Advanced stochastic processes: Part I

Jan A. Van Casteren

bookboon
The eBook company

Jan A. Van Casteren

Advanced Stochastic Processes

Part I

Advanced stochastic processes: Part I
2nd edition

ISBN 978-87-403-1115-0

The author is obliged to Department of Mathematics and Computer Science of the University of Antwerp for its material support. The author is also indebted to Freddy Delbaen (ETH, Zürich), and the late Jean Haezendock (University of Antwerp). A big part of Chapter 5 is due to these people. The author is thankful for comments and logistic support by numerous students who have taken courses based on this text. In particular, he is grateful to Lieven Smits and Johan Van Biesen (former students at the University of Antwerp) who wrote part of the Chapters 1 and 2. The author gratefully acknowledges World Scientic Publishers (Singapore) for their permission to publish the contents of Chapter 4 which also makes up a substantial portion of Chapter 1 in [144]. The author also learned a lot from the book by Stirzaker [124]. Section 1 of Chapter 2 is taken from [124], and the author is indebted to David Stirzaker to allow him to include this material in this book. The author is also grateful to the people of Bookboon, among whom Karin Hamilton Jakobsen and Ahmed Zsolt Dakroub, who assisted him in the final stages of the preparation of this book.

Contents

Preface

This book deals with several aspects of stochastic process theory: Markov chains, renewal theory, Brownian motion, Brownian motion as a Gaussian process, Brownian motion as a Markov process, Brownian motion as a martingale, stochastic calculus, Itô's formula, regularity properties, Feller-Dynkin semigroups and (strong) Markov processes. Brownian motion can also be seen as limit of normalized random walks. Another feature of the book is a thorough discussion of the Doob-Meyer decomposition theorem. It also contains some features of stochastic differential equations and the Girsanov transformation. The first chapter (Chapter 1) contains a (gentle) introduction to the theory of stochastic processes. It is more or less required to understand the main part of the book, which consists of discrete (time) probability models (Chapter 2), of continuous time models, in casu Brownian motion, Chapter 3, and of certain aspects of stochastic differential equations and Girsanov's transformation (Chapter 4). In the final chapter (Chapter 5) a number of other, but related, issues are treated. Several of these topics are explicitly used in the main text (Fourier transforms of distributions, or characteristic functions of random vectors, Lévy's continuity theorem, Kolmogorov's extension theorem, uniform integrability); some of them are treated, like the important Doob-Meyer decomposition theorem, but are not explicitly used. Of course Itô's formula implies that a C^2-function composed with a local semi-martingale is again a semi-martingale. The Doob-Meyer decomposition theorem yields that a submartingale of class (DL) is a semi-martingale. Section 1 of Chapter 5 contains several aspects of Fourier transforms of probability distributions (characteristic functions). Among other results Bochner's theorem is treated here. Section 2 contains convergence properties of positive measures. Section 3 gives some results in ergodic theory, and gives the connection with the strong law of large numbers (SLLN). Section 4 gives a proof of Kolmogorov's extension theorem (for a consistent family of probability measures on Polish spaces). In Section 5 the reader finds a short treatment of uniform integrable families of functions in an L^1-space. For example Scheffé's theorem is treated. Section 6 in Chapter 5 contains a precise description of the regularity properties (like almost sure right-continuity, almost sure existence of left limits) of stochastic processes like submartingales, Lévy processes, and others; it also contains a proof of Doob's maximal inequality for submartingales. Section 7 of the same chapter contains a description of Markov process theory starting from just one probability space instead of a whole family. The proof of the Doob-Meyer decompositon theorem is based on a result by Komlos: see Section 8. Throughout the book the reader will be exposed to martingales, and related processes.

Readership. From the description of the contents it is clear that the text is designed for students at the graduate or master level. The author believes that also Ph.D. students, and even researchers, might benefit from these notes. The reader is introduced in the following topics: Markov processes, Brownian motion and other Gaussian processes, martingale techniques, stochastic differential equations, Markov chains and renewal theory, ergodic theory and limit theorems.

CHAPTER 1

Stochastic processes: prerequisites

In this chapter we discuss a number of relevant notions related to the theory of stochastic processes. Topics include conditional expectation, distribution of Brownian motion, elements of Markov processes, and martingales. For completeness we insert the definitions of a σ-field or σ-algebra, and concepts related to measures.

1.1. Definition. A σ-algebra, or σ-field, on a set Ω is a subset $\mathcal{A}$ of the power set $\mathcal{P}(\Omega)$ with the following properties:

(i) $\Omega \in \mathcal{A}$;
(ii) $A \in \mathcal{A}$ implies $A^c := \Omega \backslash A \in \mathcal{A}$;
(iii) if $(A_n)_{n \geqslant 1}$ is a sequence in $\mathcal{A}$, then $\bigcup_{n=1}^{\infty} A_n$ belongs to $\mathcal{A}$.

Let $\mathcal{A}$ be a σ-field on Ω. Unless otherwise specified, a measure is an application $\mu : \mathcal{A} \to [0, \infty]$ with the following properties:

- $\mu(\varnothing) = 0$;
- if $(A_n)_{n \geqslant 1}$ is a mutually disjoint sequence in $\mathcal{A}$, then
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{N \to \infty} \sum_{j=1}^{N} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

If μ is measure on $\mathcal{A}$ for which $\mu(\Omega) = 1$, then μ is called a probability measure; if $\mu(\Omega) \leqslant 1$, then μ is called a sub-probability measure. If $\mu : \mathcal{A} \to [0, 1]$ is a probability space, then the triple $(\Omega, \mathcal{A}, \mu)$ is called a probability space, and the elements of $\mathcal{A}$ are called events.

Let $\mathcal{M}$ be a collection of subsets of $\mathcal{P}(\Omega)$, where Ω is some set like in Definition 1.1. The smallest σ-field containing $\mathcal{M}$ is called the σ-field generated by $\mathcal{M}$, and it is often denoted by $\sigma(\mathcal{M})$. Let $(\Omega, \mathcal{A}, \mu)$ be a sub-probability space, *i.e.* μ is a sub-probability on the σ-field $\mathcal{A}$. Then, we enlarge Ω with one point $\triangle$, and enlarge $\mathcal{A}$ to
$$\mathcal{A}^{\triangle} := \sigma(\mathcal{A} \cup \{\triangle\}) = \left\{A \in \mathcal{P}\left(\Omega^{\triangle}\right) : A \cap \Omega \in \mathcal{A}\right\}.$$
Then $\mu^{\triangle} : \mathcal{A}^{\triangle} \to [0, 1]$, defined by
$$\mu^{\triangle}(A) = \mu(A \cap \Omega) + (1 - \mu(\Omega)) 1_A(\triangle), \quad A \in \mathcal{A}^{\triangle}, \tag{1.1}$$
turns the space $\left(\Omega^{\triangle}, \mathcal{A}^{\triangle}, \mu^{\triangle}\right)$ into a probability space. Here $\Omega^{\triangle} = \Omega \cup \{\triangle\}$. This kind of construction also occurs in the context of Markov processes with

finite lifetime: see the equality (3.75) in (an outline of) the proof of Theorem 3.37. For the important relationship between Dynkin systems, or λ-systems, and σ-algebras, see Theorem 2.42.

1. Conditional expectation

1.2. Definition. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and let A and B be events in $\mathcal{A}$ such that $\mathbb{P}[B] > 0$. The quantity $\mathbb{P}\left(A \mid B\right) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ is then called the *conditional probability of the event A with respect to the event B*. We put $\mathbb{P}\left(A \mid B\right) = 0$ if $\mathbb{P}(B) = 0$.

Consider a finite partition $\{B_1, \dots, B_n\}$ of Ω with $B_j \in \mathcal{A}$ for all $j = 1, \dots, n$, and let $\mathcal{B}$ be the subfield of $\mathcal{A}$ generated by the partition $\{B_1, \dots, B_n\}$, and write

$$\mathbb{P}\left[A \mid \mathcal{B}\right] = \sum_{j=1}^{n} \mathbb{P}\left(A \mid B_j\right) \mathbf{1}_{B_j}.$$

Then $\mathbb{P}\left[A \mid \mathcal{B}\right]$ is a $\mathcal{B}$-measurable stochastic variable on Ω, and

$$\int_B \mathbb{P}\left[A \mid \mathcal{B}\right] d\mathbb{P} = \int_B \mathbf{1}_A d\mathbb{P} \quad \text{for all } B \in \mathcal{B}.$$

Conversely, if f is a $\mathcal{B}$-measurable stochastic variable on Ω with the property that for all $B \in \mathcal{B}$ the equality $\int_B f d\mathbb{P} = \int_B \mathbf{1}_A d\mathbb{P}$ holds, then $f = \mathbb{P}\left[A \mid \mathcal{B}\right]$ $\mathbb{P}$-almost surely. This is true, because $\int_B \left(f - \mathbb{P}\left[A \mid \mathcal{B}\right]\right) d\mathbb{P} = 0$ for all $B \in \mathcal{B}$. If $\mathcal{B}$ is a sub-field (more precisely a sub-σ-field, or sub-σ-algebra) generated by a finite partition of Ω, then for every $A \in \mathcal{A}$ there exists one and only one class of variables in $L^1(\Omega, \mathcal{B}, \mathbb{P})$, which we denote by $\mathbb{P}\left[A \mid \mathcal{B}\right]$, with the following property

$$\int_B \mathbb{P}\left[A \mid \mathcal{B}\right] d\mathbb{P} = \int_B \mathbf{1}_A d\mathbb{P} \quad \text{for all } B \in \mathcal{B}.$$

The variable $\sum_{j=1}^n \mathbb{P}\left(A \mid B_j\right) \mathbf{1}_{B_j}$ is an element from the class $\mathbb{P}\left[A \mid \mathcal{B}\right]$.

If we fix $B \in \mathcal{A}$ with $\mathbb{P}(B) > 0$, then the measure $A \mapsto \mathbb{P}\left(A \mid B\right)$ is a probability measure on $(\Omega, \mathcal{A})$. If $\mathbb{P}(B) = 0$, then the measure $A \mapsto \mathbb{P}\left(A \mid B\right)$ is the zero-measure.

Let X be a $\mathbb{P}$-integrable real or complex valued stochastic variable on Ω. Then X is also $\mathbb{P}\left(\cdot \mid B\right)$-integrable, and

$$\int X d\mathbb{P}\left(\cdot \mid B\right) = \frac{\mathbb{E}\left[X \mathbf{1}_B\right]}{\mathbb{P}(B)}, \quad \text{provided } \mathbb{P}(B) > 0.$$

This quantity is the average of the stochastic variable over the event B. As before, it is easy to show that if $\mathcal{B}$ is a subfield of $\mathcal{A}$ generated by a finite partition $\{B_1, \dots, B_n\}$ of Ω, then there exists, for every $\mathbb{P}$-integrable real or complex valued stochastic variable X on Ω one and only one class of functions in $L^1(\Omega, \mathcal{B}, \mathbb{P})$, which we denote by $\mathbb{E}\left[X \mid \mathcal{B}\right]$ with the property that

$$\int_B \mathbb{E}\left[X \mid \mathcal{B}\right] d\mathbb{P} = \int_B X \, d\mathbb{P} \quad \text{for all } B \in \mathcal{B}.$$

The variable $\sum_{j=1}^n \int X d\mathbb{P}\left(\cdot \mid B_j\right) \mathbf{1}_{B_j}$ is an element from the class $\mathbb{E}\left[X \mid \mathcal{B}\right]$.

The next theorem generalizes the previous properties to an arbitrary subfield (or more precisely sub-σ-field) $\mathcal{B}$ of $\mathcal{A}$.

1.3. THEOREM (Theorem and definition). *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\mathcal{B}$ be a subfield of $\mathcal{A}$. Then for every stochastic variable $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ there exists one and only one class in $L^1(\Omega, \mathcal{B}, \mathbb{P})$, which is denoted by $\mathbb{E}\left[X \mid \mathcal{B}\right]$ and which is called the conditional expectation of X with respect to $\mathcal{B}$, with the property that*

$$\int_B \mathbb{E}\left[X \mid \mathcal{B}\right] d\mathbb{P} = \int_B X \, d\mathbb{P} \quad \textit{for all } B \in \mathcal{B}.$$

If $X = \mathbf{1}_A$, with $A \in \mathcal{A}$, then we write $\mathbb{P}\left[A \mid \mathcal{B}\right]$ instead of $\mathbb{E}\left[\mathbf{1}_A \mid \mathcal{B}\right]$; if $\mathcal{B}$ is generated by just one stochastic variable Y, then we write $\mathbb{E}\left[X \mid Y\right]$ and $\mathbb{P}\left[A \mid Y\right]$ instead of respectively $\mathbb{E}\left[X \mid \sigma(Y)\right]$ and $\mathbb{P}\left[A \mid \sigma(Y)\right]$.

PROOF. Suppose that X is real-valued; if $X = \text{Re } X + i\text{Im } X$ is complex-valued, then we apply the following arguments to Re X and Im X. Upon writing the real-valued stochastic variable X as $X = X^+ - X^-$, where $X^\pm$ are non-negative stochastic variables in $L^1(\Omega, \mathcal{A}, \mathbb{P})$, without loss of generality we may and do assume that $X \geqslant 0$. Define the measure $\mu : \mathcal{A} \to [0, \infty)$ by $\mu(A) = \int_A X d\mathbb{P}$, $A \in \mathcal{A}$. Then μ is finite measure which is absolutely continuous with respect to the measure $\mathbb{P}$. We restrict μ to the measurable space $(\Omega, \mathcal{B})$; its absolute continuity with respect to $\mathbb{P}$ confined to $(\Omega, \mathcal{B})$ is preserved. From the Radon-Nikodym theorem it follows that there exists a unique class $Y \in L^1(\Omega, \mathcal{B}, \mathbb{P})$ such that, for all $B \in \mathcal{B}$, the following equality is valid:

$$\mu(B) = \int_B Y d\mathbb{P}, \quad \text{and hence} \int_B X d\mathbb{P} = \int_B Y d\mathbb{P}.$$

This proves Theorem1.3. $\square$

If $\mathcal{B}$ is generated by a countable or finite partition $\{B_j : j \in \mathbb{N}\}$, then it is fairly easy to give an explicit formula for the conditional expectation of a stochastic $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ with respect to $\mathcal{B}$:

$$\mathbb{E}\left[X \mid \mathcal{B}\right] = \sum_{j \in \mathbb{N}} \frac{\int_{B_j} X d\mathbb{P}}{\mathbb{P}(B_j)} \mathbf{1}_{B_j} = \sum_{j \in \mathbb{N}} \mathbb{E}\left[X \mid B_j\right] \mathbf{1}_{B_j}.$$

Next let $\mathcal{B}$ be an arbitrary subfield of $\mathcal{A}$, let X belong to $L^1(\Omega, \mathcal{A}, \mathbb{P})$, and let B be an atom in $\mathcal{B}$. The latter means that $\mathbb{P}(B) > 0$, and if $A \in \mathcal{B}$ is such that $A \subset B$, then either $\mathbb{P}(A) = 0$ or $\mathbb{P}(B \backslash A) = 0$. If Y represents $\mathbb{E}\left[X \mid \mathcal{B}\right]$, then $Y \mathbf{1}_B = b \mathbf{1}_B$, $\mathbb{P}$-almost surely, for some constant b. This follows from the $\mathcal{B}$-measurability of the variable Y together with the fact that B is an atom for $(\Omega, \mathcal{B}, \mathbb{P})$. So we get $\int_B X d\mathbb{P} = \int_B \mathbb{E}\left[X \mid \mathcal{B}\right] d\mathbb{P} = \int Y \mathbf{1}_B d\mathbb{P} = b\mathbb{P}(B)$, and hence $b = \dfrac{\int_B X d\mathbb{P}}{\mathbb{P}(B)}$. Consequently, on the atom B we have:

$$\mathbb{E}\left[X \mid \mathcal{B}\right] = \frac{\int_B X d\mathbb{P}}{\mathbb{P}(B)} = b, \quad \mathbb{P}\text{-almost surely.}$$

In particular, for $X = \mathbf{1}_A$, we have on the atom B the equality

$$\mathbb{P}\left[A \mid \mathcal{B}\right] = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \mathbb{P}\text{-almost surely.}$$

If B is not an atom, then the conditional expectation on B need not be constant.

In the following theorem we collect some properties of conditional expectation. For the notion of uniform integrability see Section 5.

1.4. Theorem. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and let $\mathcal{B}$ be a subfield of $\mathcal{A}$. Then the following assertions hold.*

(1) *If all events in $\mathcal{B}$ have probability 0 or 1 (in particular if $\mathcal{B}$ is the trivial field $\{\varnothing, \Omega\}$), then for all stochastic variables $X \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ the equality $\mathbb{E}\left[X \mid \mathcal{B}\right] = \mathbb{E}(X)$ is true $\mathbb{P}$-almost surely.*
(2) *If X is a stochastic variable in $L^1(\Omega, \mathcal{A}, \mathbb{P})$ such that $\mathcal{B}$ and $\sigma(X)$ are independent, then the equality $\mathbb{E}\left[X \mid \mathcal{B}\right] = \mathbb{E}(X)$ is true $\mathbb{P}$-almost surely.*

(3) *If a and b are real or complex constants, and if the stochastic variables X and Y belong to $L^1(\Omega, \mathcal{A}, \mathbb{P})$, then the equality*

$$\mathbb{E}\left[aX + bY \mid \mathcal{B}\right] = a\mathbb{E}\left[X \mid \mathcal{B}\right] + b\mathbb{E}\left[Y \mid \mathcal{B}\right] \quad \text{is true } \mathbb{P}\text{-almost surely.}$$

(4) *If X and Y are real stochastic variables in $L^1(\Omega, \mathcal{A}, \mathbb{P})$ such that $X \leqslant Y$, then the inequality $\mathbb{E}\left[X \mid \mathcal{B}\right] \leqslant \mathbb{E}\left[Y \mid \mathcal{B}\right]$ holds $\mathbb{P}$-almost surely. Hence the mapping $X \mapsto \mathbb{E}\left[X \mid \mathcal{B}\right]$ is a mapping from $L^1(\Omega, \mathcal{A}, \mathbb{P})$ onto $L^1(\Omega, \mathcal{B}, \mathbb{P})$.*

(5) (a) *If $(X_n : n \in \mathbb{N})$ is a non-decreasing sequence of stochastic variables in $L^1(\Omega, \mathcal{A}, \mathbb{P})$, then*

$$\sup_n \mathbb{E}\left[X_n \mid \mathcal{B}\right] = \mathbb{E}\left[\sup_n X_n \mid \mathcal{B}\right], \quad \mathbb{P}\text{-almost surely.}$$

(b) *If $(X_n : n \in \mathbb{N})$ is any sequence of stochastic variables in $L^1(\Omega, \mathcal{A}, \mathbb{P})$ which converges $\mathbb{P}$-almost surely to a stochastic variable X, and if there exists a stochastic variable $Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ such that $|X_n| \leqslant Y$ for all $n \in \mathbb{N}$, then*

$$\lim_{n\to\infty} \mathbb{E}\left[X_n \mid \mathcal{B}\right] = \mathbb{E}\left[\lim_{n\to\infty} X_n \mid \mathcal{B}\right], \quad \mathbb{P}\text{-almost surely, and in } L^1(\Omega, \mathcal{B}, \mathbb{P}).$$

The condition "$|X_n| \leqslant Y$ for all $n \in \mathbb{N}$ with $Y \in L^1(\Omega, \mathcal{A}, \mathbb{P})$" may be replaced with "the sequence $(X_n)_{n\in\mathbb{N}}$ is uniformly integrable in the space $L^1(\Omega, \mathcal{A}, \mathbb{P})$" and still keep the second conclusion in (5b). In order to have $\mathbb{P}$-almost sure convergence the uniform integrability condition should be replaced with the condition

$$\inf_{M>0,\, M\in\mathbb{R}} \sup_{n\in\mathbb{N}} \mathbb{E}\left[|X_n|, |X_n| > M \mid \mathcal{B}\right] = 0, \quad \mathbb{P}\text{-almost surely.} \tag{1.2}$$

(6) *If $c(x)$ is a convex continuous function from $\mathbb{R}$ to $\mathbb{R}$, and if X belongs to $L^1(\Omega, \mathcal{A}, \mathbb{P})$, then*

$$c\left(\mathbb{E}\left[X \mid \mathcal{B}\right]\right) \leqslant \mathbb{E}\left[c(X) \mid \mathcal{B}\right], \quad \mathbb{P}\text{-almost surely.}$$

(7) *Let $p \geqslant 1$, and let X be a stochastic variable in $L^p(\Omega, \mathcal{A}, \mathbb{P})$. Then the stochastic variable $\mathbb{E}\left[X \mid \mathcal{B}\right]$ belongs to $L^p(\Omega, \mathcal{B}, \mathbb{P})$, and*

$$\left\|\mathbb{E}\left[X \mid \mathcal{B}\right]\right\|_p \leqslant \|X\|_p.$$

So the linear mapping $X \mapsto \mathbb{E}\left[X \mid \mathcal{B}\right]$ is a projection from $L^p(\Omega, \mathcal{A}, \mathbb{P})$ onto $L^p(\Omega, \mathcal{B}, \mathbb{P})$.

(8) *(Tower property) Let $\mathcal{B}'$ be another subfield of $\mathcal{A}$ such that $\mathcal{B} \subseteq \mathcal{B}' \subseteq \mathcal{A}$. If X belongs to $L^1(\Omega, \mathcal{A}, \mathbb{P})$, then the equality*

$$\mathbb{E}\left[\mathbb{E}\left[X \mid \mathcal{B}'\right] \mid \mathcal{B}\right] = \mathbb{E}\left[X \mid \mathcal{B}\right] \quad \text{holds } \mathbb{P}\text{-almost surely.}$$

(9) *If X belongs to $L^1(\Omega, \mathcal{B}, \mathbb{P})$, then $\mathbb{E}\left[X \mid \mathcal{B}\right] = X$, $\mathbb{P}$-almost surely.*

(10) *If X belongs to $L^1(\Omega, \mathcal{A}, \mathbb{P})$, and if Z belongs to $L^\infty(\Omega, \mathcal{B}, \mathbb{P})$, then*

$$\mathbb{E}\left[ZX \mid \mathcal{B}\right] = Z\mathbb{E}\left[X \mid \mathcal{B}\right], \quad \mathbb{P}\text{-almost surely.}$$

(11) *If X belongs to $L^2(\Omega, \mathcal{A}, \mathbb{P})$, then $\mathbb{E}\left[Y\left(X - \mathbb{E}\left(X \mid \mathcal{B}\right)\right)\right] = 0$ for all $Y \in L^2(\Omega, \mathcal{B}, \mathbb{P})$. Hence, the mapping $X \mapsto \mathbb{E}\left[X \mid \mathcal{B}\right]$ is an orthogonal projection from $L^2(\Omega, \mathcal{A}, \mathbb{P})$ onto $L^2(\Omega, \mathcal{B}, \mathbb{P})$.*

Observe that for $\mathcal{B}$ the trivial σ-field, *i.e.* $\mathcal{B} = \{\emptyset, \Omega\}$, the condition in (1.2) is the same as saying that the sequence $(X_n)_n$ is uniformly integrable in the sense that

$$\inf_{M>0,\, M\in\mathbb{R}} \sup_{n\in\mathbb{N}} \mathbb{E}\left[|X_n|\,,\, |X_n| > M\right] = 0. \tag{1.3}$$

Proof. We successively prove the items in Theorem 1.4.

