d}t\right]\mathrm{d}t\\
&= \int_0^1 E\left[E\left[R\,(t+\mathrm{d}t)^2\middle|\,\mathcal{F}_t\right]\mathrm{d}t\right]\mathrm{d}t
\end{aligned}
$$

$$= \int_0^1 E\left[R(t+\mathrm{d}t)^2\,\mathrm{d}t\right]\mathrm{d}t$$
$$= E\left[\int_0^1 R(t+\mathrm{d}t)^2\,\mathrm{d}t\right]\mathrm{d}t.$$

Since $E\left[\int_0^1 R(t+\mathrm{d}t)^2\,\mathrm{d}t\right]$ was assumed to be limited, we deduce $E\left[\int_0^1 \mu(t)^2\mathrm{d}t\right] = \mathcal{O}(\mathrm{d}t) \simeq 0$. This, however, means—again by Theorem 2.7—that $\mu(t)(\omega) \simeq 0$ for P-a.e. $\omega \in \Omega$ and ν-a.e. $t \in \mathbf{T} \setminus \{1\}$. □

For special Itô processes one can prove their a.s. limitedness:

Lemma 3.8. *If ξ is an Itô process with respect to W, with limited initial value $\xi(0)$, with drift coefficient μ and diffusion coefficient σ. Suppose that $E\left[\int_0^1 \sigma(t)^2\mathrm{d}t\right]$ is limited and that μ is a.s. limited. Then ξ is a.s. limited.*

Proof. Since μ is a.s. limited, it follows that a.s. $\int_0^s \mu(t)\mathrm{d}t$ is limited (because a.s. $\max_{t\in\mathbf{T}\setminus\{1\}} |\mu(t)|$ is limited and $\max_{s\in\mathbf{T}} \left|\int_0^s \mu(t)\mathrm{d}t\right| \leq \max_{t\in\mathbf{T}\setminus\{1\}} |\mu(t)|$), and hence so is $\xi(0)+\int_0^s \mu(t)\mathrm{d}t$. What remains to be shown is that $\int \sigma\mathrm{d}W$ is a.s. limited. However,

$$E\left[\left|\int_0^1 \sigma\mathrm{d}W\right|^2\right] = E\left[\int_0^1 \sigma(t)^2\mathrm{d}t\right]$$

by the Itô isometry (Lemma 3.4), hence by the Cauchy–Schwarz inequality,

$$E\left[\left|\int_0^1 \sigma\mathrm{d}W\right|\right] \leq E\left[\int_0^1 \sigma(t)^2\mathrm{d}t\right]^{1/2},$$

and the right-hand side is limited by assumption. Since $\int \sigma\mathrm{d}W$ is a martingale (Theorem 3.2), we may apply the corollary to Nelson's martingale inequality (Corollary 2.13) and obtain that $\int \sigma\mathrm{d}W$ is a.s. limited. Since we have already seen that $\int \mu(t)\mathrm{d}t$ is a.s. limited, we conclude that ξ is a.s. limited. □

3.3 A Basic Radically Elementary Itô Formula

A function $f : \mathbf{R} \to \mathbf{R}$ is said to be *uniformly limited* if and only if there is some limited real C such that $|f(x)| \leq C$ for all $x \in \mathbf{R}$. f is said to be *limited* if and only if $f(x)$ is limited for all limited $x \in \mathbf{R}$.

If $\omega \in \Omega$ and ξ is a stochastic process, then $\xi(\omega)$ will also be called the *ω-trajectory of* ξ; a trajectory $\lambda : \mathbf{T}' \to \mathbf{R}$ is said to be *limited* if and only if $\lambda(t)$ is limited for all $t \in \mathbf{T}'$.

Let now $p \in \mathbf{R}$. A trajectory $\lambda : \mathbf{T}' \to \mathbf{R}$ is said to be $\mathfrak{o}((\mathrm{d}t)^p)$ (limited, respectively) if and only if $\max_{t\in\mathbf{T}'} |\lambda(t)|$ is $\mathfrak{o}((\mathrm{d}t)^p)$ (limited, respectively).

The following result, a basic radically elementary version of the Itô–Doeblin formula, is essentially due to Benoît [10, Proposition 4.6.1]. It allows to calculate the increment process of a function of a Wiener walk plus linear drift.

Lemma 3.9 (Itô–Doeblin formula for Wiener walks with additive linear drift). *Let $L(t) = \mu t + \sigma W(t)$ for all $t \in \mathbf{T}$ for limited $\mu, \sigma \in \mathbf{R}$, and let f be a thrice continuously differentiable function. Then for every $s \in \mathbf{T}$ and every ω such that the ω-trajectories of $f''(L)$ and $f'''(L)$ are $\mathfrak{o}\left((\mathrm{d}t)^{-1/2}\right)$,*

$$f(L(s)(\omega)) - f(L(0)(\omega)) \simeq \int_0^s f'(L(t)(\omega))\mathrm{d}L(t)(\omega) + \frac{\sigma^2}{2}\int_0^s f''(L(t)(\omega))\mathrm{d}t. \tag{3.4}$$

In particular, if f'' and f''' are uniformly limited, then the above formula (3.4) holds for all $\omega \in \Omega$.

Proof. Let us suppress the argument ω. Fix $t \in \mathbf{T} \setminus \{1\}$. Then, by the third-order Taylor formula,

$$\mathrm{d}f(L(t)) = f'(L(t))\mathrm{d}L(t) + \frac{1}{2}f''(L(t))\left(\mathrm{d}L(t)\right)^2 + \frac{1}{6}f'''(\zeta(t))\left(\mathrm{d}L(t)\right)^3,$$

for some $\zeta(t) \in [L(t), L(t+\mathrm{d}t)]$. By assumption on L,

$$(\mathrm{d}L(t))^2 = \mu^2(\mathrm{d}t)^2 + 2\mu\sigma\mathrm{d}t\,\mathrm{d}W(t) + \sigma^2\mathrm{d}t,$$

hence

$$(\mathrm{d}L(t))^3 = \left(\mu^2\mathrm{d}t + 2\mu\sigma\mathrm{d}W(t) + \sigma^2\right)^{3/2}(\mathrm{d}t)^{3/2}.$$

By assumption,

$$\max_{t\in\mathbf{T}\cap[0,s]}\left|f''\left(L(t)\right)\right| \vee \left|f'''\left(L(t)\right)\right| = \mathfrak{o}\left((\mathrm{d}t)^{-1/2}\right),$$

so

$$\begin{aligned}
&\left|f(L(s)) - f(L(0)) - \int_0^s f'(L(t))\mathrm{d}L(t) - \frac{\sigma^2}{2}\int_0^s f''(L(t))\mathrm{d}t\right| \\
&= \left|\sum_{t<s}\mathrm{d}f(L(t))f'(L(t))\mathrm{d}L(t) - \frac{\sigma^2}{2}f''(L(t))\mathrm{d}t\right| \\
&= \left|\sum_{t<s}\left(\frac{1}{2}f''(L(t))\left(\mu^2(\mathrm{d}t)^2 + 2\mu\sigma\mathrm{d}t\,\mathrm{d}W(t)\right)\right.\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad + \frac{1}{6} f'''(\zeta(t)) \left(\mu^2 \mathrm{d}t + 2\mu\sigma \mathrm{d}W(t) + \sigma^2\right)^{3/2} (\mathrm{d}t)^{3/2} \Big) \Big| \\
&\leq C \sum_{t<s} \frac{1}{2} \left|\mu^2 (\mathrm{d}t)^2 + 2\mu\sigma \mathrm{d}t\, \mathrm{d}W(t)\right| \\
&\quad + \frac{1}{6} \left| \left(\mu^2 \mathrm{d}t + 2\mu\sigma \mathrm{d}W(t) + \sigma^2\right)^{3/2} (\mathrm{d}t)^{3/2} \right| \\
&\leq C \frac{s}{\mathrm{d}t} \left(\frac{1}{2}\mu^2 (\mathrm{d}t)^2 + |\mu\sigma| \mathrm{d}t \sqrt{\mathrm{d}t} + \frac{1}{6} \left(\mu^2 \mathrm{d}t + 2|\mu\sigma| \sqrt{\mathrm{d}t} + \sigma^2\right)^{3/2} (\mathrm{d}t)^{3/2} \right) \\
&\leq C s \left(\frac{1}{2}\mu^2 \mathrm{d}t + |\mu\sigma| \sqrt{\mathrm{d}t} + \frac{1}{6} \left(\mu^2 \mathrm{d}t + 2|\mu\sigma| \sqrt{\mathrm{d}t} + \sigma^2\right)^{3/2} \sqrt{\mathrm{d}t} \right) \\
&\simeq 0.
\end{aligned}
$$

□

In applications, one will rather often not be able to literally apply this version of the Itô–Doeblin formula in Lemma 3.9, as it is usually not obvious how to establish sufficient upper bounds on $f''(L)$ of $f'''(L)$. Nevertheless, the proof idea—i.e. a third-order Taylor expansion—will usually be applicable even in those settings. An important example will be studied in Sect. 3.4 of Chap. 3, which is concerned with a particularly simple class of Itô processes.

3.4 Analytic Excursion: A Radically Elementary Treatment of Geometric Itô Processes with Monotone Drift

Geometric Itô processes are processes which satisfy a stochastic differential equation of the form

$$
\forall t \in \mathbf{T} \setminus \{1\} \qquad \mathrm{d}\xi(t) = \xi(t)\mu(t)\mathrm{d}t + \xi(t)\sigma(t)\mathrm{d}W(t) \tag{3.5}
$$

for some limited $\xi(0)$. For limited μ, σ, one can show that $\xi(t) > 0$ for all $t \in \mathbf{T}$. (See the proof of Lemma 3.10.) Hence, whenever $\mu(t) \geq 0$ for all $t \in \mathbf{T}$ or $\mu(t) \leq 0$ for all $t \in \mathbf{T}$, the drift coefficient of the Itô process ξ will be either monotonely increasing or decreasing in t (for every fixed $\omega \in \Omega$).

Such processes are of paramount importance in applications of Girsanov's theorem, in particular to mathematical finance, and therefore merit to be studied in some detail. (For instance, the radically elementary analogue of the stock price process of the classical Black–Scholes [18] model satisfies Eq. (3.5) for constant limited μ, σ.) However, the main parts of the book—in particular our version of Girsanov's theorem—do not depend on the results of this Sect. 3.4.

Lemma 3.10. *Let μ, σ be $\mathcal{F}$-adapted limited processes, let ξ be the process given by*

$$\forall t \in \mathbf{T} \setminus \{1\} \qquad \mathrm{d}\xi(t) = \xi(t)\mu(t)\mathrm{d}t + \xi(t)\sigma(t)\mathrm{d}W(t)$$

for some limited $\xi(0) \in \mathbf{R}_{>0}$. Suppose that either $\mu(t) \geq 0$ for all $t \in \mathbf{T}$ or $\mu(t) \leq 0$ for all $t \in \mathbf{T}$. Then, for all $s \in \mathbf{T}$, $\xi(s)$ is $L^1(P)$ with limited second moment. Moreover, with probability 1, one has $\xi(s) > 0$ for all $s \in \mathbf{T}$.

The proof uses a radically elementary analogue of the Harnack inequality.

Lemma 3.11 (Harnack inequality). *Let $\alpha, \gamma \in \mathbf{R}_{>0}$ and $v : \mathbf{T} \to \mathbf{R}$. If*

$$\forall s \in \mathbf{T} \qquad v(s) \leq \alpha + \gamma \int_0^s v(t)\mathrm{d}t,$$

then

$$\forall s \in \mathbf{T} \qquad v(s) \leq \alpha \mathrm{e}^{\gamma s}.$$

Proof of the Harnack inequality. The proof proceeds by induction on $s \in \mathbf{T}$. Let $C = \mathrm{e}^{\gamma}$ and suppose $v(s) \leq \alpha C^t$ for all $t < s$. Then, using that $\mathrm{e}^{\gamma\,\mathrm{d}t} = \sum_{n=0}^{\infty} \frac{\gamma^n (\mathrm{d}t)^n}{n!} \geq 1 + \gamma\,\mathrm{d}t$, one obtains

$$\begin{aligned}
v(s) &= \alpha + \gamma \int_0^s \alpha C^t \mathrm{d}t = \alpha + \gamma\alpha \sum_{\ell=0}^{s/\mathrm{d}t - 1} C^{\ell\,\mathrm{d}t}\mathrm{d}t \\
&= \alpha + \gamma\alpha \frac{C^s - 1}{C^{\mathrm{d}t} - 1}\mathrm{d}t = \alpha\left(1 + \gamma\,\mathrm{d}t\,\frac{\mathrm{e}^{\gamma s} - 1}{\mathrm{e}^{\gamma\,\mathrm{d}t} - 1}\right) \\
&\leq \alpha\left(1 + \gamma\,\mathrm{d}t\,\frac{\mathrm{e}^{\gamma s} - 1}{\gamma\,\mathrm{d}t}\right) \\
&= \alpha \mathrm{e}^{\gamma s} = \alpha C^s.
\end{aligned}$$

□

Proof of Lemma 3.10. Since $\mu, \sigma, \xi(0)$ are limited, there must be some limited $C \in \mathbf{R}_{>0}$ such that $|\mu(t)| \vee |\sigma(t)| \vee \xi(0) \leq C$ for all $t \in \mathbf{T}$. Combining this estimate with the fact that Itô integrals are martingales (Theorem 3.2), the Itô isometry (Lemma 3.4) and the Cauchy–Schwarz inequality, we may calculate for all $s \in \mathbf{T}$,

$$\begin{aligned}
&E\left[\xi(s)^2\right] \\
&\quad = \xi(0)^2 + 2\xi(0)\left(E\left[\int_0^s \xi(t)\mu(t)\mathrm{d}t\right] + \underbrace{E\left[\int_0^s \xi(t)\sigma(t)\mathrm{d}W(t)\right]}_{=0}\right)
\end{aligned}$$

$$
\begin{aligned}
&+E\left[\left|\int_0^s \xi(t)\mu(t)\mathrm{d}t\right|^2\right] + E\left[\left|\int_0^s \xi(t)\sigma(t)\mathrm{d}W(t)\right|^2\right] \\
&+2E\left[\left(\int_0^s \xi(t)\mu(t)\mathrm{d}t\right)\left(\int_0^s \xi(t)\sigma(t)\mathrm{d}W(t)\right)\right] \\
\leq\ &\xi(0)^2 + 2\xi(0)\ E\left[\left|\int_0^s \xi(t)\mu(t)\mathrm{d}t\right|^2\right]^{1/2} \\
&+E\left[\left|\int_0^s \xi(t)\mu(t)\mathrm{d}t\right|^2\right] + E\left[\int_0^s \xi(t)^2\sigma(t)^2\mathrm{d}t\right] \\
&+2E\left[\left|\int_0^s \xi(t)\mu(t)\mathrm{d}t\right|^2\right]^{1/2} E\left[\left|\int_0^s \xi(t)\sigma(t)\mathrm{d}W(t)\right|^2\right]^{1/2} \\
\leq\ &\xi(0)^2 + 2\xi(0)\ E\left[\left|\int_0^s \xi(t)\mu(t)\mathrm{d}t\right|^2\right]^{1/2} \\
&+E\left[\left|\int_0^s \xi(t)\mu(t)\mathrm{d}t\right|^2\right] + E\left[\int_0^s \xi(t)^2\sigma(t)^2\mathrm{d}t\right] \\
&+2E\left[\left|\int_0^s \xi(t)\mu(t)\mathrm{d}t\right|^2\right]^{1/2} E\left[\int_0^s \xi(t)^2\sigma(t)^2\mathrm{d}t\right]^{1/2}.
\end{aligned}
$$

Note that $\frac{1}{s}\int_0^s \cdot\mathrm{d}t$ defines an expectation operator on $\mathbf{T} \cap [0, s)$. Applying Jensen's inequality, we find for arbitrary η and $s \in \mathbf{T}$,

$$
\begin{aligned}
\left|\frac{1}{s}\int_0^s \eta(t)\mathrm{d}t\right|^2 &= s^2\left|\frac{1}{s}\int_0^s \eta(t)\mathrm{d}t\right|^2 \\
&\leq s^2\frac{1}{s}\int_0^s \eta(t)^2\mathrm{d}t = s\int_0^s \eta(t)^2\mathrm{d}t \\
&\leq \int_0^s \eta(t)^2\mathrm{d}t.
\end{aligned}
$$

Applying this to $\eta = \xi\mu$ in the above estimates, we obtain

$$
\begin{aligned}
E\left[\xi(s)^2\right] \leq\ &\xi(0)^2 + 2\xi(0)\ E\left[\int_0^s \xi(t)^2\mu(t)^2\mathrm{d}t\right]^{1/2} \\
&+E\left[\int_0^s \xi(t)^2\mu(t)^2\mathrm{d}t\right] + E\left[\int_0^s \xi(t)^2\sigma(t)^2\mathrm{d}t\right]
\end{aligned}
$$

$$
\begin{aligned}
&+2E\left[\int_0^s \xi(t)^2\mu(t)^2\mathrm{d}t\right]^{1/2} E\left[\int_0^s \xi(t)^2\sigma(t)^2\mathrm{d}t\right]^{1/2} \\
&\leq C^2 + 2C^2\, E\left[\int_0^s \xi(t)^2\mathrm{d}t\right]^{1/2} \\
&\quad +C^2\, E\left[\int_0^s \xi(t)^2\mathrm{d}t\right] + C^2\, E\left[\int_0^s \xi(t)^2\mathrm{d}t\right] \\
&\quad +2C^2\, E\left[\int_0^s \xi(t)^2\mathrm{d}t\right]^{1/2} E\left[\int_0^s \xi(t)^2\mathrm{d}t\right]^{1/2} \\
&\leq C^2 + 2C^2\, E\left[\int_0^s \xi(t)^2\mathrm{d}t\right]^{1/2} + 4C^2\, E\left[\int_0^s \xi(t)^2\mathrm{d}t\right].
\end{aligned}
$$

Now, clearly $x^{1/2} \leq 1 + x$ for all $x \geq 0$, whence

$$
E\left[\xi(s)^2\right] \leq 3C^2 + 6C^2\, E\left[\int_0^s \xi(t)^2\mathrm{d}t\right].
$$

Applying the Harnack inequality (Lemma 3.11) with $v : t \mapsto E\left[\xi(t)^2\right]$ and suitable α and γ, we find that $E\left[\xi(s)^2\right]$ is limited (as C is limited). Therefore, $\xi(s)$ is $L^1(P)$ by Remark 2.9, and $E\left[|\xi(s)|\right]$ is limited (by the Cauchy–Schwarz inequality).

Now one can prove that $\xi(t) > 0$ for all $t \in \mathbf{T}$. Indeed, let $\omega \in \Omega$ be such that $\{t \in \mathbf{T} : \xi(t)(\omega) \leq 0\}$ is nonempty, and let $t_\omega + \mathrm{d}t$ be its least element (which must be $\geq \mathrm{d}t$, as $\xi(0) > 0$). Then, $\xi(t_\omega)(\omega) > 0$ while $0 \geq \xi(t_\omega + \mathrm{d}t)(\omega) = \xi(t_\omega)(\omega)\,(1 + \mu(t)(\omega)\mathrm{d}t + \sigma(t_\omega)(\omega)\mathrm{d}W(t_\omega)(\omega))$, so $1 + \mu(t)(\omega)\mathrm{d}t + \sigma(t_\omega)(\omega)\mathrm{d}W(t_\omega)(\omega) \leq 0$, hence either $\sigma(t_\omega)(\omega) \leq -(1 + \mu(t)(\omega)\mathrm{d}t)\,/\sqrt{\mathrm{d}t}$ (if $\mathrm{d}W(t_\omega)(\omega) = \sqrt{\mathrm{d}t}$) or $\sigma(t_\omega)(\omega) \geq (1 + \mu(t)(\omega)\mathrm{d}t)\,/\sqrt{\mathrm{d}t}$ (if $\mathrm{d}W(t_\omega)(\omega) = -\sqrt{\mathrm{d}t}$). In either case, $\sigma(t_\omega)(\omega)$ is unlimited (as μ is limited and thus $1 + \mu(t)(\omega)\mathrm{d}t \simeq 1$). Hence the set of ω such that $\xi(t)(\omega) > 0$ for all $t \in \mathbf{T}$ is for every limited $C' > 0$ a superset of the set of all $\omega \in \Omega$ such that $|\sigma(t)(\omega)| \leq C'$, and for sufficiently large limited C', this set has probability 1, as σ is a limited process.

Therefore, since $\mu(t)$ is either nonpositive for all $t \in \mathbf{T}$ or nonnegative for all $t \in \mathbf{T}$, $\left(\int_0^s \xi(t)\mu(t)\mathrm{d}t\right)_{s\in\mathbf{T}}$ is either a decreasing or an increasing process. On the other hand, $\int \xi\sigma\mathrm{d}W$ is a martingale (by the converse of the martingale representation theorem, Theorem 3.2) as the recursive definition of ξ ensures its adaptedness, so $\xi = \xi(0) + \int \xi(t)\mu(t)\mathrm{d}t + \int \xi(t)\sigma(t)\mathrm{d}W(t)$ is a submartingale or a supermartingale. Therefore, we may apply the corollary to Nelson's super-/submartingale inequality (Corollary 2.13), which, combined with the limitedness of $\xi(0)$ and $E\left[|\xi(s)|\right]$ (see above), yields that ξ is a.s. limited. □

Lemma 3.12. *Let μ, σ be limited $\mathcal{F}$-adapted stochastic processes, and let ξ be the process defined by*

$$
\mathrm{d}\xi(t) = \xi(t)\mu(t)\,\mathrm{d}t + \xi(t)\sigma(t)\,\mathrm{d}W(t)
$$

for all $t \in \mathbf{T} \setminus \{1\}$, *wherein* $\xi(0)$ *is a limited real number* > 0. *Suppose that either* $\mu(t) \geq 0$ *for all* $t \in \mathbf{T}$ *or* $\mu(t) \leq 0$ *for all* $t \in \mathbf{T}$. *Then, a.s. for all* $s \in \mathbf{T}$,

$$\xi(s) \simeq \xi(0) \exp\left(\int_0^s \mu(t)\mathrm{d}t + \int_0^s \sigma(t)\, \mathrm{d}W(t) - \frac{1}{2}\int_0^s \sigma(t)^2 \mathrm{d}t\right). \tag{3.6}$$

Hence, if $\xi(0) \gg 0$, *then a.s. for all* $s \in \mathbf{T}$, $\xi(s) \gg 0$.

Proof. Since $\frac{1}{\xi(t)}\mathrm{d}\xi(t) = \int_0^s \mu(t)\mathrm{d}t + \int_0^s \sigma(t)\ \mathrm{d}W(t)$ (the subtrahend in the argument of the exponential function in Eq. (3.7)) it is enough to prove that

$$\ln \xi(s) - \ln \xi(0) \simeq \int_0^s \frac{1}{\xi(t)}\mathrm{d}\xi(t) - \frac{1}{2}\int_0^s \sigma(t)^2 \mathrm{d}t,$$

and since

$$\frac{1}{\xi(t)^2}(\mathrm{d}\xi(t))^2 = \mu(t)^2(\mathrm{d}t)^2 + 2\mu(t)\sigma(t)\mathrm{d}t\mathrm{d}W(t) + \sigma^2\mathrm{d}t = \sigma(t)^2\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right),$$

it is actually enough to show that

$$\ln \xi(s) - \ln \xi(0) \simeq \int_0^s \frac{1}{\xi(t)}\mathrm{d}\xi(t) - \frac{1}{2}\int_0^s \frac{1}{\xi(t)^2}(\mathrm{d}\xi(t))^2. \tag{3.7}$$

Now, since $\ln' : x \mapsto 1/x$, $\ln'' : x \mapsto -1/x^2$, $\ln''' : x \mapsto 2/x^3$, the third-order Taylor formula yields for every $t \in \mathbf{T}$

$$\mathrm{d}(\ln \xi(t)) = \frac{1}{\xi(t)}\mathrm{d}\xi(t) - \frac{1}{2}\frac{1}{\xi(t)^2}(\mathrm{d}\xi(t))^2 + \frac{1}{3}\frac{1}{\bar{\xi}(t)^3}(\mathrm{d}\xi(t))^3$$

for some $\bar{\xi}(t) \in [\xi(t), \xi(t + \mathrm{d}t)] \cup [\xi(t + \mathrm{d}t), \xi(t)]$, hence

$$\begin{aligned}\ln \xi(s) - \ln \xi(0) = \int_0^s \mathrm{d}(\ln \xi(t)) &= \int_0^s \frac{1}{\xi(t)}\mathrm{d}\xi(t) - \frac{1}{2}\int_0^s \frac{1}{\xi(t)^2}(\mathrm{d}\xi(t))^2 \\ &\quad + \frac{1}{3}\int_0^s \frac{1}{\bar{\xi}(t)^3}(\mathrm{d}\xi(t))^3\end{aligned}$$

for all $s \in \mathbf{T}$. All we need to prove therefore is that a.s. for all $s \in \mathbf{T}$,

$$\int_0^s \frac{1}{\bar{\xi}(t)^3}(\mathrm{d}\xi(t))^3 = \int_0^s \frac{\xi(t)^3}{\bar{\xi}(t)^3}(\mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t))^3 \simeq 0.$$

However, combining $\xi(t) > 0$ with the fact that $\bar{\xi}(t) \in [\xi(t), \xi(t + \mathrm{d}t)] \cup [\xi(t + \mathrm{d}t), \xi(t)]$, one gets the following uniform bound:

$$\left|\frac{\xi(t)}{\bar{\xi}(t)}\right| \leq \frac{\xi(t)}{\xi(t) \wedge \underbrace{\xi(t+\mathrm{d}t)}_{=\xi(t)+\mathrm{d}\xi(t)}} \leq 1 \vee \frac{1}{1+\underbrace{\mu(t)\mathrm{d}t+\sigma(t)\mathrm{d}W(t)}_{\simeq 0}} \simeq 1 \ll 2.$$

Moreover, $\mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t) = \mathcal{O}\left((\mathrm{d}t)^{1/2}\right)$, therefore we obtain indeed a.s.

$$\int_0^s \frac{\xi(t)^3}{\bar{\xi}(t)^3} \underbrace{(\mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t))^3}_{=\mathcal{O}\left((\mathrm{d}t)^{3/2}\right)} = \mathcal{O}\left((\mathrm{d}t)^{1/2}\right) \simeq 0.$$

□

3.5 The Radically Elementary Version of Lévy's Characterization of Wiener Processes

One of the most remarkable results in Nelson's *Radically elementary probability theory* is a single, unified theorem, called "de Moivre–Laplace–Lindeberg–Feller–Wiener–Lévy–Doob–Erdős–Kac–Donsker–Prokhorov theorem" by Nelson [60, Chap. 18], which entails:

- The necessity and sufficiency of the Lindeberg–Feller condition for the central limit theorem of de Moivre and Laplace.
- Wiener's result about the a.s. continuity of the trajectories of Wiener processes.
- Donsker's invariance principle.
- Lévy's martingale characterization of Wiener processes.

The last item (Lévy's martingale characterization of Wiener processes) is of great importance in stochastic analysis and its applications. It means that whenever a martingale (with respect to the filtration generated by a given Wiener process) has the same quadratic variation as the Wiener process, it already is the Wiener process; a related result is the theorem that the only path-continuous and square-integrable martingale which has stationary and independent increments (i.e. is a *Lévy process*[2]) is a (constant multiple of a) Wiener process.

Keeping in mind that the filtration generated by the Wiener process is a particularly simple and natural one, Lévy's martingale characterization informally asserts that any martingale which has a few desirable properties will already be, up to multiplicative constants, a Wiener process or the exponential of a Wiener process plus a linear drift term (a geometric Wiener process). As a consequence, Lévy's martingale characterization can be fruitfully applied both within pure mathematics

[2]For more on Lévy processes—from the perspective of radically elementary probability theory—see Chap. 9.

(for instance, in the proof of Girsanov's theorem, which establishes a relation between changing the probability measure and adding a linear drift term to the Wiener process) and in mathematical finance (as a mathematical rationale for the adequacy of the Samuelson–Black–Scholes model).

Nelson's unified result, which entails a radically elementary version of Lévy's martingale characterization, can be stated as follows:

Remark 3.13. (Cf. Nelson [60, Theorem 18.1, p. 75].) For a normalized martingale $(\xi(t))_{t\in\mathbf{T}}$ with $\xi(0) = 0$, the following three conditions are equivalent:

- ξ is a Wiener process,
- $\xi(1)$ is $L^2(P)$ and ξ is P-a.s. continuous,
- ξ satisfies the *(near) Lindeberg condition*, i.e.

$$E\left[\sum_{t\in\mathbf{T}\setminus\{1\}} (\mathrm{d}\xi(t))^2\right] \simeq E\left[\sum_{t\in\mathbf{T}\setminus\{1\}} \left(\mathrm{d}\xi(t)\,\chi_{\{|\mathrm{d}\xi(t)|\leq\varepsilon\}}\right)^2\right]$$

for all $\varepsilon \gg 0$.