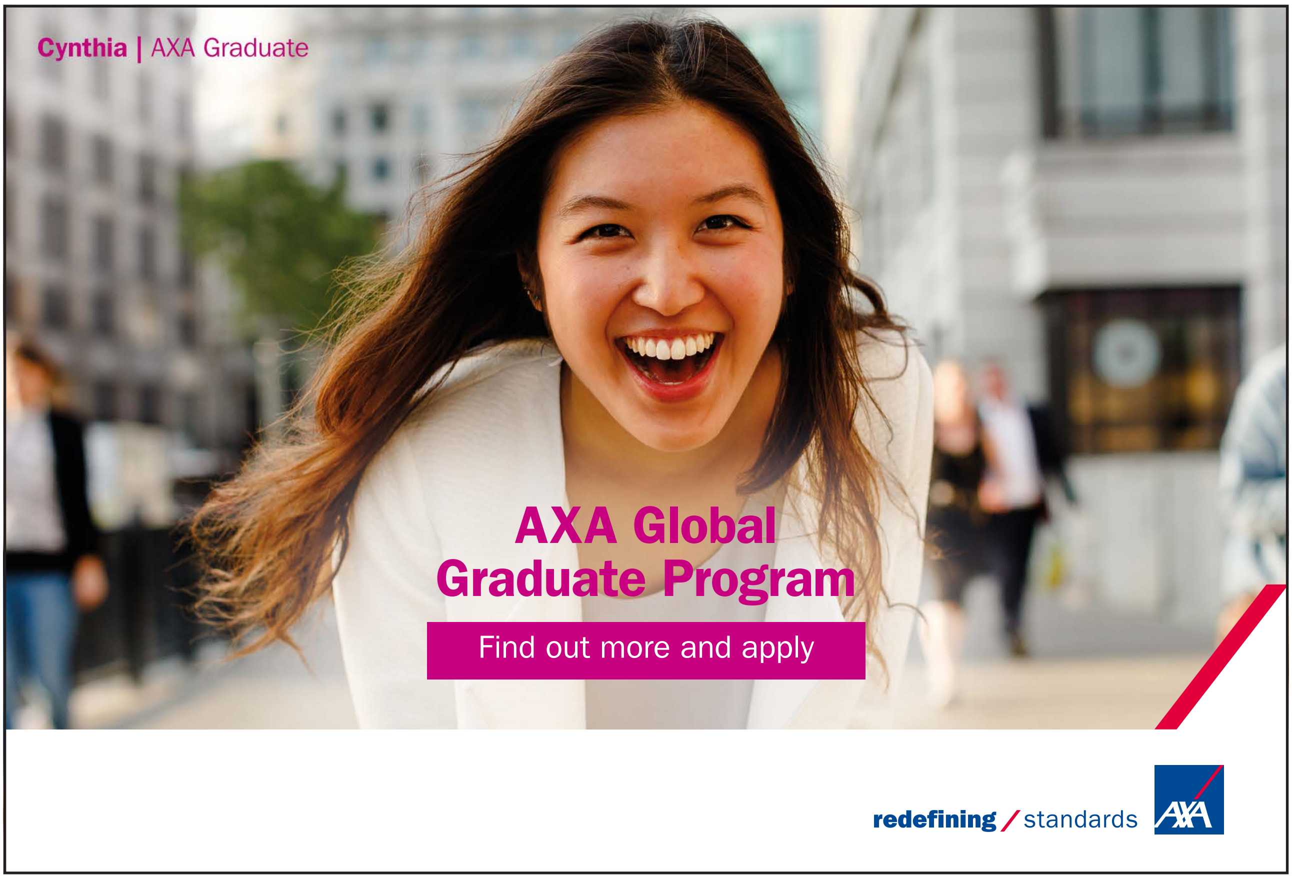
(1) For every $B \in \mathcal{B}$ we have to verify the equality:

$$\int_B X\, d\mathbb{P} = \int_B \mathbb{E}(X)\, d\mathbb{P}.$$

If $\mathbb{P}(B) = 0$, then both members are 0; if $\mathbb{P}(B) = 1$, then both members are equal to $\mathbb{E}(X)$. This proves that the constant $\mathbb{E}(X)$ can be identified with the class $\mathbb{E}\left[X \,\middle|\, \mathcal{B}\right]$.

(2) For every $B \in \mathcal{B}$ we again have to verify the equality: $\int_B X d\mathbb{P} = \int_B \mathbb{E}(X)\, d\mathbb{P}$. Employing the independence of X and $B \in \mathcal{B}$ this can be seen as follows:

$$\int_B X\, d\mathbb{P} = \int_\Omega X\mathbf{1}_B\, d\mathbb{P} = \mathbb{E}\left[X\mathbf{1}_B\right] = \mathbb{E}\left[X\right]\mathbb{E}\left[\mathbf{1}_B\right] = \int_B \mathbb{E}\left[X\right] d\mathbb{P}. \tag{1.4}$$



(3) This assertion is clear.
(4) This assertion is clear.
(5) (a) For all $B \in \mathcal{B}$ and $n \in \mathbb{N}$ we have $\int_B \mathbb{E}\left[X_n \mid \mathcal{B}\right] d\mathbb{P} = \int_B X_n \, d\mathbb{P}$. By (4) we see that the sequence of conditional expectations

$$\mathbb{E}\left[X_n \mid \mathcal{B}\right], \ n \in \mathbb{N}, \text{ increases } \mathbb{P}\text{-almost surely.}$$

The assertion in (5a) then follows from the monotone convergence theorem.

(b) Put $X_n^* = \sup_{k \geqslant n} X_k$, $X_n^{**} = \inf_{k \geqslant n} X_k$. Then we have $-Y \leqslant X_n^{**} \leqslant X_n \leqslant X_n^* \leqslant Y$, $\mathbb{P}$-almost surely. Moreover, the sequences $(Y - X_n^*)_{n\in\mathbb{N}}$ and $(Y + X_n^{**})_{n\in\mathbb{N}}$ are increasing sequences consisting of non-negative stochastic variables with $Y - \limsup_{n\to\infty} X_n$ and $Y + \liminf_{n\to\infty} X_n$ as their respective suprema. Since the sequence $(X_n)_{n\in\mathbb{N}}$ converges $\mathbb{P}$-almost surely to X, it follows by (5a) together with (4) that

$$\mathbb{E}\left[X_n^{**} \mid \mathcal{B}\right] \uparrow \mathbb{E}\left[X^{**} \mid \mathcal{B}\right] \quad \text{and} \quad \mathbb{E}\left[X_n^{*} \mid \mathcal{B}\right] \downarrow \mathbb{E}\left[X^{**} \mid \mathcal{B}\right].$$

From the pointwise inequalities $X_n^{**} \leqslant X_n \leqslant X_n^*$ it then follows that $\lim_{n\to\infty} \mathbb{E}\left[X_n \middle| \mathcal{B}\right] = \mathbb{E}\left[X \mid \mathcal{B}\right]$, $\mathbb{P}$-almost surely. Next let the uniformly integrable sequence $(X_n)_n$ in $L^1(\Omega, \mathcal{A}, \mathbb{P})$ be pointwise convergent to X. Then $\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|\right] = 0$. What we need is that

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n - X| \mid \mathcal{B}\right] = 0. \tag{1.5}$$

Under the extra hypothesis (1.2) this can be achieved as follows:

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{E}\left[|X_n - X| \mid \mathcal{B}\right] \\
&\leqslant \limsup_{n\to\infty} \mathbb{E}\left[|X_n - X|, |X_n - X| \leqslant M \mid \mathcal{B}\right] \\
&\quad + \limsup_{n\to\infty} \mathbb{E}\left[|X_n - X, |X_n - X| > M| \mid \mathcal{B}\right]
\end{aligned}$$

(apply what already has been proved in (5b), with $|X_n - X|$ instead of X_n, to the first term)

$$\leqslant \limsup_{n\to\infty} \mathbb{E}\left[|X_n - X, |X_n - X| > M| \mid \mathcal{B}\right]. \tag{1.6}$$

In (1.6) we let M tend to ∞, and employ (1.2) to conclude (1.5). This completes the proof of item (5).

(6) Write $c(x)$ as a countable supremum of affine functions

$$c(x) = \sup_{n\in\mathbb{N}} L_n(x), \tag{1.7}$$

where $L_n(z) = a_n z + b_n \leqslant c(z)$, for all those z for which $c(z) < \infty$, *i.e.* for appropriate constants a_n and b_n. Every stochastic variable $L_n(X)$ is integrable; by linearity (see (3)) we have $L_n\left(\mathbb{E}\left[X \mid \mathcal{B}\right]\right) = \mathbb{E}\left[L_n(X) \mid \mathcal{B}\right]$. Hence

$$L_n\left(\mathbb{E}\left[X \mid \mathcal{B}\right]\right) \leqslant \mathbb{E}\left[c(X) \mid \mathcal{B}\right].$$

Consequently,

$$c\left(\mathbb{E}\left[X \mid \mathcal{B}\right]\right) = \sup_{n\in\mathbb{N}} L_n\left(\mathbb{E}\left[X \mid \mathcal{B}\right]\right) \leqslant \mathbb{E}\left[c(X) \mid \mathcal{B}\right].$$

The fact that convex function can be written in the form (1.7) can be found in most books on convex analysis; see *e.g.* Chapter 3 in [**28**].

(7) It suffices to apply item (6) to the function $c(x) = |x|^p$.

(8) This assertion is clear.

(9) This assertion is also obvious.

(10) This assertion is evident if Z is a finite linear combination of indicator functions of events taken from $\mathcal{B}$. The general case follows via a limiting procedure.

(11) This assertion is clear if Y is a finite linear combination of indicator functions of events taken from $\mathcal{B}$. The general case follows via a limiting procedure.

The proof of Theorem 1.4 is now complete. □

2. Lemma of Borel-Cantelli

1.5. Definition. The *limes superior* or *upper-limit* of a sequence $(A_n)_{n\in\mathbb{N}}$ in a universe Ω is the set A of those elements $\omega \in \Omega$ with the property that ω belongs to infinitely many A_n's. In a formula:

$$A = \limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geqslant n} A_k.$$

The indicator-function $\mathbf{1}_A$ of the limes-superior of the sequence $(A_n)_{n\in\mathbb{N}}$ is equal to the lim sup of the sequence of its indicator-functions: $\mathbf{1}_A = \limsup_{n\to\infty} \mathbf{1}_{A_n}$.

The *limes inferior* or *lower-limit* of a sequence $(A_n)_{n\in\mathbb{N}}$ in a universe Ω is the set A of those elements $\omega \in \Omega$ with the property that, up to finitely many A_k's, the element (sample) ω belongs to all A_n's. In a formula:

$$A = \liminf_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} \bigcap_{k\geqslant n} A_k.$$

The indicator-function $\mathbf{1}_A$ of the limes-inferior of the sequence $(A_n)_{n\in\mathbb{N}}$ is equal to the lim inf of the sequence of its indicator-functions: $\mathbf{1}_A = \liminf_{n\to\infty} \mathbf{1}_{A_n}$.

1.6. Lemma. *Let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence of real numbers such that $0 \leqslant \alpha_n < 1$. Then $\lim_{n\to\infty} \sum_{k=1}^{n} \alpha_k < \infty$ if and only if $\lim_{n\to\infty} \prod_{k=1}^{n} (1-\alpha_k) > 0$.*

Proof. For $0 \leqslant \alpha < 1$ the following elementary inequalities hold:

$$-\frac{\alpha}{1-\alpha} \leqslant \log(1-\alpha) \leqslant -\alpha.$$

Hence we see

$$-\sum_{k=1}^{n} \frac{\alpha_k}{1-\alpha_k} \leqslant \log\left(\prod_{k=1}^{n}(1-\alpha_k)\right) \leqslant -\sum_{k=1}^{n}\alpha_k.$$

The assertion in Lemma 1.6 easily follows from these inequalities. □

1.7. Lemma (Lemma of Borel-Cantelli). *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events, and put $A = \limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geqslant n} A_k$.*

(i) *If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A) = 0$.*
(ii) *If the events A_n, $n \in \mathbb{N}$, are mutually $\mathbb{P}$-independent, then the converse statement is true as well: $\mathbb{P}(A) < 1$ implies $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$, and hence $\sum_{k=1}^{\infty} \mathbb{P}(A_k) = \infty$ if and only if $\mathbb{P}(A) = 1$.*

Proof. (i) For $\mathbb{P}(A)$ we have the following estimate:

$$\mathbb{P}(A) \leqslant \inf_{n\in\mathbb{N}} \sum_{k=n}^{\infty} \mathbb{P}(A_k). \tag{1.8}$$

Since $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, we see that the right-hand side of (1.8) is 0.

(ii) The statement in assertion (ii) is trivial if for infinitely many numbers k the equality $\mathbb{P}(A_k) = 1$ holds. So we may assume that for all $k \in \mathbb{N}$ the probability $\mathbb{P}(A_k)$ is strictly less than 1. Apply Lemma 1.6 with $\alpha_k = \mathbb{P}(A_k)$ to obtain that $\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty$ if and only if

$$0 < \lim_{n\to\infty} \prod_{k=1}^{n} (1 - \mathbb{P}(A_k)) = \lim_{n\to\infty} \prod_{k=1}^{n} \mathbb{P}(\Omega\backslash A_k)$$

(the events $(A_k)_{n\in\mathbb{N}}$ are independent)

$$= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k=1}^{n} (\Omega\backslash A_k)\right) = \lim_{n\to\infty} \mathbb{P}\left(\Omega\backslash \bigcup_{k=1}^{n} A_k\right) = 1 - \mathbb{P}(A). \tag{1.9}$$

This proves assertion (ii) of Lemma 1.7. □

3. Stochastic processes and projective systems of measures

1.8. Definition. Consider a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and an index set I. Suppose that for every $t \in I$ a measurable space $(E_t, \mathcal{E}_t)$ and an $\mathcal{A}$-$\mathcal{E}_t$-measurable mapping $X(t) : \Omega \to E_t$ are given. Such a family $\{X(t) : t \in I\}$ is called a stochastic process.

1.9. Remark. The space Ω is often called the *sample path space*, the space E_t is often called the *state space* of the *state variable* $X(t)$. The σ-field $\mathcal{A}$ is often replaced with (some completion of) the σ-field generated by the state variables $X(t)$, $t \in I$. This σ-field is written as $\mathcal{F}$. Let $(S, \mathcal{S})$ be some measurable space. An $\mathcal{F}$-$\mathcal{S}$ measurable mapping $Y : \Omega \to S$ is called an S-valued *stochastic variable*. Very often the state spaces are the same, i.e. $(E_t, \mathcal{E}_t) = (E, \mathcal{E})$, for all state variables $X(t)$, $t \in I$.

In applications the index set I is often interpreted as the time set. So I can be a finite index set, e.g. $I = \{0, 1, \ldots, n\}$, or an infinite discrete time set, like $I = \mathbb{N} = \{0, 1, \ldots\}$ or $I = \mathbb{Z}$. The set I can also be a *continuous time* set: $I = \mathbb{R}$ or $I = \mathbb{R}^+ = [0, \infty)$. In the present text, most of the time we will consider $I = [0, \infty)$. Let I be $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, or $[0, \infty)$. In the so-called *time-homogeneous* or *stationary* case we also consider mappings $\vartheta_s : \Omega \to \Omega$, $s \in I$, $s \geqslant 0$, such that $X(t) \circ \vartheta_s = X(t+s)$, $\mathbb{P}$-almost surely. It follows that these translation mappings $\vartheta_s : \Omega \to \Omega$, $s \in I$, are $\mathcal{F}_t$-$\mathcal{F}_{t-s}$-measurable, for all $t \geqslant s$. If Y is a stochastic variable, then $Y \circ \vartheta_s$ is measurable with respect to the σ-field $\sigma\{X(t) : t \geqslant s\}$. The concept of time-homogeneity of the process $(X(t) : t \in I)$ can be explained as follows. Let $Y : \Omega \to \mathbb{R}$ be a stochastic variable; e.g. $Y = \prod_{j=1}^n f_j(X(t_j))$, where $f_j : E \to \mathbb{R}$, $1 \leqslant j \leqslant n$, are bounded measurable functions. Define the transition probability $P(s, B)$ as follows: $P(s, B) = \mathbb{P}(X(s) \in B)$, $s \in I$, $B \in \mathcal{E}$. The measure $B \mapsto \mathbb{E}[Y \circ \vartheta_s, X(s) \in B]$ is absolutely continuous with respect to the measure $B \mapsto P(s, B)$, $B \in \mathcal{E}$. It follows that there exists a function $F(s, x)$, called the Radon-Nikodym derivative of the measure $B \mapsto \mathbb{E}[Y \circ \vartheta_s, X(s) \in B]$ with respect $B \mapsto P(s, B)$, such that $\mathbb{E}[Y \circ \vartheta_s, X(s) \in B] = \int F(s, x) P(s, dx)$. The function $F(s, x)$ is usually written as

$$F(s, x) = \mathbb{E}\left[Y \circ \vartheta_s \,\middle|\, X(s) \in dx\right] = \frac{\mathbb{E}[Y \circ \vartheta_s, X(s) \in dx]}{\mathbb{P}[X(s) \in dx]}.$$

1.10. Definition. The process $(X(t) : t \in I)$ is called *time-homogeneous* or *stationary in time*, provided that for all bounded stochastic variables $Y : \Omega \to \mathbb{R}$ the function

$$\mathbb{E}\left[Y \circ \vartheta_s \,\middle|\, X(s) \in dx\right] \quad \text{is independent of } s \in I,\ s \geqslant 0.$$

In practice we only have to verify the property in Definition 1.10 for Y of the form $Y = \prod_{j=1}^n f_j(X(t_j))$, where $f_j : E_{t_j} \to \mathbb{R}$, $1 \leqslant j \leqslant n$, are bounded measurable functions. Then $Y \circ \vartheta_s = \prod_{j=1}^n f_j(X(t_j + s))$. This statement is a consequence of the monotone class theorem.

3.1. Finite dimensional distributions. As above let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\{X(t) : t \in I\}$ be a stochastic process where each state variable $X(t)$ has state space $(E_t, \mathcal{E}_t)$. For every non-empty subset J of I we write $E^J = \prod_{t \in J} E_t$ and $\mathcal{E}^J = \otimes_{t \in J} \mathcal{E}_t$ denotes the product-field. We also write $X_J = \otimes_{t \in J} X_t$. So that, if $J = \{t_1, \ldots, t_n\}$, then $X_J = (X(t_1), \ldots, X(t_n))$. The mapping X_J is the product mapping from Ω to E^J. The mapping $X_J : \Omega \to E^J$

is $\mathcal{A}$-$\mathcal{E}^J$-measurable. We can use it to define the image measure $\mathbb{P}_J$:

$$\mathbb{P}_J(B) = X_J\mathbb{P}(B) = \mathbb{P}\left[X_J^{-1}B\right] = \mathbb{P}\left[\omega \in \Omega : X_J(\omega) \in B\right] = \mathbb{P}\left[X_J \in B\right],$$

where $B \in \mathcal{E}^J$. Between the different probability spaces $\left(E^J, \mathcal{E}^J, \mathbb{P}_J\right)$ there exist relatively simple relationships. Let J and H be non-empty subsets of I such that $J \subset H$, and consider the $\mathcal{E}^H$-$\mathcal{E}^J$-measurable projection mapping $p_J^H : E^H \to E^J$, which "forgets" the "coordinates" in $H\backslash J$. If $H = I$, then we write $p_J = p_J^I$. For every pair J and H with $J \subset H \subset I$ we have $X_J = p_J^H \circ X_H$, and hence we get $\mathbb{P}_J(B) = p_J^H\mathbb{P}_H(B) = \mathbb{P}_H\left[p_J^H \in B\right]$, where B belongs to $\mathcal{E}^J$. In particular if $H = I$, then $\mathbb{P}_J(B) = p_J\mathbb{P}(B) = \mathbb{P}\left[p_J \in B\right]$, where B belongs to $\mathcal{E}_J$. If $J = \{t_1, \dots, t_n\}$ is a finite set, then we have

$$\begin{aligned}\mathbb{P}_J\left[B_1 \times \cdots \times B_n\right] &= \mathbb{P}\left[X_J^{-1}\left(B_1 \times \cdots \times B_n\right)\right] \\ &= \mathbb{P}\left[X(t_1) \in B_1, \dots, X(t_n) \in B_n\right],\end{aligned}$$

with $B_j \in \mathcal{E}_{t_j}$, for $1 \leqslant j \leqslant n$.

1.11. Remark. If the process $\{X(t) : t \in I\}$ is interpreted as the movement of a particle, which at time t happens to be in the state spaces E_t, and if $J = \{t_1, \dots, t_n\}$ is a finite subset of I, then the probability measure $\mathbb{P}_J$ has the following interpretation:

> For every collection of sets $B_1 \in \mathcal{E}_{t_1}, \dots, B_n \in \mathcal{E}_{t_n}$ the number
> $$\mathbb{P}_J\left[B_1 \times \cdots \times B_n\right]$$
> is the probability that at time t_1 the particle is in B_1, at time t_2 it is in B_2, ..., and at time t_n it is in B_n.

1.12. Definition. Let $\mathcal{H}$ be the collection of all finite subsets of I. Then the family $\left\{\left(E^J, \mathcal{E}^J, \mathbb{P}_J\right) : J \in \mathcal{H}\right\}$ is called the *family of finite-dimensional distributions* of the process $\{X(t) : t \in I\}$; the one-dimensional distributions

$$\left\{\left(E_t, \mathcal{E}_t, \mathbb{P}_{\{t\}}\right) : t \in I\right\}$$

are often called the *marginals* of the process.

The family of finite-dimensional distributions is a projective or consistent family in the sense as explained in the following definition.

1.13. Definition. A family of probability spaces $\left\{\left(E^J, \mathcal{E}^J, \mathbb{P}_J\right) : J \in \mathcal{H}\right\}$ is called a *projective*, a *consistent system*, or a *cylindrical* measure provided that

$$\mathbb{P}_J(B) = p_J^H(\mathbb{P}_H)(B) = \mathbb{P}_H\left[p_J^H \in B\right]$$

for all finite subsets $J \subset H$, J, $H \in \mathcal{H}$, and for all sets $B \in \mathcal{E}^J$.

1.14. Theorem (Theorem of Kolmogorov). *Let $\left\{\left(E^J, \mathcal{E}^J, \mathbb{P}_J\right) : J \in \mathcal{H}\right\}$ be a projective system of probability spaces. Suppose that every space E_t is a σ-compact metrizable Hausdorff space. Then there exists a unique probability space $\left(E^I, \mathcal{E}^I, \mathbb{P}_I\right)$ with the property that for all finite subsets $J \in \mathcal{H}$ the equality $\mathbb{P}_J(B) = \mathbb{P}_I[p_J \in B]$ holds for all $B \in \mathcal{E}^J$.*

Theorem 5.81 is the same as Theorem 1.14, but formulated for Polish and Souslin spaces; its proof can be found in Chapter 5. Theorem 1.14 is the same as Theorem 3.1. The reason that the conclusion in Theorem 1.14 holds for σ-compact metrizable topological Hausdorff spaces is the fact that a finite Borel measure μ on a metrizable σ-compact space E is regular in the sense that

$$\mu(B) = \sup_{K \subset B,\, K \text{ compact}} \mu(K) = \inf_{U \supset K,\, U \text{ open}} \mu(U), \quad B \text{ any Borel subset } E. \tag{1.10}$$

1.15. Lemma. *Let E be a σ-compact metrizable Hausdorff space. Then the equality in (1.10) holds for all Borel subsets B of E.*

Proof. The equalities in (1.10) can be deduced by proving that the collection $\mathcal{D}$ define by

$$\begin{aligned} \mathcal{D} &= \left\{B \in \mathcal{B}_E : \sup_{K \subset B} \mu(K) = \inf_{U \supset B} \mu(U)\right\} \\ &= \left\{B \in \mathcal{B}_E : \sup_{F \subset B} \mu(F) = \inf_{U \supset B} \mu(U)\right\} \end{aligned} \tag{1.11}$$

contains the open subsets of E, is closed under taking complements, and is closed under taking mutually disjoint countable unions. The second equality holds because every closed subset of E is a countable union of compact subsets. In (1.11) the sets K are taken from the compact subsets, the sets U from the open subsets, and the sets F from the closed subsets of E. It is clear that $\mathcal{D}$ is closed under taking complements. Let $(x, y) \mapsto d(x, y)$ be a metric on E which is compatible with its topology. Let F be a closed subset of E, and define U_n by

$$U_n = \left\{ x \in E : \inf_{y \in F} d(x, y) < \frac{1}{n} \right\}.$$

Then the subset U_n is open, $U_{n+1} \supset U_n$, and $F = \bigcap U_n$. It follows that $\mu(F) = \inf_n \mu(U_n)$, and consequently, F belongs to $\mathcal{D}$. In other words the collection $\mathcal{D}$ contains the closed, and so the open subsets of E. Next let $(B_n)_n$ be a sequence of subsets in $\mathcal{D}$. Fix $\varepsilon > 0$, and choose closed subsets $F_n \subset B_n$, and open subsets $U_n \supset B_n$, such that

$$\mu(B_n \backslash F_n) \leqslant \varepsilon 2^{-n-1}, \quad \text{and} \quad \mu(U_n \backslash B_n) \leqslant \varepsilon 2^{-n}. \tag{1.12}$$

From (1.12) it follows that

$$\begin{aligned} \mu\left(\left(\bigcup_{n=1}^{\infty} U_n\right) \backslash \left(\bigcup_{n=1}^{\infty} B_n\right)\right) &\leqslant \mu\left(\bigcup_{n=1}^{\infty} (U_n \backslash B_n)\right) \\ &\leqslant \sum_{n=1}^{\infty} \mu(U_n \backslash B_n) \leqslant \varepsilon \sum_{n=1}^{\infty} 2^{-n} = \varepsilon. \end{aligned} \tag{1.13}$$

From (1.13 it follows that

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \inf\left\{\mu(U) : U \supset \bigcup_{n=1}^{\infty} B_n, \ \ U \text{ open}\right\}. \tag{1.14}$$

The same argumentation shows that

$$\mu\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \backslash \left(\bigcup_{n=1}^{\infty} F_n\right)\right) \leqslant \sum_{n=1}^{\infty} \mu(B_n \backslash F_n) \leqslant \varepsilon \sum_{n=1}^{\infty} 2^{-n-1} = \frac{1}{2}\varepsilon. \tag{1.15}$$

From (1.15) it follows that

$$\mu\left(\left(\bigcup_{n=1}^{\infty} B_n\right) \backslash \left(\bigcup_{n=1}^{N_\varepsilon} F_n\right)\right) \leqslant \varepsilon \tag{1.16}$$

for N_ε large enough. From (1.16 it follows that

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sup\left\{\mu(F) : F \subset \bigcup_{n=1}^{\infty} B_n, \ \ F \text{ closed}\right\}. \tag{1.17}$$

From (1.14) and (1.17) it follows that $\bigcup_{n=1}^{\infty} B_n$ belongs to $\mathcal{D}$. As already mentioned, since every closed subset is the countable union of compact subsets the supremum over closed subsets in (1.17) may replaced with a supremum over compact subsets. Altogether, this completes the proof of Lemma 1.15. □

It is a nice observation that a locally compact Hausdorff space is metrizable and σ-compact if and only if it is a Polish space. This is part of Theorem 5.3 (page 29) in Kechris [**68**]. This theorem reads as follows.