Chapter 4
The Radically Elementary Girsanov Theorem and the Diffusion Invariance Principle

4.1 Girsanov's Theorem

Lemma 4.1. *Let η be a limited, $\mathcal{F}$-adapted stochastic process, let ξ be the process defined by $\xi(0) = 1$ and $\mathrm{d}\xi(t) = \xi(t)\eta(t)\,\mathrm{d}W(t)$ for all $t \in \mathbf{T} \setminus \{1\}$. Then, for all $s \in \mathbf{T}$,*

(1) $\xi(s) > 0$ (with P-probability 1).
Moreover,
(2) ξ is an $(\mathcal{F}, P)$-martingale.
(3) $Q : A \mapsto \int_A \xi(1)\,\mathrm{d}P$ is a finite probability measure.
(4) $Q \restriction \mathcal{F}_t : A \mapsto \int_A \xi(t)\,\mathrm{d}P$ for all $t \in \mathbf{T}$. (ξ is the density process of Q.)

By Lemma 3.8, ξ is a.s. limited.

Proof of Lemma 4.1.

(1) A particularly simple form of the argument in the proof of Lemma 3.10 can be used here. Let $\omega \in \Omega$ be such that $\{t \in \mathbf{T} : \xi(t)(\omega) \leq 0\}$ is nonempty, and let $t_\omega + \mathrm{d}t$ be its least element (which must be $> \mathrm{d}t$, as $\xi(0) = 1$ and η is limited, whence $\xi(\mathrm{d}t) \simeq \xi(0) \gg 0$); for all other $\omega' \in \Omega$, put $t_{\omega'} = 0$. Then, $\xi(t_\omega)(\omega) > 0$ and $0 \geq \xi(t_\omega + \mathrm{d}t)(\omega) = \xi(t_\omega)(\omega)\,(1 + \eta(t_\omega)(\omega)\mathrm{d}W(t_\omega)(\omega))$, so $1 + \eta(t_\omega)(\omega)\mathrm{d}W(t_\omega)(\omega) \leq 0$, hence either $\eta(t_\omega)(\omega) \leq -1/\sqrt{\mathrm{d}t}$ (if $\mathrm{d}W(t_\omega)(\omega) = \sqrt{\mathrm{d}t}$) or $\eta(t_\omega)(\omega) \geq 1/\sqrt{\mathrm{d}t}$ (if $\mathrm{d}W(t_\omega)(\omega) = -\sqrt{\mathrm{d}t}$). In either case, $\eta(t_\omega)(\omega)$ is unlimited. Since Ω is finite, the minimum of $|\eta(t_\omega)(\omega)|$ for all those $\omega \in \Omega$ with $t_\omega > 0$ exists and must be an unlimited number, say C. Thus,

$$\bigcup_{t\in\mathbf{T}} \{\xi(t) \leq 0\} = \{\omega \in \Omega : \xi\,(t_\omega + \mathrm{d}t)\,(\omega) \leq 0\}$$
$$\subseteq \{\omega \in \Omega : |\eta(t_\omega)(\omega)| \geq C\} \subseteq \bigcup_{t\in\mathbf{T}} \{|\eta(t)| \geq C\},$$

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_4,

and the right-hand side has probability zero since η was assumed to be limited.

(2) Fix $s \in \mathbf{T}$. Note that one can prove, via an induction on t, that $\xi(s) = \xi(0) + \sum_{t<s} \xi(t)\eta(t)\mathrm{d}W(t)$ is $\mathcal{F}_s$-measurable. It follows that ξ is $\mathcal{F}$-adapted. Hence, by the converse of the martingale representation theorem (Theorem 3.2), $\int \xi\eta \mathrm{d}W$ is a martingale, and so is $\xi = \xi(0) + \int \xi\eta \mathrm{d}W$.

(3) Part 1 of the lemma says that $\xi(1) > 0$ with P-probability 1, and part 2 of the lemma allows to calculate $E[\xi(1)] = \xi(0) = 1$.

(4) Let $t \in \mathbf{T}$ and A be $\mathcal{F}_t$-measurable. Then, using that ξ is an $(\mathcal{F}, P)$-martingale (part 2 in the Lemma)

$$Q(A) = \int \chi_A \xi(1)\, \mathrm{d}P = \int \chi_A E[\xi(1)|\, \mathcal{F}_t]\, \mathrm{d}P = \int \chi_A \xi(t)\, \mathrm{d}P.$$

□

Theorem 4.2 (Girsanov's theorem). *Let η be limited and $\mathcal{F}$-adapted stochastic process, and let ξ and Q be as in Lemma 4.1. Define a process W^G by*

$$W^G(0) = 0, \qquad \forall t \in \mathbf{T} \setminus \{1\} \qquad \mathrm{d}W^G(t) = \mathrm{d}W(t) - \eta(t)\mathrm{d}t.$$

Then, W^G is a Wiener martingale on (Ω, Q).

A simpler radically elementary version of Girsanov's theorem (one which relates the distribution of W^G under Q to the distribution of W under P) was established, by means of a second-order Taylor expansion, by Benoît [10, Theorem 4.6.1].

Proof. Our proof strategy is as follows: First, we shall prove that the process W^Q, defined by

$$W^Q(0) = 0, \qquad \forall t \in \mathbf{T} \setminus \{1\} \qquad \mathrm{d}W^Q(t) = \frac{\mathrm{d}W(t) - \eta(t)\mathrm{d}t}{\sqrt{1 - \eta(t)^2 \mathrm{d}t}},$$

is a Q-a.s. continuous normalized martingale on (Ω, Q) with $W^Q(1)$ being $L^2(Q)$. Nelson's radically elementary version of Lévy's martingale characterization of Wiener processes (see Remark 3.13) allows us then to deduce that W^Q is a Wiener process on (Ω, Q). Thereafter, we will show that for all $t \in \mathbf{T}$, one has $W^G(t) \simeq W^Q(t)$. This implies, due to yet another result of Nelson's (see Remark 2.1), that W^G is nearly equivalent to W^Q—and hence a Wiener process, too. In passing, we shall see that W^G is a Q-martingale (as W^Q is a Q-martingale).

Let us first show that W^Q is a normalized martingale: Exploiting parts 3 and 4 of Lemma 4.1 and using a general form of Bayes' formula (which asserts that for every $s > t$ one has

$$\xi(t) E^Q [z|\, \mathcal{F}_t] = E[\xi(s) z|\, \mathcal{F}_t], \tag{4.1}$$

wherein z is an $\mathcal{F}_s$-measurable random variable and Q a probability measure, equivalent to P, with density process ξ)[1] we see

$$E^Q\left[\mathrm{d}W^Q(t)\middle|\,\mathcal{F}_t\right] = \frac{1}{\xi(t)} E\left[\xi(t+\mathrm{d}t)\mathrm{d}W^Q(t)\middle|\,\mathcal{F}_t\right].$$

Hence, resubstituting $\mathrm{d}W^Q(t)$ and using the definition of the increments of ξ, we obtain

$$\begin{aligned}
&E^Q\left[\mathrm{d}W^Q(t)\middle|\,\mathcal{F}_t\right]\\
&= \frac{1}{\xi(t)} E\left[\xi(t+\mathrm{d}t)\frac{\mathrm{d}W(t)-\eta(t)\mathrm{d}t}{\sqrt{1-\eta(t)^2\mathrm{d}t}}\middle|\,\mathcal{F}_t\right]\\
&= \frac{1}{\sqrt{1-\eta(t)^2\mathrm{d}t}} E\left[\frac{\overbrace{\xi(t+\mathrm{d}t)}^{=\xi(t)+\mathrm{d}\xi(t)}}{\xi(t)}\left(\mathrm{d}W(t)-\eta(t)\mathrm{d}t\right)\middle|\,\mathcal{F}_t\right]\\
&= \frac{1}{\sqrt{1-\eta(t)^2\mathrm{d}t}} E\left[\left(1+\eta(t)\mathrm{d}W(t)\right)\left(\mathrm{d}W(t)-\eta(t)\mathrm{d}t\right)\middle|\,\mathcal{F}_t\right]\\
&= \frac{1}{\sqrt{1-\eta(t)^2\mathrm{d}t}}\\
&\quad \times E\left[\mathrm{d}W(t)+\eta(t)\underbrace{\left(\mathrm{d}W(t)\right)^2}_{=\mathrm{d}t}-\eta(t)\mathrm{d}t-\eta(t)^2\mathrm{d}W(t)\mathrm{d}t\middle|\,\mathcal{F}_t\right]\\
&= \frac{1}{\sqrt{1-\eta(t)^2\mathrm{d}t}}\left(\underbrace{E\left[\mathrm{d}W(t)\middle|\,\mathcal{F}_t\right]}_{=0}-\eta(t)^2\mathrm{d}t\,\underbrace{E\left[\mathrm{d}W(t)\middle|\,\mathcal{F}_t\right]}_{=0}\right)\\
&= 0,
\end{aligned}$$

---

[1] Bayes' formula can be proven as follows: For all $s > t$, every $\mathcal{F}_s$-measurable z and every $\mathcal{F}_t$-measurable A, one has

$$\begin{aligned}
\int_A \xi(t) E^Q\left[z\middle|\,\mathcal{F}_t\right]\mathrm{d}Q &= \int_A E^Q\left[\xi(t)z\middle|\,\mathcal{F}_t\right]\mathrm{d}Q\\
&= \int_A \xi(t)z\,\mathrm{d}Q = \int_A \xi(t)z\xi(s)\,\mathrm{d}Q\\
&= \int_A \xi(t)\xi(s)z\,\mathrm{d}P = \int_A \xi(s)z\,\mathrm{d}Q\\
&= \int_A E\left[\xi(s)z\middle|\,\mathcal{F}_t\right]\mathrm{d}Q.
\end{aligned}$$

and similarly, one gets (using the definition of the increments of ξ as well as Bayes' formula, Eq. (4.1))

$$\begin{aligned}
E^Q[\mathrm{d}W(t)|\mathcal{F}_t] &= \frac{1}{\xi(t)}E[\xi(t+\mathrm{d}t)\mathrm{d}W(t)|\mathcal{F}_t] \\
&= E\left[\left.\frac{\overbrace{\xi(t+\mathrm{d}t)}^{=\xi(t)+\mathrm{d}\xi(t)}}{\xi(t)}\mathrm{d}W(t)\right|\mathcal{F}_t\right] \\
&= E[(1+\eta(t)\mathrm{d}W(t))\,\mathrm{d}W(t)|\mathcal{F}_t] \\
&= E\left[\left.\mathrm{d}W(t)+\eta(t)\underbrace{(\mathrm{d}W(t))^2}_{=\mathrm{d}t}\right|\mathcal{F}_t\right] \\
&= E[\mathrm{d}W(t)|\mathcal{F}_t]+\eta(t)\mathrm{d}t=\eta(t)\mathrm{d}t,
\end{aligned}$$

which leads to

$$\begin{aligned}
&E^Q\left[\left.\left(\mathrm{d}W^Q(t)\right)^2\right|\mathcal{F}_t\right] \\
&= \frac{1}{1-\eta(t)^2\mathrm{d}t}E^Q\left[\left.\overbrace{(\mathrm{d}W(t))^2}^{=\mathrm{d}t}-2\eta(t)\mathrm{d}t\,\mathrm{d}W(t)+\eta(t)^2(\mathrm{d}t)^2\right|\mathcal{F}_t\right] \\
&= \frac{1}{1-\eta(t)^2\mathrm{d}t}\left(\mathrm{d}t+\eta(t)^2(\mathrm{d}t)^2-2\eta(t)\mathrm{d}t\,\underbrace{E^Q[\mathrm{d}W(t)|\mathcal{F}_t]}_{=\eta(t)\mathrm{d}t}\right) \\
&= \frac{1}{1-\eta(t)^2\mathrm{d}t}\left(\mathrm{d}t-\eta(t)^2(\mathrm{d}t)^2\right)=\frac{\mathrm{d}t}{1-\eta(t)^2\mathrm{d}t}\left(1-\eta(t)^2\mathrm{d}t\right) \\
&= \mathrm{d}t.
\end{aligned}$$

Therefore, W^Q is a normalized martingale on (Ω, Q). Moreover, its increments are infinitesimal as η is limited. Hence, W^Q is Q-a.s. continuous by Remark 3.5. In addition, since $\sqrt{1-\eta(t)^2\mathrm{d}t}\,\mathrm{d}W^Q(t) = \mathrm{d}W(t)-\eta(t)\mathrm{d}t = W^G(t)$ and η is $\mathcal{F}$-adapted, we have

$$E^Q\left[\mathrm{d}W^G(t)\middle|\mathcal{F}_t\right]=\sqrt{1-\eta(t)^2\mathrm{d}t}\,E^Q\left[\mathrm{d}W^Q(t)\middle|\mathcal{F}_t\right]=0$$

for all $t\in\mathbf{T}\setminus\{1\}$, whence W^G is a martingale.

Next we shall show that $W^Q(1)$ is $L^2(Q)$, and for that purpose, we first compute $E^Q\left[W^Q(1)^4\right]$. For some standard natural numbers C_1, C_2, C_3, C_4 (viz. limited integer multiples of certain multinomial coefficients), we have

$$
\begin{aligned}
&E^Q\left[W^Q(1)^4\right] \\
&= E^Q\left[\left(\sum_{t<1} \mathrm{d}W^Q(t)\right)^4\right] \\
&= C_1 E^Q\left[\sum_{r<s<t<u<1} \mathrm{d}W^Q(r)\,\mathrm{d}W^Q(s)\,\mathrm{d}W^Q(t)\,\mathrm{d}W^Q(u)\right] \\
&\quad + C_2 E^Q\left[\sum_{r<s<t<1} \mathrm{d}W^Q(r)\,\mathrm{d}W^Q(s)\,\left(\mathrm{d}W^Q(t)\right)^2\right] \\
&\quad + C_3 E^Q\left[\sum_{s<t<1} \left(\mathrm{d}W^Q(s)\right)^2 \left(\mathrm{d}W^Q(t)\right)^2\right] \\
&\quad + C_4 E^Q\left[\sum_{r<s<1} \mathrm{d}W^Q(r)\,\left(\mathrm{d}W^Q(s)\right)^3\right] + E^Q\left[\sum_{r<1} \left(\mathrm{d}W^Q(r)\right)^4\right].
\end{aligned}
$$

Taking conditional expectations and exploiting the $\mathcal{F}_v$-linearity of $E^Q[\cdot|\mathcal{F}_v]$ for $v = t + \mathrm{d}t, s + \mathrm{d}t$, we obtain

$$
\begin{aligned}
&E^Q\left[W^Q(1)^4\right] \\
&= C_1 E^Q\left[\sum_{r<s<t<u<1} \mathrm{d}W^Q(r)\,\mathrm{d}W^Q(s)\,\mathrm{d}W^Q(t) E^Q\left[\mathrm{d}W^Q(u)\middle|\,\mathcal{F}_{t+\mathrm{d}t}\right]\right] \\
&\quad + C_2 E^Q\left[\sum_{r<s<t<1} \mathrm{d}W^Q(r)\,\mathrm{d}W^Q(s) E^Q\left[\left(\mathrm{d}W^Q(t)\right)^2\middle|\,\mathcal{F}_{s+\mathrm{d}t}\right]\right] \\
&\quad + C_3 E^Q\left[\sum_{s<t<1} \left(\mathrm{d}W^Q(s)\right)^2 E^Q\left[\left(\mathrm{d}W^Q(t)\right)^2\middle|\,\mathcal{F}_{s+\mathrm{d}t}\right]\right] \\
&\quad + C_4 E^Q\left[\sum_{r<s<1} \mathrm{d}W^Q(r)\,\left(\mathrm{d}W^Q(s)\right)^3\right] + E^Q\left[\sum_{r<1} \left(\mathrm{d}W^Q(r)\right)^4\right].
\end{aligned}
$$

Since W^Q is a normalized martingale on (Ω, Q), we have $E^Q\left[\mathrm{d}W^Q(u)\middle|\,\mathcal{F}_{t+\mathrm{d}t}\right] = 0$ and $E^Q\left[\left(\mathrm{d}W^Q(t)\right)^2\middle|\,\mathcal{F}_{s+\mathrm{d}t}\right] = \mathrm{d}t$ for all $u > t$ and $t > s$, hence

$$
\begin{aligned}
&E^Q\left[W^Q(1)^4\right] \\
&= C_2 \mathrm{d}t\, E^Q\left[\sum_{r<s<t<1} \mathrm{d}W^Q(r)\,\mathrm{d}W^Q(s)\right] + C_3 \mathrm{d}t\, E^Q\left[\sum_{s<t<1} \left(\mathrm{d}W^Q(s)\right)^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&+C_4 E^Q\left[\sum_{r<s<1} \mathrm{d}W^Q(r)\ \left(\mathrm{d}W^Q(s)\right)^3\right] + E^Q\left[\sum_{r<1}\left(\mathrm{d}W^Q(r)\right)^4\right]\\
&= C_2\mathrm{d}t\ E^Q\left[\sum_{r<s<t<1} \mathrm{d}W^Q(r) E^Q\left[\mathrm{d}W^Q(s)\middle|\, \mathcal{F}_{r+\mathrm{d}t}\right]\right]\\
&+C_3\mathrm{d}t \sum_{s<t<1} E^Q\left[\left(\mathrm{d}W^Q(s)\right)^2\right]\\
&+C_4 E^Q\left[\sum_{r<s<1} \mathrm{d}W^Q(r)\ \left(\mathrm{d}W^Q(s)\right)^3\right] + E^Q\left[\sum_{r<1}\left(\mathrm{d}W^Q(r)\right)^4\right]\\
&= C_3\mathrm{d}t \sum_{s<t<1} \mathrm{d}t\\
&+C_4 E^Q\left[\sum_{r<s<1} \mathrm{d}W^Q(r)\ \left(\mathrm{d}W^Q(s)\right)^3\right] + E^Q\left[\sum_{r<1}\left(\mathrm{d}W^Q(r)\right)^4\right]\\
&= C_3(\mathrm{d}t)^2 \sum_{s<t<1} 1\\
&+C_4 \sum_{r<s<1} E^Q\left[\mathrm{d}W^Q(r)\ \left(\mathrm{d}W^Q(s)\right)^3\right] + \sum_{r<1} E^Q\left[\left(\mathrm{d}W^Q(r)\right)^4\right].
\end{aligned}
$$

However, for every $t \in \mathbf{T}$,

$$
\begin{aligned}
\sum_{s<t<1} 1 &= \sum_{t<1}\sum_{s<t} 1 = \sum_{k=0}^{t/\mathrm{d}t-1}\sum_{i=0}^{k-1} 1\\
&= \sum_{k=0}^{t/\mathrm{d}t-1} k = \frac{1}{2}\frac{t}{\mathrm{d}t}\left(\frac{t}{\mathrm{d}t}-1\right) = \frac{1}{2(\mathrm{d}t)^2} t(t-\mathrm{d}t) < \frac{1}{2(\mathrm{d}t)^2},
\end{aligned}
$$

so $C_3(\mathrm{d}t)^2 \sum_{s<t<1} 1 < \frac{C_3}{2}$, and similarly, $\sum_{r<s<1} 1 < \frac{1}{2(\mathrm{d}t)^2}$. Also, clearly, $\sum_{r<1} 1 = \frac{1}{\mathrm{d}t}$.

Therefore, we can now estimate $E^Q\left[W^Q(1)^4\right]$ as follows:

$$
\begin{aligned}
E^Q\left[W^Q(1)^4\right] &\le \frac{C_3}{2} + \frac{C_4}{2(\mathrm{d}t)^2}\max_{r<s<1} E^Q\left[\mathrm{d}W^Q(r)\ \left(\mathrm{d}W^Q(s)\right)^3\right]\\
&+ \frac{1}{\mathrm{d}t}\max_{r<1} E^Q\left[\left(\mathrm{d}W^Q(r)\right)^4\right]\\
&\le \frac{C_3}{2} + \left(\frac{C_4}{2(\mathrm{d}t)^2} + \frac{1}{\mathrm{d}t}\right)\max_{r,s<1} E^Q\left[\mathrm{d}W^Q(r)\ \left(\mathrm{d}W^Q(s)\right)^3\right] \quad (4.2)
\end{aligned}
$$

However, since η is limited, there is some limited real number C such that

$$\forall t \in \mathbf{T} \qquad |\eta(t)| \leq C$$

(see Remark 2.10), whence we can establish an upper bound on the norm of the increments of W^Q:

$$\forall t \in \mathbf{T} \qquad \left|\mathrm{d}W^Q(t)\right| \leq \frac{|\mathrm{d}W(t)| + |\eta(t)|\,\mathrm{d}t}{\sqrt{1-\eta(t)^2\mathrm{d}t}} \leq 2(\sqrt{\mathrm{d}t} + C\,\mathrm{d}t) \leq 3\sqrt{\mathrm{d}t},$$

whence for all $r, s < 1$,

$$\left|\mathrm{d}W^Q(r)\,\left(\mathrm{d}W^Q(s)\right)^3\right| \leq 81\,(\mathrm{d}t)^2,$$

so

$$E^Q\left[\left|\mathrm{d}W^Q(r)\,\left(\mathrm{d}W^Q(s)\right)^3\right|\right] \leq 81\,(\mathrm{d}t)^2.$$

Combining this with Eq. (4.2), we see that

$$E^Q\left[W^Q(1)^4\right] \leq \frac{C_3}{2} + 81 \times \frac{C_4}{2} + 81\mathrm{d}t \ll \infty.$$

Hence, by Remark 2.9, it follows that $W^Q(1)$ is $L^2(Q)$.

Therefore, W^Q is a Q-a.s. continuous normalized martingale on (Ω, Q) with $W^Q(1)$ being $L^2(Q)$. Thus, by Nelson's radically elementary version of Lévy's martingale characterization of Wiener processes (see Remark 3.13), W^Q is a Wiener process on (Ω, Q).

Using the limitedness of η, we now show that the process W^G is infinitely close to W^Q, in the sense that $W^Q(t) \simeq W^G(t)$ for all $t \in \mathbf{T}$. Indeed,

$$\left|\mathrm{d}W^Q(t) - \mathrm{d}W^G(t)\right| = |\mathrm{d}W(t) - \eta(t)\mathrm{d}t| \left(\frac{1}{\sqrt{1-\eta(t)^2\mathrm{d}t}} - 1\right). \tag{4.3}$$

The second factor on the right-hand side of Eq. (4.3) can be estimated using a first-order Taylor expansion for the function $x \mapsto \frac{1}{\sqrt{1-x}}$ around $x_0 = 0$ at $x = \eta(t)^2\mathrm{d}t$: For all $t \in \mathbf{T}$, there exists some $\zeta(t) \in \left(-\eta(t)^2\mathrm{d}t, \eta(t)^2\mathrm{d}t\right)$ such that

$$\frac{1}{\sqrt{1-\eta(t)^2\mathrm{d}t}} - 1 = \frac{1}{2\,(1-\zeta(t))^{3/2}}\eta(t)^2\mathrm{d}t.$$

Hence, exploiting that there is some limited real number C such that $|\eta(t)| \leq C$ holds for all $t \in \mathbf{T}$, we deduce that for all $t \in \mathbf{T}$,

$$\frac{1}{\sqrt{1-\eta(t)^2 \mathrm{d}t}} - 1 \leq \frac{1}{2\left(1-C^2 \mathrm{d}t\right)^{3/2}} C^2 \mathrm{d}t.$$

In order to estimate the first factor on the right-hand side of Eq. (4.3), it is enough to note that as C limited, one has

$$|\mathrm{d}W(t) - \eta(t)\mathrm{d}t| \leq |\mathrm{d}W(t)| + |\eta(t)|\,\mathrm{d}t \leq \sqrt{\mathrm{d}t} + C\,\mathrm{d}t \leq 2\sqrt{\mathrm{d}t}$$

for all $t \in \mathbf{T}$.

Inserting these estimations into Eq. (4.3), we finally obtain

$$\begin{aligned}
\max_{t\in\mathbf{T}} \left|W^Q(t) - W^G(t)\right| &\leq \max_{t\in\mathbf{T}} \sum_{s<t} \left|\mathrm{d}W^Q(s) - \mathrm{d}W^G(s)\right| \\
&\leq \operatorname{card}(\mathbf{T}) \max_{t\in\mathbf{T}} \left|\mathrm{d}W^Q(t) - \mathrm{d}W^G(t)\right| \\
&\leq \frac{1}{\mathrm{d}t} \max_{t\in\mathbf{T}} |\mathrm{d}W(t) - \eta(t)\mathrm{d}t| \left(\frac{1}{\sqrt{1-\eta(t)^2\mathrm{d}t}} - 1\right) \\
&\leq \frac{1}{\mathrm{d}t} \frac{1}{2\left(1-C^2\mathrm{d}t\right)^{3/2}} C^2 \mathrm{d}t\, 2\sqrt{\mathrm{d}t} \\
&= \frac{C^2}{\left(1-C^2\mathrm{d}t\right)^{3/2}} \sqrt{\mathrm{d}t} \\
&\simeq 0.
\end{aligned}$$

(The last line holds since C and thus also $\frac{C^2}{(1-C\,\mathrm{d}t)^{3/2}}$ is limited.) This proves that $W^Q(t) \simeq W^G(t)$ for all $t \in \mathbf{T}$, just as claimed.

This implies (see Remark 2.1) that W^Q and W^G are equivalent. Hence, W^G is equivalent to a Wiener walk on (Ω, Q), too. In other words, W^G is a Wiener process on (Ω, Q). □

4.2 The Radically Elementary Diffusion Invariance Principle

A corollary to Girsanov's theorem is the diffusion invariance principle, which asserts that under some technical conditions, the diffusion coefficient of an Itô process remains essentially the same even when the probability measure is changed. A number of definitions is necessary to state our—even though rather basic—radically elementary version of this important result.

A probability measure Q on Ω is said to be *P-continuous* if and only if for all subsets A of Ω with $P(A) = 0$ also $Q(A) = 0$ (equivalently, if for all $\omega \in \Omega$, one has $Q\{\omega\} = 0$ whenever $P\{\omega\} = 0$). The probability measures Q and P are said to be *equivalent* if and only if Q is P-continuous and Q is P-continuous.

The *density of Q with respect to P* is the random variable

$$\frac{\mathrm{d}Q}{\mathrm{d}P} : \omega \mapsto \begin{cases} \frac{Q\{\omega\}}{P\{\omega\}}, & P\{\omega\} > 0, \\ 0, & P\{\omega\} = 0. \end{cases}$$

The *density process of Q with respect to P* is the stochastic process $\left(E\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\middle|\mathcal{F}_t\right]\right)_{t\in\mathbf{T}}$. If Q is P-continuous, then

$$Q(A) = \sum_{\omega\in A} Q\{\omega\} = \sum_{\omega\in A} \frac{\mathrm{d}Q}{\mathrm{d}P}(\omega)P\{\omega\} = E\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\chi_A\right],$$

and for all $t \in \mathbf{T}$ and $\mathcal{F}_t$-measurable A,

$$Q(A) = E\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\chi_A\right] = E\left[E\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\chi_A\middle|\mathcal{F}_t\right]\right] = E\left[E\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\middle|\mathcal{F}_t\right]\chi_A\right].$$

Remark 4.3 (Positive density process for equivalent measures). If Q is equivalent to P, then $\frac{\mathrm{d}Q}{\mathrm{d}P} > 0$ (with P-probability 1) and therefore $E\left[\frac{\mathrm{d}Q}{\mathrm{d}P}\middle|\mathcal{F}_t\right] > 0$ (with P-probability 1) for all $t \in \mathbf{T}$.

Proof. If $\frac{\mathrm{d}Q}{\mathrm{d}P}(\omega) = 0$ for some $\omega \in \Omega$, then by the equivalence of P and Q, we must have $P\{\omega\} = 0$. □

Definition 4.4. Let ξ be a stochastic process. The *quadratic-variation derivative* of ξ is the process $\langle\xi\rangle = (\langle\xi\rangle(t))_{t\in\mathbf{T}\setminus\{1\}}$ defined by

$$\langle\xi\rangle(t) = \frac{(\mathrm{d}\xi(t))^2}{\mathrm{d}t}$$

for all $t \in \mathbf{T}\setminus\{1\}$. The *relative quadratic-variation derivative* of ξ is the process $\left(\frac{\langle\xi\rangle(t)}{\xi(t)^2}\chi_{\{\xi(t)\neq 0\}}\right)_{t\in\mathbf{T}\setminus\{1\}}$.

Corollary 4.5 (Diffusion invariance principle). *Let Q be equivalent to P with a density process that has a limited relative quadratic-variation derivative. Then, there exists a Wiener martingale W^G on (Ω, Q) such that for all processes μ, σ there is some process $\mu^{\sigma,Q}$ such that for all $s \in \mathbf{T}$,*

$$\int_0^s \mu(t)\mathrm{d}t + \int_0^s \sigma(t)\mathrm{d}W(t) = \int_0^s \mu^{\sigma,Q}(t)\mathrm{d}t + \int_0^s \sigma(t)\mathrm{d}W^G(t). \tag{4.4}$$

Note that the left-hand side of Eq. (4.4) is an Itô process on (Ω, P) whilst the right-hand side is an Itô process on (Ω, Q)—however, although under different probability measures, the diffusion coefficients are the same.