1.16. Theorem. *Let E be a locally compact Hausdorff space. The following assertions are equivalent:*

(1) *The space E is second countable,* i.e. *E has a countable basis for its topology.*
(2) *The space E is metrizable and σ-compact.*
(3) *The space E has a metrizable one-point compactification (or Alexandroff compactification).*
(4) *The space E is Polish,* i.e. *E is complete metrizable and separable.*
(5) *The space E is homeomorphic to an open subset of a compact metrizable space.*

A second-countable locally-compact Hausdorff space is Polish: let $(U_i)_i$ be a countable basis of open subsets with compact closures $(K_i)_i$, and let V_i be an open subset with compact closure and containing K_i. From Urysohn's Lemma, let $0 \leqslant f_i \leqslant 1$ be continuous functions identically 0 off V_i, identically 1 on K_i, and put

$$d(x,y) = \sum_{i=1}^{\infty} 2^{-i} \left|f_i(x) - f_i(y)\right| + \left| \frac{1}{\sum_{i=1}^{\infty} 2^{-i} f_i(x)} - \frac{1}{\sum_{i=1}^{\infty} 2^{-i} f_i(y)} \right|, \quad x,\, y \in E. \tag{1.18}$$

The triangle inequality for the usual absolute value shows that this is a metric. This metric gives the same topology, and it is straightforward to verify its completeness. For this argument see Garrett [**57**].

4. A definition of Brownian motion

In this section we give a (preliminary) definition of Brownian motion.

4.1. Gaussian measures on $\mathbb{R}^d$. For every $t > 0$ we define the *Gaussian kernel* on $\mathbb{R}^d$ as the function

$$p_d(t,x,y) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x-y|^2}{2t}\right).$$

Then we have $\int p_d(t,x,z)\,dz = 1$, and

$$p_d(s,x,z)p(t,z,y) = p_d(s+t,x,y)p_d\left(\frac{st}{s+t}, \frac{sx+ty}{s+t}, z\right).$$

Hence the function $p_d(t,x,y)$ satisfies the equation of Chapman-Kolmogorov:

$$\int p_d(s,x,z)p_d(t,z,y)dz = p_d(s+t,x,y).$$

This property will enable us to consider d-dimensional Brownian motion as a Markov process. Next we calculate the finite-dimensional distributions of the Brownian motion.

4.2. Finite dimensional distributions of Brownian motion. Let $0 < t_1 < \cdots < t_n < \infty$ be a sequence of time instances in $(0,\infty)$, and fix $x_0 \in \mathbb{R}^d$. Define the probability measure $\mathbb{P}_{x_0;t_1,\dots,t_n}$ on the Borel field of $\mathbb{R}^d \times \cdots \times \mathbb{R}^d$ (n times) by ($t_0 = 0$)

$$\mathbb{P}_{x_0;t_1,\dots,t_n}[B_1 \times \cdots \times B_n] = \int_{B_1} \dots \int_{B_n} dx_n \dots dx_1 \prod_{j=1}^{n} p_d(t_j - t_{j-1}, x_{j-1}, x_j), \tag{1.19}$$

where $B_1, \dots, B_n$ are Borel subsets of $\mathbb{R}^d$. Then, with $B_k = \mathbb{R}^d$, we have

$$\begin{aligned}
&\mathbb{P}_{x_0;t_1,\dots,t_{k-1},t_k,t_{k+1},\dots,t_n}\left[B_1 \times \cdots \times B_{k-1} \times \mathbb{R}^d \times B_{k+1} \times \cdots \times B_n\right] \\
&= \int_{B_1} \dots \int_{B_{k-1}} \int_{\mathbb{R}^d} \int_{B_{k+1}} \dots \int_{B_n} dx_n \dots dx_{k+1}\, dx_k\, dx_{k-1} \dots dx_1 \\
&\quad \prod_{j=1}^{k-1} p_d(t_j - t_{j-1}, x_{j-1}, x_j) \\
&\quad p_d(t_k - t_{k-1}, x_{k-1}, x_k)\, p_d(t_{k+1} - t_k, x_k, x_{k+1}) \prod_{j=k+2}^{n} p(t_j - t_{j-1}, x_{j-1}, x_j)
\end{aligned}$$

(Chapman-Kolmogorov)

$$\begin{aligned}
&= \int_{B_1} \dots \int_{B_{k-1}} \int_{B_{k+1}} \dots \int_{B_n} dx_n \dots dx_{k+1}\, dx_{k-1} \dots dx_1 \prod_{j=1}^{k-1} p_d(t_j - t_{j-1}, x_{j-1}, x_j) \\
&\quad p_d(t_{k+1} - t_{k-1}, x_{k-1}, x_{k+1}) \prod_{j=k+2}^{n} p(t_j - t_{j-1}, x_{j-1}, x_j)
\end{aligned}$$

$$= \mathbb{P}_{x_0;t_1,\dots,t_{k-1},t_{k+1},\dots,t_n}\left[B_1 \times \cdots \times B_{k-1} \times B_{k+1} \times \cdots \times B_n\right]. \tag{1.20}$$

It follows that the family

$$\left\{ \left(\underbrace{\mathbb{R}^d \times \cdots \times \mathbb{R}^d}_{n \text{ times}}, \underbrace{\mathcal{B}^d \otimes \cdots \otimes \mathcal{B}^d}_{n \text{ times}}, \mathbb{P}_{x_0;t_1,\dots,t_n} \right); 0 < t_1 < \cdots < t_n < \infty, \, n \in \mathbb{N} \right\}$$

is a projective or consistent system. Such families are also called cylindrical measures. The extension theorem of Kolmogorov implies that in the present situation a cylindrical measure can be considered as a genuine measure on the product field of $\Omega := \left(\mathbb{R}^d\right)^{[0,\infty)}$. This is the measure corresponding to Brownian motion starting at x_0. More precisely, the theorem of Kolmogorov says that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P}_{x_0})$ and state variables $X(t) : \Omega \to \mathbb{R}^d$, $t \geqslant 0$, such that

$$\mathbb{P}_{x_0}\left[X\left(t_1\right) \in B_1, \dots, X\left(t_n\right) \in B_n\right] = \mathbb{P}_{x_0;t_1,\dots,t_n}\left[B_1 \times \cdots \times B_n\right],$$

where the subsets B_j, $1 \leqslant j \leqslant n$, belong to $\mathcal{B}^d$. It is assumed that

$$\mathbb{P}_{x_0}\left[X(0) = x_0\right] = 1.$$

5. Martingales and related processes

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{\mathcal{F}_t : t \in I\}$ be a family of subfields of $\mathcal{F}$, indexed by a totally ordered index set $(I, \leqslant)$. Suppose that the family $\{\mathcal{F}_t : t \in I\}$ is increasing in the sense that $s \leqslant t$ implies $\mathcal{F}_s \subseteq \mathcal{F}_t$. Such a family of σ-fields is called a *filtration*. A stochastic process $\{X(t) : t \in I\}$, where $X(t)$, $t \in I$, are mappings from Ω to E_t, is called *adapted*, or more precisely, *adapted* to the filtration $\{\mathcal{F}_t : t \in I\}$ if every $X(t)$ is $\mathcal{F}_t$-$\mathcal{E}_t$-measurable. For the σ-field $\mathcal{F}_t$ we often take (some completion of) the σ-field generated by $X(s)$, $s \leqslant t$: $\mathcal{F}_t = \sigma\{X(s) : s \leqslant t\}$.

1.17. Definition. An adapted process $\{X(t) : t \in I\}$ with state space $(\mathbb{R}, \mathcal{B})$ is called a *super-martingale* if every variable $X(t)$ is $\mathbb{P}$-integrable, and if $s \leqslant t$, $s,\ t \in I$, implies $\mathbb{E}\left[X(t) \mid \mathcal{F}_s\right] \leqslant X(s)$, $\mathbb{P}$-almost surely. An adapted process $\{X(t) : t \in I\}$ with state space $(\mathbb{R}, \mathcal{B})$ is called a *sub-martingale* if every variable $X(t)$ is $\mathbb{P}$-integrable, and if $s \leqslant t$, $s,\ t \in I$, implies $\mathbb{E}\left[X(t) \mid \mathcal{F}_s\right] \geqslant X(s)$, $\mathbb{P}$-almost surely. If an adapted process is at the same time a super- and a sub-martingale, then it is called a *martingale*.

The martingale in the following example is called a *closed* martingale.

1.18. Example. Let X_∞ belong to $L^1(\Omega, \mathcal{F}, \mathbb{P})$, and let $\{\mathcal{F}_t : t \in [0, \infty)\}$ be a filtration in $\mathcal{F}$. Put $X(t) = \mathbb{E}\left[X_\infty \mid \mathcal{F}_t\right]$, $t \geqslant 0$. Then the process $\{X(t) : t \geqslant 0\}$ is a martingale with respect to the filtration $\{\mathcal{F}_t : t \in [0, \infty)\}$.

The following theorem shows that uniformly integrable martingales are closed martingales.

1.19. Theorem (Doob's theorem). *Any uniformly integrable martingale*

$$\{X(t): t \geqslant 0\} \quad in \quad L^1(\Omega, \mathcal{F}, \mathbb{P})$$

converges $\mathbb{P}$-almost surely and in mean (i.e. in $L^1(\Omega, \mathcal{F}, \mathbb{P})$) to a stochastic variable X_∞ such that for every $t \geqslant 0$ the equality $X(t) = \mathbb{E}\left[X_\infty \mid \mathcal{F}_t\right]$ holds $\mathbb{P}$-almost surely.

Let F be a subset of $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then F is uniformly integrable if for every $\varepsilon > 0$ there exists a function $g \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $\int_{\{|f| \geqslant |g|\}} |f| \, d\mathbb{P} \leqslant \varepsilon$ for all $f \in F$. Since $\mathbb{P}$ is a finite positive measure we may assume that g is a (large) positive constant.

1.20. Theorem. *Sub-martingales constitute a convex cone:*

(i) *A positive linear combination of sub-martingales is again a sub-martingale; the space of sub-martingales forms a convex cone.*
(ii) *A convex function of a sub-martingale is a sub-martingale.*

Not all martingales are closed, as is shown in the following example.

1.21. Example. Fix $t > 0$, and $x, y \in \mathbb{R}^d$. Let

$$\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t), t \geqslant 0), (\vartheta_t : t \geqslant 0), (\mathbb{R}^n, \mathcal{B}^n)\}$$

be Brownian motion starting at $x \in \mathbb{R}^d$, and put, as above,

$$p_d(t, x, y) = \frac{1}{(2\pi t)^{d/2}} \exp\left(-\frac{|x-y|^2}{2t}\right).$$

The process $s \mapsto p(t-s, X(s), y)$ is $\mathbb{P}_x$-martingale on the half-open interval $[0, t)$.

5.1. Stopping times. A stochastic variable $T : \Omega \to [0, \infty]$ is called a stopping time with respect to the filtration $\{\mathcal{F}_t : t \geqslant 0\}$, if for every $t \geqslant 0$ the event $\{T \leqslant t\}$ belongs to $\mathcal{F}_t$. If T is a stopping time, the process $t \mapsto \mathbf{1}_{[T \leqslant t]}$ is adapted to $\{\mathcal{F}_t : t \geqslant 0\}$. The meaning of a stopping is the following one. The moment T is the time that some phenomena happens. If at a given time t the information contained in $\mathcal{F}_t$ suffices to conclude whether or not this phenomena occurred before time t, then T is a stopping time. Let

$$\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t), t \geqslant 0), (\vartheta_t : t \geqslant 0), (\mathbb{R}^n, \mathcal{B}^n)\}$$

be Brownian motion starting at $x \in \mathbb{R}^d$, let $p : \mathbb{R}^d \to (0, \infty)$ be a strictly positive continuous function, and O an open subset of $\mathbb{R}^d$. The first exit time from O, or the first hitting time of the complement of O, defined by

$$T = \inf\left\{t > 0 : X(t) \in \mathbb{R}^d \backslash O\right\}$$

is a (very) relevant stopping time. The time T is a so-called *terminal* stopping time: on the event $\{T > s\}$ it satisfies $s + T \circ \vartheta_s = T$. Other relevant stopping times are:

$$\tau_\xi = \inf\left\{t > 0 : \int_0^t p(X(s)) \, ds > \xi\right\}, \quad \xi \geqslant 0.$$

Such stopping times are used for (stochastic) time change:

$$\tau_\xi + \tau_\eta \circ \vartheta_{\tau_\xi} = \tau_{\xi+\eta}, \quad \xi, \quad \eta \geqslant 0.$$

Note that the mapping $\xi \mapsto \tau_\xi$ is the inverse of the mapping $t \mapsto \int_0^t p\,(X(s))\,ds$. Also note the equality: $\{\tau_\xi < t\} = \left\{\int_0^t p\,(X(s)) > \xi\right\}$, $\int_0^\infty p\,(X(s))\,ds > \xi > 0$. The mapping $\xi \mapsto \tau_\xi$ is strictly increasing from the interval $\left[0, \int_0^\infty p\,(X(s))\,ds\right)$ onto $[0,\infty)$. Arbitrary stopping times T are often approximated by "discrete" stopping times: $T = \lim_{n\to\infty} T_n$, where $T_n = 2^{-n}\left\lceil 2^n T\right\rceil$. Notice that $T \leqslant T_{n+1} \leqslant T_n \leqslant T + 2^{-n}$, and that $\{T_n = k2^{-n}\} = \{(k-1)2^{-n} < T \leqslant k2^{-n}\}$, $k \in \mathbb{N}$.

1.22. Theorem. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{\mathcal{F}_t : t \geqslant 0\}$ be a filtration in $\mathcal{F}$. The following assertions hold true:*

(1) *constant times are stopping times: for every $t \geqslant 0$ fixed the time $T \equiv t$ is a stopping time;*
(2) *if S and T are stopping times, then so are $\min(S,T)$ and $\max(S,T)$;*
(3) *If T is a stopping time, then the collection $\mathcal{F}_T$ defined by*

$$\mathcal{F}_T = \{A \in \mathcal{F} :\ A \cap \{T \leqslant t] \in \mathcal{F}_t,\ \text{for all } t \geqslant 0\}$$

is a subfield of $\mathcal{F}$;
(4) *If S and T are stopping times, then $S + T\circ\vartheta_S$ is a stopping time as well, provided the paths of the process are $\mathbb{P}$-almost surely right-continuous and the same is true for the filtration $\{\mathcal{F}_t : t \geqslant 0\}$.*

The filtration $\{\mathcal{F}_t : t \geqslant 0\}$ is right-continuous if $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$, $t \geqslant 0$. The (sample) paths $t \mapsto X(t)$ are said to be $\mathbb{P}$-almost surely right-continuous, provided for all $t \geqslant 0$ we have $X(t) = \lim_{s\downarrow t} X(s)$, $\mathbb{P}$-almost surely.

The following theorem shows that in many cases fixed times can be replaced with stopping times. In particular this is true if we study (right-continuous) sub-martingales, super-martingales or martingales.

1.23. THEOREM (Doob's optional sampling theorem). *Let $(X(t) : t \geqslant 0)$ be a uniformly integrable process in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ which is a sub-martingale with respect to the filtration $(\mathcal{F}_t : t \geqslant 0)$. Let S and T be stopping times such that $S \leqslant T$. Then $\mathbb{E}\left[X(T) \mid \mathcal{F}_S\right] \geqslant X(S)$, $\mathbb{P}$-almost surely.*

Similar statements hold for super-martingales and martingales.

Notice that $X(T)$ stands for the stochastic variable $\omega \mapsto X(T(\omega))(\omega) = X(T(\omega), \omega)$.

We conclude this introduction with a statement of the decomposition theorem of Doob-Meyer. A process $\{X(t) : t \geqslant 0\}$ is of class (DL) if for every $t > 0$ the family

$$\{X(\tau) : 0 \leqslant \tau \leqslant t, \tau \text{ is an } (\mathcal{F}_t)\text{-stopping time}\}$$

is uniformly integrable. An $\mathcal{F}_t$-martingale $\{M(t) : t \geqslant 0\}$ is of class (DL), an increasing adapted process $\{A(t) : t \geqslant 0\}$ in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ is of class (DL) and hence the sum $\{M(t) + A(t) : t \geqslant 0\}$ is of class (DL). If $\{X(t) : t \geqslant 0\}$ is a submartingale and if μ is a real number, then the process $\{\max(X(t), \mu) : t \geqslant 0\}$ is a sub-martingale of class (DL). Processes of class (DL) are important in the Doob-Meyer decomposition theorem. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\{\mathcal{F}_t : t \geqslant 0\}$ be a right-continuous filtration in $\mathcal{F}$ and let $\{X(t) : t \geqslant 0\}$ be right continuous sub-martingale of class (DL) which possesses almost sure left limits. We mention the following version of the Doob-Meyer decomposition theorem. See Remark 3.54 as well.

1.24. THEOREM. *Let $\{X(t) : t \geqslant 0\}$ be a sub-martingale of class (DL) which has $\mathbb{P}$ almost surely left limits, and which is right-continuous. Then there exists a unique predictable right continuous increasing process $\{A(t) : t \geqslant 0\}$ with $A(0) = 0$ such that the process $\{X(t) - A(t) : t \geqslant 0\}$ is an $\mathcal{F}_t$-martingale.*

A process $(\omega, t) \mapsto X(t)(\omega) = X(t, \omega)$ is *predictable* if it is measurable with respect to the σ-field generated by $\{A \times (a, b] : A \in \mathcal{F}_a, a < b\}$. For more details on càdlàg sub-martingales, see Theorem 3.77. The following proposition says that a non-negative right-continuous sub-martingale is of class (DL).

1.25. PROPOSITION. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(\mathcal{F}_t)_{t \geqslant 0}$ be a filtration of σ-fields contained in $\mathcal{F}$. Suppose that $t \mapsto X(t)$ is a right-continuous sub-martingale relative to the filtration $(\mathcal{F}_t)_{t \geqslant 0}$ attaining its values in $[0, \infty)$. Then the family $\{X(t) : t \geqslant 0\}$ is of class (DL).*

In fact it suffices to assume that there exists a real number m such that $X(t) \geqslant -m$ $\mathbb{P}$-almost surely. This follows from Proposition 1.25 by considering $X(t)+m$ instead of $X(t)$.

If $t \mapsto M(t)$ is a continuous martingale in $L^2(\Omega, \mathcal{F}, \mathbb{P})$, then $t \mapsto |M(t)|^2$ is a non-negative sub-martingale, and so it splits as the sum of a martingale $t \mapsto$

$|M(t)|^2 - \langle M, M \rangle (t)$ and an increasing process $t \mapsto \langle M < M \rangle (t)$, the quadratic variation process of $M(t)$.

Proof of Proposition 1.25. Fix $t > 0$, and let $\tau : \Omega \to [0, t]$ be a stopping time. Let for $m \in \mathbb{N}$ the stopping time $\tau_m : \Omega \to [0, \infty]$ be defined by $\tau_m = \inf \{s > 0 : X(s) > m\}$ if $X(s) > m$ for some $s < \infty$, otherwise $\tau_m = \infty$. Then the event $\{X(\tau) > m\}$ is contained in the event $\{\tau_m \leqslant \tau\}$. Hence,

$$\begin{aligned} &\mathbb{E}\left[X(\tau) : X(\tau) > m\right] \leqslant \mathbb{E}\left[X(t) : X(\tau) > m\right] \leqslant \mathbb{E}\left[X(t) : \tau_m \leqslant \tau\right] \\ &\leqslant \mathbb{E}\left[X(t) : \tau_m \leqslant t\right]. \end{aligned} \tag{1.21}$$

Since, $\mathbb{P}$-almost surely, $\tau_m \uparrow \infty$ for $m \to \infty$, it follows that

$$\lim_{m \to \infty} \sup \left\{\mathbb{E}\left[X(\tau) : X(\tau) > m\right] : \tau \in [0, t] : \tau \text{ stopping time}\right\} = 0.$$

Consequently, the sub-martingale $t \mapsto X(t)$ is of class (DL). The proof of Proposition 1.25 is complete now. □

It is perhaps useful to insert the following proposition.

1.26. Proposition. *Processes of the form $M(t) + A(t)$, with $M(t)$ a martingale and with $A(t)$ an increasing process in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ are of class (DL).*

Proof. Let $\{X(t) = M(t) + A(t) : t \geqslant 0\}$ be the decomposition of the sub-martingale $\{X(t) : t \geqslant 0\}$ in a martingale $\{M(t) : t \geqslant 0\}$ and an increasing process $\{A(t) : t \geqslant 0\}$ with $A(0) = 0$ and $0 \leqslant \tau \leqslant t$ be any $\mathcal{F}_t$-stopping time. Here t is some fixed time. For $N \in \mathbb{N}$ we have

$$\begin{aligned} \mathbb{E}\left(|X(\tau)| : |X(\tau)| \geqslant N\right) &\leqslant \mathbb{E}\left(|M(\tau)| : |X(\tau)| \geqslant N\right) + \mathbb{E}\left(A(\tau) : |X(\tau)| \geqslant N\right) \\ &\leqslant \mathbb{E}\left(|M(t)| : |X(\tau)| \geqslant N\right) + \mathbb{E}\left(A(\tau) : |X(\tau)| \geqslant N\right) \\ &\leqslant \mathbb{E}\left(|M(t)| + A(t) : |X(\tau)| \geqslant N\right) \\ &\leqslant \mathbb{E}\left(|M(t)| + A(t) : \sup_{0 \leqslant s \leqslant t} |X(s)| \geqslant N\right). \end{aligned}$$

Since, by the Doob's maximality theorem 1.28,

$$\begin{aligned} &N\mathbb{P}\left\{\sup_{0 \leqslant s \leqslant t} |X(s)| \geqslant N\right\} \leqslant N\mathbb{P}\left\{\sup_{0 \leqslant s \leqslant t} |M(s)| \geqslant \frac{N}{2}\right\} + N\mathbb{P}\left\{\sup_{0 \leqslant s \leqslant t} A(s) \geqslant \frac{N}{2}\right\} \\ &\leqslant 2\mathbb{E}\left(|M(t)| + A(t)\right), \end{aligned}$$

it follows that

$$\lim_{N \to \infty} \sup \left\{\mathbb{E}\left(|X(\tau)| : |X(\tau)| \geqslant N\right) : 0 \leqslant \tau \leqslant t, \ \tau \ \text{ stopping time}\right\} = 0.$$

This proves Proposition 1.26. □

First we formulate and prove Doob's maximal inequality for time-discrete sub-martingales. In Theorem 1.27 the sequence $i \mapsto X_i$ is defined on a filtered probability space $(\Omega, \mathcal{F}_i.\mathbb{P})_{i \in \mathbb{N}}$, and in Theorem 1.28 the process $t \mapsto X(t)$ is defined on a filtered probability space $(\Omega, \mathcal{F}_t.\mathbb{P})_{t \geqslant 0}$.