Proof. Let ξ be the density process of Q. By the tower property of conditional expectations, ξ is a martingale. The martingale representation theorem (Theorem 3.2) yields the existence of some adapted process ϕ such that

$$\mathrm{d}\xi(t) = \phi(t)\mathrm{d}W(t)$$

for all $t \in \mathbf{T} \setminus \{1\}$. By Remark 4.3, $\xi(t) > 0$ (with P-probability 1) for all $t \in \mathbf{T}$, whence we may define

$$\eta(t) = \frac{\phi(t)}{\xi(t)}$$

for all $t \in \mathbf{T}$. By assumption, there exists a limited C such that

$$\frac{(\mathrm{d}\xi(t))^2}{\xi(t)^2\mathrm{d}t} \leq C,$$

hence

$$C \geq \frac{(\phi(t)\mathrm{d}W(t))^2}{\xi(t)^2\mathrm{d}t} = \left(\frac{\phi(t)}{\xi(t)}\right)^2 = \eta(t)^2,$$

whence η is limited, whilst by definition

$$\mathrm{d}\xi(t) = \xi(t)\eta(t)\mathrm{d}W(t).$$

Therefore, Girsanov's theorem (Theorem 4.2) may be applied, and ensures that $W^G = \left(W(s) - \int_0^s \eta(t)\mathrm{d}t\right)_{s\in\mathbf{T}}$ is a Wiener martingale on (Ω, Q). Clearly,

$$\begin{aligned}\mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W(t) &= \mu(t)\mathrm{d}t + \sigma(t)\left(\mathrm{d}W^G(t) + \eta(t)\mathrm{d}t\right) \\ &= (\mu(t) + \sigma(t)\eta(t))\,\mathrm{d}t + \sigma(t)\mathrm{d}W^G(t)\end{aligned}$$

for every $t \in \mathbf{T}\setminus\{1\}$. As soon as we define $\mu^{\sigma,Q}(t) = \mu(t) + \sigma(t)\eta(t)$ for all $t \in \mathbf{T}$, Eq. (4.4) is established. □

Chapter 5
Excursion to Financial Economics: A Radically Elementary Approach to the Fundamental Theorems of Asset Pricing

What follows in this excursion is the attempt to construct a radically elementary version of continuous-time financial economics. Mathematicians sometimes confuse financial economics with mathematical finance (also known as financial mathematics) or even financial engineering. There is however, a profound difference in interest and methodology between the two: While mathematical finance and financial engineering are concerned with technical mathematical problems arising from the analysis of quantitative models of financial markets (in particular, models used at financial institutions), financial economics is a subdiscipline of economic theory and has a *conceptual* interest in understanding how financial markets work.

In this excursion, we present a radically elementary approach to the fundamental theorems of asset pricing. These theorems, while fairly elementary in the discrete-time setting, are notoriously difficult to prove in a continuous-time framework and were only established in their greatest generality in the mid-1990s by Delbaen and Schachermayer [24, 25]. Our findings, while lacking the same technical strength as the results by Delbaen-Schachermayer [24, 25], are incommensurably easier to prove and allow nevertheless for a rigorous economic justification of the martingale pricing method in very general continuous-time financial market models. The proof ideas are similar to the discussion of the first fundamental theorem in the standard textbook by Duffie [26].

First, we need to introduce some notation and terminology related to multi-dimensional stochastic processes.

An m-dimensional tuple of real-valued stochastic processes $X = (X^{(1)}, \ldots, X^{(m)})$ is called an *m-dimensional stochastic process*.

For convenience, we shall omit the dot in the scalar product: If $a, b \in \mathbf{R}^m$, we define $ab = \sum_{i=1}^{m} a_i b_i$, and more generally, if $X = \left(X^{(1)}, \ldots, X^{(m)}\right)$ and $\vartheta = \left(\vartheta^{(1)}, \ldots, \vartheta^{(m)}\right)$ are m-dimensional stochastic processes, we define ϑX to be the real-valued process $\sum_{i=1}^{m} \vartheta^{(i)} X^{(i)}$ obtained by scalar multiplication.

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_5,

The m-dimensional increment $\mathrm{d}X$ is defined as the m-tuple of real-valued stochastic processes given by applying the increment operator d componentwise:

$$\mathrm{d}X = \left(\mathrm{d}X^{(1)}, \ldots, \mathrm{d}X^{(m)}\right).$$

If $\mathcal{G}$ is a filtration, then we shall say that an m-dimensional stochastic process X is $\mathcal{G}$-*adapted* if and only if $X^{(i)}$ is $\mathcal{G}$-adapted for all $i \in \{1, \ldots, m\}$.

In the following, let $X = \left(X^{(1)}, \ldots, X^{(m)}\right)$ be a $\mathcal{G}$-adapted m-dimensional stochastic process, to be called the *process of (discounted) asset prices*.

Let ϑ be another m-dimensional stochastic process. Let us adopt the following definitions, which are ubiquitous in the asset pricing and financial economics literature (cf. e.g. Duffie [26]):

- The process $\int \vartheta \mathrm{d}X$ is called the *(discounted) gains-from-trading process* of ϑ given X. The random variable $\int_0^1 \vartheta \mathrm{d}X$ is the *(discounted) terminal gains from trading*.
- The process ϑX is called the *(discounted) value process* of ϑ given X. The *initial value* of ϑ given X is $\vartheta(0)X(0)$, and the *(discounted) terminal value* is $\vartheta(1)X(1)$.
- The process ϑ is said to be a *self-financing trading strategy with respect to X* if and only if ϑ is $\mathcal{G}$-adapted and satisfies the *intertemporal budget constraint*

  $$\mathrm{d}(\vartheta X)(t) = \vartheta(t)\mathrm{d}X(t)$$

  for all $t \in \mathbf{T}$.
- Let $\Theta(X)$ be the set of all self-financing trading strategies with respect to X. Since the increment operator d is linear, it is easy to prove that for all $\alpha, \beta \in \mathbf{R}$ and $\vartheta, \phi \in \Theta(X)$, the m-dimensional process $\alpha\vartheta + \beta\phi$ (defined componentwise) is not only $\mathcal{G}$-adapted, but also self-financing. In other words, $\Theta(X)$ is a linear space of m-dimensional processes.
- The *marketed space with respect to X*, denoted $M(X)$, is the set of random variables of the form $\vartheta(1)X(1)$ for some $\vartheta \in \Theta(X)$. Since $\Theta(X)$ is a linear space, it is immediate that $M(X)$ is a linear space of random variables.

For any real-valued random variable x, we shall write $x \succ 0$ if and only if $x \gtrsim 0$ a.s. and there exists some event A such that both $P(A) \gg 0$ and $x(\omega) \gg 0$ for all $\omega \in A$. With this definition, we can now propose a radically elementary definition of approximate arbitrage or free lunch with vanishing risk: A process $\vartheta \in \Theta(X)$ is called a *free lunch with vanishing risk* (FLVR) or *near arbitrage* if and only if either

$$\vartheta(1)X(1) \succ 0 \geq \vartheta(0)X(0)$$

or

$$\vartheta(1)X(1) \gtrsim 0 \gg \vartheta(0)X(0) \qquad \text{a.s.}$$

Lemma 5.1. *If there exists no FLVR, then the map* $\psi : M(X) \to \mathbf{R}$ *defined by*

$$\psi\left(\vartheta(1)X(1)\right) = \vartheta(0)X(0)$$

is well-defined and linear.

Proof. That ψ is well-defined can be seen by contraposition: If ψ were not well-defined, then there would exist $\vartheta, \phi \in \Theta(X)$ such that $\vartheta(1)X(1) = \phi(1)X(1)$ while $\vartheta(0)X(0) \neq \phi(0)X(0)$. Let us say $\vartheta(0)X(0) < \phi(0)X(0)$. Then

$$(\vartheta - \phi)(0)X(0) < 0 = (\vartheta - \phi)(1)X(1).$$

Hence, there exists some real $\alpha > 0$ such that

$$\alpha(\vartheta - \phi)(0)X(0) \ll 0 = \alpha(\vartheta - \phi)(1)X(1).$$

On the other hand, $\alpha(\vartheta - \phi) \in \Theta(X)$ as $\Theta(X)$ is a linear space. Therefore, $\alpha(\vartheta - \phi)$ is a FLVR.

Since $\Theta(X)$ and $M(X)$ are linear spaces, one can easily check that ψ is linear.

□

Lemma 5.2. *Suppose there is no FLVR and assume that there exists some* $k \in \{1, \ldots, m\}$ *such that* $X^{(k)}(t) = 1$ *for all* $t \in \mathbf{T}$. *Then,* ψ *is strictly increasing in the sense that for all* $x, y \in M(X)$,

- *if* $x \geq y$, *then* $\psi(x) \geq \psi(y)$, *and*
- *if* $x \succ y$, *then* $\psi(x) \gg \psi(y)$.

Since ψ is linear, it would be enough to prove this for $y = 0$, but the proof is short anyway.

Proof by contraposition. Consider $x, y \in M(X)$, and let ϑ, ϕ be such that $x = \vartheta(1)X(1)$ and $y = \phi(1)X(1)$, hence $\psi(x) = \vartheta(0)X(0)$ and $\psi(y) = \phi(0)X(0)$.

- First suppose $x \geq y$ while $\psi(x) < \psi(y)$. By the linearity of ψ, it follows that

$$\underbrace{(\vartheta - \phi)(1)X(1)}_{=x-y} \geq 0 > \underbrace{(\vartheta - \phi)(0)X(0)}_{=\psi(x)-\psi(y)=\psi(x-y)}.$$

  Then $\vartheta - \phi$ is a FLVR.
- Now suppose $x \succ y$ while $\psi(x) \lesssim \psi(y)$. Again by the linearity of ψ, it follows that

$$\underbrace{(\vartheta - \phi)(1)X(1)}_{=x-y} \succ 0 \gtrsim \underbrace{(\vartheta - \phi)(0)X(0)}_{=\psi(x)-\psi(y)=\psi(x-y)}.$$

  Let $\varepsilon = -(\vartheta - \phi)(0)X(0)$, let e_k be the m-dimensional process such that $e_k^{(i)} = 1$ (deterministic constant) if $i = k$ and $e_k^{(i)} = 0$ otherwise. It then follows that

$$(\vartheta - \phi + \varepsilon e_k)(1)X(1) = (\vartheta - \phi)(1)X(1) + \varepsilon \succ 0 = (\vartheta - \phi + \varepsilon e_k)(0)X(0),$$

whence $\vartheta - \phi + \varepsilon e_k$ is a FLVR.

□

A probability measure Q is said to be

- *near-equivalent* to P, denoted $Q \approx P$, if and only if for all events A, one has $Q(A) \gg 0$ if and only if $P(A) \gg 0$.
- a *(near-)equivalent martingale measure* for X under P (abbreviated: *near-EMM*) if and only if $Q \approx P$ and X is a $(\mathcal{G}, Q)$-martingale.

Lemma 5.3. *Suppose x is $L^1(Q)$ for some probability measure Q with $Q \approx P$.*

- *If $x \gtrsim 0$ a.s., then $E^Q[x] \gtrsim 0$.*
- *If even $x \succ 0$, then $E^Q[x] \gg 0$.*

Proof. If $x \gtrsim 0$ a.s., there exists for all standard $n \in \mathbf{N}$ some event $N^{1/n}$ such that both $P\left(N^{1/n}\right) \leq \frac{1}{n}$ and $x(\omega) \gtrsim 0$, in particular $x(\omega) \geq -\frac{1}{n}$, for all $\omega \in N^{1/n}$. Therefore, by the underspill/overspill principle in $\mathbf{N}$ (see Remark 1.1) there exists some nonstandard $n \in \mathbf{N}$ such that both $P\left(N^{1/n}\right) \leq \frac{1}{n}$ and $x(\omega) \geq -\frac{1}{n}$ for all $\omega \notin N^{1/n}$. Since $Q \approx P$, also $Q\left(N^{1/n}\right) \simeq 0$. Hence, we have Q-a.s. $x \simeq x\chi_{\Omega \setminus N^{1/n}}$, and since x is $L^1(Q)$, the truncated random variable $x\chi_{\Omega \setminus N^{1/n}}$ is $L^1(Q)$, too. Therefore, Nelson's Lebesgue theorem (Remark 2.3) may be applied, which yields

$$\begin{aligned} E^Q[x] \simeq E^Q\left[x\chi_{\Omega \setminus N^{1/n}}\right] &\geq -\frac{1}{n} Q\left(\Omega \setminus N^{1/n}\right) \\ &\geq -\frac{1}{n} \simeq 0 \end{aligned}$$

and thus establishes that $E^Q[x] \gtrsim 0$.

If even $x \succ 0$, then there exists some event A such that both $P(A) \gg 0$ and $x(\omega) \gg 0$ on $\omega \in A$. Since Ω is finite, so is A, whence $\min_{\omega \in A} x(\omega)$ exists and is $\gg 0$. Hence, we can choose a limited $c \gg 0$ such that $x(\omega) \geq c$ for all $\omega \in A$. Since $Q \approx P$, also $Q(A) \gg 0$. In this situation, we get the following chain of equations and estimates (exploiting that $x(\omega) \geq c$ for $\omega \in A$ and $x(\omega) \geq -\frac{1}{n}$ for $\omega \notin N^{1/n}$):

$$\begin{aligned} E^Q[x] &\simeq E^Q\left[x\chi_{\Omega \setminus N^{1/n}}\right] \\ &= E^Q\left[x\chi_{A \setminus N^{1/n}}\right] + E^Q\left[x\chi_{\Omega \setminus \left(N^{1/n} \cup A\right)}\right] \\ &\geq c\, Q\left(A \setminus N^{1/n}\right) - \frac{1}{n} Q\left(\Omega \setminus \left(N^{1/n} \cup A\right)\right) \end{aligned}$$

$$\geq c \left(Q(A) - \underbrace{Q\left(N^{1/n}\right)}_{\simeq 0} \right) - \frac{1}{n}$$

$$\simeq \underbrace{c}_{\gg 0} \underbrace{Q(A)}_{\gg 0} - \underbrace{\frac{1}{n}}_{\simeq 0} \gg 0.$$

□

An m-dimensional process ϑ is said to be *limited* if and only if it has limited norm, i.e. if the real-valued process $|\vartheta| = \left(\sum_{i=1}^{2} \left|\vartheta^{(i)}\right|^2\right)^{1/2}$ is limited.

The following theorem asserts that no trading strategy which satisfies a natural integrability condition can be a FLVR; thus, up to technicalities, the existence of a near-EMM is sufficient for the absence of FLVR. This result provides a rigorous economic justification of the martingale pricing method.

Theorem 5.4 (First fundamental theorem of asset pricing, sufficiency part). *Suppose there exists a near-EMM Q for X. Then there is no FLVR among those trading strategies which have Q-integrable terminal gains from trading and limited initial value (all with respect to X).*

Corollary 5.5. *If there exists a near-EMM Q for X and $X(1)$ is $L^1(Q)$, then no limited trading strategy can be a FLVR with respect to X.*

Proof of Theorem 5.4. Let ϑ be a self-financing trading strategy such that $\int_0^1 \vartheta \mathrm{d}X$ is $L^1(Q)$ and $\vartheta(0)X(0)$ is limited. We need to prove that ϑ cannot be a FLVR. Now,

$$\vartheta(1)X(1) = \int_0^1 \vartheta \mathrm{d}X + \vartheta(0)X(0)$$

as ϑ is self-financing, while $\int \vartheta \mathrm{d}X$ is a $(\mathcal{G}, Q)$-martingale (by Theorem 3.1, as $\bar{X}$ is a $(\mathcal{G}, Q)$-martingale and ϑ is $\mathcal{G}$-adapted) with initial value 0. Thus,

$$E^Q\left[\vartheta(1)X(1)\right] = \vartheta(0)X(0).$$

Hence, whenever $\vartheta(1)X(1) \succ 0$, we not only have $E^Q\left[\vartheta(1)X(1)\right] \gg 0$ (by Lemma 5.3), but even $\vartheta(0)X(0) \gg 0$, and whenever $\vartheta(1)X(1) \gtrsim 0$, we have $E^Q\left[\vartheta(1)X(1)\right] \gtrsim 0$ (by Lemma 5.3) and thus $\vartheta(0)X(0) \gtrsim 0$. Therefore, ϑ cannot be a FLVR. □

We have seen that the existence of a near-EMM implies the absence of "well-behaved" FLVR. Almost the converse is also true: The absence of FLVR implies the existence of a near-EMM, hence the existence of a near-EMM is necessary for the absence of FLVR.

Theorem 5.6 (First fundamental theorem of asset pricing, necessity part). *Suppose $X^{(k)} = 1$ for some $k \in \{1, \ldots, m\}$, and assume there is no FLVR with respect to X. Then there exists a near-EMM Q.*

The proof needs the following Lemma:

Lemma 5.7. *Suppose there is no FLVR with respect to X. Then, ψ is near-continuous in the sense that for all $x, y \in M(X)$, if $x \simeq y$, then also $\psi(x) \simeq \psi(y)$.*

Proof of Lemma 5.7. Since ψ is linear, it is enough to prove this for $y = 0$. Let $x \in M(X)$ with $x \simeq 0$ and choose $\vartheta \in M(X)$ such that $x = \vartheta(1)X(1) \simeq 0$. Suppose, for a contradiction, that $\vartheta(0)X(0) \not\simeq 0$. Then either $\vartheta(0)X(0) \ll 0$, in which case ϑ is a FLVR, or $\vartheta(0)X(0) \gg 0$, in which case $-\vartheta$ is a FLVR. Thus, in any case, there exists a FLVR, contradiction. □

Proof of Theorem 5.6. We have already remarked that the marketed space $M(X)$ is a linear subspace of $\mathbf{R}^\Omega$, and that $\psi : M(X) \to \mathbf{R}$ is a well-defined, linear, and strictly increasing map (Lemmas 5.1; 5.2). Let B be a basis of $M(X)$, and choose a basis C of the orthogonal complement $M(X)^\perp$ of $M(X)$. Without loss of generality, we can ensure that $\max_{\substack{\omega\in\Omega \\ P\{\omega\}>0}} |x(\omega)| = 1$ for all $x \in C$. Now define a map $\Psi : \mathbf{R}^\Omega \to \mathbf{R}$ such that $\Psi \restriction M(X) = \psi$ and such that $\Psi \restriction M(X)^\perp = E[\cdot]$, that is

$$\forall x \in M(X)^\perp \qquad \Psi(x) = E[x].$$

Note that Ψ is then a linear and increasing functional. By the Riesz representation theorem applied to the finite-dimensional linear space $\mathbf{R}^\Omega$, there must exist some element of $\mathbf{R}^\Omega$, henceforth denoted $\xi(1)$, such that

$$\forall x \in \mathbf{R}^\Omega \qquad \Psi(x) = \sum_{\omega\in\Omega} x(\omega)\xi(1)(\omega) = E\left[x\xi(1)\right].$$

Since ψ is strictly increasing, one has $\Psi(x) \geq 0$ for all $x \geq 0$. This entails that $\xi(1) \geq 0$. Define $Q : A \mapsto \int_A \xi(1)\mathrm{d}P$. It is clear that Q is a measure. Moreover, since $X^{(k)} = 1$, there is a self-financing trading strategy, namely e_k (the m-dimensional process whose k-th coordinate is constantly 1 and whose other coordinates are zero), such that both $e_k(1)X(1) = 1$ and $e_k(0)X(0) = 1$, therefore $\psi(1) = 1$, and thus $\Psi(1) = 1$, which means that $E\left[\xi(1)\right] = 1$. Therefore, Q is even a probability measure.

Let us next show that Q is near-equivalent to P. Let A be an event. Then there exist events A', A'' such that $\chi_{A'} \in M(X)$, $\chi_{A''} \in M(X)^\perp$ and $\chi_A = \chi_{A'} + \chi''_A$ whilst by definition $E\left[\chi_{A'}\chi_{A''}\right] = 0$, that is $A = A' \cup A''$ and $P\left(A' \cap A''\right) = 0$. Now suppose $P(A) \gg 0$, then either $P\left(A'\right) \gg 0$, in which case

$$Q(A) \geq Q(A') = \Psi\left(\chi_{A'}\right) = \psi\left(\chi_{A'}\right) \gg 0$$

since ψ is strictly increasing (Lemma 5.2) or $P\,(A'') \gg 0$, in which case

$$Q(A) \geq Q(A'') = \Psi\,(\chi_{A''}) = E\,[\chi_{A''}] = P\,\left(A''\right) \gg 0$$

by definition of Ψ on $M(X)^{\perp}$. Conversely, if $Q(A) \gg 0$, then either $Q\,(A') \gg 0$, in which case $\psi(\chi'_A) \gg 0$ (by contraposition from Lemma 5.7) or $Q\,(A'') \gg 0$, in which case

$$0 \ll E^{Q}\,[\chi_{A''}] = \Psi\,(\chi_{A''}) = E\,[\chi_{A''}] = P\,\left(A''\right).$$

This proves that $Q \approx P$.

Finally, let us show that $E^{Q}\left[X^{(i)}(\tau)\right] = X^{(i)}(0)$ for all $\mathbf{T} \cup \{+\infty\}$-valued stopping times τ. This implies then that X is a $(\mathcal{G}, Q)$-martingale. By the construction of Q from Ψ and thus ultimately from ψ, it is enough to prove that $\psi\left(X^{(i)}(\tau)\right) = X^{(i)}(0)$. Thus we only have to find a self-financing trading strategy ϑ such that $\vartheta(1)X(1) = X^{(i)}(\tau)$ and $\vartheta(0)X(0) = X^{(i)}(0)$. However, one can easily convince oneself that the strategy ϑ defined by $\vartheta(t) = \chi_{[0,\tau)}(t)e_i + \chi_{[\tau,1]}(t)\left(X^{(i)}(\tau)e_k\right)$ (i.e. hold asset i up to time τ, sell it, and invest the proceeds into asset k) meets this requirement. □

Remark 5.8 (Second fundamental theorem of asset pricing). An inspection of the proof of Theorem 5.6 shows that, given the absence of FLVR with respect to X, the choice of Ψ and thus of Q is unique if and only if X is a *complete market model* in the sense that $M(X) = \mathbf{R}^{\Omega}$.

Having developed a radically elementary version of the fundamental theory of asset pricing in continuous time, we should now at least examine whether it is applicable to the radically elementary version of the *(Samuelson-) Black–Scholes model* as well. The point of this is not to analyse the Black–Scholes model, but rather to check the adequacy of the radically elementary economic theory of continuous-time financial markets outlined in the preceding paragraphs. The following deliberations simply serve as a demonstration that the theory devised in this chapter is not vacuous, but is at least applicable to the most important model of a continuous-time financial market.

First, let us describe very briefly the (Samuelson-) Black–Scholes model in the language of radically elementary probability theory. The (radically elementary version of the) Black–Scholes [18] model is defined on a probability space (Ω, P) carrying a Wiener walk W that in turn generates a filtration $\mathcal{F}$. It models a financial market with a risky asset and a risk-free bond, and assumes that the *discounted* price process of the risky asset $X^{(2)}$ follows, under P, a geometric Itô process of the form

$$\forall t \in \mathbf{T} \setminus \{1\} \qquad \begin{aligned} \mathrm{d}X^{(2)}(t) &= \mu\, X^{(2)}(t)\mathrm{d}t + \sigma\, X^{(2)}(t)\mathrm{d}W(t) \\ &= X^{(2)}(t)\,(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W(t)) \end{aligned} \tag{5.1}$$

for some limited $\mu, \sigma \in \mathbf{R}$ with $\sigma \gg 0$ and limited $X^{(2)}(0) \gg 0$, whereas the discounted price process $X^{(1)}$ of the risk-free bond is, of course, just constantly $= 1$.

Lemma 5.9. *For the Black–Scholes model* $X = \left(X^{(1)}, X^{(2)}\right)$ *given by* $X^{(1)} = 1$ *and Eq.* (5.1)*, there exists a unique near-EMM* Q*, given by*

$$\xi(0) = 1, \qquad \forall t \in \mathbf{T} \setminus \{1\} \qquad \mathrm{d}\xi(t) = -\frac{\mu}{\sigma}\xi(t)\mathrm{d}W(t). \tag{5.2}$$

Hence, the Black–Scholes model admits no well-behaved FLVR (by Theorem 5.4) and it is a complete market model in the sense that $M(X) = \mathbf{R}^{\Omega}$ (by Remark 5.8).

Proof. For every probability measure Q, the density process ξ is a $(P, \mathcal{F})$-martingale (by the tower property of conditional expectations) with $\xi(0) = 1$ and hence, by the martingale representation theorem (Theorem 3.2), there exists an adapted process ϕ such that $\xi = 1 + \int \phi \mathrm{d}W$. Now, $X^{(2)}$ is a $(Q, \mathcal{F})$-martingale if and only if $E^Q\left[\mathrm{d}X^{(2)}(t)\middle|\,\mathcal{F}_t\right] = 0$ for all $t \in \mathbf{T} \setminus \{1\}$, and Bayes' formula (4.1) allows us to simplify the latter equation as follows:

$$\begin{aligned}
0 &= E^Q\left[\mathrm{d}X^{(2)}(t)\middle|\,\mathcal{F}_t\right] \\
&= \frac{1}{\xi(t)}E\left[\underbrace{\xi(t+\mathrm{d}t)}_{=\xi(t)+\mathrm{d}\xi(t)}\mathrm{d}X^{(2)}(t)\middle|\,\mathcal{F}_t\right] \\
&= \underbrace{E\left[\mathrm{d}X^{(2)}(t)\middle|\,\mathcal{F}_t\right]}_{=X^{(2)}(t)E[\mu\,\mathrm{d}t+\sigma\,\mathrm{d}W(t)|\mathcal{F}_t]=X^{(2)}(t)\,\mu\,\mathrm{d}t} \\
&\quad + \frac{1}{\xi(t)}E\left[\underbrace{\mathrm{d}\xi(t)}_{=\phi(t)\mathrm{d}W(t)}\underbrace{\mathrm{d}X^{(2)}(t)}_{=X^{(2)}(t)(\mu\,\mathrm{d}t+\sigma\,\mathrm{d}W(t))}\middle|\,\mathcal{F}_t\right] \\
&= X^{(2)}(t)\mu\,\mathrm{d}t \\
&\quad + \frac{1}{\xi(t)}E\left[X^{(2)}(t)\phi(t)\left(\mu\,\mathrm{d}W(t)\,\mathrm{d}t + \sigma\underbrace{(\mathrm{d}W(t))^2}_{=\mathrm{d}t}\right)\middle|\,\mathcal{F}_t\right] \\
&= \frac{X^{(2)}(t)}{\xi(t)}\left(\mu\,\xi(t)\,\mathrm{d}t + \phi(t)\left(\mu\,\mathrm{d}t\underbrace{E\left[\mathrm{d}W(t)|\,\mathcal{F}_t\right]}_{=0} + \sigma\,\mathrm{d}t\right)\right) \\
&= \frac{X^{(2)}(t)}{\xi(t)}\left(\mu\,\xi(t) + \phi(t)\sigma\right)\,\mathrm{d}t.
\end{aligned}$$

Hence, $X^{(2)}$ is a $(Q, \mathcal{F})$-martingale if and only if $\mu\,\xi(t) + \phi(t)\sigma = 0$ for all $t \in \mathbf{T} \setminus \{1\}$, hence $\phi(t) = -\frac{\mu}{\sigma}\xi(t)$ for all $t \in \mathbf{T} \setminus \{1\}$. In other words, $X^{(2)}$ is a $(Q, \mathcal{F})$-martingale if and only if the density process ξ of Q is the unique solution of the stochastic differential equation

$$\forall t \in \mathbf{T} \setminus \{1\} \qquad \mathrm{d}\xi(t) = -\frac{\mu}{\sigma}\xi(t)\mathrm{d}W(t)$$

with initial condition $\xi(0) = 1$. Hence, if there exists a near-EMM for X, it must be unique.

What remains to be shown is that the probability measure Q with the density process ξ given by Eq. (5.2) is actually a near-EMM. We have already seen that $X^{(2)}$ and hence $X = \left(1, X^{(2)}\right)$ is a $(Q, \mathcal{F})$-martingale by the choice of ξ and hence of Q; we still need to show that $Q \approx P$.

Since $\sigma \gg 0$ and μ is limited, the fraction $-\frac{\mu}{\sigma}$ is limited, hence Lemma 3.10 yields that $\xi(1)$ is $L^1(P)$, whence Nelson's Radon–Nikodym theorem (see Remark 2.2) shows that for every $A \subseteq \Omega$ with $P(A) \simeq 0$, we have $Q(A) \simeq 0$. For the converse implication, suppose that there were some $A \subseteq \Omega$ with $Q(A) \simeq 0$ but $P(A) \gg 0$. Then, we must have P-a.s. $\xi(1)\chi_A \simeq 0$, for otherwise there would be some $\lambda \gg 0$ with $P\left\{\xi(1)\chi_A \geq \lambda\right\} \gg 0$ (see Remark 2.4) and hence $Q(A) = E\left[\xi(1)\chi_A\right] \geq \lambda P\left\{\xi(1)\chi_A \geq \lambda\right\} \gg 0$. On the other hand, Lemma 3.12 can be applied (as $\frac{\mu}{\sigma}$ is limited and $\xi(0) \gg 0$) to see that P-a.s. $\xi(1) \gg 0$. Combining this with the previously observed fact that P-a.s. $\xi(1)\chi_A \simeq 0$, we see that P-a.s. $\chi_A \simeq 0$. Since χ_A is a characteristic function and only assumes values from $\{0, 1\}$, we conclude that P-a.s. $\chi_A = 0$. This implies that $P\left(\Omega \setminus A\right) = P\left\{\chi_A = 0\right\} \simeq 1$, hence $P(A) \simeq 0$ (see footnote 2 on page 8). □

Chapter 6
Excursion to Financial Engineering: Volatility Invariance in the Black–Scholes Model

It is well-known that the diffusion invariance principle conveys an important insight for financial economics: when (logarithmic) Itô processes are used as models of stock prices, the drift coefficient of the logarithmic price process is interpreted as a measure for the expected return, and its diffusion coefficient is interpreted as a measure for the volatility. In this context, the diffusion invariance principle asserts roughly that under an equivalent change of probability measure, only the expected returns will be affected, but not the volatilities. In particular, a price process will have the same volatility under the real-world probability measure as under an equivalent risk-neutral (i.e. arbitrage-free) probability measure; changing the probability measure corresponds to changing the expected return (and vice versa).