1.27. Theorem (Doob's maximal inequality). *Let* $(X_i)_{i\in\mathbb{N}}$ *be a sub-martingale w.r.t. a filtration* $(\mathcal{F}_i)_{i\in\mathbb{N}}$. *Let* $S_n = \max_{1\leqslant i\leqslant n} X_i$ *be the running maximum of* X_i. *Then for any* $\ell > 0$,

$$\mathbb{P}\left[S_n \geqslant \ell\right] \leqslant \frac{1}{\ell}\mathbb{E}\left[X_n^+ \mathbf{1}_{\{S_n\geqslant \ell\}}\right] \leqslant \frac{1}{\ell}\mathbb{E}\left[X_n^+\right], \tag{1.22}$$

where $X_n^+ = X_n \vee 0$. *In particular, if* X_i *is a martingale and* $M_n = \max\limits_{1\leqslant i\leqslant n} |X_i|$, *then*

$$\mathbb{P}\left[M_n \geqslant \ell\right] \leqslant \frac{1}{\ell}\mathbb{E}\left[|X_n|\, \mathbf{1}_{\{M_n\geqslant \ell\}}\right] \leqslant \frac{1}{\ell}\mathbb{E}\left[|X_n|\right]. \tag{1.23}$$

Proof. Let $\tau_\ell = \inf\{i \geqslant 1 : X_i \geqslant \ell\}$. Then $\mathbb{P}\left[S_n \geqslant \ell\right] = \sum_{i=1}^n \mathbb{P}\left[\tau_\ell = i\right]$. For each $1 \leqslant i \leqslant n$,

$$\mathbb{P}\left[\tau_\ell = i\right] = \mathbb{E}\left[\mathbf{1}_{\{X_i\geqslant \ell\}}\mathbf{1}_{\{\tau_\ell = i\}}\right] \leqslant \frac{1}{\ell}\mathbb{E}\left[X_i^+\mathbf{1}_{\{\tau_\ell = i\}}\right]. \tag{1.24}$$

Note that $\{\tau_\ell = i\} \in \mathcal{F}_i$, and X_i^+ is a sub-martingale because X_i itself is a sub-martingale while $\varphi(x) = x^+ = x \vee 0 = \max(x, 0)$ is an increasing convex function. Therefore

$$\mathbb{E}\left[X_n^+\mathbf{1}_{\{\tau_\ell = i\}} \mid \mathcal{F}_i\right] = \mathbf{1}_{\{\tau_\ell = i\}}\mathbb{E}\left[X_n^+ \mid \mathcal{F}_i\right] \geqslant \mathbf{1}_{\{\tau_\ell = i\}}\left(\mathbb{E}\left[X_n \mid \mathcal{F}_i\right]\right)^+ \geqslant \mathbf{1}_{\{\tau_\ell = i\}}X_i^+,$$

and hence $\mathbb{E}\left[X_i^+\mathbf{1}_{\{\tau_\ell = i\}}\right] \leqslant \mathbb{E}\left[X_n^+\mathbf{1}_{\{\tau_\ell = i\}}\right]$. Substituting this inequality into (1.24) and then summing over $1 \leqslant i \leqslant n$ then yields (1.22). The inequality in (1.23) follows by applying (1.22 to the sub-martingale $|X_i|$. $\square$

Next we formulate and prove Doob's maximal inequality for continuous time sub-martingales.

1.28. THEOREM (Doob's maximal inequality). *Let $(X(t))_{t\geqslant 0}$ be a sub-martingale w.r.t. a filtration $(\mathcal{F}_t)_{t\geqslant 0}$. Let $S(t) = \sup_{0\leqslant s\leqslant t} X(s)$ be the running maximum of $X(t)$. Suppose that the process $t \mapsto X(t)$ is $\mathbb{P}$-almost surely continuous from the right (and possesses left limits $\mathbb{P}$-almost surely). Then for any $\ell > 0$,*

$$\mathbb{P}\left[S(t) \geqslant \ell\right] \leqslant \frac{1}{\ell}\mathbb{E}\left[X(t)^{+}\mathbf{1}_{\{S(t)\geqslant \ell\}}\right] \leqslant \frac{1}{\ell}\mathbb{E}\left[X^{+}(t)\right], \tag{1.25}$$

where $X^{+}(t) = X_(t) \vee 0 = \max\left(X(t), 0\right)$. In particular, if $t \mapsto X(t)$ is a martingale and $M(t) = \sup\limits_{0\leqslant s\leqslant t} |X(t)|$, then

$$\mathbb{P}\left[M(t) \geqslant \ell\right] \leqslant \frac{1}{\ell}\mathbb{E}\left[|X(t)|\,\mathbf{1}_{\{M(t)\geqslant \ell\}}\right] \leqslant \frac{1}{\ell}\mathbb{E}\left[|X(t)|\right]. \tag{1.26}$$

PROOF. Let, for every $N \in \mathbb{N}$, τ_N be the $(\mathcal{F}_t)_{t\geqslant 0}$-stopping time defined by $\tau_N = \inf\{t > 0 : X(t)^{+} \geqslant N\}$. In addition define the double sequence of processes $X_{n.N}(t)$ by

$$X_{n,N}(t) = X\left(2^{-n}\left\lceil 2^{n}t\right\rceil \wedge \tau_N\right).$$

Theorem 1.28 follows from Theorem 1.27 by applying it the processes $t \mapsto X_{n,N}(t)$, $n \in \mathbb{N}$, $N \in \mathbb{N}$. As a consequence of Theorem 1.27 we see that Theorem 1.28 is true for the double sequence $t \mapsto X_{n,N}(t)$, because, essentially speaking, these processes are discrete-time processes with the property that the processes $(n,t) \mapsto X_{n,N}(t)^{+}$ attain $\mathbb{P}$-almost surely their values in the interval $[0, N]$. Then we let $n \to \infty$ to obtain Theorem 1.28 for the processes $t \mapsto X\left(t \wedge \tau_N\right)$, $N \in \mathbb{N}$. Finally we let $N \to \infty$ to obtain the full result in Theorem 1.28. □

5.2. Additive processes. In this final section we introduce the notion of additive and multiplicative processes. Let E be a second countable locally compact Hausdorff space. In the non-time-homogeneous case we consider real-valued processes which depend on two time parameters: $(t_1, t_2) \mapsto Z\left(t_1, t_2\right)$, $0 \leqslant t_1 \leqslant t_2 \leqslant T$. It is assumed that for all $0 \leqslant t_1 \leqslant t_2 \leqslant T$, the variable $Z\left(t_1, t_2\right)$ only depends, or is measurable with respect to, $\sigma\left\{X(s) : t_1 \leqslant s \leqslant t_2\right\}$. Such a process is called *additive* if

$$Z\left(t_1, t_2\right) = Z\left(t_1, t\right) + Z\left(t, t_2\right), \quad t_1 \leqslant t \leqslant t_2.$$

The process Z is called *multiplicative* if

$$Z\left(t_1, t_2\right) = Z\left(t_1, t\right) \cdot Z\left(t, t_2\right), \quad t_1 \leqslant t \leqslant t_2.$$

Let $p : [0, T] \times E \to \mathbb{R}$ be a continuous function, and let $\{X(t) : 0 \leqslant t \leqslant T\}$ be an E-valued process which has left limits in E, and which is right-continuous (i.e. it is càdlàg). Put $Z\left(t_1, t_2\right) = \int_{t_1}^{t_2} p\left(s, X(s)\right) ds$. Then the process $(t_1, t_2) \mapsto Z\left(t_1, t_2\right)$, $0 \leqslant t_1 \leqslant t_2 \leqslant T$ is additive, and the process $(t_1, t_2) \mapsto \exp\left(Z\left(t_1, t_2\right)\right)$, $0 \leqslant t_1 \leqslant t_2 \leqslant T$, is multiplicative.

Next we consider the particular case that we deal with time-homogeneous processes like Brownian motion:

$$\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t), t \geqslant 0), (\vartheta_t : t \geqslant 0), (\mathbb{R}^n, \mathcal{B}^n)\},$$

which represents Brownian motion starting at $x \in \mathbb{R}^d$. An adapted process $t \mapsto Z(t)$ is called additive if $Z(s+t) = Z(s) + Z(t) \circ \vartheta_s$, $\mathbb{P}_x$-almost surely, for all $s, t \geqslant 0$. It is called multiplicative provided $Z(s+t) = Z(s) \cdot Z(t) \circ \vartheta_s$, $\mathbb{P}_x$-almost surely, for all $s, t \geqslant 0$. Examples of additive processes are integrals of the form $Z(t) = \int_0^{t_2} p(X(s))\, ds$, where $x \mapsto p(x)$ is a continuous (or Borel) function on $\mathbb{R}^d$, or stochastic integrals (Itô, Stratonovich integrals) of the form $Z(t) = \int_0^t p(X(s))\, dX(s)$. Such integrals have to be interpreted in some L^2-sense. More details will be given in Section 6. If $t \mapsto Z(t)$ is an additive process, then its exponent $t \mapsto \exp(Z(t))$ is a multiplicative process. If T is a terminal stopping time, then the process $t \mapsto \mathbf{1}_{\{T>t\}}$ is a multiplicative process.

Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of non-negative i.i.d. random variables each of which has density $f_1 \geqslant 0$. Suppose that f_n is the density of the distribution of $\sum_{j=1}^n X_j$. Note "i.i.d." means "independent, identically distributed". Then $\mathbb{P}\left[\sum_{j=1}^n X_j \leqslant t\right] = \int_0^t f_n(s)ds$, and hence

$$\int_0^t f_{n+1}(s)ds = \mathbb{P}\left[\sum_{j=1}^{n+1} X_j \leqslant t\right] = \mathbb{P}\left[\sum_{j=1}^{n} X_j + X_{n+1} \leqslant t\right]$$
$$= \int_0^t f_n(\rho) f_1(t-\rho) d\rho.$$

It follows that

$$\begin{aligned}
&\int_0^t f_n(s)ds - \int_0^t f_{n+1}(\rho)d\rho = \int_0^t f_n(s)ds - \int_0^t \int_0^\rho f_n(s) f_1(\rho - s)ds\, d\rho \\
&= \int_0^t f_n(s)ds - \int_0^t f_n(s) \int_s^t f_1(\rho - s)d\rho\, ds \\
&= \int_0^t f_n(s) \left(1 - \int_s^t f_1(\rho - s)d\rho\right) ds \\
&= \int_0^t f_n(s) \int_t^\infty f_1(\rho - s)d\rho\, ds \\
&= \int_0^t f_n(s) \int_{t-s}^\infty f_1(\rho)d\rho\, ds. \qquad (1.27)
\end{aligned}$$

If $f_1(s) = \lambda e^{-\lambda s}$, then $f_n(s) = \dfrac{\lambda^n s^{n-1}}{(n-1)!} e^{-\lambda s}$. This follows by induction.

5.3. Continuous time discrete processes. Here we suppose that the process

$$\{(\Omega, \mathcal{F}, \mathbb{P}), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (S, \mathcal{S})\}$$

is governed by a time-homogeneous or stationary transition probabilities:

$$p_{j,i}(t) = \mathbb{P}\left[X(t) = j \mid X(0) = i\right] = \mathbb{P}\left[X(t+s) = j \mid X(s) = i\right], \quad i,\, j \in S, \tag{1.28}$$

for all $s \geqslant 0$. Here, S is a discrete state space, e.g. $S = \mathbb{Z}$, $S = \mathbb{Z}^n$, $S = \mathbb{N}$, or $S = \{0, N\}$. The measurable space $(\Omega, \mathcal{F})$ is called the sample or sample path space. Its elements $\omega \in \Omega$ are called realizations. The mappings $X(t) : \Omega \to S$ are called the state variables; the application $t \mapsto X(t)(\omega)$ is called a sample path or realization. The translation operators ϑ_t, $t \geqslant 0$, are mappings from Ω to Ω with the property that: $X(s) \circ \vartheta_t = X(s+t)$, $\mathbb{P}$-almost surely. For the time being these operators will not be used; they are very convenient to express the Markov property in the time-homogeneous case. We assume that the Chapman-Kolmogorov conditions are satisfied:

$$p_{j,i}\left(s+t\right) = \sum_{k \in S} p_{j,k}(s) p_{k,i}(t), \; i,\, j \in S,\; s,\, t \geqslant 0. \tag{1.29}$$

In fact the Markov property is a consequence of the Chapman-Kolmogorov identity (1.29). From the Chapman-Kolmogorov (1.29) the following important identity follows:

$$P\left(s+t\right) = P(s)P(t), \quad s,\, t \geqslant 0. \tag{1.30}$$

The identity in (1.30) is called the semigroup property; the identity has to be interpreted as matrix multiplication. Suppose that the functions $t \mapsto p_{j,i}(t)$, j, $i \in S$, are right differentiable at $t = 0$. The latter means that the following limits exist:

$$q_{j,i} = \lim_{\triangle \downarrow 0} \frac{p_{j,i}\left(\triangle\right) - p_{j,i}(0)}{\triangle}, \quad i,\; j \in S.$$

We assume that $p_{j,i}(0) = \delta_{j,i}$, where $\delta_{j,i}$ is the Dirac delta function: $\delta_{j,i} = 0$ if $j \neq i$, and $\delta_{j,j} = 1$. Put $Q = (q_{j,i})_{i,j\in S}$. Then the matrix Q is a Kolmogorov matrix in the sense that $q_{j,i} \geqslant 0$ for $j \neq i$ and $\sum_{j\in S} q_{j,i} = 0$. It follows that $q_{i,i} = -\sum_{j\in S, j\neq i} q_{j,i} \leqslant 0$. The reason that the off-diagonal entries $q_{j,i}$, $j \neq i$, are non-negative is due to the fact that for $j \neq i$ we have

$$q_{j,i} = \lim_{t\downarrow 0} \frac{p_{j,i}(t) - p_{j,i}(0)}{t} = \lim_{t\downarrow 0} \frac{p_{j,i}(t) - \delta_{j,i}(0)}{t} = \lim_{t\downarrow 0} \frac{p_{j,i}(t)}{t} \geqslant 0.$$

In addition, we have

$$\begin{aligned}\sum_{j\in S} q_{j,i} &= \sum_{j\in S} \lim_{t\downarrow 0} \frac{p_{j,i}(t) - p_{j,i}(0)}{t} = \lim_{t\downarrow 0} \sum_{j\in S} \frac{p_{j,i}(t) - p_{j,i}(0)}{t} \\ &= \lim_{t\downarrow 0} \frac{\sum_{j\in S} p_{j,i}(t) - \sum_{j\in S} p_{j,i}(0)}{t} = \lim_{t\downarrow 0} \frac{1-1}{t} = 0, \end{aligned} \tag{1.31}$$

provided we may interchange the summation and the limit. Finally we have the following general fact. Let $t \mapsto P(t)$ be the matrix function $t \mapsto (p_{j,i}(t))_{i,j\in S}$. Then $P(t)$ satisfies the Kolmogorov backward and forward differential equation:

$$\frac{dP(t)}{dt} = QP(t) = P(t)Q, \quad t \geqslant 0. \tag{1.32}$$

The first equality in (1.32) is called the Kolmogorov forward equation, and the second one the Kolmogorov backward equation. The solution of this matrix-valued differential equation is given by $P(t) = e^{tQ}P(0)$. But since $P(0) = (p_{j,i}(0))_{j,i\in S} = (\delta_{j,i})_{i,j\in S}$ is the identity matrix, it follows that $P(t) = e^{tQ}$. The equalities in (1.32) hold true, because by the semigroup property (1.30) we have:

$$\frac{P(t+\triangle(t)) - P(t)}{\triangle(t)} = \frac{P(\triangle(t)) - P(0)}{\triangle(t)} P(t) = P(t) \frac{P(\triangle(t)) - P(0)}{\triangle(t)}. \tag{1.33}$$

Then we let $\triangle(t)$ tend to 0 in (1.33) to obtain (1.32).

5.4. Poisson process. We begin with a formal definition.

1.29. Definition. A Poisson process

$$\{(\Omega, \mathcal{F}, \mathbb{P}), (X(t), t \geqslant 0), (\vartheta_t, t \geqslant 0), (\mathbb{N}, \mathcal{N})\}$$

(see (1.46) below) is a continuous time process $X(t)$, $t \geqslant 0$, with values in $\mathbb{N} = \{0, 1, \ldots\}$ which possesses the following properties:

(a) For $\Delta t > 0$ sufficiently small the transition probabilities satisfy:

$$\begin{aligned} p_{i+1,i}(\Delta t) &= \mathbb{P}\left[X(t+\Delta t) = i+1 \mid X(t) = i\right] = \lambda \Delta t + o(\Delta t); \\ p_{i,i}(\Delta t) &= \mathbb{P}\left[X(t+\Delta t) = i \mid X(t) = i\right] = 1 - \lambda \Delta t + o(\Delta t); \\ p_{j,i}(\Delta t) &= \mathbb{P}\left[X(t+\Delta t) = j \mid X(t) = i\right] = o(\Delta t); \\ p_{j,i}(\Delta t) &= 0, \quad j < i. \end{aligned} \tag{1.34}$$

(b) The probability transitions $(s, i; t, j) \mapsto \mathbb{P}\left[X(t) = j \mid X(s) = i\right]$, $t > s$, only depend on $t - s$ and $j - i$.

(c) The process $\{X(t) : t \geqslant 0\}$ has the Markov property.

Item (b) says that the Poisson process is homogeneous in time and in space: (b) is implicitly used in (a). Note that a Poisson process is not continuous, because when it moves it makes a jump. Put

$$p_i(t) = p_{i0}(t) = p_{j+i,j}(t) = \mathbb{P}\left[X(t) = j + i \mid X(0) = i\right], \quad i,\ j \in \mathbb{N}. \tag{1.35}$$

1.30. Proposition. *Let the process*

$$\{(\Omega, \mathcal{F}, \mathbb{P}), (X(t), t \geqslant 0), (\vartheta_t, t \geqslant 0), (\mathbb{N}, \mathcal{N})\}$$

possess properties (a) and (b) in Definition 1.29. Then the following equality holds for all $t \geqslant 0$ and $i \in \mathbb{N}$:

$$p_i(t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}. \tag{1.36}$$

1.31. Remark. It is noticed that the equalities in (1.42), (1.40), and (1.44) only depend on properties (a) and (b) in Definition 1.29. So that from (a), and (b) we obtain

$$\frac{d}{dt} p_i(t) + \lambda p_i(t) = \lambda \frac{(\lambda t)^{i-1}}{(i-1)!} e^{-\lambda t} = \lambda p_{i-1}(t), \quad i \geqslant 1, \tag{1.37}$$

and hence

$$p_{j,i}(t) = p_{j-i}(t) = \mathbb{P}\left[X(t) = j \mid X(0) = i\right] = \frac{(\lambda t)^{j-i}}{(j-i)!} e^{-\lambda t}, \quad j \geqslant i. \tag{1.38}$$

If $0 \leqslant j < i$, then $p_{j,i}(t) = 0$.

Proof. By definition we see that $p_j(0) = \mathbb{P}\left[X(0) = j \mid X(0) = 0\right] = \delta_{0,j}$, and so $p_0(0) = 1$ and $p_j(0) = 0$ for $j \neq 0$. Let us first prove that the functions $t \mapsto p_i(t)$, $i \geqslant 1$, satisfy the differential equation in (1.45) below. First suppose that $i \geqslant 2$, and we consider:

$$\begin{aligned}
&p_i(t + \Delta t) - p_i(t) = \mathbb{P}\left[X(t + \Delta t) = i\right] - p_i(t) \\
&= \sum_{k=0}^{i} \mathbb{P}\left[X(t + \Delta t) = i,\ X(t) = k\right] - p_i(t) \\
&= \sum_{k=0}^{i} \mathbb{P}\left[X(t + \Delta t) = i \mid X(t) = k\right] \mathbb{P}\left[X(t) = k\right] - p_i(t) \\
&= \mathbb{P}\left[X(t + \Delta t) = i \mid X(t) = i\right] p_i(t) + \mathbb{P}\left[X(t + \Delta t) = i \mid X(t) = i - 1\right] p_{i-1}(t) \\
&\quad + \sum_{k=0}^{i-2} \mathbb{P}\left[X(t + \Delta t) = i \mid X(t) = k\right] p_k(t) - p_i(t) \\
&= (1 - \lambda \Delta t + o(\Delta t))\, p_i(t) + (\lambda \Delta t + o(\Delta t))\, p_{i-1}(t) + \sum_{k=0}^{i-2} p_k(t) o(\Delta t) - p_i(t) \\
&= -\lambda \Delta t p_i(t) + \lambda \Delta t p_{i-1}(t) + \sum_{k=0}^{i} p_k(t) o(\Delta t).
\end{aligned} \tag{1.39}$$

From (1.39) we obtain

$$\frac{d}{dt}p_i(t) = -\lambda p_i(t) + \lambda p_{i-1}(t). \tag{1.40}$$

Next we consider $i = 0$:

$$\begin{aligned}
&p_0(t+\Delta t) - p_0(t) = \mathbb{P}\left[X(t+\Delta t) = 0\right] - p_0(t) \\
&= \mathbb{P}\left[X(t+\Delta t) = 0 \mid X(t) = 0\right]\mathbb{P}\left[X(t) = 0\right] - p_0(t) \\
&= \mathbb{P}\left[X(t+\Delta t) = 0 \mid X(t) = 0\right] p_0(t) - p_0(t) = \left(-\lambda\Delta t + o(\Delta t)\right) p_0(t).
\end{aligned} \tag{1.41}$$

From (1.41) we get the equation

$$\frac{d}{dt}p_0(t) = -\lambda p_0(t). \tag{1.42}$$

For $i = 1$ we have:

$$\begin{aligned}
&p_1(t+\Delta t) - p_1(t) = \mathbb{P}\left[X(t+\Delta t) = 1\right] - p_1(t) \\
&= \mathbb{P}\left[X(t+\Delta t) = 1 \mid X(t) = 1\right]\mathbb{P}\left[X(t) = 1\right] - p_1(t) \\
&\quad + \mathbb{P}\left[X(t+\Delta t) = 1 \mid X(t) = 0\right]\mathbb{P}\left[X(t) = 0\right] \\
&= \mathbb{P}\left[X(t+\Delta t) = 1 \mid X(t) = 1\right] p_1(t) - p_1(t) \\
&\quad + \mathbb{P}\left[X(t+\Delta t) = 1 \mid X(t) = 0\right] p_0(t) \\
&= \left(-\lambda\Delta t + o(\Delta t)\right) p_1(t) + \left(\lambda\Delta t + o(\Delta t)\right) p_0(t).
\end{aligned} \tag{1.43}$$

From (1.43) we obtain:

$$\frac{d}{dt}p_1(t) = -\lambda p_1(t) + \lambda p_0(t). \tag{1.44}$$

By definition we see that $p_j(0) = \mathbb{P}\left[X(0) = j \mid X(0) = 0\right] = \delta_{0,j}$, and so $p_0(0) = 1$ and $p_j(0) = 0$ for $j \neq 0$. From (1.42) we get $p_0(t) = e^{-\lambda t}$. From (1.40) and (1.44) we obtain

$$\frac{d}{dt}\left(e^{\lambda t}p_i(t)\right) = \lambda e^{\lambda t}p_{i-1}(t), \quad i \geqslant 1. \tag{1.45}$$

By induction it follows that $p_i(t) = \dfrac{(\lambda t)^i}{i!}e^{-\lambda t}$. This completes the proof of Proposition 1.30. $\square$

In the Proposition 1.33 below we show that a process

$$\{(\Omega, \mathcal{F}, \mathbb{P}), (X(t), t \geqslant 0), (\vartheta_t, t \geqslant 0), (\mathbb{N}, \mathcal{N})\} \tag{1.46}$$

which satisfies (a) and (b) of Definition 1.29 is a time-homogeneous Markov process if and only if its increments are $\mathbb{P}$-independent. First we prove a lemma, which is of independent interest.

1.32. Lemma. *Let the functions $p_i(t)$ be defined as in (1.35). Then the equality*

$$p_i(t) = \mathbb{P}\left[X(s+t) - X(s) = i\right] \tag{1.47}$$

holds for all $i \in \mathbb{N}$ and all s, $t \geqslant 0$.

Proof. Using the space and time invariance properties of the process $X(t)$ shows:

$$\begin{aligned}\mathbb{P}\left[X(s+t) - X(s) = i\right] &= \sum_{k=0}^{\infty}\mathbb{P}\left[X(s+t) - X(s) = i,\, X(s) = k\right] \\ &= \sum_{k=0}^{\infty}\mathbb{P}\left[X(s+t) = i+k,\, X(s) = k\right] \\ &= \sum_{k=0}^{\infty}\mathbb{P}\left[X(s+t) = i+k \mid X(s) = k\right]\mathbb{P}\left[X(s) = k\right]\end{aligned}$$

(space and time invariance properties of $p_i(t)$)

$$= \sum_{k=0}^{\infty}p_i(t)\mathbb{P}\left[X(s) = k\right] = p_i(t). \tag{1.48}$$

The conclusion in Lemma 1.32 follows from (1.48). $\square$

The following proposition says that a time and space-homogeneous process satisfying the equalities in (1.34) of Definition 1.29 is a Poisson process if and only if its increments are $\mathbb{P}$-independent.

1.33. Proposition. *The process $\{X(t) : t \geqslant 0\}$ possessing properties (a) and (b) of Definition 1.29 possesses the Markov property if and only if its increments are $\mathbb{P}$-independent. Moreover, the equalities*

$$\begin{aligned}\mathbb{P}\left[X(t)-X(s)=j-i\right] &= \mathbb{P}\left[X(t)=j \;\middle|\; X(s)=i\right] \\ &= p_{j-i}(t-s)=\frac{(\lambda(t-s))^{j-i}}{(j-i)!}e^{-\lambda(t-s)} \end{aligned} \tag{1.49}$$

hold for all $t \geqslant s \geqslant 0$ and for all $j \geqslant i$, i, $j \in \mathbb{N}$.