If the logarithmic price process is just a multiple of the Wiener walk plus linear drift (Black–Scholes [18] model; compare Eq. (5.1) with Lemma 3.12), then a result with the same financial interpretation can be proved by purely elementary estimates, as a consequence of the following theorem.

Theorem 6.1. *Let $\sigma \gg 0$ be limited. Then, for every limited $\mu \in \mathbf{R}$, there is a unique $q \in (0,1)$ and a unique probability measure Q on Ω such that the increments of W are independent under Q with $Q\left\{\mathrm{d}W(t) = \sqrt{\mathrm{d}t}\right\} = q$ for all $t \in \mathbf{T} \setminus \{1\}$ and*

$$E^Q\left[\mathrm{e}^{\sigma \mathrm{d}W(t)}\right] = \mathrm{e}^{\mu\,\mathrm{d}t}, \qquad \frac{1}{\mathrm{d}t} Var^Q\left[\mathrm{e}^{\sigma \mathrm{d}W(t)}\right] \simeq \sigma^2 \simeq \frac{1}{\mathrm{d}t} Var\left[\mathrm{e}^{\sigma \mathrm{d}W(t)}\right]$$

for all $t \in \mathbf{T} \setminus \{1\}$. Q is equivalent to P, the density being given by

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = (4q(1-q))^{\frac{1}{2\mathrm{d}t}} \left(\frac{q}{1-q}\right)^{\frac{1}{2\sqrt{\mathrm{d}t}} W(1)} \tag{6.1}$$

$$\simeq \exp\left(\eta W(1) - \frac{\eta^2}{2}\right), \tag{6.2}$$

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_6,

wherein

$$\eta = \frac{1}{2\sqrt{\mathrm{d}t}}\left(\ln(q) - \ln(1-q)\right) \tag{6.3}$$
$$= \left(\frac{\mu}{\sigma} - \frac{\sigma}{2}\right)\sqrt{\mathrm{d}t} + \frac{\sigma^2}{12}\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right),$$

$$q = \frac{\mathrm{e}^{\mu\,\mathrm{d}t} - \mathrm{e}^{-\sigma\sqrt{\mathrm{d}t}}}{\mathrm{e}^{\sigma\sqrt{\mathrm{d}t}} - \mathrm{e}^{-\sigma\sqrt{\mathrm{d}t}}} \tag{6.4}$$
$$= \frac{1}{2} + \left(\frac{\mu}{2\sigma} - \frac{\sigma}{4}\right)\sqrt{\mathrm{d}t} + \frac{\sigma^2}{12}\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right).$$

Proof. Let us first consider an arbitrary $q \in (0,1)$. Let Q be a probability measure on Ω for which the increments of W are independent and distributed according to $Q\left\{\mathrm{d}W(t) = \sqrt{\mathrm{d}t}\right\} = q$ for all $t \in \mathbf{T} \setminus \{1\}$. (Such a Q can be constructed by the same procedure as in Remark 2.15.) First, we shall establish the density formula for Q in Eq. (6.1) (which entails the uniqueness of Q given q). Let $\lambda : \mathbf{T} \to \mathbf{R}$ be a trajectory of W. Then, $P\{W = \lambda\} = 2^{-1/\mathrm{d}t}$ and

$$Q\{W = \lambda\} = q^{\mathrm{card}\left\{t\in\mathbf{T}\setminus\{1\}\,:\,\mathrm{d}\lambda(t)=\sqrt{\mathrm{d}t}\right\}}(1-q)^{\mathrm{card}\left\{t\in\mathbf{T}\setminus\{1\}\,:\,\mathrm{d}\lambda(t)=-\sqrt{\mathrm{d}t}\right\}}.$$

However,

$$\mathrm{card}\left\{t \in \mathbf{T}\setminus\{1\} : \mathrm{d}\lambda(t) = \sqrt{\mathrm{d}t}\right\}$$
$$+\mathrm{card}\left\{t \in \mathbf{T}\setminus\{1\} : \mathrm{d}\lambda(t) = -\sqrt{\mathrm{d}t}\right\} = \frac{1}{\mathrm{d}t}$$

and

$$\sqrt{\mathrm{d}t}\,\mathrm{card}\left\{t \in \mathbf{T}\setminus\{1\} : \mathrm{d}\lambda(t) = \sqrt{\mathrm{d}t}\right\}$$
$$-\sqrt{\mathrm{d}t}\,\mathrm{card}\left\{t \in \mathbf{T}\setminus\{1\} : \mathrm{d}\lambda(t) = -\sqrt{\mathrm{d}t}\right\} = \lambda(1),$$

whence

$$\mathrm{card}\left\{t \in \mathbf{T}\setminus\{1\} : \mathrm{d}\lambda(t) = \sqrt{\mathrm{d}t}\right\} = \frac{1}{2\mathrm{d}t} + \frac{\lambda(1)}{2\sqrt{\mathrm{d}t}}$$

and

$$\mathrm{card}\left\{t \in \mathbf{T}\setminus\{1\} : \mathrm{d}\lambda(t) = -\sqrt{\mathrm{d}t}\right\} = \frac{1}{2\mathrm{d}t} - \frac{\lambda(1)}{2\sqrt{\mathrm{d}t}}.$$

It follows that

$$
\begin{aligned}
\frac{Q\{W=\lambda\}}{P\{W=\lambda\}} &= 2^{\frac{1}{\mathrm{d}t}} q^{\frac{1}{2\mathrm{d}t}+\frac{\lambda(1)}{\sqrt{\mathrm{d}t}}} (1-q)^{\frac{1}{2\mathrm{d}t}-\frac{\lambda(1)}{2\sqrt{\mathrm{d}t}}} \\
&= (4q(1-q))^{\frac{1}{2\mathrm{d}t}} q^{\frac{\lambda(1)}{\sqrt{\mathrm{d}t}}} (1-q)^{-\frac{\lambda(1)}{2\sqrt{\mathrm{d}t}}} \\
&= (4q(1-q))^{\frac{1}{2\mathrm{d}t}} \left(\frac{q}{1-q}\right)^{\frac{\lambda(1)}{2\sqrt{\mathrm{d}t}}},
\end{aligned}
$$

hence

$$
\frac{\mathrm{d}Q}{\mathrm{d}P} = (4q(1-q))^{\frac{1}{2\mathrm{d}t}} \left(\frac{q}{1-q}\right)^{\frac{W(1)}{2\sqrt{\mathrm{d}t}}}.
$$

In order to obtain the infinitesimal approximation in Eq. (6.2), we first rewrite the exact formula for $\frac{\mathrm{d}Q}{\mathrm{d}P}$ as

$$
\begin{aligned}
\frac{\mathrm{d}Q}{\mathrm{d}P} &= \exp\begin{pmatrix} \frac{1}{2\sqrt{\mathrm{d}t}} (\ln(q) - \ln(1-q))\, W(1) \\ +\frac{1}{2\mathrm{d}t} (\ln(q) + \ln(1-q) + 2\ln 2) \end{pmatrix} \\
&= \exp\left(\eta W(1) + \frac{1}{2}\left(\frac{1}{\mathrm{d}t}(\ln(q) + \ln(1-q) + 2\ln 2)\right)\right)
\end{aligned}
$$

and we note that for $q = \frac{1}{2} + \gamma\sqrt{\mathrm{d}t} + \delta\,\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right)$ (for any limited γ, δ), a third-order Taylor expansion of the logarithm function around $\frac{1}{2}$ yields

$$
\begin{aligned}
\ln(q) &= -\ln 2 + 2\gamma\sqrt{\mathrm{d}t} + \delta\,\mathrm{d}t - 2\gamma^2\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right), \\
\ln(1-q) &= -\ln 2 - 2\gamma\sqrt{\mathrm{d}t} - \delta\,\mathrm{d}t - 2\gamma^2\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right),
\end{aligned}
$$

so

$$
\begin{aligned}
\eta = \frac{\ln(q) - \ln(1-q)}{2\sqrt{\mathrm{d}t}} &= 2\gamma + \delta\sqrt{\mathrm{d}t} + \mathcal{O}(\mathrm{d}t), \\
\frac{\ln(q) + \ln(1-q) + 2\ln 2}{\mathrm{d}t} &= 4\gamma^2 + \mathcal{O}\left((\mathrm{d}t)^{1/2}\right),
\end{aligned}
$$

and therefore

$$
\eta^2 - \frac{\ln(q) + \ln(1-q) + 2\ln 2}{\mathrm{d}t} = \mathcal{O}\left((\mathrm{d}t)^{1/2}\right).
$$

Put $u = \mathrm{e}^{\sigma\sqrt{\mathrm{d}t}}$ and $d = \mathrm{e}^{-\sigma\sqrt{\mathrm{d}t}}$ and let $t \in \mathbf{T}$. Since the distribution of $\mathrm{d}W(t)$ under Q does not depend on t, clearly

$$
E^Q\left[\mathrm{e}^{\sigma\mathrm{d}W(t)}\right] = qu + (1-q)d = q(u-d) + d
$$

and

$$Var^Q\left[e^{\sigma dW(t)}\right] = q(1-q)(u-d)^2$$

(the formula for the variance uses $qu^2 + (1-q)\,d^2 - (qu+(1-q)d)^2 = q\,(1-q)\left(u^2 - 2ud + d^2\right)$), in particular (for $q = 1/2$)

$$Var\left[e^{\sigma dW(t)}\right] = \frac{(u-d)^2}{4},$$

which (through a third-order Taylor expansion of the exponential function around 0) becomes

$$\begin{aligned}
&Var\left[e^{\sigma dW(t)}\right] \\
&= \frac{1}{4}\left(1 + \sigma\sqrt{dt} + \frac{\sigma^2 dt}{2} - 1 + \sigma\sqrt{dt} - \frac{\sigma^2 dt}{2} + \mathcal{O}\left((dt)^{3/2}\right)\right)^2 \\
&= \frac{1}{4}\left(2\sigma\sqrt{dt} + \mathcal{O}\left((dt)^{3/2}\right)\right)^2 = \sigma^2 dt + \mathcal{O}\left((dt)^{3/2}\right)
\end{aligned}$$

Let us now fix $q = \frac{e^{\mu\,dt}-d}{u-d}$ as in the Theorem. By the above formula for $E^Q\left[e^{\sigma dW(t)}\right]$, this is the unique q that leads to $E^Q\left[e^{\sigma dW(t)}\right] = e^{\mu\,dt}$. Moreover, for such q, we have

$$\begin{aligned}
Var^Q\left[e^{\sigma dW(t)}\right] &= q(1-q)(u-d)^2 = \frac{e^{\mu\,dt}-d}{u-d}\,\frac{u-e^{\mu\,dt}}{u-d}(u-d)^2 \\
&= \left(e^{\mu\,dt}-d\right)\left(u-e^{\mu\,dt}\right) = e^{\mu\,dt}\,(u+d) - \underbrace{ud}_{=1} - e^{2\mu\,dt} \\
&= e^{\mu\,dt}\left(e^{\sigma\sqrt{dt}} + e^{-\sigma\sqrt{dt}}\right) - 1 - e^{2\mu\,dt}
\end{aligned}$$

A third-order Taylor expansion of the exponential function around 0 yields

$$\begin{aligned}
&Var^Q\left[e^{\sigma dW(t)}\right] \\
&= \left(1 + \mu\,dt + \mathcal{O}\left((dt)^2\right)\right) \\
&\quad \times\left(1 + \sigma\sqrt{dt} + \frac{\sigma^2 dt}{2} + 1 - \sigma\sqrt{dt} + \frac{\sigma^2 dt}{2} + \mathcal{O}\left((dt)^{3/2}\right)\right) \\
&\quad -1 - \left(1 + 2\mu\,dt + \mathcal{O}\left((dt)^2\right)\right) \\
&= \left(1 + \mu\,dt + \mathcal{O}\left((dt)^2\right)\right)\left(2 + \sigma^2 dt + \mathcal{O}\left((dt)^{3/2}\right)\right) \\
&\quad -2 - 2\mu\,dt + \mathcal{O}\left((dt)^2\right)
\end{aligned}$$

$$
\begin{aligned}
&= 2 + 2\mu\,\mathrm{d}t + \sigma^2\mathrm{d}t - 2 - 2\mu\,\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right) \\
&= \sigma^2\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right)
\end{aligned}
$$

and a fourth-order Taylor expansion of the exponential function around 0 yields

$$
\begin{aligned}
q &= \frac{1 + \mu\,\mathrm{d}t - 1 + \sigma\sqrt{\mathrm{d}t} - \frac{\sigma^2}{2}\mathrm{d}t + \frac{\sigma^3}{6}(\mathrm{d}t)^{3/2} + \mathcal{O}\left((\mathrm{d}t)^2\right)}{1 + \sigma\sqrt{\mathrm{d}t} + \frac{\sigma^2\mathrm{d}t}{2} - 1 + \sigma\sqrt{\mathrm{d}t} - \frac{\sigma^2\mathrm{d}t}{2} + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right)} \\
&= \frac{\sigma\sqrt{\mathrm{d}t} + (\mu - \sigma^2/2)\mathrm{d}t + \frac{\sigma^3}{6}(\mathrm{d}t)^{3/2} + \mathcal{O}\left((\mathrm{d}t)^2\right)}{2\sigma\sqrt{\mathrm{d}t} + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right)} \\
&= \frac{1}{2} + \left(\frac{\mu}{2\sigma} - \frac{\sigma}{4}\right)\sqrt{\mathrm{d}t} + \frac{\sigma^2}{12}\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right).
\end{aligned}
$$

□

Chapter 7
A Radically Elementary Theory of Itô Diffusions and Associated Partial Differential Equations

7.1 Itô Diffusions

Definition 7.1 ((Time-homogeneous) Itô diffusions). Let $t_0 \in \mathbf{T} \setminus \{1\}$ and $a, b : \mathbf{R} \times [0,1] \to \mathbf{R}$. A family $\xi = (\xi(t))_{t \in \mathbf{T} \cap [t_0,1]}$ is called an *Itô diffusion with drift coefficient function a and diffusion coefficient function b starting at time t_0* if and only if

$$\mathrm{d}\xi(t) = a(\xi(t), t)\,\mathrm{d}t + b(\xi(t), t)\,\mathrm{d}W(t) \tag{7.1}$$

for all $t \in \mathbf{T} \cap [t_0, 1)$. If a, b are constant in the second argument, then ξ is called a *time-homogeneous Itô diffusion*.

Equation (7.1) is called the *diffusion equation* of ξ.

Definition 7.2. For $x \in \mathbf{R}$, the unique Itô diffusion ξ with drift coefficient function a and diffusion coefficient function b starting at time t_0 with initial value $\xi(t_0) = x$ is denoted by $\xi^{t_0,x}$. If $\mathcal{G}$ is any subalgebra of the finite set 2^Ω and $f : \mathbf{R}^m \to \mathbf{R}$, we shall use the abbreviation

$$E^{t_0,x}\left[f(\xi(t_1), \ldots, \xi(t_m)) \middle| \mathcal{G}\right] := E\left[f\left(\xi^{t_0,x}(t_1), \ldots, \xi^{t_0,x}(t_m)\right) \middle| \mathcal{G}\right].$$

If $x : \Omega \to \mathbf{R}$, then $E^{t_0,x}\left[f(\xi(t_1), \ldots, \xi(t_m))\right]$ shall denote the random variable $\omega \mapsto E^{t_0,x(\omega)}\left[f(\xi(t_1), \ldots, \xi(t_m))\right]$.

If $t_0 = 0$, we shall suppress the first superscript, thus writing E^x and ξ^x instead of $E^{0,x}$ and $\xi^{0,x}$, respectively.

The following result is another radically elementary version of the Itô–Doeblin formula (cf. Benoît [10, Proposition 4.6.1] for a special case); it is a straightforward generalization of Lemma 3.9, but its statement and proof need some notational effort.

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_7,

Lemma 7.3 (Itô–Doeblin formula). *Let ξ be an Itô diffusion with drift coefficient function a and diffusion coefficient function b starting at time t_0, let $f : \mathbf{R}\times[t_0, 1] \to \mathbf{R}$ be thrice continuously differentiable, and let $\omega \in \Omega$ be such that the stochastic processes $a(\xi, \cdot)$, $b(\xi, \cdot)$, $f''(\xi, \cdot)$, $f'''(\xi, \cdot)$ have limited ω-trajectories. Then for all $s \in \mathbf{T} \cap [t_0, 1]$, suppressing the argument ω,*

$$f(\xi(s), s) - f(\xi(t_0), t_0) \simeq \int_{t_0}^{s} \partial_1 f(\xi(t), t)\,\mathrm{d}\xi(t) + \int_{t_0}^{s} \partial_2 f(\xi(t), t)\,\mathrm{d}t + \frac{1}{2}\int_{t_0}^{s} \partial_{1,1} f(\xi(t), t)\, b(\xi(t), t)^2\,\mathrm{d}t$$

Proof. Let $t \in \mathbf{T} \cap [t_0, 1)$. Note that $|\mathrm{d}W(t)| = \sqrt{\mathrm{d}t}$, whence

$$|\mathrm{d}\xi(t)| \leq |a(\xi(t), t)|\,\mathrm{d}t + |b(\xi(t), t)|\,|\mathrm{d}W(t)| = \mathcal{O}\left((\mathrm{d}t)^{1/2}\right).$$

and thus

$$\mathrm{d}\xi(t)\mathrm{d}t = \mathcal{O}\left((\mathrm{d}t)^{3/2}\right), \quad |\mathrm{d}\xi(t) + \mathrm{d}t|^3 = \mathcal{O}\left((\mathrm{d}t)^{3/2}\right).$$

Moreover,

$$(\mathrm{d}\xi(t))^2 = a(\xi(t), t)^2 (\mathrm{d}t)^2 + 2a(\xi(t), t)\, b(\xi(t), t)\, \underbrace{\mathrm{d}W(t)\mathrm{d}t}_{=\mathcal{O}\left((\mathrm{d}t)^{3/2}\right)} + b(\xi(t), t)^2 \underbrace{(\mathrm{d}W(t))^2}_{=\mathrm{d}t},$$

so

$$(\mathrm{d}\xi(t))^2 = b(\xi(t), t)^2\,\mathrm{d}t + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right).$$

Therefore, using the third-order Taylor formula and the symmetry of the Hessian (Schwarz's theorem),

$$\begin{aligned}
&\mathrm{d}f(\xi(t), t) \\
&= {}^{\top}\begin{pmatrix} \mathrm{d}\xi(t) \\ \mathrm{d}t \end{pmatrix} \begin{pmatrix} \partial_1 f(\xi(t), t) \\ \partial_2 f(\xi(t), t) \end{pmatrix} \\
&\quad + \frac{1}{2}{}^{\top}\begin{pmatrix} \mathrm{d}\xi(t) \\ \mathrm{d}t \end{pmatrix} \begin{pmatrix} \partial_{1,1} f(\xi(t), t) & \partial_{1,2} f(\xi(t), t) \\ \underbrace{\partial_{2,1} f(\xi(t), t)}_{=\partial_{1,2} f(\xi(t), t)} & \partial_{2,2} f(\xi(t), t) \end{pmatrix} \begin{pmatrix} \mathrm{d}\xi(t) \\ \mathrm{d}t \end{pmatrix} \\
&\quad + \mathcal{O}\left(|\mathrm{d}\xi(t) + \mathrm{d}t|^3\right) \\
&= \partial_1 f(\xi(t), t)\,\mathrm{d}\xi(t) + \partial_2 f(\xi(t), t)\,\mathrm{d}t
\end{aligned}$$

$$+\frac{1}{2}\left(\partial_{1,1} f\left(\xi(t), t\right) \underbrace{(\mathrm{d}\xi(t))^2}_{=b(\xi(t),t)^2 \mathrm{d}t+\mathcal{O}\left((\mathrm{d}t)^{3/2}\right)} +\partial_{2,2} f\left(\xi(t), t\right)(\mathrm{d}t)^2\right)$$

$$+\partial_{1,2} f\left(\xi(t), t\right) \mathrm{d}\xi(t)\mathrm{d}t+\mathcal{O}\left(|\mathrm{d}\xi(t)+\mathrm{d}t|^3\right)$$

$$=\partial_1 f\left(\xi(t), t\right) \mathrm{d}\xi(t)+\partial_2 f\left(\xi(t), t\right) \mathrm{d}t+\frac{1}{2} \partial_{1,1} f\left(\xi(t), t\right)(\mathrm{d}\xi(t))^2$$

$$+\mathcal{O}\left((\mathrm{d}t)^{3/2}\right)$$

$$=\partial_1 f\left(\xi(t), t\right) \mathrm{d}\xi(t)+\partial_2 f\left(\xi(t), t\right) \mathrm{d}t+\frac{1}{2} \partial_{1,1} f\left(\xi(t), t\right) b\left(\xi(t), t\right)^2 \mathrm{d}t$$

$$+\mathcal{O}\left((\mathrm{d}t)^{3/2}\right).$$

Integrating this while noting that $\int_{t_0}^s \mathcal{O}\left((\mathrm{d}t)^{3/2}\right) \mathrm{d}t=\frac{s-t_0}{\mathrm{d}t}\mathcal{O}\left((\mathrm{d}t)^{3/2}\right)=\mathcal{O}\left((\mathrm{d}t)^{1/2}\right) \simeq 0$ (as $\mathbf{T} \cap [0, s)$ is finite) for all $s \in \mathbf{T}$, one obtains

$$\begin{aligned}
f\left(\xi(s), s\right)-f\left(\xi(t_0), t_0\right) &=\int_{t_0}^s \mathrm{d} f\left(\xi(t), t\right) \\
&\simeq \int_{t_0}^s \partial_1 f\left(\xi(t), t\right) \mathrm{d}\xi(t)+\int_{t_0}^s \partial_2 f\left(\xi(t), t\right) \mathrm{d}t \\
&\quad +\frac{1}{2} \int_{t_0}^s \partial_{1,1} f\left(\xi(t), t\right) b\left(\xi(t), t\right)^2 \mathrm{d}t
\end{aligned}$$

for every $s \in \mathbf{T}$, which already is the Itô–Doeblin formula. □

As a corollary one obtains:

Corollary 7.4 (Dynkin's formula). *Let ξ be an Itô diffusion with drift coefficient function a and diffusion coefficient function b starting at time t_0 with $\xi(t_0)=x$, assume that the stochastic processes $a(\xi,\cdot)$ and $b(\xi,\cdot)$ are limited, and let $f:\mathbf{R}\times[t_0,1]\to\mathbf{R}$ be thrice continuously differentiable with uniformly limited second- and third-order derivatives. Then,*

$$\begin{aligned}
\frac{1}{\mathrm{d}t} E\left[\mathrm{d} f\left(\xi(t_0), t_0\right)\right] \simeq\ & \partial_1 f\left(x, t_0\right) a\left(x, t_0\right)+\partial_2 f\left(x, t_0\right) \\
& +\frac{1}{2} \partial_{1,1} f\left(x, t_0\right) b\left(x, t_0\right)^2 .
\end{aligned}$$

In the special case where $f:\mathbf{R}\to\mathbf{R}$,

$$\frac{1}{\mathrm{d}t} E\left[\mathrm{d} f\left(\xi(t_0)\right)\right] \simeq f'(x)\, a\left(x, t_0\right)+\frac{1}{2} f''(x)\, b\left(x, t_0\right)^2 .$$

Proof. From the Itô–Doeblin formula, we obtain

$$\begin{aligned} E\left[\mathrm{d}f\left(\xi(t_0),t_0\right)\right] = {} & \partial_1 f\left(\xi(t_0),t_0\right) E\left[\mathrm{d}\xi(t_0)\right] + \partial_2 f\left(\xi(t_0),t_0\right)\mathrm{d}t \\ & + \frac{1}{2}\partial_{1,1} f\left(\xi(t_0),t_0\right) b\left(\xi(t_0),t_0\right)^2 \mathrm{d}t \\ & + \mathcal{O}\left((\mathrm{d}t)^{3/2}\right). \end{aligned}$$

This leads us, after noting that

$$E\left[\mathrm{d}\xi(t_0)\right] = a\left(\xi(t_0),t_0\right)\mathrm{d}t + b\left(\xi(t),t\right)\underbrace{E\left[\mathrm{d}W(t_0)\right]}_{=0} = a\left(\xi(t_0),t_0\right)\mathrm{d}t$$

and that $\xi(t_0) = x$, to the desired result. □

7.2 The Markov Property of Itô Diffusions and the Feynman–Kac Formula

Whenever $a, b : \mathbf{R} \to \mathbf{R}$, we can view a and b as functions from $\mathbf{R} \times [0,1] \to \mathbf{R}$ which are constant in the second argument. Then, any Itô diffusion with drift coefficient function a and diffusion coefficient function b will be time-homogeneous as defined in Definition 7.1.

The next result shows that time-homogeneous Itô diffusions also have another property which might also be termed time-homogeneity.

Lemma 7.5 (Distribution-wise time-homogeneity). *Let $x \in \mathbf{R}$, $s \in \mathbf{T} \setminus \{1\}$ and $a, b : \mathbf{R} \to \mathbf{R}$ (so that ξ^x and $\xi^{s,x}$ are time-homogeneous). Then, the joint P-distribution of $\xi^x(0), \ldots, \xi^x(1-s)$ equals the joint P-distribution of $\xi^{s,x}(s), \ldots, \xi^{s,x}(1)$.*

Proof. Define a new process $\tilde{W}$ through

$$\tilde{W}(t) = W(s+t) - W(s)$$

for all $t \in \mathbf{T} \cap [0, 1-s]$, so that $\tilde{W}(0) := 0$ and $\mathrm{d}\tilde{W}(t) = \mathrm{d}W(s+t)$ for all $t \in \mathbf{T} \cap [0, 1-s)$.

Inductively in $t \in \mathbf{T} \cap [0, 1-s]$, we shall prove that the joint P-distribution of $\xi^x(0), \ldots, \xi^x(t), W(0), \ldots, W(1-s)$ equals the joint P-distribution of $\xi^{s,x}(s), \ldots, \xi^{s,x}(s+t), \tilde{W}(0), \ldots, \tilde{W}(1-s)$. For the basis step ($t=0$), one only has to verify that $W(0), \ldots, W(1-s)$ have the same joint distribution as $\tilde{W}(0), \ldots, \tilde{W}(1-s)$. For the induction step, fix an arbitrary $t \in \mathbf{T} \cap [0, 1-s)$, and assume that the joint distribution of $\xi^x(0), \ldots, \xi^x(t), W(0), \ldots, W(1-s)$ equals the joint distribution of $\xi^{s,x}(s), \ldots, \xi^{s,x}(s+t), \tilde{W}(0), \ldots, \tilde{W}(1-s)$. Let F be the function

$$F : \mathbf{R}^3 \to \mathbf{R}, \quad (y, z, w) \mapsto y + a(y)\mathrm{d}t + b(y)(z - w).$$

Then, applying the diffusion equation of ξ, we have

$$\begin{aligned}
&\xi^{s,x}(s + t + \mathrm{d}t) \\
&\quad = \xi^{s,x}(s + t) + \mathrm{d}\xi^{s,x}(s + t) \\
&\quad = \xi^{s,x}(s + t) + a\left(\xi^{s,x}(s + t)\right)\mathrm{d}t + b\left(\xi^{s,x}(s + t)\right)\mathrm{d}W(s + t) \\
&\quad = F\left(\xi^{s,x}(s + t), W(s + t + \mathrm{d}t), W(s + t)\right) \\
&\quad = F\left(\xi^{s,x}(s + t), W(s + t + \mathrm{d}t) - W(s), W(s + t) - W(s)\right) \\
&\quad = F\left(\xi^{s,x}(s + t), \tilde{W}(t + \mathrm{d}t), \tilde{W}(t)\right)
\end{aligned}$$

and similarly

$$\begin{aligned}
\xi^{x}(t + \mathrm{d}t) &= \xi^{x}(t) + \mathrm{d}\xi^{x}(t) \\
&= \xi^{x}(t) + a\left(\xi^{x}(t)\right)\mathrm{d}t + b\left(\xi^{x}(t)\right)\mathrm{d}W(t) \\
&= F\left(\xi^{x}(t), W(t + \mathrm{d}t), W(t)\right).
\end{aligned}$$

Our induction hypothesis implies that the joint distribution of $\xi^x(0), \ldots, \xi^x(t)$, $W(0), \ldots, W(1 - s), F(\xi^x(t), W(t + \mathrm{d}t), W(t))$ equals the joint distribution of $\xi^{s,x}(s), \ldots, \xi^{s,x}(s + t), \tilde{W}(0), \ldots, \tilde{W}(1 - s), F(\xi^{s,x}(s + t), \tilde{W}(t + \mathrm{d}t), \tilde{W}(t))$. This completes the induction step, whence we have established that for every $t \in \mathbf{T} \cap [0, 1 - s]$, the joint distribution of $\xi^x(0), \ldots, \xi^x(t), W(0), \ldots, W(1 - s)$ equals the joint distribution of $\xi^{s,x}(s), \ldots, \xi^{s,x}(s + t), \tilde{W}(0), \ldots, \tilde{W}(1 - s)$. This vacuously implies the result asserted in the lemma. □

Theorem 7.6 (Markov property of time-homogeneous Itô diffusions). *Let $x \in \mathbf{R}$ and $a, b : \mathbf{R} \to \mathbf{R}$. For all $s \in \mathbf{T} \setminus \{1\}$, $t \in \mathbf{T} \cap [0, 1 - s]$, and $f : \mathbf{R} \to \mathbf{R}$*

$$E^x\left[f\left(\xi(s + t)\right) \middle| \mathcal{F}_s\right] = E^{\xi^x(s)}\left[f\left(\xi(t)\right)\right]. \tag{7.2}$$

Proof. Define a function g by

$$g : \mathbf{R} \times \Omega \to \mathbf{R}, \quad (x, \omega) \mapsto f\left(\xi^{s,x}(s + t)(\omega)\right).$$

The set Ω being finite, one can find functions ϕ_k, ψ_k, $k \in \{1, \ldots, m\}$ for $m \in \mathbf{N}$, such that

$$g(x, \omega) = \sum_{k=1}^{m} \phi_k(x)\psi_k(\omega)$$

for all $(x, \omega) \in \mathbf{R} \times \Omega$. (Simply choose $m = \mathrm{card}(\Omega)$, and letting $\Omega = \{\omega_1, \ldots, \omega_m\}$, define $\psi_k(\omega_\ell) = 1$ if $\ell = k$ and $\psi_k(\omega_\ell) = 0$ otherwise, and define $\phi_k(x) = g(x, \omega_k)$ for all k and x.)