Proof. First assume that the process in (1.46) has the Markov property. Let $t_{n+1} > t_n > \cdots > t_1 > t_0 = 0$, and let i_k, $1 \leqslant k \leqslant n+1$, be nonnegative integers. Then by induction we have

$$\begin{aligned}
&\mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, 1\leqslant \ell\leqslant n+1\right] \\
&= \sum_{k=0}^{\infty}\mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, 1\leqslant \ell\leqslant n+1,\, X(t_n)=k\right] \\
&= \sum_{k=0}^{\infty}\frac{\mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, 1\leqslant \ell\leqslant n+1,\, X(t_n)=k\right]}{\mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, 1\leqslant \ell\leqslant n,\, X(t_n)=k\right]} \\
&\qquad \times \mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, 1\leqslant \ell\leqslant n,\, X(t_n)=k\right] \\
&= \sum_{k=0}^{\infty}\mathbb{P}\left[X(t_{n+1})-X(t_n)=i_{n+1} \;\middle|\; X(t_\ell)-X(t_{\ell-1})=i_\ell,\, X(t_n)=k,\right. \\
&\qquad\quad \left. 1\leqslant \ell\leqslant n\right] \\
&\qquad \times \mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, X(t_n)=k,\, 1\leqslant \ell\leqslant n,\,\right]
\end{aligned}$$

(Markov property)

$$\begin{aligned}
&= \sum_{k=0}^{\infty}\mathbb{P}\left[X(t_{n+1})-X(t_n)=i_{n+1} \;\middle|\; X(t_n)=k\right] \\
&\qquad \times \mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, 1\leqslant \ell\leqslant n,\, X(t_n)=k\right] \\
&= \sum_{k=0}^{\infty}\mathbb{P}\left[X(t_{n+1})=i_{n+1}+k \;\middle|\; X(t_n)=k\right] \\
&\qquad \times \mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, 1\leqslant \ell\leqslant n,\, X(t_n)=k\right]
\end{aligned}$$

(homogeneity in space and time of the function $t \mapsto p_{i_{n+1}}(t)$)

$$= \sum_{k=0}^{\infty}p_{i_{n+1}}(t_{n+1}-t_n)\,\mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, 1\leqslant \ell\leqslant n,\, X(t_n)=k\right]$$

(apply equality (1.47) in Lemma 1.29)

$$\begin{aligned}
&= \sum_{k=0}^{\infty}\mathbb{P}\left[X(t_{n+1})-X(t_n)=i_{n+1}\right] \\
&\qquad \mathbb{P}\left[X(t_\ell)-X(t_{\ell-1})=i_\ell,\, 1\leqslant \ell\leqslant n,\, X(t_n)=k\right]
\end{aligned}$$

$$= \mathbb{P}\left[X\left(t_{n+1}\right) - X\left(t_n\right) = i_{n+1}\right] \mathbb{P}\left[X\left(t_\ell\right) - X\left(t_{\ell-1}\right) = i_\ell,\ 1 \leqslant \ell \leqslant n\right]. \tag{1.50}$$

By induction and employing (1.50) it follows that

$$\begin{aligned}\mathbb{P}\left[X\left(t_\ell\right) - X\left(t_{\ell-1}\right) = i_\ell,\ 1 \leqslant \ell \leqslant n\right] &= \prod_{\ell=1}^{n} \mathbb{P}\left[X\left(t_\ell\right) - X\left(t_{\ell-1}\right) = i_\ell\right] \\ &= \prod_{\ell=1}^{n} p_{i_\ell}\left(t_\ell - t_{\ell-1}\right). \end{aligned} \tag{1.51}$$

We still have to prove the converse statement, i.e. to prove that if the increments of the process $X(t)$ are $\mathbb{P}$-independent, then the process $X(t)$ has the Markov property. Therefore we take states $0 = i_0,\ i_1, \dots, i_n,\ i_{n+1}$, and times $0 = t_0 < t_1 < \cdots < t_n < t_{n+1}$, and we consider the conditional probability:

$$\begin{aligned}&\mathbb{P}\left[X\left(t_{n+1}\right) = i_{n+1} \,\middle|\, X\left(t_0\right) = i_0, \dots, X\left(t_n\right) = i_n\right] \\ &= \frac{\mathbb{P}\left[X\left(t_{n+1}\right) = i_{n+1},\ X\left(t_0\right) = i_0, \dots, X\left(t_n\right) = i_n\right]}{\mathbb{P}\left[X\left(t_0\right) = i_0, \dots, X\left(t_n\right) = i_n\right]} \\ &= \frac{\mathbb{P}\left[X\left(t_0\right) = i_0,\ X\left(t_\ell\right) - X\left(t_{\ell-1}\right) = i_\ell - i_{\ell-1},\ 1 \leqslant \ell \leqslant n+1\right]}{\mathbb{P}\left[X\left(t_0\right) = i_0,\ X\left(t_\ell\right) - X\left(t_{\ell-1}\right) = i_\ell - i_{\ell-1},\ 1 \leqslant \ell \leqslant n\right]}\end{aligned}$$

(increments are $\mathbb{P}$-independent)

$$\begin{aligned}&= \mathbb{P}\left[X\left(t_{n+1}\right) - X\left(t_n\right) = i_{n+1} - i_n\right] \\ &= \mathbb{P}\left[X\left(t_{n+1}\right) = i_{n+1} \,\middle|\, X\left(t_n\right) = i_n\right]. \end{aligned} \tag{1.52}$$

The final equality in (1.52) follows by invoking another application of the fact that increments are $\mathbb{P}$-independent. More precisely, since $X(t_{n+1}) - X(t_n)$ and $X(t_n) - X(0)$ are $\mathbb{P}$-independent we have

$$\begin{aligned}
&\mathbb{P}\left[X(t_{n+1}) = i_{n+1} \;\middle|\; X(t_n) = i_n\right] \\
&= \frac{\mathbb{P}\left[X(t_{n+1}) - X(t_n) = i_{n+1} - i_n,\ X(t_n) - X(0) = i_n\right]}{\mathbb{P}\left[X(t_n) - X(0) = i_n\right]} \\
&= \mathbb{P}\left[X(t_{n+1}) - X(t_n) = i_{n+1} - i_n\right]. \qquad (1.53)
\end{aligned}$$

The equalities in (1.49) follow from equality (1.47) in Lemma 1.32, from (1.53), from the definition of the function $p_i(t)$ (see equality (1.37)), and from the explicit value of $p_i(t)$ (see (1.36) in Proposition 1.30). This completes the proof of Proposition 1.33. □

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let the process $t \mapsto N(t)$ and the probability measures $\mathbb{P}_j$, $j \in \mathbb{N}$ in $\left\{(\Omega, \mathcal{F}, \mathbb{P}_j)_{j\in\mathbb{N}}, (N(t) : t \geqslant 0), (\vartheta_s : s \geqslant 0), (\mathbb{N}, \mathcal{N})\right\}$ have the following properties:

(a) It has independent increments: $N(t+h) - N(t)$ is independent of
$$\mathcal{F}_t^0 = \sigma\left(N(s) - N(0) : 0 \leqslant s \leqslant t\right).$$
(b) Constant intensity: the chance of arrival in any interval of length h is the same:
$$\mathbb{P}\left[N(t+h) - N(t) \geqslant 1\right] = \lambda h + o(h).$$
(c) Rarity of jumps $\geqslant 2$:
$$\mathbb{P}\left[N(t+h) - N(t) \geqslant 2\right] = o(h).$$
(d) the measures $\mathbb{P}_j$, $j \geqslant 1$, are defined by: $\mathbb{P}_j[A] = \mathbb{P}\left[A \;\middle|\; N(0) = j\right]$; moreover, it is assumed that $\mathbb{P}_0[N(0) = 0] = 1$.

The following theorem and its proof are taken from Stirzaker [**126**] Theorem (13) page 74.

1.34. THEOREM. *Suppose that the process $N(t)$ and the probability measures satisfy (a), (b), (c) and (d). Then the process $N(t)$ is a Poisson process and*

$$\mathbb{P}_j\left[N(t) = k\right] = \frac{(\lambda t)^{k-j}}{(k-j)!} e^{-\lambda(k-j)}, \quad k \geqslant j. \qquad (1.54)$$

PROOF. In view of Proposition 1.33 it suffices to prove the identity in (1.54). To this end we put

$$f_n(t) = \mathbb{P}_0\left[N(t) = n\right] = \mathbb{P}\left[N(t) = n \;\middle|\; N(0) = 0\right] = \mathbb{P}\left[N(t) - N(0) = n\right].$$

Then we have, for $n \geqslant 2$ fixed,

$$f_n(t+h) = \mathbb{P}_0\left[N(t+h) = n\right] = \sum_{k=0}^{n} \mathbb{P}_0\left[N(t+h) - N(t) = k,\ N(t) = n-k\right]$$

(the variables $N(t+h) - N(t)$ and $N(t)$ are $\mathbb{P}_0$-independent)

$$\begin{aligned}
&= \sum_{k=0}^{n} \mathbb{P}_0 \left[N(t+h) - N(t) = k \right] \times \mathbb{P}_0 \left[N(t) = n-k \right] \\
&= \mathbb{P}_0 \left[N(t+h) - N(t) = 0 \right] \times \mathbb{P}_0 \left[N(t) = n \right] \\
&\quad + \mathbb{P}_0 \left[N(t+h) - N(t) = 1 \right] \times \mathbb{P}_0 \left[N(t) = n-1 \right] \\
&\quad + \sum_{k=2}^{n} \mathbb{P}_0 \left[N(t+h) - N(t) = k \right] \times \mathbb{P}_0 \left[N(t) = n-k \right] \\
&= \left(1 - \mathbb{P}_0 \left[N(t+h) - N(t) \geqslant 1 \right]\right) \times \mathbb{P}_0 \left[N(t) = n \right] \\
&\quad + \mathbb{P}_0 \left[N(t+h) - N(t) \geqslant 1 \right] \times \mathbb{P}_0 \left[N(t) = n-1 \right] \\
&\quad - \mathbb{P}_0 \left[N(t+h) - N(t) \geqslant 2 \right] \times \mathbb{P}_0 \left[N(t) = n-1 \right] \\
&\quad + \sum_{k=2}^{n} \mathbb{P}_0 \left[N(t+h) - N(t) = k \right] \times \mathbb{P}_0 \left[N(t) = n-k \right] \\
&= \left(1 - \lambda h + o(h)\right) \times f_n(t) + \left(\lambda h + o(h)\right) f_{n-1}(t) + o(h) \sum_{k=1}^{n} f_{n-k}(t) \\
&= \left(1 - \lambda h\right) f_n(t) + \lambda h f_{n-1}(t) + o(h). \qquad (1.55)
\end{aligned}$$

Observe that a similar argument yields

$$f_1(t+h) = \left(1 - \lambda h\right) f_1(t) + \lambda h f_0(t) + o(h), \qquad (1.56)$$

and also

$$f_0(t+h) = \left(1 - \lambda h\right) f_0(t) + o(h). \qquad (1.57)$$

From (1.55), (1.56) and (1.57) we obtain by rearranging, dividing by h and allowing $h \downarrow 0$:

$$\begin{aligned}
f_n'(t) &= -\lambda f_n(t) + \lambda f_{n-1}(t), \quad n \geqslant 1, \\
f_0'(t) &= -\lambda f(t).
\end{aligned}$$

These equations can be solved by induction relative to n. A alternative way is to consider the generating function

$$G(s,t) := \mathbb{E}_0 \left[e^{sN(t)} \right] = \sum_{n=0}^{\infty} s^n \mathbb{P}_0 \left[N(t) = n \right] = \sum_{n=0}^{\infty} s^n f_n(t).$$

Then $\dfrac{\partial G(s,t)}{\partial t} = \lambda(s-1)G(s,t)$, and so $G(s,t) = e^{\lambda t(s-1)}$. It follows that $\mathbb{P}_0 \left[N(t) = n \right] = f_n(t) = e^{-\lambda t} \dfrac{(\lambda t)^n}{n!}$. Consequently, for $k \geqslant j$ we obtain

$$\begin{aligned}
\mathbb{P}_j \left[N(t) = k \right] &= \mathbb{P} \left[N(t) = k \mid N(0) = j \right] = \mathbb{P} \left[N(t) - N(0) = k - j \mid N(0) = j \right] \\
&= \mathbb{P} \left[N(t) - N(0) = k - j \right] = e^{-\lambda(k-j)} \frac{(\lambda t)^{k-j}}{(k-j)!} = \text{(RHS of (1.54)}.
\end{aligned}$$

This completes the proof of Theorem 1.34. $\square$

CHAPTER 2

Renewal theory and Markov chains

Our main topic in this chapter is a discussion on renewal theory, classification properties of irreducible Markov chains, and a discussion on invariant measures. Its contents is mainly taken from Stirzaker [**126**].

1. Renewal theory

Let $(X_r)_{r\in\mathbb{N}}$ be a sequence of independent identically distributed random variables with the property that $\mathbb{P}[X_r > 0] > 0$. Put $S_n = \sum_{r=1}^n X_r$, $S_0 = 0$, and define the renewal process $N(t)$ by $N(t) = \max\{n : S_n \leqslant t\}$, $t \geqslant 0$. The mean $m(t) = \mathbb{E}[N(t)]$ is called the renewal function. We have $N(t) \geqslant n$ if and only if $S_n \leqslant t$, and hence

$$\mathbb{P}[N(t) = n] = \mathbb{P}[S_n \leqslant t] - \mathbb{P}[S_{n+1} \leqslant t], \quad \text{and} \tag{2.1}$$

$$\mathbb{E}[N(t)] = \sum_{r=1}^{\infty} \mathbb{P}[N(t) \geqslant r] = \sum_{r=1}^{\infty} \mathbb{P}[S_r \leqslant t]. \tag{2.2}$$

For more details see e.g. [**4**] (for birth-death processes) and [**126**] (for renewal theory).

2.1. Theorem. *If $\mathbb{E}[X_r] > 0$, then $N(t)$ has finite moments for all $t < \infty$.*

Proof. Since $\mathbb{E}[X_r] > 0$ there exists $\varepsilon > 0$ such that $\mathbb{P}[X_r \geqslant \varepsilon] \geqslant \varepsilon$. Put $M(t) = \max\left\{n : \varepsilon \sum_{r=1}^n \mathbf{1}_{\{X_r \geqslant \varepsilon\}} \leqslant t\right\}$. Since $\varepsilon \sum_{r=1}^n \mathbf{1}_{\{X_r \geqslant \varepsilon\}} \leqslant \sum_{r=1}^n X_r$ it follows that $N(t) \leqslant M(t)$, and hence, with $m = \lfloor t\varepsilon^{-1} \rfloor$,

$$\begin{aligned}
\mathbb{E}[N(t)] &\leqslant \mathbb{E}[M(t)] = \sum_{n=1}^{\infty} \mathbb{P}[M(t) \geqslant n] = \sum_{n=1}^{\infty} \mathbb{P}\left[\varepsilon \sum_{r=1}^{n} \mathbf{1}_{\{X_r \geqslant \varepsilon\}} \leqslant t\right] \\
&= \sum_{n=1}^{\infty} \sum_{\Lambda\subset\{1,\dots,n\},\, \#\Lambda\leqslant m} \mathbb{P}[X_j \geqslant \varepsilon,\, j \in \Lambda,\, X_j < \varepsilon,\, j \notin \Lambda] \\
&= \sum_{n=1}^{\infty} \sum_{\Lambda\subset\{1,\dots,n\},\, \#\Lambda\leqslant m} \mathbb{P}[X_1 \geqslant \varepsilon]^{\#\Lambda} (1 - \mathbb{P}[X_1 \geqslant \varepsilon])^{n-\#\Lambda} \\
&= \sum_{n=1}^{\infty} \sum_{k=0}^{n\wedge m} \binom{n}{k} \mathbb{P}[X_1 \geqslant \varepsilon]^{k} (1 - \mathbb{P}[X_1 \geqslant \varepsilon])^{n-k} \\
&\leqslant \sum_{k=0}^{m} \mathbb{P}[X_1 \geqslant \varepsilon]^{k} \sum_{n=k}^{\infty} \binom{n}{k} (1 - \mathbb{P}[X_1 \geqslant \varepsilon])^{n-k}
\end{aligned}$$

$$= \sum_{k=0}^{m} \mathbb{P}\left[X_1 \geqslant \varepsilon\right]^k \sum_{n=0}^{\infty} \binom{n+k}{n} \left(1 - \mathbb{P}\left[X_1 \geqslant \varepsilon\right]\right)^n$$
$$= \frac{1}{\mathbb{P}\left[X_r \geqslant \varepsilon\right\}} \left(\left\lfloor \frac{t}{\varepsilon} \right\rfloor + 1 \right). \tag{2.3}$$

In the final equality in (2.3) we used the equality: $\sum_{n=0}^{\infty} \binom{n+k}{n} z^k = \frac{1}{(1-z)^{k+1}}$ for $|z| < 1$. The inequality in (2.3) shows Theorem 2.1. □

It follows that $\mathbb{E}\left[N(t)\right]$ is finite whenever $\mathbb{E}\left[X_r\right]$ is strictly positive. This fact will be used in Theorem 2.2.

2.2. THEOREM. *The following equality is valid:*

$$\mathbb{E}\left[S_{N(t)+1}\right] = \mathbb{E}\left[X_1\right] \mathbb{E}\left[N(t)+1\right].$$

The equality in Theorem 2.2 is called *Wald's equation.*

PROOF. The time $N(t)+1$ is a stopping time with respect to the filtration

$$\mathcal{F}_n = \sigma\left(X_r : 0 \leqslant r \leqslant n\right) = \sigma\left(S_r - r\mathbb{E}\left[X_1\right] : 0 \leqslant r \leqslant n\right).$$

Notice that the process $n \mapsto S_n - n\mathbb{E}\left[X_1\right]$ is a martingale, and hence

$$\mathbb{E}\left[S_{(N(t)+1)\wedge n} - \left((N(t)+1) \wedge n\right) \mathbb{E}\left[X_1\right]\right]$$
$$= \mathbb{E}\left[S_{(N(t)+1)\wedge 0} - \left((N(t)+1) \wedge 0\right) \mathbb{E}\left[X_1\right]\right] = 0. \tag{2.4}$$

Since $\mathbb{E}\left[N(t)\right]$ is finite, from (2.4) we get by letting n tend to ∞:

$$0 = \lim_{n\to\infty} \mathbb{E}\left[S_{(N(t)+1)\wedge n} - \left((N(t)+1) \wedge n\right) \mathbb{E}\left[X_1\right]\right]$$
$$= \mathbb{E}\left[S_{(N(t)+1)} - \left((N(t)+1)\right) \mathbb{E}\left[X_1\right]\right]. \tag{2.5}$$

Consequently, the conclusion in Theorem 2.2 follows. □

2.3. THEOREM. *Let $(X_r)_{r\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables such that $\mathbb{P}\left[X_r = 0\right] = 0$. Put $S_0 = 0$ and $S_n = \sum_{r=1}^{n} X_r$. Let the process $N(t)$ be defined as in (2.2). Let $F(t)$ be the distribution function of the variable X_r. Put $m(t) = \mathbb{E}\left[N(t)\right]$. Then $m(t)$ satisfies the renewal equation:*

$$m(t) = F(t) + \int_0^t m(t-s)dF(s) = \sum_{k=1}^{\infty} \left(\mu_F^*\right)^k [0,t], \tag{2.6}$$

where $\mu_F(a,b] = F(b) - F(a)$, and $\mu_1 * \mu_2(a,b] = \int_0^\infty \int_0^\infty \mathbf{1}_{(a,b]}(s+t)d\mu_1(s)d\mu_2(t)$, $0 \leqslant a < b$ (i.e. convolution product of the measures μ_1 and μ_2). Moreover,

$$\left(1 - \int_0^\infty e^{-\lambda s} dF(s)\right) \times \lambda \int_0^\infty e^{-\lambda t} m(t)\, dt = \int_0^\infty e^{-\lambda s} dF(s).$$

If X_r are independent exponentially distributed random variables, and thus the process $(N(t) : t \geqslant 0)$ is Poisson of parameter $\lambda > 0$, then $m(t) = \lambda t$.

Proof. On the event $\{X_1 > t\}$ we have $N(t) = 0$, and hence by using conditional expectation we see

$$\begin{aligned} m(t) &= \mathbb{E}\left[N(t)\right] = \mathbb{E}\left[N(t)\mathbf{1}_{\{X_1 \leqslant t\}}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[N(t)\mathbf{1}_{\{X_1 \leqslant t\}} \mid \sigma\left(X_1\right)\right]\right] \\ &= \mathbb{E}\left[\mathbf{1}_{\{X_1 \leqslant t\}}\mathbb{E}\left[N(t) - N\left(X_1\right) \mid \sigma\left(X_1\right)\right]\right] + \mathbb{E}\left[\mathbf{1}_{\{X_1 \leqslant t\}}\mathbb{E}\left[N\left(X_1\right) \mid \sigma\left(X_1\right)\right]\right] \end{aligned}$$

(on the event $\{X_1 \leqslant t\}$ we have $N\left(X_1\right) = 1$)

$$= \mathbb{E}\left[\mathbf{1}_{\{X_1 \leqslant t\}}\mathbb{E}\left[N(t) - N\left(X_1\right) \mid \sigma\left(X_1\right)\right]\right] + \mathbb{E}\left[\mathbf{1}_{\{X_1 \leqslant t\}}\mathbb{E}\left[\mathbf{1} \mid \sigma\left(X_1\right)\right]\right]$$

(the distribution of $N(t) - N(s)$, $t > s$, is the same as the distribution of $N(t-s)$)

$$\begin{aligned} &= \mathbb{E}\left[\mathbf{1}_{\{X_1 \leqslant t\}}\mathbb{E}\left[N\left(t - X_1\right) \mid \sigma\left(X_1\right)\right]\right] + \mathbb{E}\left[\mathbf{1}_{\{X_1 \leqslant t\}}\mathbb{E}\left[\mathbf{1} \mid \sigma\left(X_1\right)\right]\right] \\ &= \mathbb{E}\left[N\left(t - X_1\right)\mathbf{1}_{\{X_1 \leqslant t\}}\right] + \mathbb{E}\left[\mathbf{1}_{\{X_1 \leqslant t\}}\right] \\ &= \int_0^t m\left(t - x\right) dF(x) + F(t). \qquad (2.7) \end{aligned}$$

This completes the proof of Theorem 2.3. $\square$

2.4. LEMMA. *Suppose* $\mathbb{P}[X_r < \infty] = 1$. *Then*

$$\lim_{t\to\infty} N(t) = \infty, \quad \mathbb{P}\text{-almost surely.} \tag{2.8}$$

PROOF. Put $Z = \lim_{t\to\infty} N(t) = \sup_{t\geqslant 0} N(t)$. Observe that $\sum_{k=1}^{N(t)+1} X_k \geqslant t$, and hence by letting $t \to \infty$, the event $\{Z < \infty\}$ is contained in $\bigcup_{r=1}^{Z+1} \{X_r = \infty\}$, and thus

$$\begin{aligned}\mathbb{P}[Z < \infty] &= \mathbb{P}\left[\bigcup_{r=1}^{Z+1} \{X_r = \infty\}, Z < \infty\right] \leqslant \mathbb{P}\left[\bigcup_{r=1}^{\infty} \{X_r = \infty\}\right] \\ &\leqslant \sum_{r=1}^{\infty} \mathbb{P}[X_r = \infty] = 0. \end{aligned} \tag{2.9}$$

The result in Lemma 2.4 follows from (2.9). □

Since $\lim_{t\to\infty} N(t) = \infty$ $\mathbb{P}$-almost surely, we have $\lim_{t\to\infty} \frac{N(t)+1}{N(t)} = 1$ $\mathbb{P}$-almost surely. The following proposition follows from the strong "law" of large numbers (SSLN).

2.5. PROPOSITION. *Let* $(X_r)_{r\in\mathbb{N}}$ *be a sequence of non-negative independent, identically distributed random variables in* $L^1(\Omega, \mathcal{F}, \mathbb{P})$ *such that* $\mathbb{P}[X_r < \infty] = 1$. *Then*

$$\lim_{t\to\infty} \frac{S_{N(t)+1}}{N(t)+1} = \lim_{t\to\infty} \frac{S_{N(t)}}{N(t)} = \mathbb{E}[X_1], \quad \mathbb{P}\text{-almost surely.} \tag{2.10}$$

2.6. THEOREM (First renewal theorem). *Let the hypotheses be as in Proposition 2.5. Then*

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mathbb{E}[X_1]}, \quad \mathbb{P}\text{-almost surely.} \tag{2.11}$$

PROOF. By definition we have $S_{N(t)} \leqslant t < S_{N(t)+1}$, therefore

$$\frac{S_{N(t)}}{N(t)} \leqslant \frac{t}{N(t)} \leqslant \frac{N(t)+1}{N(t)} \frac{S_{N(t)+1}}{N(t)+1}. \tag{2.12}$$

The result in (2.11) now follows from (2.12) in conjunction with (2.8) and (2.10). This proves Theorem 2.6. □

The proof of the following theorem is somewhat more intricate.

2.7. THEOREM (Elementary renewal theorem). *Let the hypotheses be as in Proposition 2.5. As above, put* $m(t) = \mathbb{E}[N(t)]$. *Then*

$$\lim_{t\to\infty} \frac{m(t)}{t} = \frac{1}{\mathbb{E}[X_1]}. \tag{2.13}$$

2.8. Remark. From Theorem 2.6 and 2.7 it follows that the family

$$\left\{\frac{N(t)}{t} : t \geqslant 0\right\}$$

is uniformly integrable. Here we use Scheffé's theorem.