Note that by definition, $\xi^x(s+t)(\omega) = \xi^{\xi^x(s)(\omega)}(t)(\omega)$ for all $\omega \in \Omega$ and therefore

$$E^x\left[f\left(\xi(s+t)\right)\middle|\,\mathfrak{F}_s\right] = E\left[g\left(\xi^x(s),\cdot\right)\middle|\,\mathfrak{F}_s\right],$$

wherein $g\left(\xi^x(s),\cdot\right)$ denotes, of course, the random variable $\omega \mapsto g\left(\xi^x(s)(\omega),\omega\right)$.

It then follows from the choice of the functions ϕ, ψ and from the $\mathfrak{F}_s$-measurability of $\xi^x(s)$ that on the one hand

$$\begin{aligned}
&E^x\left[f\left(\xi(s+t)\right)\middle|\,\mathfrak{F}_s\right]\\
&\quad= E\left[g\left(\xi^x(s),\cdot\right)\middle|\,\mathfrak{F}_s\right] = E\left[\sum_{k=1}^m \phi_k\left(\xi^x(s)\right)\psi_k\middle|\,\mathfrak{F}_s\right]\\
&\quad= \sum_{k=1}^m \phi_k\left(\xi^x(s)\right)E\left[\psi_k\middle|\,\mathfrak{F}_s\right] = \sum_{k=1}^m \phi_k(y)\,E\left[\psi_k\middle|\,\mathfrak{F}_s\right]\Big|_{y=\xi^x(s)}\\
&\quad= \sum_{k=1}^m \phi_k(y)\,E\left[\psi_k\middle|\,\mathfrak{F}_s\right]\Big|_{y=\xi^x(s)} = E\left[\sum_{k=1}^m \phi_k(y)\,\psi_k\middle|\,\mathfrak{F}_s\right]\Big|_{y=\xi^x(s)}\\
&\quad= E\left[g(y,\cdot)\middle|\,\mathfrak{F}_s\right]\Big|_{y=\xi^x(s)}.
\end{aligned}$$

On the other hand, $\xi^{s,y}(s+t)$ is, for every fixed $y \in \mathbf{R}$, a function of $\mathrm{d}W(s), \ldots, \mathrm{d}W(s+t-\mathrm{d}t)$, all of which are independent of $\mathfrak{F}_s$; therefore, $\xi^{s,y}(s+t)$ and hence $f\left(\xi^{s,y}(s+t)\right)$ are independent of $\mathfrak{F}_s$, too. Moreover, the distribution of $\xi^{s,y}(s+t)$ equals the distribution of $\xi^y(t)$ by the time-homogeneity result of Lemma 7.5. Therefore,

$$\begin{aligned}
E\left[g(y,\cdot)\middle|\,\mathfrak{F}_s\right] &= E\left[f\left(\xi^{s,y}(s+t)\right)\middle|\,\mathfrak{F}_s\right] = E\left[f\left(\xi^{s,y}(s+t)\right)\right]\\
&= E\left[f\left(\xi^y(t)\right)\right].
\end{aligned}$$

Combining this with the previously established equation $E^x\left[f\left(\xi(s+t)\right)\middle|\,\mathfrak{F}_s\right] = E\left[g(y,\cdot)\middle|\,\mathfrak{F}_s\right]\Big|_{y=\xi^x(s)}$, we finally arrive at

$$\begin{aligned}
E^x\left[f\left(\xi(s+t)\right)\middle|\,\mathfrak{F}_s\right] &= E\left[g(y,\cdot)\middle|\,\mathfrak{F}_s\right]\Big|_{y=\xi^x(s)}\\
&= E\left[f\left(\xi^{s,y}(s+t)\right)\right]\Big|_{y=\xi^x(s)}\\
&= E\left[f\left(\xi^y(t)\right)\right]\Big|_{y=\xi^x(s)},
\end{aligned}$$

hence

$$E^x\left[f\left(\xi(s+t)\right)\middle|\,\mathfrak{F}_s\right] = E^{\xi^x(s)}\left[f\left(\xi(t)\right)\right].$$

$\square$

If $u : \mathbf{R} \times \mathbf{T} \to \mathbf{R}$, define—in analogy to the increment process ξ—for all $(x,t) \in \mathbf{R} \times (\mathbf{T} \setminus \{1\})$,

$$\mathrm{d}u(x,t) = u(x,t+\mathrm{d}t) - u(x,t).$$

If $u : \mathbf{R}\times[0,1] \to \mathbf{R}$ is a function which is differentiable in the second argument and can be defined internally without parameters, then one can verify that $\mathrm{d}u(x,t)/\mathrm{d}t \simeq \partial_2 u(x,t)$ for all $t \in \mathbf{T}\setminus\{1\}$ and limited $x \in \mathbf{R}$.

Theorem 7.7 (Feynman–Kac formula). *Let $a, b : \mathbf{R} \to \mathbf{R}$, let $g : \mathbf{R} \to \mathbf{R}$ be thrice continuously differentiable with uniformly limited second- and third-order derivatives, and let u be the function*

$$u : \mathbf{R}\times\mathbf{T} \to \mathbf{R}, \qquad (x,t) \mapsto E^x\left[g\left(\xi(t)\right)\right].$$

Then for all $(x,t) \in \mathbf{R}\times(\mathbf{T}\setminus\{1\})$,

$$\frac{\mathrm{d}u(x,t)}{\mathrm{d}t} \simeq \partial_1 u\left(x,t\right) a\left(x\right) + \frac{1}{2}\partial_{1,1} u\left(x,t\right) b\left(x\right)^2. \tag{7.3}$$

Proof. Fix $t \in \mathbf{T}\setminus\{1\}$. Since g is thrice continuously differentiable with uniformly limited second- and third-order derivatives and Ω is finite, it is not difficult to prove that $u(\cdot,t)$ must also be thrice continuously differentiable with uniformly limited second- and third-order derivatives $\partial_1 u(\cdot,t)$, $\partial_{1,1}u(\cdot,t)$, $\partial_{1,1,1}u(\cdot,t)$. Hence, we may apply Corollary 7.4 (Dynkin's formula) to the function $u(\cdot,t)$ and obtain for all $x \in \mathbf{R}$,

$$\begin{aligned}
&\frac{1}{\mathrm{d}t}E^x\left[u\left(\xi(\mathrm{d}t),t\right) - u\left(\xi(0),t\right)\right] \\
&\quad\simeq \partial_1 u\left(x,t\right) a\left(x\right) + \frac{1}{2}\partial_{1,1}u\left(x,t\right) b\left(x\right)^2.
\end{aligned}$$

However, on the other hand, the Markov property of the Itô diffusion ξ and the definition of u allow us to calculate

$$\begin{aligned}
E^x\left[u\left(\xi(\mathrm{d}t),t\right)\right] &= E^x\left[E^{\xi(\mathrm{d}t)}\left[g\left(\xi(t)\right)\right]\right] = E\left[E^{\xi^x(\mathrm{d}t)}\left[g\left(\xi(t)\right)\right]\right] \\
&= E\left[E\left[g\left(\xi^x(t+\mathrm{d}t)\right)\middle|\,\mathcal{F}_{\mathrm{d}t}\right]\right] \\
&= E\left[g\left(\xi^x(t+\mathrm{d}t)\right)\right] = u(x,t+\mathrm{d}t),
\end{aligned}$$

hence

$$E^x\left[u\left(\xi(\mathrm{d}t),t\right) - u\left(\xi(0),t\right)\right] = u(x,t+\mathrm{d}t) - u(x,t) = \mathrm{d}u(x,t)$$

for all $x \in \mathbf{R}$. Therefore, for all $x \in \mathbf{R}$,

$$\frac{1}{\mathrm{d}t}\mathrm{d}u(x,t) \simeq \partial_1 u\left(x,t\right) a\left(x\right) + \frac{1}{2}\partial_{1,1}u\left(x,t\right) b\left(x\right)^2.$$

□

Our Feynman–Kac formula (7.3), whilst not being a classical partial differential equation, constitutes for every fixed $x \in \mathbf{R}$ a difference equation of infinitesimal spacing in t (up to an infinitesimal). Such difference equations have been studied extensively in the framework of Internal Set Theory (and successfully linked to classical ordinary differential equations) by the Alsatian school of nonstandard analysis, following an initiative of Reeb and Callot; cf. Sari [69, 70] and van den Berg [12, 14, 15].

The Feynman–Kac formula has proven to be fruitful in quantum mechanics. Consider the semigroup $\left(\mathrm{e}^{-t\left(-\frac{\hbar^2}{2m}\Delta+V\right)}\right)_{t\in[0,1]}$ generated by the Hamiltonian $-\frac{\hbar^2}{2m}\Delta+V$ of a particle of mass m moving in a potential V. Given any $f : \mathbf{R} \to \mathbf{R}$ in the domain of the operator, the function $(x,t) \mapsto \mathrm{e}^{-t\left(-\frac{\hbar^2}{2m}\Delta+V\right)} f(x)$ is the solution u to the initial value problem

$$\partial_2 u = -\left(-\frac{\hbar^2}{2m}\Delta + V\right)u = \frac{\hbar^2}{2m}\Delta u - Vu.$$

The following Corollary 7.8 to the Feynman–Kac formula provides a probabilistic interpretation of (infinitesimal approximations of) such partial differential equations—and thus of the semigroup generated by the Hamiltonian—for the case of dimension one.

Corollary 7.8 (Feynman integral). *Let $m \in \mathbf{R}_{>0}$, let $V : \mathbf{R} \to \mathbf{R}$ be limited, let $f : \mathbf{R} \to \mathbf{R}$ be thrice continuously differentiable with uniformly limited second- and third-order derivatives. For any $x \in \mathbf{R}$, let ξ^x be the Itô diffusion given by $\xi^x(t) = x + \frac{\hbar}{\sqrt{m}}W(t)$ for all $t \in \mathbf{T}$, and let u be the function*

$$u : \mathbf{R} \times \mathbf{T} \to \mathbf{R}, \qquad (x,t) \mapsto \mathrm{e}^{-tV(x)} E^x\left[f\left(\xi(t)\right)\right].$$

Then for all $t \in \mathbf{T} \setminus \{1\}$ and all limited $x \in \mathbf{R}$,

$$\frac{\mathrm{d}u(x,t)}{\mathrm{d}t} \simeq \frac{\hbar^2}{2m}\partial_{1,1}u\left(x,t\right) - V(x)u\left(x,t\right). \tag{7.4}$$

The proof of Corollary 7.8 uses the following radically elementary product rule of stochastic differentiation, which is interesting in its own right: If $f, g : Y \times \mathbf{T} \to \mathbf{R}$ for some set Y, then for all $y \in Y$ and $t \in \mathbf{T} \setminus \{1\}$,

$$\begin{aligned} &\mathrm{d}(fg)(y,t) \\ &\quad = \mathrm{d}f\left(y,t\right)\mathrm{d}g\left(y,t\right) + \mathrm{d}f\left(y,t\right)g\left(y,t\right) + f\left(y,t\right)\mathrm{d}g\left(y,t+\mathrm{d}t\right) \end{aligned} \tag{7.5}$$

as an easy algebraic calculation shows.[1]

In addition, the proof of Corollary 7.8 uses the fact that whenever $V : \mathbf{R} \to \mathbf{R}$, then for all $t \in \mathbf{T} \setminus \{1\}$, one has

$$\frac{\mathrm{d}e^{-tV(x)}}{\mathrm{d}t} \simeq -V(x)e^{-tV(x)} \tag{7.6}$$

for all $x \in \mathbf{R}$ such that $V(x)$ is limited, which can be easily proven through a second-order Taylor expansion of the exponential function.[2]

Proof. First, note that f is limited, on account of the uniform limitedness of its second derivatives, as a second-order Taylor expansion shows.

Let $\hat{u}$ be the function $(x,t) \mapsto E^x\left[f\left(\xi(t)\right)\right]$, so that $u(x,t) = e^{-tV(x)}\hat{u}(x,t)$ for all x, t. Let $x \in \mathbf{R}$ be limited and $t \in \mathbf{T} \setminus \{1\}$. By Eq. (7.3) and the definition of ξ, one has $\frac{\mathrm{d}\hat{u}(x,t)}{\mathrm{d}t} \simeq \frac{\hbar^2}{2m}\partial_{1,1}\hat{u}(x,t)$ and by the limitedness of $V(x)$ combined with Eq. (7.6) also $\frac{\mathrm{d}e^{-tV(x)}}{\mathrm{d}t} \simeq -V(x)e^{-tV(x)}$. Note that this implies that $\frac{\mathrm{d}\hat{u}(x,t)}{\mathrm{d}t}$ is limited, even uniformly in t (because $\partial_{1,1}\hat{u}$ is uniformly limited as f has uniformly limited second-order derivatives), and $\frac{\mathrm{d}e^{-tV(x)}}{\mathrm{d}t}$ is limited, too (because $V(x)$ is limited), whence $\hat{u}(x,t)$ is limited and moreover $\frac{\mathrm{d}\hat{u}(x,t)\mathrm{d}e^{-tV(x)}}{\mathrm{d}t} = \frac{\mathrm{d}\hat{u}(x,t)}{\mathrm{d}t}\frac{\mathrm{d}e^{-tV(x)}}{\mathrm{d}t}\mathrm{d}t \simeq 0$. Clearly, $e^{-tV(x)}$ is limited as $V(x)$ is limited. Hence, the product rule (7.5) yields

---

[1] Indeed, by adding both $0 = -f(y,t)\,g(y,t+\mathrm{d}t) + f(y,t)\,g(y,t+\mathrm{d}t)$ and $0 = -\mathrm{d}f(y,t)\,g(y,t) + \mathrm{d}f(y,t)\,g(y,t)$ on each side of the equation, one obtains

$$\begin{aligned}
\mathrm{d}(fg)(y,t) &= f(y,t+\mathrm{d}t)\,g(y,t+\mathrm{d}t) - f(y,t)\,g(y,t) \\
&= f(y,t+\mathrm{d}t)\,g(y,t+\mathrm{d}t) - f(y,t)\,g(y,t+\mathrm{d}t) \\
&\quad + f(y,t)\,g(y,t+\mathrm{d}t) - f(y,t)\,g(y,t) \\
&= \mathrm{d}f(y,t)\,g(y,t+\mathrm{d}t) + f(y,t)\,\mathrm{d}g(y,t+\mathrm{d}t) \\
&= \mathrm{d}f(y,t)\,\mathrm{d}g(y,t) + \mathrm{d}f(y,t)\,g(y,t) + f(y,t)\,\mathrm{d}g(y,t+\mathrm{d}t)\,.
\end{aligned}$$

[2] Indeed,

$$\begin{aligned}
\frac{\mathrm{d}e^{-tV(x)}}{\mathrm{d}t} &= \left(e^{-tV(x)}e^{-\mathrm{d}tV(x)} - e^{-tV(x)}\right)/\mathrm{d}t \\
&= e^{-tV(x)}\left(\underbrace{e^{-\mathrm{d}tV(x)} - 1}_{=-\mathrm{d}tV(X)+\mathcal{O}\left((\mathrm{d}t)^2\right)}\right)/\mathrm{d}t \\
&= -e^{-tV(x)}V(x).
\end{aligned}$$

$$\begin{aligned}\frac{\mathrm{d}u(x,t)}{\mathrm{d}t} &= \frac{\mathrm{d}\hat{u}(x,t)}{\mathrm{d}t}\mathrm{e}^{-tV(x)} + \hat{u}(x,t)\frac{\mathrm{d}\mathrm{e}^{-tV(x)}}{\mathrm{d}t} + \frac{\mathrm{d}\hat{u}(x,t)\mathrm{d}\mathrm{e}^{-tV(x)}}{\mathrm{d}t}\\
&\simeq \mathrm{e}^{-tV(x)}\frac{\hbar^2}{2m}\partial_{1,1}\hat{u}\,(x,t) - V(x)\mathrm{e}^{-tV(x)}\hat{u}(x,t)\\
&= \frac{\hbar^2}{2m}\partial_{1,1}\left(\mathrm{e}^{-tV(x)}\hat{u}\,(x,t)\right) - V(x)\mathrm{e}^{-tV(x)}\hat{u}(x,t).\end{aligned}$$

□

All results in this work, including the Feynman–Kac formula, can easily be generalized to multi-dimensional Itô diffusions, even with respect to multi-dimensional Wiener walks. For instance, a *d-dimensional Wiener walk* can be defined as a d-tuple $\left(W^{(1)}, \ldots, W^{(d)}\right)$ of independent Wiener walks; equivalently, $\left(W^{(1)}, \ldots, W^{(d)}\right)$ is a d-dimensional Wiener walk if and only if

$$\mathrm{d}W^{(1)}(0), \ldots, \mathrm{d}W^{(1)}(1-\mathrm{d}t),\ \ldots\ \ldots\ \ldots\ , \mathrm{d}W^{(d)}(0), \ldots, \mathrm{d}W^{(d)}(1-\mathrm{d}t)$$

is a finite sequence of independent random variables. Generalizing the Feynman–Kac formula from dimension 1 to dimension d also allows for a d-dimensional generalization of Corollary 7.8; this result will then be an infinitesimal approximate probabilistic interpretation of the semigroup generated by the Hamiltonian of a particle of mass m moving in a potential V on $\mathbf{R}^d$.

Chapter 8
Excursion to Mathematical Physics: A Radically Elementary Definition of Feynman Path Integrals

In this excursion, which is inspired by Albeverio et al. [3, Sect. 6.6] and the classical article by Nelson [58], we give another demonstration of the usefulness of radically elementary mathematics in mathematical physics, by providing a rigorous, radically elementary definition of Feynman path integrals in Minimal Internal Set Theory. A summary of these ideas—combined with a brief introduction to radically elementary mathematics for mathematical physicists and some references to previous attempts at formalizing the Feynman path integral by means of nonstandard analysis—can be found in [35].

Consider a particle of mass m moving in a potential V on $\mathbf{R}^k$, with the initial state (time 0) described by φ_0, and let ψ be the particle's wave function. Mathematically, this means that ψ is the solution of the Schrödinger equation:

$$\mathrm{i}\hbar\partial_{k+1}\psi = H\psi, \qquad \psi(\cdot, 0) = \varphi_0$$

with H being the Hamiltonian operator

$$H = H_0 + V, \qquad H_0 = -\frac{\hbar^2}{2m}\Delta$$

(wherein we assume V to be Lebesgue measurable and φ_0 to be Lebesgue square-integrable). The physical interpretation of ψ rests on the fact that $|\psi|^2$ is a probability density and views—following the Copenhagen school of quantum mechanics—$|\psi(x,t)|^2$ as the likelihood for the particle to be in x at time t.

The quest for a mathematically rigorous definition of the Feynman path integral concerns the problem whether one can find, for any $t > 0$ and $x \in \mathbf{R}^k$, a measure μ_x^t on some space Γ of $\mathbf{R}^k$-valued trajectories such that $\psi(x,t)$ is given by $\int_{\{\gamma\in\Gamma\,:\,\gamma(t)=x\}} \exp\left(\frac{\mathrm{i}}{\hbar} S_t(\gamma)\right) \varphi_0(\gamma(0))\, \mathrm{d}\mu_x^t(\gamma)$, wherein S_t is the energy functional on the path space Γ. If that were possible, the solution of the Schrödinger equation at time t could be written, up to a constant, as the expected value of a very simple

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_8,

term with a straightforward physical interpretation—the mean being with respect to a distribution on paths.

Since it is well known that ψ is, as a function of time, given by the semigroup generated by the Hamiltonian, i.e.

$$\forall t > 0 \qquad \psi(\cdot, t) = \exp\left(-\frac{\mathrm{i}}{\hbar} tH\right) \varphi_0,$$

the search for the Feynman path integral is tantamount to the problem of writing $\exp\left(-\frac{\mathrm{i}}{\hbar}tH\right)\varphi$, at least for sufficiently regular φ, in the form $\int_{\{\gamma\in\Gamma\,:\,\gamma(t)=x\}} \exp\left(\frac{\mathrm{i}}{\hbar}S_t(\gamma)\right)\varphi(\gamma(0))\,\mathrm{d}\mu_x^t(\gamma)$.

Let us fix $t > 0$ and, for simplicity, let us consider a smooth φ with compact support. By the Lie–Trotter product formula,

$$\exp\left(-\frac{\mathrm{i}}{\hbar} tH\right) \varphi = \lim_{n\to\infty} \left(\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n}V\right)\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n}H_0\right)\right)^n \varphi.$$

Therefore, for sufficiently large (nonstandard) n,[1] we have

$$\exp\left(-\frac{\mathrm{i}}{\hbar} tH\right) \varphi \simeq \left(\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n}V\right)\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n}H_0\right)\right)^n \varphi. \tag{8.1}$$

On the other hand for all $x \in \mathbf{R}^k$

$$\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n}H_0\right)\varphi(x) = \left(2\pi\frac{\mathrm{i}\hbar}{m}\frac{t}{n}\right)^{-k/2}\int_{\mathbf{R}^k}\exp\left(\frac{\mathrm{i}m}{2\hbar t/n}|x-y|^2\right)\varphi(y)\mathrm{d}y,$$

if the complex fractional power function $z \mapsto z^{-k/2}$ is appropriately determined. Thus for all $x \in \mathbf{R}^k$,

$$\begin{aligned}
&\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n}V\right)\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n}H_0\right)\varphi(x)\\
&= \left(2\pi\frac{\mathrm{i}\hbar}{m}\frac{t}{n}\right)^{-k/2}\int_{\mathbf{R}^k}\exp\left(\frac{\mathrm{i}}{\hbar}\frac{t}{n}\left(\frac{m}{2(t/n)^2}|x-y|^2 - V(x)\right)\right)\varphi(y)\mathrm{d}y
\end{aligned}$$

[1] An analysis of the proof of the Lie–Trotter product formula (cf. e.g. Nelson [58, Appendix B, proof of Theorem 9]) shows that the condition which n has to meet is

$$\frac{\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n}H_0\right)\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n}V\right)\psi - \psi + \frac{\mathrm{i}}{\hbar}\frac{t}{n}H\psi}{t/n} \simeq 0.$$

At least when we impose limited bounds on m, V and ψ, this holds for all nonstandard n.

and therefore, after defining

$$S\,(x_n, \ldots, x_0) = \sum_{j=1}^{n} \left(\frac{m}{2(t/n)^2} |x_j - x_{j-1}|^2 - V(x_j) \right) \frac{t}{n}$$

for all $x_0, \ldots, x_n \in \mathbf{R}^k$, we obtain

$$\begin{aligned} &\left(\exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n} V \right) \exp\left(-\frac{\mathrm{i}}{\hbar}\frac{t}{n} H_0 \right) \right)^n \varphi(x) \\ &\quad = \left(2\pi \frac{\mathrm{i}\hbar}{m}\frac{t}{n} \right)^{-kn/2} \int_{\mathbf{R}^k} \cdots \int_{\mathbf{R}^k} \exp\left(\frac{\mathrm{i}}{\hbar} S\,(x, x_{n-1} \ldots, x_0) \right) \\ &\quad\quad \varphi(x_0)\mathrm{d}x_0 \cdots \mathrm{d}x_{n-1}. \end{aligned} \tag{8.2}$$

Combining Eq. (8.2) with the approximate identity (8.1), we arrive at

$$\begin{aligned} &\exp\left(-\frac{\mathrm{i}}{\hbar} t H \right) \varphi(x) \\ &\quad \simeq \left(2\pi \frac{\mathrm{i}\hbar}{m}\frac{t}{n} \right)^{-kn/2} \int_{\mathbf{R}^k} \cdots \int_{\mathbf{R}^k} \exp\left(\frac{\mathrm{i}}{\hbar} S\,(x, x_{n-1} \ldots, x_0) \right) \\ &\quad\quad \varphi(x_0)\mathrm{d}x_0 \cdots \mathrm{d}x_{n-1}. \end{aligned} \tag{8.3}$$

Let us fix $x \in \mathbf{R}^k$ and define a measure $\nu^{(t,x)}$ on $\mathbf{R}^{(n+1)\times k}$ by

$$\mathrm{d}\nu^{(t,x)}(y, x_{n-1} \ldots, x_0) = \left(2\pi \frac{\mathrm{i}\hbar}{m}\frac{t}{n} \right)^{-kn/2} \mathrm{d}x_0 \cdots \mathrm{d}x_{n-1}\, \delta_x(\mathrm{d}y)$$

(wherein δ_x is the Dirac measure concentrated on $\{x\}$) and introduce the convention that $\vec{\gamma}$ always denotes a $\mathbf{R}^{(n+1)\times k}$ matrix with columns $\gamma_n, \cdots, \gamma_0$ (i.e., $\vec{\gamma} = (\gamma_n, \cdots, \gamma_0)$). Then we can rewrite Eq. (8.3) as

$$\begin{aligned} \exp\left(-\frac{\mathrm{i}}{\hbar} t H \right) \varphi(x) &\simeq \int_{\mathbf{R}^{(n+1)\times k}} \exp\left(\frac{\mathrm{i}}{\hbar} S\,(\vec{\gamma}) \right) \varphi(\gamma_0)\mathrm{d}\nu^{(t,x)}(\vec{\gamma}) \\ &= \int_{\{x\}\times\mathbf{R}^{n\times k}} \exp\left(\frac{\mathrm{i}}{\hbar} S\,(\vec{\gamma}) \right) \varphi(\gamma_0)\mathrm{d}\nu^{(t,x)}(\vec{\gamma}). \end{aligned} \tag{8.4}$$

Observe that one can interpret $\mathbf{R}^{(n+1)\times k}$ as a space of trajectories—with time-line of cardinality $n+1$—in $\mathbf{R}^k$ and S as an energy functional on those paths. Next define

$$\Gamma = \left\{ \gamma : [0, t] \to \mathbf{R}^k : \begin{array}{c} \gamma \text{ continuous}, \\ \forall j \in \{1, \ldots, n\} \quad \gamma \upharpoonright \left[\frac{(j-1)t}{n}, \frac{jt}{n} \right) \text{ affine-linear} \end{array} \right\}$$

and

$$\Gamma_x = \{\gamma \in \Gamma \; : \; \gamma(t) = x\} \, .$$

Then $\Gamma \cong \mathbf{R}^{(n+1)\times k}$, the isomorphism being

$$\gamma \mapsto (\gamma(t), \gamma(t - t/n), \cdots, \gamma(t/n), \gamma(0)) \, ,$$

and under this isomorphism, Γ_x is mapped to $\{x\} \times \mathbf{R}^{n\times k}$. Denote the image of $\gamma \in \Gamma$ under this isomorphism by $\vec{\gamma}$, again denoting its columns $\gamma_n, \cdots, \gamma_0$, and define $S_t(\gamma) = S(\vec{\gamma})$ for all $\gamma \in \Gamma_x$. Thereby, S_t can be viewed as an energy functional on Γ_x, for:

$$\begin{aligned}
S_t(\gamma) &= \sum_{j=1}^{n} \left(\frac{m}{2(t/n)^2} |\gamma_j - \gamma_{j-1}|^2 - V(\gamma_j) \right) \frac{t}{n} \\
&= \sum_{j=1}^{n} \left(\frac{m}{2(t/n)^2} |\gamma(jt/n) - \gamma\left((j-1)t/n\right)|^2 - V\left(\gamma(jt/n)\right) \right) \frac{t}{n} \\
&= \frac{m}{2} \frac{t}{n} \sum_{j=1}^{n} \left| \frac{\gamma(jt/n) - \gamma\left((j-1)t/n\right)}{t/n} \right|^2 - \frac{t}{n} \sum_{j=1}^{n} V\left(\gamma(jt/n)\right) \\
&= \frac{m}{2} \int_0^t |\dot{\gamma}(s)|^2 \, \mathrm{d}s - \int_0^t V\left(\gamma(s)\right) \mathrm{d}s,
\end{aligned}$$

wherein $\frac{m}{2} \int_0^t |\dot{\gamma}(s)|^2 \, \mathrm{d}s$ is the kinetic energy associated with the path γ and $-\int_0^t V\left(\gamma(s)\right) \mathrm{d}s$ is its potential energy. (Note that, unlike in those parts of this monograph that deal with stochastic analysis proper, we do employ *conventional* Riemann integrals in the present excursion.)