Proof of Theorem 2.7. This equality has to be considered as two inequalities. First we have $t < S_{N(t)+1}$, and hence by Theorem 2.6 we see

$$t < \mathbb{E}\left[S_{N(t)+1}\right] = \mathbb{E}\left[X_1\right]\left(\mathbb{E}\left[N(t)\right] + 1\right) = \mathbb{E}\left[X_1\right]\left(m(t) + 1\right). \tag{2.14}$$

The inequality in (2.14) is equivalent to

$$\frac{m(t)}{t} \geqslant \frac{1}{\mathbb{E}\left[X_1\right]} - \frac{1}{t}. \tag{2.15}$$

From (2.15) we see

$$\liminf_{t\to\infty} \frac{m(t)}{t} \geqslant \liminf_{t\to\infty} \left(\frac{1}{\mathbb{E}\left[X_1\right]} - \frac{1}{t}\right) = \frac{1}{\mathbb{E}\left[X_1\right]}. \tag{2.16}$$

For the second inequality we proceed as follows. Fix a strictly positive real number a, and put $N_a(t) = \max\left\{n \in \mathbb{N} : \sum_{r=1}^{n} \min\left(a, X_r\right) \leqslant t\right\}$. Then $N(t) \leqslant N_a(t)$. Moreover, by Theorem 2.2 we have

$$\begin{aligned} t &\geqslant \mathbb{E}\left[S_{N_a(t)}\right] = \mathbb{E}\left[S_{N_a(t)+1} - \min\left(a, X_{N_a(t)+1}\right)\right] \\ &= \mathbb{E}\left[\min\left(a, X_1\right)\right]\mathbb{E}\left[N_a(t) + 1\right] - \mathbb{E}\left[\min\left(a, X_{N_a(t)+1}\right)\right] \\ &\geqslant \mathbb{E}\left[\min\left(a, X_1\right)\right]\mathbb{E}\left[N(t) + 1\right] - a = \left(m(t) + 1\right)\mathbb{E}\left[\min\left(a, X_1\right)\right] - a. \end{aligned} \tag{2.17}$$

Hence, from (2.17) we obtain:

$$\frac{m(t)}{t} \leqslant \frac{1}{\mathbb{E}\left[\min\left(a, X_1\right)\right]} + \frac{a - \mathbb{E}\left[\min\left(a, X_1\right)\right]}{t\mathbb{E}\left[\min\left(a, X_1\right)\right]}. \tag{2.18}$$

From (2.18) we deduce:

$$\limsup_{t\to\infty} \frac{m(t)}{t} \leqslant \frac{1}{\mathbb{E}\left[\min\left(a, X_1\right)\right]}, \quad \text{for all large } a > 0. \tag{2.19}$$

By letting $a \to \infty$ in (2.19) we see

$$\limsup_{t\to\infty} \frac{m(t)}{t} \leqslant \frac{1}{\mathbb{E}\left[X_1\right]}. \tag{2.20}$$

A combination of the inequalities (2.16) and (2.20) yields the result in Theorem 2.7. □

Next we extend these renewal theorems a little bit, by introducing a *renewal-reward* process $(R_n)_{n\in\mathbb{N}}$, where "costs" are considered as negative rewards. We are also interested in the cumulative reward up to time t: $C(t)$ (the reward is collected at the end of any interval); $C_i(t)$ (the reward is collected at the start of any interval); $C_P(t)$ (the reward accrues during any given time interval). More precisely we have:

$$C(t) = \sum_{j=1}^{N(t)} R_j, \quad \text{terminal reward at the end of time interval,} \tag{2.21}$$

$$C_i(t) = \sum_{j=1}^{N(t)+1} R_j, \quad \text{initial reward at the beginning of time interval,} \tag{2.22}$$

$$C_P(t) = \sum_{j=1}^{N(t)} R_j + P_{N(t)+1}, \quad \text{partial rewards during time interval.} \tag{2.23}$$

For the corresponding reward functions we write

$$c(t) = \mathbb{E}\left[C(t)\right], \quad c_i(t) = \mathbb{E}\left[C_i(t)\right] \quad \text{and} \quad c_p(t) = \mathbb{E}\left[C_P(t)\right]. \tag{2.24}$$

We are interested in the rates of reward: $\dfrac{C(t)}{t}$, $\dfrac{C_i(t)}{t}$, and $\dfrac{C_P(t)}{t}$. It is assumed that the renewal process $N(t)$ is defined by inter-arrival times X_r, $r \in \mathbb{N}$. As above these inter-arrival times are non-negative, independent and identically distributed on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. It is also assumed that the renewal-reward process R_n, $n \in \mathbb{N}$, consists of independent and identically distributed random variables in the space $L^1(\Omega, \mathcal{F}, \mathbb{P})$.

The following theorem will be proved.

2.9. Theorem (Renewal-reward theorem). *Suppose that* $0 < \mathbb{E}[X_1] < \infty$, $\mathbb{E}[|R_1|] < \infty$, *and that the sequence* $(n^{-1}P_n)_{n\in\mathbb{N}}$ *is uniformly bounded in* $n \in \mathbb{N}$ *and* ω, *and has the property that* $\lim_{n\to\infty} n^{-1}P_n = 0$, $\mathbb{P}$*-almost surely. Let the notation be as in (2.21), (2.22), (2.23), and (2.24). Then the following time average limits exist* $\mathbb{P}$*-almost surely and they are identified as* $\dfrac{\mathbb{E}[R_1]}{\mathbb{E}[X_1]}$:

$$\lim_{t\to\infty}\frac{C(t)}{t} = \lim_{t\to\infty}\frac{C_i(t)}{t} = \lim_{t\to\infty}\frac{C_P(t)}{t} = \frac{\mathbb{E}[R_1]}{\mathbb{E}[X_1]}, \quad \mathbb{P}\text{-almost surely.} \tag{2.25}$$

The following equalities hold as well:

$$\lim_{t\to\infty}\frac{c(t)}{t} = \lim_{t\to\infty}\frac{c_i(t)}{t} = \lim_{t\to\infty}\frac{c_P(t)}{t} = \frac{\mathbb{E}[R_1]}{\mathbb{E}[X_1]}. \tag{2.26}$$

Observe that the quotient $\dfrac{\mathbb{E}[R_1]}{\mathbb{E}[X_1]}$ can be interpreted as the "expected reward accruing in a cycle" divided by "expected duration of a cycle".

Other conditions on the sequence $(P_n : n \in \mathbb{N})$ can be given while retaining the conclusion in Theorem 2.9. For example the following conditions could be imposed. The sequence $(P_n : n \in \mathbb{N})$ is $\mathbb{P}$-independent and identically distributed, or there are finite deterministic constants c_1 and c_2 such that $|P_n| \leqslant c_1 n + c_2 |R_n|$ and $\lim\limits_{n\to\infty}\dfrac{P_n}{n} = 0$. In these cases the sequence $\left(\dfrac{P_n}{n} : n \in \mathbb{N}\right)$ is uniformly integrable and $\lim\limits_{n\to\infty}\dfrac{P_n}{n} = 0$ $\mathbb{P}$-almost surely.

Proof. By employing Theorem 2.2 and the strong law of large numbers we have

$$\lim_{t\to\infty}\frac{C(t)}{t} = \lim_{t\to\infty}\frac{\sum_{k=1}^{N(t)} R_k}{N(t)}\frac{N(t)}{t} = \frac{\mathbb{E}[R_1]}{\mathbb{E}[X_1]}. \tag{2.27}$$

In exactly the same manner, with $N(t)+1$ replacing $N(t)$, we see $\lim\limits_{t\to\infty}\dfrac{C_i(t)}{t} = \dfrac{\mathbb{E}[R_1]}{\mathbb{E}[X_1]}$. By hypothesis we know that $\lim\limits_{n\to\infty}\dfrac{P_n}{n} = 0$ $\mathbb{P}$-almost surely. Since

$$\lim_{t\to\infty} N(t) = \infty \quad \mathbb{P}\text{-almost surely}$$

we see that $\lim\limits_{t\to\infty}\dfrac{P_{N(t)+1}}{N(t)+1} = 0$. This together with (2.27) shows that

$$\lim_{t\to\infty}\frac{C_P(t)}{t} = \frac{\mathbb{E}[R_1]}{\mathbb{E}[X_1]}.$$

These arguments take care of the $\mathbb{P}$-almost sure convergence.

Next we consider the convergence of the time averaged expected values. For convergence of time average of the reward function $c_i(t) = \mathbb{E}[C_i(t)]$ we use

Wald's equation (see Theorem 2.2) and the elementary renewal Theorem 2.7. More precisely we have:

$$c_i(t) = \mathbb{E}\left[C_i(t)\right] = \mathbb{E}\left[\sum_{j=1}^{N(t)+1} R_j\right] = \mathbb{E}\left[R_1\right]\left(\mathbb{E}\left[N(t)\right]+1\right). \tag{2.28}$$

Then we divide by t, take the limit in (2.28) as t tends to ∞. An appeal to Theorem 2.6 then shows the existence of the limit $\lim_{t\to\infty}\frac{c_i(t)}{t} = \frac{\mathbb{E}\left[R_1\right]}{\mathbb{E}\left[X_1\right]}$ which is the second part of (2.26) in Theorem 2.9. First observe that $\lim_{n\to\infty}\frac{R_n}{n} = 0$ $\mathbb{P}$-almost surely. This can be seen by an appeal to the Borel-Cantelli lemma. In fact we have

$$\sum_{n=1}^{\infty}\mathbb{P}\left[\frac{|R_n|}{n} > \varepsilon\right] = \sum_{n=1}^{\infty}\mathbb{P}\left[\frac{|R_1|}{\varepsilon} > n\right] \leqslant \int_0^{\infty}\mathbb{P}\left[\frac{|R_1|}{\varepsilon} > x\right]dx \leqslant \mathbb{E}\left[\frac{|R_1|}{\varepsilon}\right]. \tag{2.29}$$

From (2.29) together with the Borel-Cantelli lemma it follows that $\lim_{n\to\infty}\frac{R_n}{n} = 0$ $\mathbb{P}$-almost surely. Consequently, the sequence $\left\{\frac{R_n}{n} : n \in \mathbb{N}\right\}$ is $\mathbb{P}$-uniformly integrable. Then we have

$$\frac{\left|R_{N(t)+1}\right|}{t} \leqslant \frac{N(t)+1}{t}\frac{\sum_{k=1}^{N(t)+1}|R_n|}{N(t)+1}, \tag{2.30}$$

and hence by Wald's equality

$$\mathbb{E}\left[\frac{\left|R_{N(t)+1}\right|}{t}\right] \leqslant \mathbb{E}\left[\frac{N(t)+1}{t}\frac{\sum_{k=1}^{N(t)+1}|R_n|}{N(t)+1}\right] = \frac{m(t)+1}{t}\mathbb{E}\left[|R_1|\right]. \tag{2.31}$$

By the strong law of large numbers and by the elementary renewal theorem 2.7 we see that the families of random variables

$$\frac{\sum_{k=1}^{N(t)+1}|R_n|}{t} = \frac{N(t)+1}{t}\frac{\sum_{k=1}^{N(t)+1}|R_n|}{N(t)+1}, \quad t > 0, \tag{2.32}$$

is uniformly integrable. Consequently the family $\left\{\frac{R_{N(t)+1}}{t} : t > 0\right\}$ is uniformly integrable, and hence it converges pointwise and in $L^1\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ to 0. Since

$$\frac{c(t)}{t} = \frac{1}{t}\left(\mathbb{E}\left[\sum_{k=1}^{N(t)+1} R_k\right] - \mathbb{E}\left[R_{N(t)+1}\right]\right) = \frac{m(t)+1}{t}\mathbb{E}\left[R_1\right] - \frac{\mathbb{E}\left[R_{N(t)+1}\right]}{t}. \tag{2.33}$$

The right-hand side of (2.33) converges to $\frac{\mathbb{E}\left[R_1\right]}{\mathbb{E}\left[X_1\right]}$. This proves the first part of (2.26) in Theorem 2.9. In order to prove the third part we need the uniform integrability of the family $\left\{\frac{P_{N(t)+1}}{t} : t \geqslant 1\right\}$. This fact is not entirely trivial.

Let the finite constant C be such that $|P_{n+1}| \leqslant C(n+1)$ for all $n \in \mathbb{N}$ and $\mathbb{P}$-almost surely; by hypothesis such a constant exists. From Remark 2.8 it follows that the family $\left\{\dfrac{N(t)+1}{t} : t \geqslant 1\right\}$ is uniformly integrable. Since

$$\frac{\left|P_{N(t)+1}\right|}{t} = \frac{\left|P_{N(t)+1}\right|}{N(t)+1}\frac{N(t)+1}{t} \leqslant C\frac{N(t)+1}{t} \tag{2.34}$$

it follows that the family $\left\{\dfrac{P_{N(t)+1}}{t} : t \geqslant 1\right\}$ is uniformly integrable as well. If $t \uparrow \infty$, then $N(t) \uparrow \infty$, and $\lim\limits_{t\to\infty} \dfrac{N(t)}{t} = \dfrac{1}{\mathbb{E}\left[X_1\right]}$ in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ as well as $\mathbb{P}$-almost surely: see Lemma 2.4, theorems 2.6, 2.7, and Remark 2.8. From (2.34) it follows that $\lim\limits_{t\to\infty} \dfrac{\mathbb{E}\left[\left|P_{N(t)+1}\right|\right]}{t} = 0$, which concludes the proof of Theorem 2.9. $\square$

1.1. Renewal theory and Markov chains. Next we consider this renewal theory in the context of strong Markov chains. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X_m, $m \in \mathbb{N}$, be a Markov chain on $(\Omega, \mathcal{F}, \mathbb{P})$ with state space $(S, \mathcal{S})$. Fix two states j and $k \in S$. Define the sequence of stopping times $T_k^{(r)}$, $r \in \mathbb{N}$, as follows:

$$T_k^{(r+1)} = \min\left\{n > T_k^{(r)} : X_n = k\right\}, \quad T_k^{(0)} = 0. \tag{2.35}$$

If $X_n \neq k$ for $n > T_k^{(r)}$, then we put $T_k^{(r+1)} = \infty$. The sequence of differences $T_k^{(r)} - T_k^{(r-1)}$, $r \geqslant 1$, are $\mathbb{P}_j$-independent and identically distributed.

2.10. Theorem. *Let $f : [0, \infty] \times S \to \mathbb{R}$ be a bounded measurable function. Then*

$$T_k^{(r+s)} = T_k^{(r)} + T_k^{(s)} \circ \vartheta_{T_k^{(r)}} \quad \text{on } \left\{T_k^{(r)} < \infty\right\}, \text{ and} \tag{2.36}$$

$$\begin{aligned}
&\mathbb{E}_j\left[f\left(T_k^{(r+s)}, X_{T_k^{(r+s)}}\right)\mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}} \mid \mathcal{F}_{T_k^{(r)}}\right] \\
&= \mathbb{E}_j\left[f\left(T_k^{(r+s)}, X_{T_k^{(r+s)}}\right)\mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}} \mid \sigma\left(T_k^{(r)}, X_{T_k^{(r)}}\right)\right] \\
&= \mathbb{E}_j\left[f\left(T_k^{(r+s)}, X_{T_k^{(r+s)}}\right)\mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}} \mid \sigma\left(T_k^{(r)}\right)\right] \\
&= \omega \mapsto \mathbb{E}_{X_{T_k^{(r)}(\omega)}(\omega)}\left[\omega' \mapsto f\left(T_k^{(r)}(\omega) + T_k^{(s)}(\omega'), X_{T_k^{(s)}(\omega')}(\omega')\right)\right]\mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}}(\omega) \\
&= \omega \mapsto \mathbb{E}_{k}\left[\omega' \mapsto f\left(T_k^{(r)}(\omega) + T_k^{(s)}(\omega'), X_{T_k^{(s)}(\omega')}(\omega')\right)\right]\mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}}(\omega).
\end{aligned} \tag{2.37}$$

Consequently, conditioned on the event $\left\{T_k^{(r)} < \infty\right\}$ the stochastic variable

$$T_k^{(r+s)} - T_k^{(r)}$$

and the σ-field $\mathcal{F}_{T_k^{(r)}}$ are $\mathbb{P}_j$-independent.

Suppose that $\mathbb{P}_k\left[T_k^{(r)} < \infty\right] = 1$ and $\mathbb{P}_j\left[T_k^{(1)} < \infty\right] > 0$. Then

$$\mathbb{P}_j\left[T_k^{(r+1)} < \infty\right] = \mathbb{P}_j\left[T_k^{(1)} < \infty\right],$$

and the variables $T_k^{(r+1)} - T_k^{(r)}$, $r \in \mathbb{N}$, have the same distribution with respect to the probability measure $A \mapsto \mathbb{P}_j\left[A \mid T_k^{(1)} < \infty\right]$.

Here $\mathbb{P}_j(A) = \mathbb{P}\left[A \mid X_0 = j\right]$, $A \in \mathcal{F}$, $j \in S$. Theorem 2.10 is a consequence of the strong Markov property.

Proof. First we prove (2.36). On the event $\left\{T_k^{(r)} < \infty\right\}$ we have

$$\begin{aligned}
T_k^{(r+1)} &= \min\left\{n > T_k^{(r)} : X_n = k\right\} = \min\left\{n > T_k^{(r)} : X_{n-T_k^{(r)}} \circ \vartheta_{T_k^{(r)}} = k\right\} \\
&= T_k^{(r)} + \min\left\{n - T_k^{(r)} \geqslant 1 : X_{n-T_k^{(r)}} \circ \vartheta_{T_k^{(r)}} = k\right\} \\
&= T_k^{(r)} + \min\left\{m \geqslant 1 : X_m \circ \vartheta_{T_k^{(r)}} = k\right\}
\end{aligned}$$

$$= T_k^{(r)} + T_k^{(1)} \circ \vartheta_{T_k^{(r)}}. \tag{2.38}$$

The equality in (2.38) shows (2.36) in case $s = 1$. We use (2.38) with s respectively $r + s$ instead of r to obtain (2.36) by induction on s. More precisely we have

$$\begin{aligned} T_k^{(r)} + T_k^{(s+1)} \circ \vartheta_{T_k^{(r)}} &= T_k^{(r)} + \left(T_k^{(s)} + T_k^{(1)} \circ \vartheta_{T_k^{(s)}} \right) \circ \vartheta_{T_k^{(r)}} \\ &= T_k^{(r)} + T_k^{(s)} \circ \vartheta_{T_k^{(r)}} + T_k^{(1)} \circ \vartheta_{T_k^{(r)} + T_k^{(s)} \circ \vartheta_{T_k^{(r)}}} \end{aligned} \tag{2.39}$$

(induction hypothesis)

$$= T_k^{(r+s)} + T_k^{(1)} \circ \vartheta_{T_k^{(r+s)}} = T_k^{(r+s+1)}, \tag{2.40}$$

where in (2.39) we employed (2.38) with s instead of r and in (2.40) we used $r+s$ instead of r. The equality in (2.40) shows (2.36) for $s+1$ assuming that it is true for s. Since by (2.38) the equality in (2.36) is true for $s = 1$, induction shows the equality in (2.36).

Next we will prove the equality in (2.37). From equality (2.36) we get

$$\begin{aligned} &\mathbb{E}_j \left[f\left(T_k^{(r+s)}, X_{T_k^{(r+s)}} \right) \mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}} \mid \mathcal{F}_{T_k^{(r)}} \right] \\ &= \mathbb{E}_j \left[f\left(T_k^{(r)} + T_k^{(s)} \circ \vartheta_{T_k^{(r)}}, X_{T_k^{(r)} + T_k^{(s)} \circ \vartheta_{T_k^{(r)}}} \right) \mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}} \mid \mathcal{F}_{T_k^{(r)}} \right] \\ &= \mathbb{E}_j \left[f\left(T_k^{(r)} + T_k^{(s)} \circ \vartheta_{T_k^{(r)}}, X_{T_k^{(s)}} \circ \vartheta_{T_k^{(r)}} \right) \mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}} \mid \mathcal{F}_{T_k^{(r)}} \right] \end{aligned}$$

(the variable $T_k^{(r)}$ is $\mathcal{F}_{T_k^{(r)}}$-measurable in combination with the strong Markov property)

$$\begin{aligned} &= \omega \mapsto \mathbb{E}_{X_{T_k^{(r)}(\omega)}(\omega)} \left[f\left(T_k^{(r)}(\omega) + T_k^{(s)}, X_{T_k^{(s)}} \right) \right] \mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}}(\omega) \\ &= \omega \mapsto \mathbb{E}_k \left[f\left(T_k^{(r)}(\omega) + T_k^{(s)}, X_{T_k^{(s)}} \right) \right] \mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}}(\omega). \end{aligned} \tag{2.41}$$

The equalities in (2.41) show that the first, penultimate and ultimate quantity in (2.37) are equal. Another appeal to the strong Markov property shows that the first and second quantity in (2.37) coincide. Since on the event $\left\{T_k^{(r)} < \infty\right\}$ the equality $X_{T_k^{(r)}} = k$ holds, the second and third quantity in (2.37) are equal as well. This proves that all quantities in (2.37) in Theorem 2.10 are the same. We still have to prove that on the event $\left\{T_k^{(r)} < \infty\right\}$ the stochastic variable $T_k^{(r+s)} - T_k^{(r)}$ and the σ-field $\mathcal{F}_{T_k^{(r)}}$ are $\mathbb{P}_j$-independent. This can be achieved as follows. Let the event A be $\mathcal{F}_{T_k^{(r)}}$-measurable and let $g : [0, \infty] \to \mathbb{R}$ be a bounded measurable function. Then we have

$$\begin{aligned} &\mathbb{E}_j \left[g\left(T_k^{(r+s)} - T_k^{(r)} \right) \mathbf{1}_A \mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}} \right] = \mathbb{E}_j \left[g\left(T_k^{(s)} \circ \vartheta_{T_k^{(r)}} \right) \mathbf{1}_A \mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}} \right] \\ &= \mathbb{E}_j \left[\mathbb{E}_j \left[g\left(T_k^{(s)} \circ \vartheta_{T_k^{(r)}} \right) \mathbf{1}_A \mid \mathcal{F}_{T_k^{(r)}} \right] \mathbf{1}_{\left\{T_k^{(r)}<\infty\right\}} \right] \end{aligned} \tag{2.42}$$

$$= \mathbb{E}_j \left[\mathbb{E}_j \left[g\left(T_k^{(s)} \circ \vartheta_{T_k^{(r)}} \right) \mid \mathcal{F}_{T_k^{(r)}} \right] \mathbf{1}_A \mathbf{1}_{\{T_k^{(r)} < \infty\}} \right]$$

(strong Markov property: (2.37))

$$= \mathbb{E}_j \left[\mathbb{E}_{X_{T_k^{(r)}}} \left[g\left(T_k^{(s)} \right) \right] \mathbf{1}_A \mathbf{1}_{\{T_k^{(r)} < \infty\}} \right] = \mathbb{E}_j \left[\mathbb{E}_k \left[g\left(T_k^{(s)} \right) \right] \mathbf{1}_A \mathbf{1}_{\{T_k^{(r)} < \infty\}} \right]$$

$$= \mathbb{E}_k \left[g\left(T_k^{(s)} \right) \right] \mathbb{E}_j \left[\mathbf{1}_A \mathbf{1}_{\{T_k^{(r)} < \infty\}} \right] \tag{2.43}$$

(another appeal to the strong Markov property: (2.37))

$$= \mathbb{E}_j \left[\mathbb{E}_j \left[g\left(T_k^{(s)} \right) \circ \vartheta_{T_k^{(r)}} \mathbf{1}_{\{T_k^{(r)} < \infty\}} \mid \mathcal{F}_{T_k^{(r)}} \right] \right] \frac{\mathbb{E}_j \left[\mathbf{1}_A \mathbf{1}_{\{T_k^{(r)} < \infty\}} \right]}{\mathbb{P}_j \left[T_k^{(r)} < \infty \right]}$$

$$= \mathbb{E}_j \left[\mathbb{E}_j \left[g\left(T_k^{(s)} \circ \vartheta_{T_k^{(r)}} \right) \mathbf{1}_{\{T_k^{(r)} < \infty\}} \mid \mathcal{F}_{T_k^{(r)}} \right] \right] \mathbb{E}_j \left[\mathbf{1}_A \mid T_k^{(r)} < \infty \right]$$

(use (2.36))

$$= \mathbb{E}_j \left[\mathbb{E}_j \left[g\left(T_k^{(r+s)} - T_k^{(r)} \right) \mathbf{1}_{\{T_k^{(r)} < \infty\}} \mid \mathcal{F}_{T_k^{(r)}} \right] \right] \mathbb{E}_j \left[\mathbf{1}_A \mid T_k^{(r)} < \infty \right]$$

$$= \mathbb{E}_j \left[g\left(T_k^{(r+s)} - T_k^{(r)} \right) \mathbf{1}_{\{T_k^{(r)} < \infty\}} \right] \mathbb{E}_j \left[\mathbf{1}_A \mid T_k^{(r)} < \infty \right]. \tag{2.44}$$

From (2.44) the $\mathbb{P}_j$-independence of $T_k^{(r+s)} - T_k^{(r)}$ and the σ-field $\mathcal{F}_{T_k^{(r)}}$ conditioned on the event $\left\{T_k^{(r)} < \infty\right\}$ follows. Since the expressions in (2.42) and (2.43) are equal it follows that the $\mathbb{P}_j\left[\cdot \mid T_k^{(1)} < \infty\right]$-distribution of the variable $T_k^{(r+1)} - T_k^{(r)}$ does not depend on r, provided that

$$\mathbb{P}_j\left[\left\{T_k^1 < \infty\right\} \setminus \left\{T_k^{(r)} < \infty\right\}\right] = 0. \tag{2.45}$$

By the strong Markov property it follows that

$$\mathbb{P}_j\left[T_k^{(r+1)} < \infty\right] = \mathbb{P}_j\left[T_k^{(1)} < \infty\right]\mathbb{P}_k\left[T_k^{(1)} < \infty\right]^r. \tag{2.46}$$

Since, by assumption, $\mathbb{P}_k\left[T_k^{(1)} < \infty\right] = 1$, (2.46) implies that the probabilities $\mathbb{P}_j\left[T_k^{(r)} < \infty\right]$ do not depend on $r \in \mathbb{N}$, and hence (2.45) follows. This proves Theorem 2.10. $\square$

2.11. Definition. Let $j \in S$. If $\mathbb{P}_j\left[T_j^{(1)} < \infty\right] = 1$, then j is called recurrent (or persistent). If $\mathbb{P}_j\left[T_j^{(1)} < \infty\right] < 1$, then j is called a transient state. A recurrent state for which $\mathbb{E}_j\left[T_j^{(1)}\right] = \infty$ is called a null state. A recurrent state for which $\mathbb{E}_j\left[T_j^{(1)}\right] < \infty$ is called a non-null or positive state.

From (2.46) it follows that $\mathbb{P}_j\left[T_j^{(r)} < \infty\right] = \mathbb{P}_j\left[T_j^{(1)} < \infty\right]^r$, and hence if a state j is recurrent it is expected to be visited infinitely many times. Let $N^k = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = k\}}$ be the number of visits to the state k, and put $\nu_j^k = \mathbb{E}_j\left[N^k\right] = \mathbb{E}\left[N^k \mid X_0 = j\right]$. Then

$$\nu_j^k = \sum_{n=1}^{\infty} p_{jk}^{(n)} = \sum_{n=1}^{\infty} \mathbb{P}\left[X_n = k \mid X_0 = j\right]. \tag{2.47}$$

We also have $\left\{N^k \geqslant r+1\right\} = \left\{T_k^{(r+1)} < \infty\right\}$ and hence by (2.46) we get

$$\begin{aligned}\mathbb{P}_j\left[T_k^{(r+1)} < \infty\right] &= \mathbb{P}_j\left[N^k \geqslant r+1\right] = \mathbb{P}_j\left[N^k > 0\right]\mathbb{P}_k\left[N^k > 0\right]^r \\ &= \mathbb{P}_j\left[T_k^{(1)} < \infty\right]\mathbb{P}_k\left[T_k^{(1)} < \infty\right]^r.\end{aligned} \tag{2.48}$$

From (2.47) and (2.48) it follows that

$$\nu_j^k = \sum_{n=1}^{\infty} p_{jk}^{(n)} = \sum_{n=1}^{\infty} \mathbb{P}\left[X_n = k \mid X_0 = j\right] = \mathbb{P}_j\left[N^k > 0\right]\sum_{r=1}^{\infty} \mathbb{P}_k\left[N^k > 0\right]^r. \tag{2.49}$$

Suppose that the state j communicates with k, i.e. suppose that $p_{jk}^{(n)} > 0$ for some integer $n \geqslant 1$. From (2.49) it follows that the state k is recurrent if and only if $\sum_{n=1}^{\infty} p_{jk}^{(n)} = \infty$. The state k is transient if and only if $\sum_{n=1}^{\infty} p_{jk}^{(n)} < \infty$.

2.12. THEOREM. *Suppose that the states j and k intercommunicate. Then either both states are recurrent or both states are transient.*

PROOF. Since the states j and k intercommunicate the exist positive integers m and n such that $p_{jk}^{(m)} > 0$ and $p_{kj}^{(n)} > 0$. For any positive integer r we then have

$$p_{jj}^{(m+r+n)} \geqslant p_{jk}^{(m)} p_{kk}^{(r)} p_{kj}^{(n)} \tag{2.50}$$

By summing over r in (2.50) we see that $\sum_{r=1}^{\infty} p_{jj}^{(r)} < \infty$ if and only if $\sum_{r=1}^{\infty} p_{kk}^{(r)} < \infty$. From this fact together with (2.49) the statement in Theorem 2.12 follows. □

2.13. DEFINITION. A Markov chain with state space S is called irreducible of all states communicate, *i.e.* for every $j,\ k \in S$ there exists $n \in \mathbb{N}$ such that $p_{j,k}^{(n)} > 0$. If X is irreducible and all states and one, and so all states, are recurrent, then X is called recurrent.

2.14. THEOREM. *Let X be a recurrent and irreducible Markov chain. Put*

$$v_j^k = \mathbb{E}\left[\sum_{u=1}^{T_k^{(1)}} \mathbf{1}_{\{X_u=j\}} \;\Big|\; X_0 = k\right] = \mathbb{E}_k\left[\sum_{u=1}^{T_k^{(1)}} \mathbf{1}_{\{X_u=j\}}\right]. \tag{2.51}$$

Then $0 < v_j^k < \infty$, $j,\ k \in S$, and $v_j^k = \sum_{i\in S} v_i^k p_{ij}$. In other words the vector $\left(v_j^k : j \in S\right)$ is an invariant measure for X.

PROOF. First we prove that $0 < v_j^k < \infty$. Therefore we notice that

$$\begin{aligned} v_j^k &= \mathbb{E}_k\left[\sum_{u=1}^{T_k^{(1)}} \mathbf{1}_{\{X_u=j\}}\right] = \sum_{r=0}^{\infty} \mathbb{P}_k\left[T_k^{(1)} \geqslant T_j^{(r+1)}\right] \\ &= \sum_{r=0}^{\infty} \mathbb{P}_k\left[T_k^{(1)} \geqslant T_j^{(1)}\right]\left(\mathbb{P}_j\left[T_k^{(1)} \geqslant T_j^{(1)}\right]\right)^r, \end{aligned} \tag{2.52}$$

where we used the equality:

$$\mathbb{P}_k\left[T_k^{(1)} \geqslant T_j^{(r+1)}\right] = \mathbb{P}_k\left[T_k^{(1)} \geqslant T_j^{(1)}\right]\left(\mathbb{P}_j\left[T_k^{(1)} \geqslant T_j^{(1)}\right]\right)^r. \tag{2.53}$$

Suppose $j \neq k$; for $j = k$ we have $v_k^k = 1$. The equality in (2.53) follows from the strong Markov property as follows. For $r = 0$ the equality is clear. For $r \geqslant 1$ we have

$$\begin{aligned} &\mathbb{P}_k\left[T_k^{(1)} \geqslant T_j^{(r+1)}\right] = \mathbb{P}_k\left[T_k^{(1)} \geqslant T_j^{(r+1)},\, T_k^{(1)} \geqslant T_j^{(r)} + 1\right] \\ &= \mathbb{P}_k\left[T_j^{(r)} + T_k^{(1)} \circ \vartheta_{T_j^{(r)}} \geqslant T_j^{(r)} + T_j^{(1)} \circ \vartheta_{T_j^{(r)}},\, T_k^{(1)} \geqslant T_j^{(r)} + 1\right] \\ &= \mathbb{E}_k\left[\mathbb{P}_k\left[T_k^{(1)} \circ \vartheta_{T_j^{(r)}} \geqslant T_j^{(1)} \circ \vartheta_{T_j^{(r)}} \mid \mathcal{F}_{T_j^{(r)}}\right],\, T_k^{(1)} \geqslant T_j^{(r)} + 1\right] \end{aligned}$$

(strong Markov property)

$$= \mathbb{E}_k\left[\mathbb{P}_{X_{T_j^{(r)}}}\left[T_k^{(1)} \geqslant T_j^{(1)}\right],\, T_k^{(1)} \geqslant T_j^{(r)} + 1\right]$$

$$= \mathbb{E}_k \left[\mathbb{P}_j \left[T_k^{(1)} \geqslant T_j^{(1)} \right], \, T_k^{(1)} \geqslant T_j^{(r)} + 1 \right]$$
$$= \mathbb{P}_j \left[T_k^{(1)} \geqslant T_j^{(1)} \right] \mathbb{P}_k \left[T_k^{(1)} \geqslant T_j^{(r)} + 1 \right]$$

(induction with respect to r)

$$= \left(\mathbb{P}_j \left[T_k^{(1)} \geqslant T_j^{(1)} \right] \right)^r \mathbb{P}_k \left[T_k^{(1)} \geqslant T_j^{(1)} \right]. \tag{2.54}$$

Since $\mathbb{P}_k \left[T_k^{(1)} \geqslant T_j^{(1)} \right] > 0$ it follows by (2.52) that $v_j^k > 0$. By the same equality and using the fact that $\mathbb{P}_j \left[T_k^{(1)} \geqslant T_j^{(1)} \right] < 1$ we see $v_j^k < \infty$.

Next we prove the equality: $v_j^k = \sum_{i \in S} v_i^k p_{ij}$. Therefore we write

$$v_j^k = \mathbb{E}_k \left[\sum_{n=1}^{T_k^{(1)}} \mathbf{1}_{\{X_n = j\}} \right] = \sum_{n=1}^{\infty} \mathbb{P}_k \left[X_n = j, \, T_k^{(1)} \geqslant n \right]$$
$$= \sum_{i \in S} \sum_{n=1}^{\infty} \mathbb{P}_k \left[X_n = j, \, X_{n-1} = i, \, T_k^{(1)} \geqslant n \right]$$
$$= \sum_{i \in S} \sum_{n=1}^{\infty} \mathbb{E}_k \left[\mathbb{P}_k \left[X_n = j \mid \mathcal{F}_{n-1} \right], \, X_{n-1} = i, \, T_k^{(1)} \geqslant n \right]$$

(Markov property)

$$= \sum_{i \in S} \sum_{n=1}^{\infty} \mathbb{E}_k \left[\mathbb{P}_{X_{n-1}} \left[X_1 = j \right], \, X_{n-1} = i, \, T_k^{(1)} \geqslant n \right]$$
$$= \sum_{i \in S} \mathbb{P}_i \left[X_1 = j \right] \sum_{n=0}^{\infty} \mathbb{P}_k \left[X_n = i, \, T_k^{(1)} - 1 \geqslant n \right]$$
$$= \sum_{i \in S} \mathbb{P}_i \left[X_1 = j \right] \mathbb{E}_k \left[\sum_{n=0}^{T_k^{(1)}-1} \mathbf{1}_{\{X_n = i\}} \right]$$
$$= \sum_{i \in S} \mathbb{P}_i \left[X_1 = j \right] \mathbb{E}_k \left[\sum_{n=1}^{T_k^{(1)}} \mathbf{1}_{\{X_n = i\}} \right]. \tag{2.55}$$

In the last equality of (2.55) we used the equality

$$\mathbb{E}_k \left[\sum_{n=0}^{T_k^{(1)}-1} \mathbf{1}_{\{X_n = i\}} \right] = \mathbb{E}_k \left[\sum_{n=1}^{T_k^{(1)}} \mathbf{1}_{\{X_n = i\}} \right],$$

which is evident for $i \neq k$ and both are equal to 1 for $i = k$. As a consequence from (2.55) we see that $v_j^k = \sum_{i \in S} v_i^k p_{ij}$. □

2.15. Corollary. *Let the row vector* $\mathbf{v}^k := \left(v_j^k : j \in S \right)$ *be as in equality (2.51) of Theorem 2.14. Then* $\mathbf{v}^k$ *is minimal invariant measure in the sense that if*

$\mathbf{x} = (x_j : j \in S)$ *is another invariant measure such that* $x_k = 1$. *Then* $x_j \geqslant v_j^k$, $j \in S$.

PROOF. We write:

$$
\begin{aligned}
x_j &= p_{kj} + \sum_{s \in S,\, s \neq k} x_s p_{sj} \\
&= \mathbb{P}_k\left[X_1 = j,\, T_k^{(1)} \geqslant 1\right] + \sum_{s_1 \in S,\, s_1 \neq k} \sum_{s_2 \in S} x_{s_2} p_{s_2 s_1} p_{s_1 j} \\
&= \mathbb{P}_k\left[X_1 = j,\, T_k^{(1)} \geqslant 1\right] + \sum_{s_1 \in S,\, s_1 \neq k} p_{k s_1} p_{s_1 j} + \sum_{s_1 \in S,\, s_1 \neq k} \sum_{s_2 \in S,\, s_2 \neq k} x_{s_2} p_{s_2 s_1} p_{s_1 j} \\
&\geqslant \mathbb{P}_k\left[X_1 = j,\, T_k^{(1)} \geqslant 1\right] + \mathbb{P}_k\left[X_2 = j,\, T_k^{(1)} \geqslant 2\right] \\
&\quad + \cdots + \mathbb{P}_k\left[X_n = j,\, T_k^{(1)} \geqslant n\right]. \qquad (2.56)
\end{aligned}
$$

Upon letting n tend to ∞ in (2.56) we see that $x_j \geqslant v_j^k$. This proves Corollary 2.15. $\square$

2.16. Theorem. *Let X be a irreducible Markov chain with transition matrix*

$$\mathbf{P} = (p_{ij})_{(i,j)\in S\times S}\,.$$

The following assertions hold:

(a) *If any state is non-null recurrent, then all states are.*
(b) *The chain is non-null recurrent if and only if there exists a stationary distribution π or invariant measure. If this is the case, then*

$$\pi_k = \frac{1}{\mathbb{E}_k\left[T_k^{(1)}\right]} \quad \textit{and} \quad v_j^k = \frac{\pi_j}{\pi_k} \quad \textit{(see (2.51))}. \tag{2.57}$$

As a consequence of (2.57) stationary distributions are unique.

Proof. (a) The proof of assertion will follow from the proof of assertion (b).

(b) Let k be a state which is non-null recurrent. By Theorem 2.12 it follows that the chain is recurrent. By Theorem 2.14 the vector $\left(v_j^k : j \in S\right)$ as defined in (2.51) is an invariant vector. Since k is non-null recurrent it follows that $0 < \mathbb{E}_k\left[T_k^{(1)}\right] < \infty$, and that the vector $\left(\dfrac{v_j^k}{\pi_k} : j \in S\right)$, with $\pi_k = \dfrac{1}{\mathbb{E}_k\left[T_k^{(1)}\right]}$, is a stationary vector. It follows that if the irreducible chain X contains at least one non-null recurrent state, then there exists a stationary distribution. Next suppose that there exists a stationary distribution $\pi := (\pi_j : j \in S)$. Then $\pi_k = \sum_{j\in S} \pi_j p_{jk}^{(n)}$ for all $n \in \mathbb{N}$. Since the chain is irreducible, and the vector is a probability vector, at least one $\pi_{j_0} \neq 0$. By irreducibility there exists $n \in \mathbb{N}$ such that $p_{j_0 k}^{(n)} \neq 0$, and hence $\pi_k \neq 0$, $k \in S$. Consider for any given $k \in S$ the vector $\mathbf{x} = \left(\dfrac{\pi_j}{\pi_k} : j \in S\right)$. Then $x_k = 1$ and by Corollary 2.15 $x_j \geqslant v_j^k$ for all $j \in S$. It follows that

$$\mathbb{E}_k\left[T_k^{(1)}\right] = \sum_{j\in S} v_j^k \leqslant \sum_{j\in S} \frac{\pi_j}{\pi_k} = \frac{1}{\pi_k}. \tag{2.58}$$

Therefore k is non-null recurrent for all $k \in S$. It follows that if there exists one non-null recurrent vector $k \in S$, then all states in S are non-null recurrent. Altogether this proves assertion (a), and also a large part of (b). From (2.58) the first equality in (2.57) follows. Finally we will show the second equality in (2.57). The vector $\mathbf{x} - \mathbf{v}^k$ is invariant and, by Corollary 2.15 $\dfrac{\pi_j}{\pi_k} - v_j^k \geqslant 0$. Hence we obtain, for all positive integers n,

$$0 = 1 - v_k^k = \sum_{i\in S}\left(\frac{\pi_i}{\pi_k} - v_i^k\right) p_{ik}^{(n)} \geqslant \left(\frac{\pi_j}{\pi_k} - v_j^k\right) p_{jk}^{(n)}. \tag{2.59}$$

In (2.59) we choose n in such a way that $p_{jk}^{(n)} > 0$. By irreducibility this is possible. It follows that (see (2.51))

$$v_j^k = \mathbb{E}\left[\sum_{u=1}^{T_k^{(1)}} \mathbf{1}_{\{X_u=j\}} \;\middle|\; X_0 = k\right] = \mathbb{E}_k\left[\sum_{u=1}^{T_k^{(1)}} \mathbf{1}_{\{X_u=j\}}\right] = \frac{\pi_j}{\pi_k} = \frac{\mathbb{E}_k\left[T_k^{(1)}\right]}{\mathbb{E}_j\left[T_j^{(1)}\right]}. \quad (2.60)$$

The second equality in (2.57) is the same as (2.60). This concludes the proof of Theorem 2.16. □

Let k be a non-null recurrent state, and suppose that the state is j intercommunicates with k. Then both states are non-null or positive recurrent. Next we define the renewal process $N_k(n)$, $n \in \mathbb{N}$, as follows:

$$N_k(n) = \max\left\{r : T_k^{(r)} \leqslant n\right\}. \quad (2.61)$$

Notice the inequalities:

$$T_k^{(N_k(n))} \leqslant n < T_k^{(N_k(n+1))},$$

and hence $T_k^{(m)} = n$ if $m = N_k(n)$. We are interested in the following type of limits:

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} p_{kj}^{(\ell)} &= \lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} \mathbb{P}\left[X_\ell = j \;\middle|\; X_0 = k\right] \\ &= \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}\sum_{\ell=1}^{n} \mathbf{1}_{\{X_\ell=j\}} \;\middle|\; X_0 = k\right]. \quad (2.62)\end{aligned}$$

Put $m = N_k(n)$. Then $T_k^{(m)} = n$, and consequently we see

$$\frac{1}{n}\sum_{\ell=1}^{n} \mathbf{1}_{\{X_\ell=j\}} = \frac{N_k(n)}{n}\frac{1}{m}\sum_{u=1}^{m} \sum_{\ell=T_k^{(u-1)}+1}^{T_k^{(u)}} \mathbf{1}_{\{X_\ell=j\}}. \quad (2.63)$$

Notice that for $j = k$ we have

$$\frac{1}{m}\sum_{u=1}^{m} \sum_{\ell=T_k^{(u-1)}+1}^{T_k^{(u)}} \mathbf{1}_{\{X_\ell=j\}} = 1,$$

and consequently,

$$\frac{1}{n}\sum_{\ell=1}^{n} \mathbf{1}_{\{X_\ell=k\}} = \frac{N_k(n)}{n}.$$

Hence, we observe that (see Theorem 2.6)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{\ell=1}^{n} \mathbf{1}_{\{X_\ell=k\}} = \lim_{n\to\infty} \frac{N_k(n)}{n} = \frac{1}{\mathbb{E}\left[T_k^{(1)} \;\middle|\; X_0 = k\right]} = \frac{1}{\mu_k}, \ \mathbb{P}_k\text{-almost surely.}$$

(2.64)

We also see that

$$\frac{N_j(n)}{n} = \frac{1}{n}\sum_{\ell=1}^{n} \mathbf{1}_{\{X_\ell = j\}} = \frac{N_k(n)}{n}\frac{1}{m}\sum_{u=1}^{m}\sum_{\ell=T_k^{(u-1)}+1}^{T_k^{(u)}} \mathbf{1}_{\{X_\ell = j\}}. \tag{2.65}$$

From the strong law of large numbers we get:

$$\lim_{m\to\infty}\frac{1}{m}\sum_{u=1}^{m}\sum_{\ell=T_k^{(u-1)}+1}^{T_k^{(u)}} \mathbf{1}_{\{X_\ell = j\}} = \mathbb{E}\left[\sum_{u=1}^{T_k^{(1)}} \mathbf{1}_{\{X_u = j\}} \mid X_0 = k\right] = v_j^k. \tag{2.66}$$

From (2.63), (2.64), and (2.65) we obtain:

$$\begin{aligned}\frac{1}{\mathbb{E}\left[T_j^{(1)} \mid X_0 = j\right]} &= \lim_{n\to\infty}\frac{N_j(n)}{n} = \lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n}\mathbf{1}_{\{X_\ell = j\}}\\ &= \lim_{n\to\infty}\frac{N_k(n)}{n}\frac{1}{m}\sum_{u=1}^{m}\sum_{\ell=T_k^{(u-1)}+1}^{T_k^{(u)}} \mathbf{1}_{\{X_\ell = j\}} = \frac{v_j^k}{\mathbb{E}\left[T_k^{(1)} \mid X_0 = k\right]}.\end{aligned} \tag{2.67}$$

The equality in (2.67) together with Theorem 2.9 shows the following theorem.

2.17. Theorem. *Let the sequence of stopping times $\left(T_k^{(r)}\right)_{r\in\mathbb{N}}$ be defined as in (2.35), and let v_j^k be defined as in (2.66). Suppose that the states j and k intercommunicate and that one of them is non-null recurrent, then the other is also non-null recurrent. Moreover,*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n}\mathbf{1}_{\{X_\ell = j\}} = \frac{1}{\mathbb{E}\left[T_j^{(1)} \mid X_0 = j\right]} = \frac{v_j^k}{\mathbb{E}\left[T_k^{(1)} \mid X_0 = k\right]}, \quad \text{and} \tag{2.68}$$

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n} p_{kj}^{(\ell)} &= \lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n}\mathbb{P}\left[X_\ell = j \mid X_0 = k\right]\\ &= \frac{1}{\mathbb{E}\left[T_j^{(1)} \mid X_0 = j\right]} = \frac{v_j^k}{\mathbb{E}\left[T_k^{(1)} \mid X_0 = k\right]}.\end{aligned} \tag{2.69}$$

Hence, with $\pi_j = \lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n} p_{kj}^{(\ell)}$, $\mu_j = \mathbb{E}\left[T_j^{(1)} \mid X_0 = j\right]$, and v_j^k as in equality (2.66) we have $\pi_j = \dfrac{1}{\mu_j} = \dfrac{v_j^k}{\mu_k}$.

1.1.1. *Random walks.* In this example the state space is $\mathbb{Z}$, and the process X_n, $n \in \mathbb{N}$, has a transition probability matrix with the following entries: $p_{i-1,i} = q$, $p_{i+1,i} = p$, $0 < p = 1 - q < 1$, and $p_{j,i} = 0$, $j \neq i \pm 1$. Such a random walk can be realized by putting $X_n = \sum_{k=0}^{n} S_k$, where S_0 is the initial state (which may be random), the variables S_k, $k \in \mathbb{N}$, $k \geqslant 1$, are $\mathbb{P}$-independent of each other and are also $\mathbb{P}$-independent of S_0. Moreover, each variable S_k, $k \geqslant 1$, is a Bernoulli variable taking the value $+1$ with probability p and the value -1 with probability q. This Markov chain is irreducible: every state communicates with every other one. The set of states is closed. The corresponding infinite transition matrix looks as follows:

$$\begin{pmatrix} \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \ddots & q & 0 & p & 0 & \dots \\ \dots & 0 & q & 0 & p & \dots \\ \dots & 0 & 0 & q & 0 & \ddots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}.$$

The state 0 has period two $p_{00}^{(2n+1)} = 0$, and $p_{00}^{(2n)} = \binom{2n}{n} p^n q^n$. In order to check transiency (or recurrence) we need to calculate

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} = \sum_{n=0}^{\infty} \binom{2n}{n} p^n q^n. \tag{2.70}$$

By Stirling's formula we have $n! \sim \sqrt{2\pi n} n^n e^{-n}$, which means that

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} = 1.$$

Since

$$\binom{2n}{n} = \frac{(2n)!}{(n!)^2} \sim \frac{\sqrt{4\pi n}\,(2n)^{2n}\,e^{-2n}}{2\pi n^{2n+1}e^{-2n}} = \frac{4^n}{\sqrt{\pi n}} \tag{2.71}$$

the sum in (2.70) is finite if and only if the sum

$$\sum_{n=1}^{\infty} \frac{(4pq)^n}{\sqrt{\pi n}} < \infty. \tag{2.72}$$

If $p = 1 - q \neq \frac{1}{2}$, then $4pq < 1$, and hence the sum in (2.72) is finite, and so the unrestricted asymmetric random walk in $\mathbb{Z}$ is transient. However, if $p = q = \frac{1}{2}$, then $4pq = 1$ and the sum in (2.72) diverges, and so the symmetric unrestricted random walk in $\mathbb{Z}$ is recurrent. One may also do similar calculations for symmetric random walks in $\mathbb{Z}^2$, $\mathbb{Z}^3$, and $\mathbb{Z}^d$, $d \geqslant 4$. It turns out that in $\mathbb{Z}^2$ the $2n$-th symmetric transition probability $p_{00}^{(2n)}$ satisfies (for $n \to \infty$) $p_{00}^{(2n)} \sim \dfrac{1}{\pi n}$, and hence the sum $\sum_{n=0}^{\infty} p_{00}^{(2n)} = \infty$. It follows that the symmetric random walk in $\mathbb{Z}^2$ is recurrent. The corresponding return probability $p_{00}^{(2n)}$ for the symmetric random walk in $\mathbb{Z}^3$ possesses the following asymptotic behavior:

$$p_{00}^{(2n)} \sim \frac{1}{2}\left(\frac{3}{\pi}\right)^{3/2} \frac{1}{n^{3/2}}.$$

Hence the sum $\sum_{n=0}^{\infty} p_{00}^{(2n)} < \infty$, and so the state 0 is transient. The $2n$-th return probabilities of the symmetric random walk in $\mathbb{Z}^d$ satisfies

$$p_{00}^{(2n)} \sim \frac{c_d}{n^{d/2}}, \quad n \to \infty,$$

for some constant c_d and hence the sum $\sum_{n=0}^{\infty} p_{00}^{(2n)} < \infty$ in dimensions $d \geqslant 3$. So in dimensions $d \geqslant 3$ the symmetric random walk is transient, and in the dimension $d = 1, 2$, the symmetric random walk is recurrent.