Furthermore, the measure $\nu^{(t,x)}$ on $\{x\} \times \mathbf{R}^{n\times k}$ induces a measure μ_x^t on Γ_x under the isomorphism $\gamma \mapsto \vec{\gamma}$. Then finally, the approximate equation (8.4) can be paraphrased

$$\begin{aligned}
&\exp\left(-\frac{\mathrm{i}}{\hbar} tH\right) \varphi(x) \\
&\quad \simeq \int_{\Gamma_x} \exp\left(\frac{\mathrm{i}}{\hbar} S_t\left(\gamma\right)\right) \varphi\left(\gamma(0)\right) \mathrm{d}\mu_x^t(\gamma) \\
&\quad = \int_{\{\gamma \in \Gamma \,:\, \gamma(t) = x\}} \exp\left(\frac{\mathrm{i}}{\hbar} \left(\frac{m}{2} \int_0^t |\dot{\gamma}(s)|^2 \, \mathrm{d}s - \int_0^t V\left(\gamma(s)\right) \mathrm{d}s\right)\right) \\
&\qquad \varphi\left(\gamma(0)\right) \mathrm{d}\mu_x^t(\gamma),
\end{aligned} \tag{8.5}$$

and $\int_{\{\gamma\in\Gamma\,:\,\gamma(t)=x\}} \cdots \mathrm{d}\mu_x$ is an actual measure-theoretic integral (with respect to the measure μ_x) over all paths in Γ that satisfy $\gamma(t) = x$. Thus, Eq. (8.5) provides a rigorous definition of the Feynman path integral—based on Minimal Internal Set Theory.

Chapter 9
A Radically Elementary Theory of Lévy Processes

9.1 Random Walks and Lévy Walks

Classically, a Lévy process is defined as a stochastically continuous process, pinned to the origin, with independent and stationary increments (cf. e.g. Sato [71]). Lévy processes have also been studied by the means of Robinsonian nonstandard analysis (in particular by Lindstrøm [49] with sequels by Albeverio and Herzberg [2], Lindstrøm [50], Albeverio et al. [1, Chap. 5], Herzberg and Lindstrøm [36] and [33]). The nonstandard analogue of a Lévy process is, following Lindstrøm [49], a *hyperfinite Lévy process*, that is a *hyperfinite random walk* which almost surely does not become unlimited. Drawing on the ideas developed by Lindstrøm's [49], we shall now formulate a theory of random walks and Lévy walks in the setting of radically elementary probability theory.

A random walk is a stochastic processes whose sequence of increments is independent and such that all increments have the same distribution, given by a finite (possibly unlimited) set of increments and a probability distribution on it.

Definition 9.1. A stochastic process $L : \mathbf{T} \to \mathbf{R}^{\Omega}$ is a *random walk* on (Ω, P) if and only if it satisfies all of the following:

(1) *Pin to the origin.* $L(0) = 0$.
(2) *Independence of increments.* $\mathrm{d}L_0, \ldots, \mathrm{d}L_{1-\mathrm{d}t}$ are independent random variables.
(3) *Stationarity of increments.* For all $s, t \in \mathbf{T} \setminus \{1\}$ and all $b \in \mathbf{R}$,

$$P\{\mathrm{d}L(s) = b\} = P\{\mathrm{d}L(t) = b\}.$$

Remark 9.2 (and Definition). If L is a random walk, then there exists a finite set $A \subseteq \mathbf{R}$ and a family $(p_a)_{a\in A} \in [0,1]^A$ such that $\sum_{a\in A} p_a = 1$ and

$$\forall t \in \mathbf{T} \setminus \{1\} \qquad \forall a \in A \qquad P\{\mathrm{d}L(t) = a\} = p_a.$$

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_9,

The set A is called the *increment set* and $(p_a)_{a\in A}$ the family of *transition probabilities* of L. If A consists only of limited elements, then L is said to have *limited increments*.

Proof. We define $A = \{a \in \mathbf{R} \, : \, P\{\mathrm{d}L(0) = a\} > 0\}$. This set is finite: For, given any $a \in \mathbf{R}$, we can only have $P\{\mathrm{d}L(0) = a\} > 0$ if there exists some $\omega \in \Omega$ such that $\mathrm{d}L(0)(\omega) = a$, and the set of such a is finite as Ω is finite. For all $a \in A$, we define

$$p_a = P\{\mathrm{d}L(0) = a\} > 0.$$

□

Remark 9.3. Given any finite set $A \subseteq \mathbf{R}$ and a family $(p_a)_{a\in A} \in [0,1]^A$, there exists a probability space (Ω, P) and a Lévy walk L on (Ω, P) with increment set A and transition probabilities $(p_a)_{a\in A}$.

Proof. The same construction as in Remarks 2.15 and 2.16 can be used: Let P_0 be the probability measure on A defined by $P_0\{a\} = p_a$ for all $a \in A$, put $\Omega = A^{\mathbf{T}\setminus\{1\}}$, let P be the product probability measure $P = \bigotimes_{t\in\mathbf{T}\setminus\{1\}} P_0$, and let

$$\forall s \in \mathbf{T} \qquad L(s) = \sum_{t<s} \pi(t),$$

wherein $\pi(t) : A^{\mathbf{T}\setminus\{1\}} \to A$, for any $t \in \mathbf{T} \setminus \{1\}$, is the projection onto the t-th Cartesian factor in $A^{\mathbf{T}\setminus\{1\}}$. □

A Lévy walk is a random walk which is almost surely limited:

Definition 9.4. A *Lévy walk* is a random walk which is a.s. limited.

Examples are the Wiener walk or the Poisson walk; for these processes, the a.s. limitedness follows from the corollary to Nelson's radically elementary martingale inequality (Corollary 2.13).

An alternative characterization of Lévy walks will be provided later on in Theorem 9.9.

Remark 9.5. Obviously, a random walk L is a Lévy walk if and only if

$$P\left\{\max_{t\in\mathbf{T}} |L(t)| \geq k\right\} \simeq 0$$

for all unlimited $k > 0$.

Proof. Indeed, if $\max_{t\in\mathbf{T}} |L(t)|$ is a.s. limited, then for every $\varepsilon \gg 0$ there exists some set N with $P\{\max_{t\in\mathbf{T}} |L(t)| \geq k\} \leq P(N) \leq \varepsilon$. Conversely, if $P\{\max_{t\in\mathbf{T}} |L(t)| \geq k\} \simeq 0$ for all unlimited $k > 0$, then for every $\varepsilon > 0$, the internal formula $P\{\max_{t\in\mathbf{T}} |L(t)| \geq k\} < \varepsilon$ is satisfied for all unlimited k and thus, by underspill (Remark 1.1) also for some limited k_ε. The event $N_\varepsilon = \{\max_{t\in\mathbf{T}} |L(t)| \geq k\}$ has then probability $< \varepsilon$, but $\max_{t\in\mathbf{T}} |L(t)|$ is limited on its complement. □

Let us now fix a random walk L with increment set A and transition probabilities $(p_a)_{a\in A}$ which is not constantly $= 0$.

Definition 9.6. The *drift coefficient* μ_L and the *diffusion coefficient* σ_L of L are defined by

$$\mu_L = \frac{1}{\mathrm{d}t}\sum_{a\in A} a p_a = \frac{1}{\mathrm{d}t} E\left[\mathrm{d}L(0)\right],$$

$$\sigma_L = \left(\frac{1}{\mathrm{d}t} E\left[|\mathrm{d}L(0)|^2\right]\right)^{1/2} = \left(\frac{1}{\mathrm{d}t}\sum_{a\in A} a^2 p_a\right)^{1/2}.$$

The following lemma translates a result by Lindstrøm [49, Lemma 1.2]; its proof can essentially be copied directly from Lindstrøm's original proof:

Lemma 9.7. *For every $s \in \mathbf{T}$,*

$$E\left[L(s)^2\right] = \sigma_L^2 s + \mu_L^2 s(s - \mathrm{d}t).$$

(Cf. Lindstrøm [49, Lemma 1.2]).

Proof. Fix $s \in \mathbf{T}$. Clearly,

$$\begin{aligned} L(s)^2 &= \left(\sum_{t<s}\mathrm{d}L(t)\right)\left(\sum_{u<s}\mathrm{d}L(t)\right) \\ &= \sum_{t<s}|\mathrm{d}L(t)|^2 + 2\sum_{u<t<s}\mathrm{d}L(t)\mathrm{d}L(u). \end{aligned}$$

However, $\mathrm{d}L(t)$ and $\mathrm{d}L(u)$ are independent for all $u < t$ and have the same distribution as $\mathrm{d}L(0)$, so

$$E\left[\mathrm{d}L(t)\mathrm{d}L(u)\right] = E\left[\mathrm{d}L(t)\right]E\left[\mathrm{d}L(u)\right] = E\left[\mathrm{d}L(0)\right]^2 = \mu_L^2(\mathrm{d}t)^2$$

and

$$E\left[|\mathrm{d}L(t)|^2\right] = E\left[|\mathrm{d}L(0)|^2\right] = \sigma_L^2\mathrm{d}t.$$

Therefore,

$$\begin{aligned} E\left[L(s)^2\right] &= \sum_{t<s}\sigma_L^2\mathrm{d}t + 2\sum_{u<t<s}\mu_L^2(\mathrm{d}t)^2 \\ &= \frac{s}{\mathrm{d}t}\sigma_L^2\mathrm{d}t + 2\frac{(s/dt)(s/\mathrm{d}t-1)}{2}\mu_L^2(\mathrm{d}t)^2. \end{aligned}$$

□

9.2 Integrability of Lévy Walks with Limited Increments

We have the following integrability result for Lévy walks with limited increments, which corresponds to Lindstrøm's integrability theorem for hyperfinite Lévy processes [49, Theorem 2.3]. The proof is by and large a translation of Lindstrøm's arguments into the setting of radically elementary probability theory. We will denote by $\mathcal{G}$ the filtration generated by L.

Theorem 9.8. *If L is a Lévy walk with limited increments, then $L(t)$ is L^p for all $t \in \mathbf{T}$ and all limited $p > 0$. (Cf. Lindstrøm [49, Theorem 2.3].)*

Proof of Theorem 9.8. For all $K \in \mathbf{N}$, define

$$\tau_K := \min\{t \in \mathbf{T} \,:\, |L(t)| \geq K\}$$

with the usual convention $\min \varnothing = \infty$. (Then τ_K is a *stopping time*: For all $t \in \mathbf{T}$, the event $\{\tau_K \leq t\}$ is $\mathcal{G}_t$-measurable.) Since L is a Lévy walk, we have $P\{\tau_K = \infty\} \simeq 1$. Therefore,

$$P\left\{\tau_K > \frac{1}{2}\right\} > \frac{1}{2}$$

for all nonstandard K, hence by Underspill in $\mathbf{N}$ (Remark 1.1) even for all sufficiently large standard K, in particular for some K which is $> \max_{a \in A} |a|$. Fix such a K and define

$$\alpha = E\left[\mathrm{e}^{-\tau_K}\right]$$

(with the convention that $\mathrm{e}^{-\infty} = 0$). The choice of K means that $\alpha \ll 1$.

Now define a sequence $(\rho_n)_{n \leq N}$ of stopping times recursively by $\rho_0 = 0$ and

$$\forall n < N \qquad \rho_{n+1} = \min\{t \in \mathbf{T} \,:\, t > \rho_n, \qquad |L(t) - L(\rho_n)| \geq K\},$$

so that $\rho_1 = \tau_K$. From the fact that L has independent and stationary increments, one can show that the random variables $\rho_1 - \rho_0, \ldots, \rho_N - \rho_{N-1}$ are independent and all have the same distribution as $\rho_1 - \rho_0$, that is τ_K. Therefore, for all $n \leq N$,

$$\begin{aligned}
E\left[\mathrm{e}^{-\rho_n}\right] &= E\left[\exp\left(-(\rho_n - \rho_0)\right)\right] = E\left[\exp\left(-\sum_{\ell=0}^{n-1} \rho_{\ell+1} - \rho_\ell\right)\right] \\
&= E\left[\prod_{\ell=0}^{n-1} \exp\left(-(\rho_{\ell+1} - \rho_\ell)\right)\right] \\
&= \prod_{\ell=0}^{n-1} E\left[\exp\left(-(\rho_{\ell+1} - \rho_\ell)\right)\right] = \prod_{\ell=0}^{n-1} E\left[\mathrm{e}^{-\tau_K}\right],
\end{aligned}$$

in other words,

$$E\left[\mathrm{e}^{-\rho_n}\right] = \alpha^n. \tag{9.1}$$

On the other hand, K was chosen such that $K > \max_{a\in A}|a| = \max|\mathrm{d}L(0)|$ and thus (for any $n < N$) the *minimal* $t > \rho_n$ satisfying $|L(t) - L(\rho_n)| \geq K$ must also satisfy $|L(t) - L(\rho_n)| < 2K$ (otherwise there would be a smaller $t > \rho_n$ such that $|L(t) - L(\rho_n)| \geq K$). Hence, for all $n < N$

$$|L(\rho_{n+1}) - L(\rho_n)| < 2K,$$

whence $\{|L(t)| \geq 2nK\} \subseteq \{t > \rho_n\}$ for any $t \in \mathbf{T}$ and $n \leq N$ and therefore

$$\mathrm{e}^{-t}\, P\{|L(t)| \geq 2nK\} \leq \mathrm{e}^{-t}\, P\{\rho_n < t\} \leq E\left[\mathrm{e}^{-\rho_n}\right].$$

By Eq. (9.1), this yields

$$P\{|L(t)| \geq 2nK\} \leq \mathrm{e}^{t}\alpha^n. \tag{9.2}$$

Since $N = 1/\mathrm{d}t$ and K was chosen to be $> \max_{a\in A}|a|$, we have

$$|L(t)| \leq \frac{t}{\mathrm{d}t}\max_{a\in A}|\mathrm{d}L(t)| \leq \frac{1}{\mathrm{d}t}\max_{a\in A}|a| \leq K\,N.$$

Moreover, since K was chosen to be limited and $\alpha \ll 1$, we can find some $\varepsilon \gg 0$ such that $\alpha\mathrm{e}^{2K\varepsilon} \ll 1$. Then, estimate (9.2) allows us to calculate for all $t \in \mathbf{T}$,

$$\begin{aligned}
E\left[\mathrm{e}^{\varepsilon|L(t)|}\right] &\leq \sum_{n<N} E\left[\chi_{\{2nK\leq|L(t)|<2(n+1)K\}}\mathrm{e}^{\varepsilon|L(t)|}\right]\\
&\leq \sum_{n<N} P\{2nK \leq |L(t)| < 2(n+1)K\}\,\mathrm{e}^{2(n+1)K\varepsilon}\\
&\leq \sum_{n<N} P\{2nK \leq |L(t)|\}\,\mathrm{e}^{2(n+1)K\varepsilon}\\
&\leq \sum_{n<N} \mathrm{e}^{t}\alpha^n\mathrm{e}^{2(n+1)K\varepsilon} \leq \mathrm{e}^{t}\mathrm{e}^{2K\varepsilon}\sum_{n<N}\left(\alpha\mathrm{e}^{2K\varepsilon}\right)^n\\
&\leq \mathrm{e}^{t}\mathrm{e}^{2K\varepsilon}\frac{1-\left(\alpha\mathrm{e}^{2K\varepsilon}\right)^N}{1-\alpha\mathrm{e}^{2K\varepsilon}},
\end{aligned}$$

and the right-hand side is limited since ε chosen such that $\alpha\mathrm{e}^{2K\varepsilon} \ll 1$. Hence, we already have $E\left[\mathrm{e}^{\varepsilon|L(t)|}\right] \ll \infty$.

Now, for every $p \in \mathbf{N}$, the Taylor expansion of order p of the exponential function around 0,

$$\mathrm{e}^{\varepsilon|L(t)|} \geq 1 + \frac{\varepsilon^p}{p!}|L(t)|^p.$$

Since we have already seen that $E\left[e^{\varepsilon|L(t)|}\right] \ll \infty$, we may concluded that $E\left[|L(t)|^p\right] \ll \infty$ for all standard $p \in \mathbf{N}$. Therefore, by Remark 2.9, $L(t)$ is $L^q(P)$ for every limited $q > 0$ and every $t \in \mathbf{T}$. □

9.3 Lindstrøm's Characterization of Lévy Walks

We shall now turn to a characterization of Lévy walks which can be used to examine whether a given random walk L is a Lévy walk (Theorem 9.9). This result is a direct translation of a theorem by Lindstrøm [49, Theorem 4.3] into our framework of radically elementary probability theory. Most of the proof and the auxiliary results that prepare it can be directly adapted from Lindstrøm's paper [49], as they only involve combinatorics, some elementary, although non-trivial estimates, and the underspill principle, all of which can be done in **minIST**. The only significant exception is Auxiliary Lemma 9.10 (the analogue of Lindstrøm's [49, Corollary 2.4]) which depends on martingale theory.

In order to state the radically elementary analogue of Lindstrøm's characterization of Lévy walks, let us fix some notation. For all $k \in \mathbf{R}$, it will be helpful to define truncated processes $L^{\leq k}$ and $L^{>k}$ as follows.

$$\forall s \in \mathbf{T} \qquad L^{\leq k}(s) = \sum_{\substack{t<s \\ |\mathrm{d}L(t)|\leq k}} \mathrm{d}L(t),$$

$$L^{>k}(s) = \sum_{\substack{t<s \\ |\mathrm{d}L(t)|>k}} \mathrm{d}L(t) = L(s) - L^{\leq k}(s).$$

Clearly, $L^{\leq k}$ and $L^{>k}$ are random walks. In addition, we define

$$q_k = \frac{1}{\mathrm{d}t} \sum_{\substack{|a|>k \\ a\in A}} p_a.$$

Moreover, for all $m \geq k \in \mathbf{R}$, we define

$$\forall s \in \mathbf{T} \qquad L^{(k,m]}(s) = \sum_{\substack{t<s \\ k<|\mathrm{d}L(t)|\leq m}} \mathrm{d}L(t).$$

Theorem 9.9 (Lindstrøm's characterization of Lévy walks). *L is a Lévy walk if and only if the following conditions are met:*

- *For all limited, yet non-infinitesimal $k \in \mathbf{R}$,*

$$\frac{1}{\mathrm{d}t}\left|\sum_{\substack{|a|\le k\\ a\in A}} a p_a\right| \ll \infty.$$

- *For all limited $k \in \mathbf{R}$,*

$$\frac{1}{\mathrm{d}t}\sum_{\substack{|a|\le k\\ a\in A}} |a|^2 p_a \ll \infty.$$

- *For all non-infinitesimal $\varepsilon > 0$, there is a standard $n \in \mathbf{N}$ such that for every $k \ge n$, $q_k < \varepsilon$. (Equivalently: $q_k \simeq 0$ for all unlimited $k \in \mathbf{R}$.)*

Now we shall give our translation of Lindstrøm's proof into the setting of radically elementary probability theory. Just as in Lindstrøm's original proof, it is helpful to first establish a series of auxiliary statements (cf. Lindstrøm [49, Corollary 2.4; Lemma 3.1; Lemma 3.2; Corollary 3.3; Corollary 4.2]). Only in the proof of Auxiliary Lemma 9.10 will we have to deviate from Lindstrøm's reasoning, as we cannot invoke the martingale theory of Robinsonian nonstandard analysis; unsurprisingly, we will employ Nelson's [60] radically elementary martingale theory instead.

Auxiliary Lemma 9.10. *If L has limited increments, then L is a Lévy walk if and only if both μ_L and σ_L are limited. (Cf. Lindstrøm [49, Corollary 2.4].)*

Proof. If L is a Lévy walk with limited increments, then $L(1)$ is L^2 by Theorem 9.8, whence $E\left[L(1)^2\right]$ is limited by Nelson's radically elementary Radon–Nikodym theorem (Remark 2.2). The formula of Lemma 9.7 implies that both μ_L and σ_L are limited.

Conversely, suppose L has limited increments and both μ_L and σ_L are limited. The process M, defined via

$$\forall s \in \mathbf{T} \qquad M(s) = L(s) - \mu_L s$$

is a martingale with respect to the filtration $\mathcal{G}$ generated by the increments of L. Indeed, since the increments are independent and stationary, $\mathrm{d}L(t)$ is independent of $\mathcal{G}_t$, hence

$$E\left[\mathrm{d}M(t)|\,\mathcal{G}_t\right] = E\left[\mathrm{d}L(t)|\,\mathcal{G}_t\right] - E\left[\mathrm{d}L(0)\right] = E\left[\mathrm{d}L(t)\right] - E\left[\mathrm{d}L(0)\right] = 0$$

for all $t \in \mathbf{T}\setminus\{1\}$. Moreover, the formula of Lemma 9.7 implies that $E\left[M(s)^2\right] = \sigma_L^2 s$, hence $E\left[M(s)^2\right]$ is limited, and so is (by Jensen's inequality) $E\left[|M(s)|\right]$. Therefore, applying the corollary to Nelson's radically elementary martingale inequality (Corollary 2.13), we find that M is a.s. limited. Since μ_L is limited, this implies that L is a.s. limited. □

Auxiliary Lemma 9.11. *If L is a Lévy walk, then for all non-infinitesimal $\varepsilon > 0$, there is a standard $n \in \mathbf{N}$ such that for every $k \geq n$, $q_k < \varepsilon$. (Cf. Lindstrøm [49, Lemma 3.1].)*

Proof. Suppose, for a contradiction, that there were some non-infinitesimal $\varepsilon > 0$ such that $q_k \geq \varepsilon$ for all limited $k \in \mathbf{R}_{>0}$. Then, by Remark 1.1 (underspill/overspill), there exists some unlimited $K \in \mathbf{R}_{>0}$ such that $q_K \geq \varepsilon$. Then, for every non-infinitesimal $s \in \mathbf{T} \setminus \{1\}$, the probability that $|\mathrm{d}L(t)| \leq K$ for all $t < s$ is given by $(1 - q_K \mathrm{d}t)^{s/\mathrm{d}t} \leq (1 - \varepsilon \mathrm{d}t)^{s/\mathrm{d}t}$. However, a Taylor expansion of order $s/\mathrm{d}t$ for the exponential function around 0 yields that

$$(1 - \varepsilon \mathrm{d}t)^{s/\mathrm{d}t} \simeq \exp(-\varepsilon s) \ll 1.$$

Hence, $1 - (1 - q_K \mathrm{d}t)^{s/\mathrm{d}t}$, which is the probability that $|\mathrm{d}L(t)| \geq K$ for some $t < s$, is non-infinitesimal. This, however, contradicts the assumption that L is a Lévy walk. □

Auxiliary Lemma 9.12. *(Cf. Lindstrøm [49, Corollary 3.3].) Suppose for every $\varepsilon > 0$ there exists a standard $n \in \mathbf{N}$ such that for all $m \geq n$, $q_m < \varepsilon$ and $L^{\leq m}$ is a Lévy walk. Then L is a Lévy walk.*

Proof. If L were not a Lévy walk, then by Remark 9.5 would be some nonstandard n such that the probability p, defined by

$$p = P\left\{\max_{t \in \mathbf{T}} |L(t)| > n\right\}$$

is non-infinitesimal. On the other hand,

$$P\left[\bigcap_{t \in \mathbf{T}} \{L^{\leq m}(t) = L(t)\}\right] = (1 - q_m \mathrm{d}t)^{1/\mathrm{d}t},$$

hence by a Taylor expansion of order $N = 1/\mathrm{d}t$ of the exponential function around 0,

$$P\left[\bigcap_{t \in \mathbf{T}} \{L^{\leq m}(t) = L(t)\}\right] \simeq \mathrm{e}^{-q_m}.$$

By assumption on the sequence $(q_m)_{m \in \mathbf{N}}$, for sufficiently large, but standard m, $q_m \ll -\ln(1 - p)$, so

$$P\left[\bigcap_{t \in \mathbf{T}} \{L^{\leq m}(t) = L(t)\}\right] \simeq \mathrm{e}^{-q_m} \gg 1 - p = 1 - P\left\{\max_{t \in \mathbf{T}} |L(t)| > n\right\},$$

whence the event

$$\left\{\max_{t\in\mathbf{T}}|L(t)|>n\right\}\cap\bigcap_{t\in\mathbf{T}}\left\{L^{\leq m}(t)=L(t)\right\}$$

has positive probability. However, this event is actually empty, since it entails that

$$n<|L(t)|\left|L^{\leq m}(t)\right|\leq m<n.$$

Contradiction. □

Auxiliary Lemma 9.13. *If L is a Lévy walk, then for all sufficiently large limited k, the processes $L^{\leq k}$ and $L^{>k}$ are Lévy walks, too. (Cf. Lindstrøm [49, Lemma 3.2].)*

Proof. First, we shall show that $L^{(k,m]}$ is a Lévy walk for all limited m. For, in order for $\max_{t\in\mathbf{T}}\left|L^{(k,m]}(t)\right|$ to be greater in norm than a given unlimited K, the number of times t at which $\left|\mathrm{d}L^{(k,m]}(t)\right|>k$ must be at least K/k, hence nonstandard. However, a simple combinatorial argument shows that for every nonstandard n_0, the probability that the number of such times is greater or equal n_0, is $\simeq 0$: First, note that $\operatorname{card}\left\{t\in\mathbf{T}\setminus\{1\}\ :\ \left|\mathrm{d}L^{(k,m]}(t)\right|>k\right\} = \operatorname{card}\{t\in\mathbf{T}\setminus\{1\}\ :\ |\mathrm{d}L(t)|>k\}$, whence

$$\begin{aligned}&P\left\{\operatorname{card}\{t\in\mathbf{T}\setminus\{1\}\ :\ |\mathrm{d}L(t)|>k\}=n\right\}\\&=\binom{1/\mathrm{d}t}{n}(1-q_k\mathrm{d}t)^{1/\mathrm{d}t-n}(q_k\mathrm{d}t)^n.\end{aligned}$$

On the other hand, by a Taylor expansion of order $N=1/\mathrm{d}t$ of the exponential function around 0,

$$\binom{1/\mathrm{d}t}{n}(1-q_k\mathrm{d}t)^{1/\mathrm{d}t-n}(q_k\mathrm{d}t)^n\simeq\frac{1}{(\mathrm{d}t)^n}\frac{1}{n!}\mathrm{e}^{-q_k}(q_k\mathrm{d}t)^n\simeq\mathrm{e}^{-q_k}\frac{1}{n!}{q_k}^n$$

for all standard n. Therefore, for all standard n_0,

$$P\left\{\operatorname{card}\{t\in\mathbf{T}\setminus\{1\}\ :\ |\mathrm{d}L(t)|>k\}\geq n_0\right\}\simeq\mathrm{e}^{-q_k}\sum_{n=0}^{n_0}\frac{1}{n!}{q_k}^n,$$

and the right-hand side is $\simeq 1$ for nonstandard n_0. By our initial deliberations, the probability of $\max_{t\in\mathbf{T}}\left|L^{(k,m]}(t)\right|$ being greater in norm than a given unlimited K is $\simeq 0$. Therefore, $L^{(k,m]}$ is a Lévy walk for all limited m.

Applying Auxiliary Lemma 9.12 and Auxiliary Lemma 9.11, we conclude that $L^{>k}$ is a Lévy walk. Since $L^{\leq k}=L-L^{>k}$ and the difference of two Lévy walks is itself a Lévy walk, it follows that $L^{\leq k}$ is a Lévy walk, too. □

Auxiliary Lemma 9.14. *If L is a Lévy walk, then for every non-infinitesimal and limited $k \in \mathbf{R}$, the processes $L^{\leq k}$ and $L^{>k}$ are Lévy walks (Cf. Lindstrøm [49, Corollary 4.2])*

Proof. With exactly the same arguments as in the proof of Auxiliary Lemma 9.13, we can prove that $L^{>k}$ is a Lévy walk. In order to perform this argument, however, we need that q_k is limited for non-infinitesimal k. We will prove this presently: one only has to note that $q_m < 1$ for sufficiently large standard m and thus, using Auxiliary Lemma 9.10 and the fact that $L^{\leq m}$ is a Lévy walk (from Auxiliary Lemma 9.13), we already have

$$q_k \leq 1 + \frac{1}{\mathrm{d}t} \sum_{k<|a|\leq m} p_a \leq 1 + \frac{1}{\mathrm{d}t}\frac{1}{k^2} \sum_{k<|a|\leq m} |a|^2 p_a \leq 1 + \frac{1}{k^2}\sigma^2_{L^{\leq m}} \ll \infty.$$

Since $L^{\leq k} = L - L^{>k}$ and the difference of two Lévy walks is itself a Lévy walk, it follows that $L^{\leq k}$ also is a Lévy walk. □

Proof of Theorem 9.9. If L satisfies the first two conditions of the Theorem, then for every non-infinitesimal limited k, $\mu_{L^{\leq k}}$ and $\sigma_{L^{\leq k}}$ are limited, whence $L^{\leq k}$ is a Lévy walk by Auxiliary Lemma 9.10. From here, the third condition in the Theorem can be combined with Auxiliary Lemma 9.12 to establish that L itself is a Lévy walk.

Conversely, if L is a Lévy walk, then Auxiliary Lemma 9.11 says that the third condition is satisfied. Consider now a non-infinitesimal, limited k. By Auxiliary Lemma 9.14, $L^{\leq k}$ is a Lévy walk with limited increments, hence both $\mu_{L^{\leq k}}$ and $\sigma_{L^{\leq k}}$ are limited by Auxiliary Lemma 9.10. Since $\sigma_{L^{\leq k}}$ is increasing in k, we conclude that $\mu_{L^{\leq k}}$ is limited for all non-infinitesimal, limited k (the first condition in the Theorem) and that $\sigma_{L^{\leq k}}$ for all limited k (the second condition in the Theorem). □

9.4 A Radically Elementary Itô–Doeblin Formula for Lévy Walks with Limited-Variation Jump Part

Our goal in this chapter is to prove a version of the Itô–Doeblin formula in a radically elementary setting. For this purpose, we restrict ourselves to the case of those Lévy walks which do not have arbitrarily small jumps and use some kind of jump-diffusion decomposition, motivated by Lindstrøm's approach [49].[1]

Definition 9.15. An element $a \in A$ is called a *jump* of L if and only if both $p_a \neq 0$ and $|a|/\sqrt{\mathrm{d}t}$ is unlimited. L is said to be a (pure) *jump process* if and only if every $a \in A$ with $p_a \neq 0$ is a jump. A jump a is called *observable* if and only if $|a| \not\simeq 0$.

[1]In the classical setting, these processes correspond to Lévy walks whose Lévy measure is concentrated on a set that is bounded from below in norm.

L is said to be an *observable jump process* if and only if every $a \in A$ with $p_a \neq 0$ is an observable jump. If L is a Lévy walk and a jump process (observable jump process), it is called a *jump Lévy walk* (observable-jump Lévy walk).

Remark 9.16. If all jumps of L are observable, then there exists some $\eta \gg 0$ such that $|a|/\sqrt{\mathrm{d}t}$ is limited for each $a \in A \cap \mathbf{R}_{\leq \eta}$ with $p_a \neq 0$. We define recursively a new process L^- by

$$L^-(0) = L(0), \qquad \forall t \in \mathbf{T} \setminus \{1\} \quad L^-(t + \mathrm{d}t) = L^-(t) + \mathrm{d}L(s)\chi_{\{|\mathrm{d}L(s)| \leq \eta\}}.$$

Proof. By assumption, for every both $p_a \neq 0$, if $\frac{|a|}{\sqrt{\mathrm{d}t}}$ is unlimited, then already $|a| \not\simeq 0$. Thus, the set

$$\left\{k \in \mathbf{N} \,:\, \forall a \in A \quad \left(\frac{|a|}{\sqrt{\mathrm{d}t}} \geq k \Rightarrow |a| > \frac{1}{k}\right)\right\}$$

contains all nonstandard natural numbers. Therefore, by underspill (see Remark1.1), it must contain a standard natural number n. Put $\eta := \frac{1}{n} \gg 0$. Then

$$\forall a \in A \quad \left(\frac{|a|}{\sqrt{\mathrm{d}t}} \geq \frac{1}{\eta} \Rightarrow |a| > \eta\right),$$

hence

$$\forall a \in A \quad \left(|a| \leq \eta \Rightarrow \frac{|a|}{\sqrt{\mathrm{d}t}} < \frac{1}{\eta}\right).$$

□

First, we note that if L has no jumps, then it is a multiple of the Wiener process plus constant drift.

Lemma 9.17. *Suppose L is a Lévy walk which has no jumps and $\sigma_L \not\simeq 0$. Then the process M, defined by $M(t) = \frac{L(t)-\mu_L t}{\sigma_L}$ is a normalized martingale and even a Wiener process.*

Proof. First, since L is not constantly $= 0$, we have $\sigma_L \neq 0$, whence M is well-defined. Since L has independent increments, one can easily check that M is a martingale. A simple calculation shows that M is even a normalized martingale and that $M(0) = 0$. Since L is a Lévy walk without jumps, Theorem 9.9 says that both μ_L and σ_L^2 are limited. Also, since L has no jumps, there is a limited C such that $|\mathrm{d}L(t)| \leq C\sqrt{\mathrm{d}t} \simeq 0$ for all $t \in \mathbf{T}$. Now, since $\sigma_L \not\simeq 0$, we have $\sqrt{\mathrm{d}t}/\sigma_L \simeq 0$, so

$$|\mathrm{d}M(t)| \leq \frac{C\sqrt{\mathrm{d}t} + \mu_L \mathrm{d}t}{\sigma_L} \leq 2C\sqrt{\mathrm{d}t}/\sigma_L \simeq 0.$$

Thus, M has no jumps and $\mathrm{d}M(t) = \mathrm{d}M(t)\chi_{\{|\mathrm{d}M(t)| \leq \varepsilon\}}$ for all $t \in \mathbf{T}$ and $\varepsilon \gg 0$. This entails that M satisfies the (near) Lindeberg condition [60, p. 57].

By Nelson's unified "de Moivre–Laplace–Lindeberg–Feller–Wiener–Lévy–Doob–Erdős–Kac–Donsker–Prokhorov theorem" (see Remark 3.13), we find that M must be a Wiener process. □

If $\sigma_L \simeq 0$, then Lemma 9.7 says that

$$E\left[|L(t)-\mu_L t|^2\right] = \sigma_L^2 t + \mu_L^2 t(t-\mathrm{d}t) - 2\mu_L t \underbrace{E\left[L(t)\right]}_{=\frac{t}{\mathrm{d}t}E[\mathrm{d}L(0)]=\mu_L t} + \mu_L^2 t^2$$
$$\simeq \sigma_L^2 t \simeq 0,$$

therefore $L(t) \simeq \mu_L t$ a.s. for all $t \in \mathbf{T}$.

For the remainder of this chapter, we will only consider Lévy walks which are of the form $L(t) = \mu_L t + \sigma_L W(t) + J(t)$, wherein J is a random walk whose increment set consists of jumps only.

We will now prove a radically elementary Itô–Doeblin formula for Lévy walks with observable jump part. The proof idea is borrowed from a relatively recent paper [33] which—using Lindstrøm's theory of hyperfinite Lévy processes—gives a simple nonstandard proof of the Itô–Doeblin formula for Lévy walks with finite-variation jump part.

Theorem 9.18 (Itô–Doeblin formula for Lévy walks). *Suppose $L(t) = \mu t + \sigma W(t) + J(t)$ for all $t \in \mathbf{T}$ for some Wiener walk W, some observable jump Lévy walk J and limited μ, σ, and let f be a thrice continuously differentiable function. Consider some ω such that $f''(L)$ and $f'''(L)$ have limited ω-trajectories. Then one has for all $s \in \mathbf{T}$, suppressing ω,*

$$\begin{aligned} &f(L(s)) - f(L(0)) \\ &\simeq \int_0^s f'(L^-(t))\mathrm{d}L(t) + \frac{\sigma_L^2}{2}\int_0^s f''(L^-(t))\mathrm{d}t \\ &\quad + \sum_{t<s} f(L(t)) - f(L^-(t)) - (L(t) - L^-(t))\, f'(L^-(t))). \end{aligned}$$

The proof needs the following result, a corollary to Theorem 9.9:

Corollary 9.19. *Define processes L_+ and L_- by*

$$\forall s \in \mathbf{T} \qquad L_+(s) = \sum_{t<s} \mathrm{d}L(t)\chi_{\{\mathrm{d}L(t)\geq 0\}},$$
$$L_-(s) = \sum_{t<s} \mathrm{d}L(t)\chi_{\{\mathrm{d}L(t)<0\}}.$$

If L is a Lévy walk, then L_+ and L_- are Lévy walks.

Proof of Corollary 9.19. Clearly, L_+ and L_- are random walks with increment sets $A_+ = A \cap \mathbf{R}_{>0} \cup \{0\}$ and $A_- \cap \mathbf{R}_{<0} \cup \{0\}$, respectively and transition functions $(p_{a,+})_{a\in A+}$ and $(p_{a,-})_{a\in A_-}$, respectively, wherein $p_{a,+} = p_a$ for all $a \in A_+ \setminus \{0\}$ and $p_{a,-} = p_a$ for all $a \in A_- \setminus \{0\}$. Clearly, L_+ and L_- satisfy the conditions of Theorem 9.9 and thus are Lévy walks. □

L^- is the analogue of a pathwise left limit.

Proof of Theorem 9.18. By Remark 9.16, there exists some $\eta \gg 0$ such that $|a|/\sqrt{\mathrm{d}t}$ is limited for each $a \in A \cap \mathbf{R}_{\leq\eta}$. We now define recursively a sequence of stopping times as follows:

$$\tau_0 := 0, \qquad \tau_n := \min\{t \in \mathbf{T} \ : \ t > \tau_{n-1}, \quad |\mathrm{d}L(t - \mathrm{d}t)| \geq \eta\} \wedge 1.$$

Exploiting that $\mathbf{T}$ has finite cardinality, we may define an $\mathbf{N}$-valued random variable $M : \Omega \to \mathbf{N}$ by

$$M := \operatorname{card}\{t \in \mathbf{T} \ : \ |\mathrm{d}L(t)| \geq \eta\}.$$

Next we show that for all nonstandard $k \in \mathbf{N}$,

$$P\{M \geq k\} \simeq 0. \tag{9.3}$$

Indeed, if $M(\omega) \geq k$ for any k, then either $\max_{t\in\mathbf{T}} |L_+(t)(\omega)| \geq \frac{\eta}{2}k$ or $\max_{t\in\mathbf{T}} |L_-(t)(\omega)| \geq \frac{\eta}{2}k$, hence either $L_+(1)(\omega) \geq \frac{\eta}{2}k$ or $L_-(1)(\omega) \leq -\frac{\eta}{2}k$. However, L_+ and L_- are Lévy walks by Corollary 9.19, whence $P\left\{L_+(1) \geq \frac{\eta}{2}k\right\} \simeq 0$ and $P\left\{L_-(1) \leq -\frac{\eta}{2}k\right\} \simeq 0$ for all unlimited k. Therefore,

$$P\{M \geq k\} \leq P\left[\left\{L_+(1) \geq \frac{\eta}{2}k\right\} \cup \left\{L_-(1) \leq -\frac{\eta}{2}k\right\}\right] \simeq 0,$$

which establishes the approximate equation (9.3).

Hence, for any $\varepsilon \gg 0$, the set $\{k \in \mathbf{N} \ : \ P\{M \geq k\} \leq \varepsilon\}$ contains all nonstandard $k \in \mathbf{N}$, and thus by underspill (see Remark 1.1) must contain some standard k_ε, such that $P\{M \geq k_\varepsilon\} \leq \varepsilon$. Therefore, M is limited a.s. For this reason, the set $\left\{\tau_0(\omega), \ldots, \tau_{M(\omega)}(\omega)\right\}$ is limited for almost every ω. Moreover, clearly $\tau_M = 1$.

By assumption on L, $A \cap [-\varepsilon, \varepsilon] \cup \{0\} = \left\{0, \mu\mathrm{d}t, \sigma\sqrt{\mathrm{d}t}, -\sigma\sqrt{\mathrm{d}t}\right\}$ for any $\varepsilon \gg 0$. Combining this fact with the choice of $\eta \gg 0$, we find that for all $t \in \mathbf{T}$ with $\tau_i < t < \tau_{i+1}$, we must have $L(t) = L(\tau_i) + \mu t + \sigma W(t)$ and hence by Itô's formula (Lemma 3.9)

$$\begin{aligned}\sum_{\tau_i < t < \tau_{i+1}} \mathrm{d}f(L(t)) &\simeq \sum_{\tau_i < t < \tau_{i+1}} f'(L(t))\,\mathrm{d}L(t) + \frac{\sigma_L^2}{2} f''(L(t))\,\mathrm{d}t \\ &\simeq \sum_{\tau_i < t < \tau_{i+1}} f'(L^-(t))\,\mathrm{d}L^-(t) + \frac{\sigma_L^2}{2} f''(L^-(t))\,\mathrm{d}t\end{aligned}$$

$$\simeq \sum_{\tau_i \le t < \tau_{i+1}} f'\left(L^-(t)\right) \mathrm{d}L^-(t) + \frac{\sigma_L^2}{2} f''\left(L^-(t)\right) \mathrm{d}t.$$

Therefore, using $\tau_M = 1$,

$$\begin{aligned}
f\left(L(s)\right) - f\left(L(0)\right) &= \sum_{t<s} \mathrm{d}f\left(L(t)\right) \\
&= \sum_{i=0}^{M-1} \sum_{\tau_i < t < \tau_{i+1}} \mathrm{d}f\left(L(t)\right) + \sum_{i=0}^{M} \mathrm{d}f\left(L(\tau_i)\right) \\
&\simeq \sum_{i=0}^{M-1} \sum_{\tau_i \le t < \tau_{i+1}} f'\left(L^-(t)\right) \mathrm{d}L^-(t) + \frac{\sigma_L^2}{2} f''\left(L^-(t)\right) \mathrm{d}t \\
&\quad + \sum_{i=0}^{M} f\left(L(\tau_i)\right) - f\left(L^-(\tau_i)\right) \\
&= \sum_{0 \le t < \tau_M} f'\left(L^-(t)\right) \mathrm{d}L^-(t) + \frac{\sigma_L^2}{2} f''(L^-(t))\mathrm{d}t \\
&\quad + \sum_{i=0}^{M} f(L(\tau_i)) - f(L^-(\tau_i)) \\
&= \sum_{0 \le t < 1} f'\left(L^-(t)\right) \mathrm{d}L^-(t) + \frac{\sigma_L^2}{2} f''\left(L^-(t)\right) \mathrm{d}t \\
&\quad + \sum_{t<s} f\left(L(t)\right) - f\left(L^-(t)\right) \\
&= \sum_{0 \le t < 1} f'\left(L^-(t)\right) \mathrm{d}L(t) + \frac{\sigma_L^2}{2} f''\left(L^-(t)\right) \mathrm{d}t \\
&\quad + \sum_{t<s} f\left(L(t)\right) - f\left(L^-(t)\right) + \sum_{t<s} \left(L(t) - L^-(t)\right) f'(L^-(t)).
\end{aligned}$$

□

9.5 A Brief Look at Lévy Finance

Since the mid-1980s and in particular during the last years of the twentieth century, Lévy processes have gained great popularity among (academic) researchers in mathematical finance, the reason being that they have a number of mathematically very pleasant properties (including a beautiful parametrization, through the *Lévy–Khintchine formula*), but on the other hand constitute a relatively rich class of stochastic processes which include examples of pure-jump processes, of diffusions and of jump-diffusions. In the field now known as *Lévy finance*, risky asset price processes are modelled, after appropriate scaling, as exponential Lévy processes.

The earliest and most popular models in Lévy finance take simple Lévy processes other than the Wiener process to model logarithmic asset price processes. Historically, the first example of such a model was proposed by Nobel laureate Robert C. Merton [57] as early as 1976, only three years after the publication of Fischer Black's and Myron S. Scholes' famous paper [18] and his own seminal work on the "theory of rational option pricing" [56]. Lévy finance as a research area in its own right did, however, not come into existence until the circulation of the papers by Madan and Seneta [53–55].

Merton's [57] jump-diffusion model—the earliest and most well-known model of Lévy finance—assumes that after scaling, the price process X of a risky asset can be modelled as

$$X(t) = \exp\left(\mu t + \sigma W(t) + \lambda \zeta(t)\right)$$

for all t, where W and ζ are Wiener and Poisson processes, respectively, and σ, μ, λ are real constants ($\sigma > 0$).

It is an easy exercise in mathematical finance to verify that given any interest rate r, there will be infinitely (indeed continuum-many) equivalent martingale measures for the discounted price process $(\mathrm{e}^{-rt}X(t))_{t\geq 0}$ associated with Merton's jump-diffusion model. Thus, in light of the First and Second Fundamental Theorems of Asset Pricing (for radically elementary versions, see Theorem 5.4, Theorem 5.6 and Remark 5.8), we not only have in Merton's model (i) the absence of well-behaved free lunches with vanishing risk, but also (ii) *market incompleteness* in the sense that the marketed space does not contain all conceivable contingent claims. This property of incompleteness is shared by all exponential Lévy-process models in mathematical finance—except for the Black–Scholes model, of course.

On the one hand, this feature of Lévy market models makes them quite attractive from an empirical point of view, because they provide mathematically beautiful and analytically (fairly) tractable models of phenomena (market incompleteness, jumps, non-Gaussian distributions) which one observes on real-world financial markets. On the other hand, incomplete models have the disadvantage that they no longer produce—at least not in a straightforward manner—strategies for hedging derivatives. Incomplete financial markets call for much more sophisticated approaches to hedging than just replication of contingent claims—but this is "merely" a practical problem and should in many contexts not affect modelling choices.

For an overview of the subject and more references, one may refer to the volume edited by Barndorff-Nielsen et al. [8], to Applebaum's survey article on Lévy processes [5] or to textbooks such as Boyarchenko and Levendorskiĭ [20], Schoutens [73] and Applebaum [6].

Very recently, a fundamental concern about Lévy financial market models has surfaced, which is that they might—except for the Black–Scholes model—lack a foundation in terms of general equilibrium theory (cf. [34]). The mathematical treatment of this question, by the way, relies heavily on (Robinsonian) nonstandard analysis in general and on Loeb probability theory in particular. (See Appendix B for a brief discussion of the relation of this book to Robinsonian nonstandard analysis.)

Chapter 10
Final Remarks

Nelson [60, Appendix] has proved that in a rigorous sense the concepts of radically elementary probability theory are equivalent to the concepts of the classical theory of stochastic processes. For this reason, the radically elementary approach to stochastic calculus as presented in the present work has the same scientific content as the usual approach to the subject.

As we have seen, radically elementary probability theory allows for elementary proofs of many results in stochastic analysis, including Itô's formula, Girsanov's theorem, the Feynman–Kac formula, and even stochastic calculus for Lévy walks with finite-variation jump part. The only prerequisites to teach basic stochastic analysis in this framework are finite probability theory, basic real analysis, and the fact that the Peano axioms do not completely characterize the natural numbers. For this reason, the radically elementary approach to stochastic analysis seems ideally suited for introductory courses on stochastic calculus in the mathematics curricula of quantitative finance, engineering, and physics programmes.

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7_10,

Appendix A
Excursion to Logic: Some Remarks on the Metamathematics of Minimal Internal Set Theory

A.1 An Alternative Road to Minimal Internal Set Theory

When we introduced Minimal Internal Set Theory in Chap. 1, we have tacitly assumed that most readers of this book will find it more intuitive to conceive of Minimal Internal Set Theory as an axiom system which describes an *extended universe*. Some readers, however, might be more comfortable with the idea that an "appropriate" mathematical model of the real numbers should contain infinitesimals and infinitely large numbers anyway. On this account, it would be more intuitive to simply *extend the language* of conventional mathematics by a new predicate, e.g. "... is a standard natural number", and impose additional axioms regulating the use of this predicate—in order to allow for a consistent and fruitful use of infinitesimals.

Of course, the choice of the axioms requires care, as the resulting axiom system should be consistent,[1] simple and powerful enough to permit a productive use of these axioms for infinitesimal calculus. In order to motivate our choice of an axiom system (which is inherited from Nelson's *Radically Elementary Probability Theory* [60]), we could have pointed to the relatively well-known fact that the Peano axiomatization of the natural numbers does not characterize the set of natural numbers completely.[2] For example, any model of the Peano axioms can be elementarily embedded as a proper subset into some other model of the Peano axioms. This observation already suffices to motivate the consistency of axiom systems with a modified principle of mathematical induction for the standard natural numbers. The axiom system in Nelson's *Radically elementary probability theory* [60] is exactly of such a kind.

[1] At least relative to the consistency of conventional mathematics, which because of Gödel's second incompleteness theorem [28] admits no consistency proof.

[2] More precisely, indeed, by Gödel's first incompleteness theorem [28], no extension of the Peano axioms could provide such a unique characterization up to isomorphism.

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7,

On this syntactic account of Minimal Internal Set Theory, the presentation of the axiom system only needs to be prefaced (as in Nelson's monograph [60]) by observing that the language of conventional mathematics does not use the word "standard", whence one may without hesitation introduce a new unary predicate for natural numbers with that name, i.e. a predicate of the form ". . . is a standard natural number". Having thus extended the language of conventional mathematics,[3] all that is left to do is to specify rules that govern the use of that predicate.

On this account, one may note[4] that the introduction of Minimal Internal Set Theory did not *per se* involve the addition of any new mathematical objects (be it atoms or sets). One may take the view that the universe of mathematical objects has remained the same, and only the language has been extended—by adjoining a new predicate which allows us to distinguish between standard natural numbers and nonstandard natural numbers. As one can gather from the axioms of Minimal Internal Set Theory and the fact (provable by External Induction, see below) that any nonstandard natural number is greater than every standard natural number, the correct interpretation of "standard" is "not extremely large".

Readers with an interest in the foundations of mathematics will observe that (i) the axiom system **minIST** would obviously be inconsistent if the Peano axioms characterized the natural numbers completely, and (ii) conversely, the incompleteness of the Peano axioms readily suggests that the axiom system **minIST** is consistent. In any case, it can be rigorously shown that **minIST** only proves those internal formulae can already be proved in conventional mathematics: **minIST** is a conservative extension of conventional mathematics and thus—in light of *ex falso quodlibet*—(relatively) consistent. The reason for the conservativity of **minIST** lies in the fact that it can be seen as a subsystem of Nelson's [59] (cf. Nelson [60, Appendix, p. 80]) which itself is a conservative extension of conventional mathematics.

We close this section with a few more technical comments on the axioms of **minIST**. First, the term "conventional mathematics" in the first axiom of Minimal Internal Set Theory is, of course, context-dependent; at present, most mathematicians would understand the term "conventional mathematics" to refer to Zermelo–Fraenkel set theory plus the Axiom of Choice (**ZFC**). In the following, we will side with the majority and view Minimal Internal Set Theory as an extension of **ZFC** by definition. We note, however, that radically elementary probability theory and radically elementary stochastic analysis certainly do not use **ZFC** to its fullest strength. Therefore, they might continue to be acceptable even when the consistency of **ZFC** should some day be subject to considerable doubt. (Edward Nelson for instance is less than convinced that Peano Arithmetic is consistent [63].)

---

[3] In fact, as Cantor, Frege, Russell, Whitehead and others had shown by the early 1900s, all of conventional mathematics may be reduced to set theory, so "the language of conventional mathematics" comes down to all that can be expressed with the $\in$-relation.

[4] This applies especially to those readers who already have come into loose contact with nonstandard analysis.

The additional axioms beyond **ZFC** are theorems of Nelson's [59] Internal Set Theory (cf. Nelson [60, Appendix, p. 80]), which itself is a conservative extension of **ZFC** (Nelson [59, Theorem 8.8, in part due to William C. Powell]) and thus consistent relative to **ZFC**. Hence, **minIST**$^+$ also is consistent relative to **ZFC** and every internal theorem of **minIST**$^+$ can also be proved in **ZFC**. It might be possible to develop Nelson's [60] radically elementary probability theory, at least partially, even when one replaces **ZFC** in our definitions of **minIST**$^+$ or **minIST** (or the even weaker system **minIST**$^-$ of Appendix A) by a weaker set-theoretic axiom system. This would be an interesting question for future research.[5]

A.2 A Simple Relative Consistency Proof for a Substantial Subsystem of minIST

Nelson [60, Appendix, p. 80] has shown, invoking the saturation principle of Internal Set Theory (cf. Nelson [61]), that the axioms of **minIST**$^+$ follow from **IST**, and since Nelson has also shown that **IST** is a conservative extension of **ZFC** [59, Theorem 8.8, in part due to William C. Powell], it follows that so is **minIST**$^+$.

The proof of the fact that **IST** is a conservative extension of **ZFC**, however, is a sophisticated argument using so-called adequate ultrapowers and ultralimits. For pedagogical reasons, one would wish to find a simple proof at least for the consistency of some subsystem of **minIST**$^+$ in which a substantial part of radically elementary probability theory can be developed. This is what we will now aim at. Consider the subsystem, henceforth denoted **minIST**$^-$, of **minIST** which one obtains through replacing the External Induction principle by the following two axioms:

- (**Unlimitedness of nonstandard numbers**) If $n \in \mathbf{N}$ is nonstandard, then $n > k$ for all standard $k \in \mathbf{N}$.
- (**Standard Induction**) Let $A(n)$ be a formula which is of the form $Q_1^{\mathbf{st}} v_1 \dots Q_m^{\mathbf{st}} v_m \varphi(p_1, \dots, p_\ell, v_1, \dots, v_m, n)$, wherein $Q_i^{\mathbf{st}} v_i$ is a quantification either of the form "for all standard $v_i \in \mathbf{N}$" (abbreviated $\forall^{\mathrm{st}} v_i$) or of the form "there exists a standard $v_i \in \mathbf{N}$" (abbreviated $\exists^{\mathrm{st}} v_i$), $p_1, \dots, p_\ell$ are standard natural numbers and φ is a formula of set theory with $\ell + m + 1$ free variables (and no parameters). Assume that $A(0)$ holds and that $A(n)$ entails $A(n+1)$ for all standard n. Then $A(n)$ readily holds for all standard n.

Note that the most important proof principle of **minIST**, viz. the underspill/overspill principle (Remark 1.1), still holds in **minIST**$^-$. For example, in order to prove that there is no set which consists only of the standard natural numbers,

[5] In fact, Henson and Keisler [30] have shown that adding nonstandard elements to certain relatively weak axiom systems of set theory may result in a stronger, i.e. non-conservative extension of the original weak axiom system.

we only have to remark that if there were such a set, one could prove by internal induction (i.e. induction in **N**, which of course holds in **minIST**$^-$ as it extends **ZFC**) that this set is the whole of **N**, contradicting the existence of nonstandard natural numbers (which continues to hold in **minIST**$^-$).

Furthermore, the External Induction principle can be replaced by the Standard Induction principle in proving a number of basic results of radically elementary mathematics. We give some examples for results which Nelson [60, p. 17] proves with the External Induction principle and which can also be proved in **minIST**$^-$ through the Standard Induction principle.

Lemma A.1 (minIST$^-$).

(1) If m and n are standard natural numbers, then so is $m + n$.
(2) If m and n are standard natural numbers, then so is mn.
(3) If n is a standard natural number and $a > 0$ is limited, then a^n is limited.
(4) For all $n \in \mathbf{N}$, n is standard if and only if it is limited.
(5) If x is infinitesimal and y is limited, then xy is infinitesimal.
(6) If $x \simeq y$ and $y \simeq z$, then $x \simeq z$.
(7) Let $n \in \mathbf{N}$ be standard and $(x_i)_{i<n}, (y_i)_{i<n} \in \mathbf{R}^n$. If $x_i \simeq y_i$ for all $i < n$, then $\sum_{i<n} x_i \simeq \sum_{i<n} y_i$.

Proof. (1) Let $m \in \mathbf{N}$ be standard. An inspection of the definition of ordinal addition and the proof of the ordinal recursion theorem shows that there exists a formula of set theory, denoted $\psi_+(m, n, k)$, whose only parameters are m, n, k and such that for all $m, n, k \in \mathbf{N}$,

$$m + n = k \Leftrightarrow \psi_+(m, n, k).$$

Let us hence apply Standard Induction to the formula

$$\exists^{\mathrm{st}} k \qquad m + n = k.$$

The base step of the induction is tautological. For the induction step, it suffices to remark that if $m + n$ is standard (induction hypothesis), then $m + n + 1$ is standard.

(2) Let $m \in \mathbf{N}$ be standard. Again, an inspection of the definition of ordinal addition and the proof of the ordinal recursion theorem shows that there exists a formula of set theory $\psi_\times(m, n, k)$ whose only parameters are m, n, k and such that for all $m, n, k \in \mathbf{N}$,

$$mn = k \Leftrightarrow \psi_\times(m, n, k).$$

We apply Standard Induction to the formula

$$\exists^{\mathrm{st}} k \qquad mn = k.$$

Base step: $m0 = 0$ is standard. Induction step: Suppose mn is standard (induction hypothesis). Then $m(n+1) = mn + n$ is the sum of two standard numbers and thus itself standard by part 1 of the present lemma.

(3) Of course, $a^n > 0$. Since a is limited, there exists some standard $m \in \mathbf{N}$ such that $a < m$. It is enough to verify the formula

$$\exists^{\mathrm{st}} k \qquad m^n = k.$$

An inspection of the definition of ordinal addition and the proof of the ordinal recursion theorem shows that there exists a formula of set theory $\psi_{\exp}(m, n, k)$ whose only parameters are m, n, k and such that for all $m, n, k \in \mathbf{N}$,

$$m^n = k \Leftrightarrow \psi_{\exp}(m, n, k).$$

Hence, we may apply Standard Induction to prove that $\exists^{\mathrm{st}} k \quad m^n = k$. Base step: $m^0 = 1$ is standard. Induction step: Suppose there is a standard k such that $m^n = k$ (induction hypothesis). Then $m^{n+1} = km$, which is the product of two standard numbers and thus itself standard by part 2 of the present lemma.

(4) If n is standard, then obviously limited (by the trivial estimate $n \leq n$). The converse follows from the unlimitedness of nonstandard numbers, an axiom of **minIST**$^-$.

(5) Fix a standard $m \in \mathbf{N}$. We have to prove $|xy| \leq 1/m$. Choose a standard $n \in \mathbf{N}$ such that $|y| \leq n$. By part 2 of the present lemma, mn is standard, whence

$$|xy| = |x|\,|y| \leq \frac{1}{mn} n \leq \frac{1}{m}.$$

(6) Fix a standard $m \in \mathbf{N}$. We have to prove $|x - z| \leq 1/m$. By part 2 of the present lemma, $2m$ is standard (as $2 = 0 + 1 + 1$ is standard), whence

$$|x - z| \leq |x - y| + |y - z| \leq \frac{1}{2m} + \frac{1}{2m} = \frac{1}{m}.$$

(7) Fix a standard $m \in \mathbf{N}$. We need to prove $\left|\sum_{i<n}(x_i - y_i)\right| \leq \frac{1}{m}$. However, mn is standard (by part 2 of the present lemma), so

$$\left|\sum_{i<n}(x_i - y_i)\right| \leq \sum_{i<n}|x_i - y_i| \leq \sum_{i<n}\frac{1}{mn} = \frac{1}{m}.$$

□

An advantage of **minIST** over **minIST**$^-$ is that its axioms are simpler and shorter to formulate; what speaks for **minIST**$^-$ is that it admits a short proof of its relative consistency.

Theorem A.2. *The axiom system* ***minIST***$^-$ *is a conservative extension of* ***ZFC***.

As an immediate corollary, **minIST**$^-$ is consistent relative to **ZFC**.

Proof. Let ψ be a formula of set theory which is not provable in **ZFC**. We shall construct a model ${}^*V$ of **minIST**$^-$ in which ψ fails. By the compactness theorem, let V be a set-size, transitive model of **ZFC**, called *ground model*, which models $\neg\psi$. Let $\mathbf{N}^V$ be the set of natural numbers as recognized by V, and let $\in^V$ denote the element-relation as recognized by V. Let I be an infinite set and let $\mathcal{U}$ be a non-principal ultrafilter on I. Consider the ultrapower ${}^*V = V^I/\mathcal{U}$, into which V can be canonically embedded, through ${}^* : v \mapsto [(v)_{i\in I}]_{\mathcal{U}}$. By Łoś's theorem, this is an elementary embedding: ${}^* : V \prec {}^*V$.

Let ${}^*\mathbf{N}$ be the set of natural numbers as recognized by ${}^*V$, and let ${}^*\in^V$ denote the element-relation as recognized by ${}^*V$. Call an element n of ${}^*\mathbf{N}$ *standard* (denoted $\mathbf{st}(n)$) if and only if it is of the form ${}^*n_0$ for some $n_0 \in \mathbf{N}^V$.

We now have to prove that $\left({}^*V, {}^*\in^V, \mathbf{st}\right)$ is a model of **minIST**$^-$ and of $\neg\psi$. Indeed, ${}^*V$ is a model of **ZFC** and of $\neg\psi$ since $V \prec {}^*V$. Moreover,

$$0^{*V} = \varnothing^{*V} = {}^*\varnothing = {}^*0$$

and for all $n_0 \in \mathbf{N}^V$, one has

$${}^*n_0 {}^*{+} {}^*1 = {}^*(n_0 + 1).$$

Therefore, 0^{*V} is standard and for every standard n, $n {}^*{+} 1$ is standard, too.

Consider next some $k \in {}^*\mathbf{N}$ with $k {}^*{\leq} n$ for some standard $n = {}^*n_0$. Let $k = [(k_i)_{i\in I}]_{\mathcal{U}}$, then $\{i \in I \ : \ k_i \leq n_0\} \in \mathcal{U}$ by Łoś's Theorem. Since $\mathcal{U}$ is non-principal and $\{i \in I \ : \ k_i \leq n_0\} = \bigcup_{j=0}^{n_0} \{i \in I \ : \ k_i = j\}$ for some finite number n_0, we must have $\{i \in I \ : \ k_i = j_0\} \in \mathcal{U}$ for some $j_0 \leq n_0$. But then $k = {}^*j_0$, whence k is standard.

Finally, we prove the Standard Induction principle in ${}^*V$. Let

$$A(n) = Q_1^{\mathbf{st}} v_1 \dots Q_m^{\mathbf{st}} v_m \, \varphi\left({}^*p_1, \dots, {}^*p_\ell, v_1, \dots, v_m, n\right),$$

wherein $p_1, \dots, p_\ell \in \mathbf{N}^V$ and φ is a formula of set theory without parameters, and define

$$A^V(n) = Q_1 v_1 \dots Q_m v_m \, \varphi\left(p_1, \dots, p_\ell, v_1, \dots, v_m, n\right).$$

Inductively in m (the number of external quantifiers in A) one can prove that for every $n_0 \in \mathbf{N}^V$,

$$\left({}^*V, {}^*\in^V, \mathbf{st}\right) \models A({}^*n_0) \Leftrightarrow (V, \in^V) \models A^V(n_0). \tag{A.34}$$

(The base step of the induction uses that $V \prec {}^*V$.) Therefore, the assumptions in the Standard Induction principle mean that $A^V(0)$ and $A^V(n) \Rightarrow A^V(n+1)$ hold for every $n \in \mathbf{N}^V$, whence $A^V(n)$ must hold for all $n \in \mathbf{N}^V$ (by induction in V). Therefore, again by equivalence (A.34) we have that $A({}^*n)$ holds for all $n \in \mathbf{N}^V$ and thus $A(n)$ holds for all standard n. □

A.3 Definable Models for (Minimal) Nonstandard Analysis

The consistency proofs for **IST** (cf. Nelson [59]) or Robinsonian nonstandard analysis (cf. Robinson [67])—and also our simple consistency proof for **minIST**$^{-}$—use ultrapower constructions and thus rest on the existence of non-principal ultrafilters, typically obtained from Zorn's Lemma. This, however, does not mean that the Axiom of Choice is an indispensable ingredient of these consistency proofs, since the ultrafilter existence theorem is in fact strictly weaker than the Axiom of Choice (cf. Halpern and Levy [29] and Banaschewski [7] for a discussion of the strength of the ultrafilter existence theorem).

Based on a technique developed by Kanovei and Shelah [40], Kanovei and Reeken [39] have shown that a slightly stronger set-theoretic axiom system than **ZFC** implies the existence of definable models of **IST** and thus of **minIST**$^{+}$. The definable nonstandard enlargement constructed in [31, 32] is obviously a model of a significant subsystem of **minIST**$^{-}$, viz. the subsystem obtained by removing those set-theoretic axiom scheme instances which do not hold for superstructures (such as **Extensionality** for atoms, in this case the reals). Moreover, by applying Kanovei and Shelah's [40] technique one can produce a countably saturated, definable, ultrapower-like extension of a set universe. In a similar manner as in the proof of our consistency result (Theorem A.2) one can then verify that this definable structure is a model of **minIST**$^{-}$.

Appendix B
Robinsonian vs. Minimal Nonstandard Analysis

The point of this book was to present a different approach to stochastic analysis, one that—for the sake of accessibility to mathematics undergraduates and students of other disciplines—avoids the use of measure theory and functional analysis which the classical approach requires and instead invokes a small axiom system, which might just be dubbed *minimal nonstandard analysis*,[1] but is a fragment of Internal Set Theory and thus called Minimal Internal Set Theory. Contrary to this intention, Robinsonian [67] nonstandard analysis has the express purpose to be just an additional tool in the hands of any research mathematician, so that any "nonstandard arguments" should yield standard theorems. For instance, the seminal result of nonstandard probability theory is the "conversion from nonstandard to standard measure spaces" [51] now known as the Loeb construction in honor of its inventor (or discoverer, depending on one's belief or disbelief in mathematical Platonism), Professor Peter A. Loeb.

From a more technical perspective, the two approaches also differ substantially: Internal Set Theory extends the *syntax* of conventional mathematics and views, say, the set of natural numbers as containing some (hitherto unclassified) nonstandard numbers—this is also the point of view taken in Nelson's *Radically Elementary Probability Theory* [60], where Minimal Internal Set Theory is derived from.

Robinsonian nonstandard analysis, however, operates *semantically*: It starts from (what may be seen as) a model of a modified fragment of Zermelo–Fraenkel set theory (with the real numbers as atoms or urelements) which is just sufficient for analysis in its broad sense—a *superstructure* over the real numbers. This is then extended to a *nonstandard universe*, viz. a superstructure over an extended set of real numbers, the *hyperreal numbers* (which is a real ordered field including infinitesimals and unlimited numbers), which also contains an extended set of natural numbers (including unlimited numbers), called *hypernatural numbers*.

---

[1]This term was suggested by Nelson in a more recent paper [62].

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7,

The extension is constructed in such a way that (among other properties) the canonical embedding is well-behaved with respect to the $\in$-relation.[2] Images of elements of the original superstructure under the canonical embedding are called *standard*, elements of standard sets are called *internal*, all other sets are called *external*.

As our motivation of Minimal Internal Set Theory in Chap. 1 already suggests, one does not need to view Minimal Internal Set Theory merely as a fragment of Internal Set Theory. Instead, Minimal Internal Set Theory can also be linked to Robinsonian nonstandard analysis relatively easily—for instance, by noting that the nonstandard universe can be viewed as a model of **minIST**: If one takes (i) the set of hypernatural numbers to be the interpretation of the constant **N** in the language of **minIST** and (ii) the class of all those hypernatural numbers which were already present in the original superstructure to be the interpretation of the predicate "…is a standard natural number" in the language of **minIST**, then the axioms of **minIST** are satisfied, and the internal sets of the superstructure are just those sets which can be defined by internal formulae (possibly with parameters) in **minIST**.

This last observation permits a new reading of the present work from the perspective of Robinsonian nonstandard analysis: The content of this book is an analysis, frequently using external formulae, of certain internal sets which intuitively[3] correspond to objects of conventional stochastic analysis. In many instances, the results of Robinsonian nonstandard analysis applied to probability theory in general and to stochastic analysis in particular (cf. e.g. Loeb [51], Anderson [4], Lindstrøm [45–48], Keisler [41], Hoover and Perkins [37, 38], Stroyan and Bayod [74], Capiński and Cutland [21–23] as well as Albeverio et al. [3] or Osswald and Y. Sun [65] and the references therein) imply that the corresponding conventional ("standard") objects of stochastic analysis can be viewed as the *standard part* of our (internal) objects in a deep, well-defined, rigorous and topologically meaningful sense: Our external notions usually correspond to the so-called S-notions of Robinsonian nonstandard analysis; for example, our definition of continuity for trajectories is known as S-continuity in the Robinsonian framework, our notion of integrability is known as S-integrability, etc.

When the present book is viewed in this light, one finds that (1) the event-wise standard part (in the topology of the real line) of any of our probability measures is—by a celebrated theorem of Loeb's [51]—always a probability measure in the conventional sense, (2) the standard part of a Wiener walk (with respect to a natural path-space topology) is—by virtue of Anderson's [4] results—a Wiener process in the sense of conventional probability theory, (3) the right standard part of our Lévy

[2] The usual method to achieve this is to define the field of hyperreals as the ultrapower of the reals with respect to a non-principal ultrafilter, and then to use some kind of $\in$-recursion in order to embed the superstructure over the reals into the superstructure over the hyperreals. The result is also known as a *bounded ultrapower* construction, cf. e.g. Albeverio et al. [3, Sect. 1.2].

[3] And, as Nelson [60, Appendix] has shown for the objects of his radically elementary probability theory, even in a formal, rigorous sense.

processes (again with respect to a natural path-space topology) is—as we know through Lindstrøm's work [49]—a Lévy process as the term is used in conventional probability theory.

A systematic, historically as well as philosophically informed comparison of Robinsonian nonstandard analysis and (subsystems of) Internal Set Theory would be beyond the scope of this book and can be found in other works such as the monographs by Kusraev and Kutateladze [42] and, in particular, Vakil [75]. Any graduate student with an interest in mathematical logic (in particular, model theory) as well as in stochastic analysis should feel encouraged to study Robinsonian nonstandard probability theory and its very interesting applications by the authors cited above, their co-authors, and many others. Hopefully the brief explanations in this section will make the transition from radically elementary stochastic analysis to stochastic nonstandard analysis in the Robinson–Loeb–Anderson setting—and to standard stochastic analysis—a little bit easier. (The mere possibility of such a transition on the basis of radically elementary stochastic analysis also is an advantage over a rival infinitesimal approach to the theory of continuous-time stochastic processes due to Benci et al. [9].)

In any case, the present book shows how to formulate an accessible, yet rigorous introduction to stochastic calculus with infinitesimals that does not require acquaintance with model theory, measure theory or functional analysis.

References

1. Albeverio, S., Fan, R., Herzberg, F.: Hyperfinite Dirichlet forms and stochastic processes. Lecture Notes of the Unione Matematica Italiana, vol. 10. Springer, Berlin (2011)
2. Albeverio, S., Herzberg, F.: A combinatorial infinitesimal representation of Lévy processes and an application to incomplete markets. Stochastics **78**(5), 301–325 (2006)
3. Albeverio, S., Høegh-Krohn, R., Fenstad, J., Lindstrøm, T.: Nonstandard methods in stochastic analysis and mathematical physics. Pure and Applied Mathematics, vol. 122. Academic, Orlando, FL (1986)
4. Anderson, R.: A non-standard representation for Brownian motion and Itô integration. Israel J. Math. **25**(1–2), 15–46 (1976)
5. Applebaum, D.: Lévy processes – from finance to probability and quantum groups. Not. Am. Math. Soc. **51**(11), 1336–1347 (2004)
6. Applebaum, D.: Lévy processes and stochastic calculus. Cambridge Studies in Advanced Mathematics, vol. 93. Cambridge University Press, Cambridge (2004)
7. Banaschewski, B.: The power of the ultrafilter theorem. J. Lond. Math. Soc. Second Series **27**(2), 193–202 (1983)
8. Barndorff-Nielsen, O., Mikosch, T., Resnick, S. (eds.): Lévy Processes: Theory and Applications. Birkhäuser, Boston, MA (2001)
9. Benci, V., Galatolo, S., Ghimenti, M.: An elementary approach to stochastic differential equations using the infinitesimals. In: Ultrafilters Across Mathematics. Contemporary Mathematics, vol. 530, pp. 1–22. American Mathematical Society, Providence, RI (2010)
10. Benoît, E.: Random walks and stochastic differential equations. In: Diener, F., Diener, M. (eds.) Nonstandard Analysis In Practice. Universitext, pp. 71–90. Springer, Berlin (1995)
11. Berg, I.v.d.: An external probability order theorem with applications. In: Nonstandard Analysis In Practice. Universitext, pp. 171–183. Springer, Berlin (1995)
12. Berg, I.v.d.: On the relation between elementary partial difference equations and partial differential equations. Ann. Pure Appl. Logic **92**(3), 235–265 (1998)
13. Berg, I.v.d.: Principles of Infinitesimal Stochastic and Financial Analysis. World Scientific, Singapore (2000)
14. Berg, I.v.d.: Asymptotic solutions of nonlinear difference equations. Annales de la Faculté des Sciences de Toulouse. Mathématiques. Série 6 **17**(3), 635–660 (2008)
15. Berg, I.v.d.: Asymptotics of families of solutions of nonlinear difference equations. Logic Anal. **1**(2), 153–185 (2008)
16. Berg, I.v.d., Amaro, E.: Nearly recombining processes and the calculation of expectations. ARIMA. Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées **9**, 389–417 (2008)

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7,

17. Bernstein, A., Wattenberg, F.: Nonstandard measure theory. In: Luxemburg, W. (ed.) Applications of Model Theory to Algebra, Analysis, and Probability, pp. 171–185. Holt, Rinehart and Winston, New York (1969)
18. Black, F., Scholes, M.: The pricing of options and corporate liabilities. J. Polit. Econ. **81**, 637–654 (1973)
19. Błasczcyk, P., Katz, M., Sherry, D.: Ten misconceptions from the history of analysis and their debunking. Found. Sci. (forthcoming) (2012). doi: 10.1007/s10699-012-9285-8
20. Boyarchenko, S., Levendorskiĭ, S.: Non-Gaussian Merton–Black–Scholes Theory. World Scientific, Singapore (2002)
21. Capiński, M., Cutland, N.: Stochastic Navier-Stokes equations. Acta Appl. Math. **25**(1), 59–85 (1991)
22. Capiński, M., Cutland, N.: Existence of global stochastic flow and attractors for Navier-Stokes equations. Probab. Theor Relat. Field **115**(1), 121–151 (1999)
23. Cutland, N.: Loeb space methods for stochastic Navier-Stokes equations. In: Cutland, N., Di Nasso, M., Ross, D. (eds.) Nonstandard Methods and Applications in Mathematics. Lecture Notes in Logic, vol. 25, pp. 195–223. Association for Symbolic Logic, La Jolla, CA (2006)
24. Delbaen, F., Schachermayer, W.: A general version of the fundamental theorem of asset pricing. Math. Ann. **300**(3), 463–520 (1994). doi: 10.1007/BF01450498
25. Delbaen, F., Schachermayer, W.: The fundamental theorem of asset pricing for unbounded stochastic processes. Math. Ann. **312**(2), 215–250 (1998). doi: 10.1007/s002080050220
26. Duffie, D.: Dynamic Asset Pricing Theory, 3rd edn. Princeton University Press, Princeton, NJ (2001)
27. Girsanov, I.: On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. Theor Probab. Appl. **5**, 285–301 (1960)
28. Gödel, K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme. I. Monatshefte für Mathematik und Physik **38**, 173–198 (1931)
29. Halpern, J., Levy, A.: The Boolean prime ideal theorem does not imply the axiom of choice. In: Scott, D. (ed.) Axiomatic Set Theory. Proceedings of Symposia in Pure Mathematics, vol. XIII. Part 1, pp. 83–134. American Mathematical Society, Providence, RI (1971)
30. Henson, C., Keisler, H.: On the strength of nonstandard analysis. J. Symbolic Logic **51**(2), 377–386 (1986)
31. Herzberg, F.: A definable nonstandard enlargement. Math. Log. Quart. **54**(2), 167–175 (2008)
32. Herzberg, F.: Addendum to "A definable nonstandard enlargement". Math. Log. Quart. **54**(6), 666–667 (2008)
33. Herzberg, F.: Hyperfinite stochastic integration for Lévy processes with finite-variation jump part. Bull. Sci. Math. **134**(4), 423–445 (2010)
34. Herzberg, F.: First steps towards an equilibrium theory for Lévy financial markets. Ann. Finance (forthcoming) (2012). doi: 10.1007/s10436-012-0202-5
35. Herzberg, F.: Radically elementary mathematics and the Feynman path integral. Manuscript (2012)
36. Herzberg, F., Lindstrøm, T.: Corrigendum and addendum to "Hyperfinite Lévy processes" (Stochastics and Stochastics Reports, 76(6):517–548, 2004). Stochastics **81**(6), 567–570 (2009)
37. Hoover, D., Perkins, E.: Nonstandard construction of the stochastic integral and applications to stochastic differential equations. I. Trans. Am. Math. Soc. **275**, 1–29 (1983)
38. Hoover, D., Perkins, E.: Nonstandard construction of the stochastic integral and applications to stochastic differential equations. II. Trans. Am. Math. Soc. **275**, 30–58 (1983)
39. Kanovei, V., Reeken, M.: Nonstandard analysis, axiomatically. Springer Monographs in Mathematics. Springer, Berlin (2004)
40. Kanovei, V., Shelah, S.: A definable nonstandard model of the reals. J. Symbolic Logic **69**(1), 159–164 (2004)
41. Keisler, H.: An infinitesimal approach to stochastic analysis. Mem. Am. Math. Soc. **297** (1984)
42. Kusraev, A., Kutateladze, S.: Nonstandard methods of analysis. Mathematics and Its Applications, vol. 291. Kluwer, Dordrecht (1994)

43. Lawler, G.: Internal Set Theory and infinitesimal random walks. In: Faris, W. (ed.) Diffusion, quantum theory, and radically elementary mathematics. Mathematical Notes, vol. 47, pp. 157–181. Princeton University Press, Princeton, NJ (2006)
44. Lévy, P.: Processus stochastiques et mouvement Brownien. Gauthier-Villars, Paris (1948)
45. Lindstrøm, T.: Addendum to "Hyperfinite stochastic integration III". Math. Scand. **46**(2), 332–333 (1980)
46. Lindstrøm, T.: Hyperfinite stochastic integration. I: The nonstandard theory. Math. Scand. **46**(2), 265–292 (1980)
47. Lindstrøm, T.: Hyperfinite stochastic integration. II: Comparison with the standard theory. Math. Scand. **46**(2), 293–314 (1980)
48. Lindstrøm, T.: Hyperfinite stochastic integration. III: Hyperfinite representations of standard martingales. Math. Scand. **46**(2), 315–331 (1980)
49. Lindstrøm, T.: Hyperfinite Lévy processes. Stochast. Stochast. Rep. **76**(6), 517–548 (2004)
50. Lindstrøm, T.: Nonlinear stochastic integrals for hyperfinite Lévy processes. Logic Anal. **1**(2), 91–129 (2008)
51. Loeb, P.: Conversion from nonstandard to standard measure spaces and applications in probability theory. Trans. Am. Math. Soc. **211**, 113–122 (1975)
52. Łoś, J.: Quelques remarques, théorèmes et problèmes sur les classes définissables d'algèbres. In: Skolem, T., Hasenjaeger, G., Kreisel, G., Robinson, A., Wang, H., Henkin, L., Łoś, J. (eds.) Mathematical Interpretation of Formal Systems. Studies in Logic and the Foundations of Mathematics, vol. 16, pp. 98–113. North-Holland, Amsterdam (1955)
53. Madan, D., Seneta, E.: Chebyshev polynomial approximations and characteristic function estimation. J. R. Stat. Soc. Ser. B Stat. Methodol. **49**(2), 163–169 (1987)
54. Madan, D., Seneta, E.: Chebyshev polynomial approximations for characteristic function estimation: some theoretical supplements. J. R. Stat. Soc. Ser. B Stat. Methodol. **51**(2), 281–285 (1989)
55. Madan, D., Seneta, E.: The variance gamma (V.G.) model for share market returns. J. Bus. **63**(4), 511–524 (1990)
56. Merton, R.: Theory of rational option pricing. Bell J. Econ. Manag. Sci. **4**(1), 141–183 (1973)
57. Merton, R.: Option pricing when underlying stock returns are discontinuous. J. Financ. Econ. **3**(1–2), 125–144 (1976)
58. Nelson, E.: Feynman integrals and the Schrödinger equation. J. Math. Phys. **5**(3), 332–343 (1964)
59. Nelson, E.: Internal set theory: A new approach to nonstandard analysis. Bull. Am. Math. Soc. **83**(6), 1165–1198 (1977)
60. Nelson, E.: Radically elementary probability theory. Annals of Mathematics Studies, vol. 117. Princeton University Press, Princeton, NJ (1987)
61. Nelson, E.: The syntax of nonstandard analysis. Ann. Pure Appl. Logic **38**(2), 123–134 (1988). doi: 10.1016/0168-0072(88)90050-4
62. Nelson, E.: The virtue of simplicity. In: Berg, I.v.d., Neves, V. (eds.) The Strength of Nonstandard Analysis, pp. 27–32. Springer, Vienna (2007)
63. Nelson, E.: Warning signs of a possible collapse of contemporary mathematics. In: Infinity, pp. 76–85. Cambridge University Press, Cambridge (2011)
64. Osswald, H.: Malliavin calculus in abstract Wiener space using infinitesimals. Adv. Math. **176**(1), 1–37 (2003)
65. Osswald, H., Sun, Y.: Measure and probability theory and applications. In: Nonstandard analysis for the working mathematician. Mathematics and its Applications, vol. 510, pp. 137–257. Kluwer, Dordrecht (2000)
66. Robinson, A.: Non-standard analysis. Nederlandse Akademie van Wetenschappen. Proceedings. Series A. Indagat. Math. **64**, 432–440 (1961)
67. Robinson, A.: Non-standard Analysis. North-Holland, Amsterdam (1966)

68. Robinson, A., Zakon, E.: A set-theoretical characterization of enlargements. In: Luxemburg, W. (ed.) Applications of Model Theory to Algebra, Analysis, and Probability (International Symposium, Pasadena, California, 1967), pp. 109–122. Holt, Rinehart and Winston, New York (1969)
69. Sari, T.: Petite histoire de la stroboscopie. In: Colloque Trajectorien à la Mémoire de Georges Reeb et Jean-Louis Callot (Strasbourg-Obernai, 1995), no. 1995/13 in Prépublications de l'Institut de Recherche Mathématique Avancée, pp. 5–15. Université Louis Pasteur, Strasbourg (1995)
70. Sari, T.: Stroboscopy and averaging. In: Colloque Trajectorien à la Mémoire de Georges Reeb et Jean-Louis Callot (Strasbourg-Obernai, 1995), no. 1995/13 in Prépublications de l'Institut de Recherche Mathématique Avancée, pp. 95–124. Université Louis Pasteur, Strasbourg (1995)
71. Sato, K.I.: Lévy Processes and Infinitely Divisible Distributions. Cambridge University Press, Cambridge (1999)
72. Schmieden, C., Laugwitz, D.: Eine Erweiterung der Infinitesimalrechnung. Math. Z. **69**, 1–39 (1958)
73. Schoutens, W.: Lévy processes in finance: Pricing financial derivatices. Wiley Series in Probability and Statistics. Wiley, Chichester (2003)
74. Stroyan, K., Bayod, J.: Foundations of infinitesimal stochastic analysis. Studies in Logic and the Foundations of Mathematics, vol. 119. North-Holland, Amsterdam (1986)
75. Vakil, N.: Real analysis through modern infinitesimals. Encyclopedia of Mathematics and Its Applications, vol. 140. Cambridge University Press, Cambridge (2011)

Index

F. S. Herzberg, *Stochastic Calculus with Infinitesimals*, Lecture Notes in Mathematics 2067, DOI 10.1007/978-3-642-33149-7,

LECTURE NOTES IN MATHEMATICS

Edited by J.-M. Morel, B. Teissier; P.K. Maini

Editorial Policy (for the publication of monographs)

1. Lecture Notes aim to report new developments in all areas of mathematics and their applications - quickly, informally and at a high level. Mathematical texts analysing new developments in modelling and numerical simulation are welcome.
   Monograph manuscripts should be reasonably self-contained and rounded off. Thus they may, and often will, present not only results of the author but also related work by other people. They may be based on specialised lecture courses. Furthermore, the manuscripts should provide sufficient motivation, examples and applications. This clearly distinguishes Lecture Notes from journal articles or technical reports which normally are very concise. Articles intended for a journal but too long to be accepted by most journals, usually do not have this "lecture notes" character. For similar reasons it is unusual for doctoral theses to be accepted for the Lecture Notes series, though habilitation theses may be appropriate.

2. Manuscripts should be submitted either online at www.editorialmanager.com/lnm to Springer's mathematics editorial in Heidelberg, or to one of the series editors. In general, manuscripts will be sent out to 2 external referees for evaluation. If a decision cannot yet be reached on the basis of the first 2 reports, further referees may be contacted: The author will be informed of this. A final decision to publish can be made only on the basis of the complete manuscript, however a refereeing process leading to a preliminary decision can be based on a pre-final or incomplete manuscript. The strict minimum amount of material that will be considered should include a detailed outline describing the planned contents of each chapter, a bibliography and several sample chapters.
   Authors should be aware that incomplete or insufficiently close to final manuscripts almost always result in longer refereeing times and nevertheless unclear referees' recommendations, making further refereeing of a final draft necessary.
   Authors should also be aware that parallel submission of their manuscript to another publisher while under consideration for LNM will in general lead to immediate rejection.

3. Manuscripts should in general be submitted in English. Final manuscripts should contain at least 100 pages of mathematical text and should always include

   – a table of contents;
   – an informative introduction, with adequate motivation and perhaps some historical remarks: it should be accessible to a reader not intimately familiar with the topic treated;
   – a subject index: as a rule this is genuinely helpful for the reader.

   For evaluation purposes, manuscripts may be submitted in print or electronic form (print form is still preferred by most referees), in the latter case preferably as pdf- or zipped psfiles. Lecture Notes volumes are, as a rule, printed digitally from the authors' files. To ensure best results, authors are asked to use the LaTeX2e style files available from Springer's web-server at:

   ftp://ftp.springer.de/pub/tex/latex/svmonot1/ (for monographs) and
   ftp://ftp.springer.de/pub/tex/latex/svmultt1/ (for summer schools/tutorials).

Additional technical instructions, if necessary, are available on request from lnm@springer.com.

4. Careful preparation of the manuscripts will help keep production time short besides ensuring satisfactory appearance of the finished book in print and online. After acceptance of the manuscript authors will be asked to prepare the final LaTeX source files and also the corresponding dvi-, pdf- or zipped ps-file. The LaTeX source files are essential for producing the full-text online version of the book (see http://www.springerlink.com/openurl.asp?genre=journal&issn=0075-8434 for the existing online volumes of LNM). The actual production of a Lecture Notes volume takes approximately 12 weeks.

5. Authors receive a total of 50 free copies of their volume, but no royalties. They are entitled to a discount of 33.3 % on the price of Springer books purchased for their personal use, if ordering directly from Springer.

6. Commitment to publish is made by letter of intent rather than by signing a formal contract. Springer-Verlag secures the copyright for each volume. Authors are free to reuse material contained in their LNM volumes in later publications: a brief written (or e-mail) request for formal permission is sufficient.

Addresses:

Professor J.-M. Morel, CMLA,
École Normale Supérieure de Cachan,
61 Avenue du Président Wilson, 94235 Cachan Cedex, France
E-mail: morel@cmla.ens-cachan.fr

Professor B. Teissier, Institut Mathématique de Jussieu,
UMR 7586 du CNRS, Équipe “Géométrie et Dynamique”,
175 rue du Chevaleret
75013 Paris, France
E-mail: teissier@math.jussieu.fr

For the “Mathematical Biosciences Subseries” of LNM:

Professor P. K. Maini, Center for Mathematical Biology,
Mathematical Institute, 24-29 St Giles,
Oxford OX1 3LP, UK
E-mail: maini@maths.ox.ac.uk

Springer, Mathematics Editorial, Tiergartenstr. 17,
69121 Heidelberg, Germany,
Tel.: +49 (6221) 4876-8259

Fax: +49 (6221) 4876-8259
E-mail: lnm@springer.com