We come back to the one-dimensional situation, and we reconsider the return times to a state $k \in \mathbb{Z}$: $T_k^{(1)} = \inf\{n \geqslant 1 : X_n = k\}$. Notice that S_k, $k \in \mathbb{N}$, are the step sizes which are ± 1. Also observe that $X_k = \sum_{j=0}^{k} S_j$. We consider the moment generating function $G_{j,k}(s) = \mathbb{E}_j\left[s^{T_k^{(1)}}\right]$, $0 \leqslant s < 1$. Observe that on the event $\left\{T_k^{(1)} = \infty\right\}$ the quantity $s^{T_k^{(1)}}$ has to be interpreted as 0. In addition, we have $\mathbb{P}_j\left[T_k^{(1)} > T_{k-1}^{(1)}\right] = \mathbb{P}_j\left[T_{k-1}^{(1)} < \infty\right]$ for $k > j$, k, $j \in \mathbb{Z}$. Then it follows that $T_k^{(1)} = T_{k-1}^{(1)} + T_k^{(1)} \circ \vartheta_{T_{k-1}^{(1)}}$, $\mathbb{P}_j$-almost surely, $k > j$, k, $j \in \mathbb{Z}$. Then by the strong Markov property we get

$$\begin{aligned}
G_{j,k}(s) &= \mathbb{E}_j\left[s^{T_k^{(1)}}\right] = \mathbb{E}_j\left[s^{T_{k-1}^{(1)} + T_k^{(1)} \circ \vartheta_{T_{k-1}^{(1)}}},\, T_{k-1}^{(1)} < \infty\right] \\
&= \mathbb{E}_j\left[s^{T_{k-1}^{(1)}} \mathbb{E}_j\left[s^{T_k^{(1)} \circ \vartheta_{T_{k-1}^{(1)}}} \mid \mathcal{G}_{T_{k-1}^{(1)}}\right],\, T_{k-1}^{(1)} < \infty\right]
\end{aligned}$$

(Markov property)

$$
\begin{aligned}
&= \mathbb{E}_j\left[s^{T_{k-1}^{(1)}}\mathbb{E}_{X_{T_{k-1}^{(1)}}}\left[s^{T_k^{(1)}}\right],\, T_{k-1}^{(1)}<\infty\right] \\
&= \mathbb{E}_j\left[s^{T_{k-1}^{(1)}}\mathbb{E}_{k-1}\left[s^{T_k^{(1)}}\right]\right] = \mathbb{E}_j\left[s^{T_{k-1}^{(1)}}\right]\mathbb{E}_{k-1}\left[s^{T_k^{(1)}}\right]. \qquad (2.73)
\end{aligned}
$$

From (2.73) we see by induction with respect to k that

$$
G_{j,k}(s) = \prod_{\ell=j}^{k-1} G_{\ell,\ell+1}(s) = G_{0,1}(s)^{k-j}. \qquad (2.74)
$$

In the final step of (2.74) we used the fact that the $\mathbb{P}_\ell$-distribution of $T_{\ell+1}^{(1)}$ is the same as the $\mathbb{P}_0$-distribution of $T_1^{(1)}$. This follows from the fact that the variables S_m, $m \in \mathbb{N}$, $m \geqslant 1$, are independent identically (Bernoulli) distributed random variables. Notice that $T_1^{(1)} = 1 + T_1^{(0)} \circ \vartheta_1$, $\mathbb{P}_0$-almost surely. Here $T_1^{(0)} = \inf\{n \geqslant 0 :\ X_n = 1\}$. Again we use the Markov property to obtain:

$$
\begin{aligned}
G_{0,1}(s) &= \mathbb{E}_0\left[e^{sT_1^{(1)}}\right] = \mathbb{E}_0\left[s^{1+T_1^{(0)}\circ\vartheta_1}\right] \\
&= s\mathbb{E}_0\left[\mathbb{E}_0\left[s^{T_1^{(1)}\circ\vartheta_1} \mid \mathcal{G}_1\right]\right] = s\mathbb{E}_0\left[\mathbb{E}_{X_1}\left[s^{T_1^{(0)}}\right]\right] \\
&= s\mathbb{E}_0\left[\mathbb{E}_{X_1}\left[s^{T_1^{(0)}}\right],\, X_1 = 1]\right] + s\mathbb{E}_0\left[\mathbb{E}_{X_1}\left[s^{T_1^{(0)}}\right],\, X_1 = -1\right] \\
&= s\mathbb{E}_0\left[\mathbb{E}_1\left[s^{T_1^{(0)}}\right],\, X_1 = 1\right] + s\mathbb{E}_0\left[\mathbb{E}_{-1}\left[s^{T_1^{(0)}}\right],\, X_1 = -1\right]
\end{aligned}
$$

(notice that $T_1^{(0)} = 0$ $\mathbb{P}_1$-almost surely, and $T_1^{(0)} = T_1^{(1)}$ $\mathbb{P}_{-1}$-almost surely)

$$
\begin{aligned}
&= sp + sq\mathbb{E}_{-1}\left[s^{T_1^{(1)}}\right] = sp + sqG_{-1,1}(s) \\
&= sp + sqG_{0,2}(s) = sp + sqG_{0,1}(s)^2. \qquad (2.75)
\end{aligned}
$$

In the final step of (2.75) we employed (2.74) with $j = -1$ and $k = 1$. From (2.75) we infer

$$
G_{0,1}(s) = \frac{1-\left(1-4pqs^2\right)^{1/2}}{2qs}. \qquad (2.76)
$$

By a similar token (i.e. by interchanging p and q) we also get

$$
G_{1,0}(s) = \frac{1-\left(1-4pqs^2\right)^{1/2}}{2ps}. \qquad (2.77)
$$

Next we rewrite $G_{0,0}(s)$:

$$
G_{0,0}(s) = \mathbb{E}_0\left[s^{T_0^{(1)}}\right] = \mathbb{E}_0\left[s^{1+T_0^{(0)}\circ\vartheta_1}\right] = s\mathbb{E}_0\left[\mathbb{E}_0\left[s^{T_0^{(0)}\circ\vartheta_1} \mid \mathcal{G}_1\right]\right]
$$

$$
= s\mathbb{E}_0\left[\mathbb{E}_{X_1}\left[s^{T_0^{(1)}}\right]\right]
$$

(on the event $\{X_1 = \pm 1\}$ the equality $T_0^{(0)} = T_0^{(1)}$ holds $\mathbb{P}_{X_1}$-almost surely)

$$
= s\mathbb{E}_0\left[\mathbb{E}_{X_1}\left[s^{T_0^{(1)}}\right],\, X_1 = 1\right] + s\mathbb{E}_0\left[\mathbb{E}_{X_1}\left[s^{T_0^{(1)}}\right],\, X_1 = -1\right]
$$

$$
\begin{aligned}
&= s\mathbb{E}_0\left[\mathbb{E}_1\left[s^{T_0^{(1)}}\right],\, X_1 = 1\right] + s\mathbb{E}_0\left[\mathbb{E}_{-1}\left[s^{T_0^{(1)}}\right],\, X_1 = -1\right] \\
&= sp\mathbb{E}_1\left[s^{T_0^{(1)}}\right] + sq\mathbb{E}_{-1}\left[s^{T_0^{(1)}}\right]
\end{aligned}
$$

((space) translation invariance)

$$
= sp\mathbb{E}_1\left[s^{T_0^{(1)}}\right] + sq\mathbb{E}_0\left[s^{T_1^{(1)}}\right] = spG_{1,0}(s) + sqG_{0,1}(s)
$$

(employ the equalities in (2.76) and (2.77))

$$
\begin{aligned}
&= sp\frac{1-\left(1-4pqs^2\right)^{1/2}}{2ps} + sq\frac{1-\left(1-4pqs^2\right)^{1/2}}{2qs} \\
&= 1 - \left(1-4pqs^2\right)^{1/2}.
\end{aligned}
$$

Then we infer

$$
\begin{aligned}
\mathbb{P}_0\left[T_0^{(1)} < \infty\right] &= \lim_{s\uparrow 1,\, s<1} G_{0,0}(s) = 1 - (1-4pq)^{1/2} \\
&= 1 - |1-2p| = 1 - |q-p|\,.
\end{aligned}
$$

As a consequence we see that the non-symmetric random walk, i.e. the one with $q \neq p$, is transient, and that the symmetric random walk (i.e. $p = q = \frac{1}{2}$) is recurrent. However, since

$$
\mathbb{E}_0\left[T_0^{(1)}\right] = \lim_{s\uparrow 1, s<1} G_{0,0}'(s) = \lim_{s\uparrow 1, s<1} \frac{s}{\left(1-s^2\right)^{3/2}} = \infty,
$$

it follows that the symmetric random walk is not positive recurrent.

In the following lemma we prove some of the relevant equalities concerning stopping times and one-dimensional random walks.

2.18. LEMMA. *Employing the above notation and hypotheses yields the following equalities:*

(i) *The equality* $T_j^{(1)} = 1 + T_j^{(0)} \circ \vartheta_1$ *holds* $\mathbb{P}_k$*-almost surely for all states* k, $j \in \mathbb{Z}$.
(ii) *For* $k > j$, $k, j \in \mathbb{Z}$ *the equality* $T_k^{(1)} = T_{k-1}^{(1)} + T_k^{(1)} \circ \vartheta_{T_{k-1}^{(1)}}$ *holds* $\mathbb{P}_j$*-almost surely.*

PROOF. First let us prove assertion (i). Let j and k be states in $\mathbb{Z}$. Then $\mathbb{P}_k$-almost surely we have

$$\begin{aligned} &1 + T_j^{(0)} \circ \vartheta_1 = 1 + \inf\{n \geqslant 0 : X_n \circ \vartheta_1 = j\} \\ &= \inf\{n + 1 : n \geqslant 0, X_{n+1} = j\} = \inf\{n \geqslant 1\, X_n = j\} = T_j^{(1)}. \end{aligned} \tag{2.78}$$

The equality in (2.78) shows assertion (i). Next we prove the somewhat more difficult equality in (ii). As remarked above we have

$$\mathbb{P}_j\left[T_k^{(1)} > T_{k-1}^{(1)}\right] = \mathbb{P}_j\left[T_{k-1}^{(1)} < \infty\right].$$

Indeed, in order to visit the state $k > j$ the process X_n, starting from j has to visit the state $k-1$, and hence $\mathbb{P}_j\left[T_{k-1}^{(1)} < T_k^{(1)}\right] = \mathbb{P}_j\left[T_{k-1}^{(1)} < \infty\right]$. Without loss of generality we may and shall assume that in the following arguments we consider the process X_n on the event $\left\{T_{k-1}^{(1)} < \infty\right\}$. (Otherwise we would automatically have $T_{k-1}^{(1)} + T_k^{(1)} \circ \vartheta_{T_{k-1}^{(1)}} = T_k^{(1)}$.) Next on the event $\left\{T_{k-1}^{(1)} < \infty\right\}$ we write:

$$\begin{aligned} T_{k-1}^{(1)} + T_k^{(1)} \circ \vartheta_{T_{k-1}^{(1)}} &= T_{k-1}^{(1)} + \inf\left\{n \geqslant 1 : X_n \circ \vartheta_{T_{k-1}^{(1)}} = k\right\} \\ &= \inf\left\{T_{k-1}^{(1)} + n : n \geqslant 1, X_{n+T_{k-1}^{(1)}} = k\right\} \\ &= \inf\left\{n > T_{k-1}^{(1)} : X_n = k\right\} = T_k^{(1)}. \end{aligned} \tag{2.79}$$

Assertion (ii) follows from (2.79). Altogether this proves Lemma 2.18. □

Perhaps it is useful to prove in an explicit manner that the one-dimensional random walk is a Markov chain. This is the content of the next lemma.

2.19. LEMMA. *Let* $\{S_n : n \in \mathbb{N}, n \geqslant 1\}$ *be independent identically distributed random variables taking their values in* $\mathbb{Z}$. *Put* $X_0 = S_0$ *be the initial value of the process* X_n, $n \in \mathbb{N}$, *where* $X_n = \sum_{m=0}^{n} S_m$. *Put*

$$\mathbb{E}_j\left[f\left(X_0, X_1, \ldots, X_n\right)\right] = \mathbb{E}\left[f\left(X_0 + j, X_1 + j, \ldots, X_n + j\right)\right] \tag{2.80}$$

for all bounded functions $f : \mathbb{Z}^{n+1} \to \mathbb{R}$, $j \in \mathbb{Z}$, $n \in \mathbb{N}$. *Then the equality in (2.80) expresses the fact that the process*

$$\left\{\left(\Omega, \mathcal{G}, \mathbb{P}_j\right)_{j\in\mathbb{Z}}, \left(X_n, n \in \mathbb{N}\right), \left(\vartheta_n, n \in \mathbb{N}\right), \mathbb{Z}\right\}$$

is a space homogeneous process, with the property that the distribution of the process $(X_{n+1} - X_n)_{n\in\mathbb{N}}$ *does not depend on the initial value* j. *It is also a time-homogeneous Markov chain.*

Proof. The first equality say that the finite dimensional distributions of the process $(X_n)_{n\in\mathbb{N}}$ are homogeneous in space. If

$$f(x_0, x_1, \dots, x_n) = g(x_1 - x_0, x_2 - x_1, \dots, x_n - x_{n-1}),$$

then from (2.80) we see that

$$\begin{aligned}\mathbb{E}_j\left[f(X_0, X_1, \dots, X_n)\right] &= \mathbb{E}\left[f(X_0 + j, X_1 + j, \dots, X_n + j)\right]\\ &= \mathbb{E}\left[g(X_1 - X_0, X_2 - X_1, \dots, X_n - X_{n-1})\right]\\ &= \mathbb{E}_j\left[g(X_1 - X_0, X_2 - X_1, \dots, X_n - X_{n-1})\right]. \qquad (2.81)\end{aligned}$$

The equality in (2.81) shows that the distribution of the process

$$(X_{n+1} - X_n)_{n\in\mathbb{N}}$$

does not depend on the initial value j. Next we prove the Markov property. Let f be any bounded function on $\mathbb{Z}$. To this end we consider:

$$\begin{aligned}&\mathbb{E}_j\left[f(X_{n+1}) \mid \mathcal{G}_n\right] = \mathbb{E}\left[f(X_{n+1} + j) \mid \mathcal{G}_n\right]\\ &= \mathbb{E}\left[f(X_{n+1} - X_n + X_n + j) \mid \mathcal{G}_n\right]\end{aligned}$$

(the variable $X_{n+1} - X_n$ and the σ-field $\mathcal{G}_n$ are $\mathbb{P}$-independent)

$$= \omega \mapsto \mathbb{E}\left[\omega' \mapsto f(X_{n+1}(\omega') - X_n(\omega') + X_n(\omega) + j)\right]$$

(the $\mathbb{P}$-distribution of $X_{n+1}(\omega') - X_n(\omega')$ neither depends on n nor on the initial value $X_0(\omega')$)

$$= \omega \mapsto \mathbb{E}\left[\omega' \mapsto f(X_1(\omega') - X_0(\omega') + X_n(\omega) + j)\right]$$

(choose $X_0(\omega') = j$)

$$\begin{aligned}&= \omega \mapsto \mathbb{E}\left[\omega' \mapsto f(X_1(\omega') - j + X_n(\omega) + j)\right]\\ &= \omega \mapsto \mathbb{E}\left[\omega' \mapsto f(X_1(\omega') + X_n(\omega))\right]\\ &= \omega \mapsto \mathbb{E}_{X_n(\omega)}\left[\omega' \mapsto f(X_1(\omega'))\right] = \mathbb{E}_{X_n}\left[f(X_1)\right]. \qquad (2.82)\end{aligned}$$

The equality in (2.82) proves Lemma 2.19. □

2.20. Remark. It would have been sufficient to take f of the form $f = \mathbf{1}_k$, $k \in \mathbb{Z}$.

2.21. Remark. Lemma 2.19 proves more than just the Markov property for a random walk. It only uses the fact that the increments S_n are identically $\mathbb{P}$-distributed and independent, and that the process $(X_n)_{n\in\mathbb{N}}$ possesses the same $\mathbb{P}_j$-distribution as the $\mathbb{P}$-distribution of the process $(X_n + j)_{n\in\mathbb{N}}$, $j \in \mathbb{Z}$. The proof only simplifies a little bit if one uses the random walk properties in an explicit manner.

1.1.2. *Some remarks.* From a historic point of view the references [**6, 47, 151**] are quite relevant. The references [**10, 43**] are relatively speaking good accessible. The reference [**153**] gives a detailed treatment of martingale theory. Citations, like [**54, 140, 142, 145**] establish a precise relationship between Feller operators, Markov processes, and solutions to the martingale problem. The references [**49, 50**] establish a relationship between hedging strategies (in mathematical finance) and (backward) stochastic differential equations.

2. Some additional comments on Markov processes

In this section we will discuss some topics related to Markov chains and Markov processes. We consider a quadruple

$$\{(\Omega, \mathcal{F}, \mathbb{P}), (X_n, n \in \mathbb{N}), (\vartheta_n, n \in \mathbb{N}), (S, \mathcal{S})\}. \tag{2.83}$$

In (2.83) the triple $(\Omega, \mathcal{F}, \mathbb{P})$ stands for a probability space, Ω is called the "sample space", $\mathcal{F}$ is a σ-field on Ω, and $\mathbb{P}$ is a probability measure on $\mathcal{F}$. The σ-field $\mathcal{F}$ is called the σ-field of events. The symbol $(S, \mathcal{S})$ stands for the state space of our process $(X_n : n \in \mathbb{N})$. In the present situation the state space S is discrete and countable with the discrete σ-field $\mathcal{S}$.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X_j : \Omega \to S$, $j \in \mathbb{N} = \{0, 1, 2, \ldots\}$, be *state variables*. It is assumed that the σ-field $\mathcal{F}$ is generated by the state variables X_j, $j \in \mathbb{N}$. Let $\vartheta_k : \Omega \to \Omega$, $k \in \mathbb{N}$, be *time shift* operators, which are also called *time translation operators*: $X_j \circ \vartheta_k = X_{j+k}$, j, $k \in \mathbb{N}$. For a bounded $\sigma(X_j : j \in \mathbb{N})$-measurable stochastic variable $F : \Omega \to \mathbb{R}$ and $x \in S$ we write

$$\mathbb{E}_x[F] = \mathbb{E}\left[F \mid X_0 = x\right] = \frac{\mathbb{E}[F, X_0 = x]}{\mathbb{P}[X_0 = x]}.$$

We also write

$$T_{x,y} = \frac{\mathbb{P}[X_1 = y, X_0 = x]}{\mathbb{P}[X_0 = x]} = \mathbb{P}_x[X_1 = y] = \mathbb{P}\left[X_1 = y \mid X_0 = x\right].$$

Here x has the interpretation of state at time 0, and y is the state at time 1. Let $\mathcal{G}_n$, $n \in \mathbb{N}$, be the internal memory up to the moment n. Hence $\mathcal{G}_n = \sigma(X_j : 0 \leqslant j \leqslant n)$.

2.22. Theorem. *Suppose that $(X_n : n \in \mathbb{N})$ is a stochastic process with values in a discrete countable state space S with the discrete σ-field $\mathcal{S}$. The state variables X_n, $n \in \mathbb{N}$, are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Write, as above, $T_{x,y} = \mathbb{P}\left[X_1 = y \mid X_0 = x\right]$, x, $y \in S$. Then the following assertions are equivalent:*

(1) *For all finite sequences of states $(s_0, \ldots, s_{n+1})$ in S the following identity holds:*

$$\mathbb{P}\left(X_{n+1} = s_{n+1} \mid X_0 = s_0, \ldots, X_n = s_n\right) = T_{s_n, s_{n+1}}; \tag{2.84}$$

(2) *For all bounded functions $f : S \to \mathbb{R}$ and for all times $n \in \mathbb{N}$ the following equality holds:*

$$\mathbb{E}\left[f(X_{n+1}) \mid \mathcal{G}_n\right] = \mathbb{E}_{X_n}[f(X_1)] \ \mathbb{P}\text{-almost surely}; \tag{2.85}$$

(3) *For all bounded functions $f_0, \ldots, f_k$ op S and for all times $n \in \mathbb{N}$ the following equality holds $\mathbb{P}$-almost surely:*

$$\mathbb{E}\left[f_0(X_n) f_1(X_{n+1}) \ldots f_k(X_{n+k}) \mid \mathcal{G}_n\right] = \mathbb{E}_{X_n}[f_0(X_0) f_1(X_1) \ldots f_k(X_k)]; \tag{2.86}$$

(4) *For all bounded measurable functions $F : (\Omega, \mathcal{F}) \to \mathbb{R}$ (stochastic variables) and for all $n \in \mathbb{N}$ the following identity holds:*

$$\mathbb{E}\left[F \circ \vartheta_n \mid \mathcal{G}_n\right] = \mathbb{E}_{X_n}[F] \ \mathbb{P}\text{-almost surely}; \tag{2.87}$$

(5) *For all bounded functions* $f : S \to \mathbb{R}$ *and for all* $(\mathcal{G}_n)_{n\in\mathbb{N}}$*-stopping times* $\tau : \Omega \to [0, \infty]$ *the following equality holds:*

$$\mathbb{E}\left[f(X_{\tau+1}) \mid \mathcal{G}_\tau\right] = \mathbb{E}_{X_\tau}\left[f(X_1)\right] \quad \mathbb{P}\text{-almost surely on the event} \quad \{\tau < \infty\}\,; \tag{2.88}$$

(6) *For all bounded measurable functions* $F : (\Omega, \mathcal{F}) \to \mathbb{R}$ *(random variables) and for all stopping times* τ *the following identity holds:*

$$\mathbb{E}\left[F \circ \vartheta_\tau \mid \mathcal{G}_\tau\right] = \mathbb{E}_{X_\tau}\left[F\right] \quad \mathbb{P}\text{-almost surely on the event} \quad \{\tau < \infty\}\,. \tag{2.89}$$

Before we prove Theorem 2.22 we make some remarks and give some explanation.

2.23. Remark. Let $\Omega = S^{\mathbb{N}}$, equipped with the product σ-field, and let $X_j : \Omega \to S$ be defined by $X_j(\omega) = \omega_j$ where $\omega = (\omega_0, \ldots, w_j, \ldots)$ belongs to Ω. If $\vartheta_k : \Omega \to \Omega$ is defined by $\vartheta_k(\omega_0, \ldots, \omega_j, \ldots) = (\omega_k, \ldots, \omega_{j+k}, \ldots)$, then it follows that $X_j \circ \vartheta_k = X_{j+k}$.

2.24. Remark. Instead of one probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we often consider a family of probability spaces $(\Omega, \mathcal{F}, \mathbb{P}_x)_{x\in S}$. The probabilities $\mathbb{P}_x$, $x \in S$, are determined by

$$\mathbb{E}_x\left[F\right] = \mathbb{E}\left[F \mid X_0 = x\right] = \frac{\mathbb{E}\left[F,\, X_0 = x\right]}{\mathbb{P}\left[X_0 = x\right]}, \quad x \in S. \tag{2.90}$$

Here $F : \Omega \to \mathbb{R}$ is $\mathcal{F}$-$\mathcal{B}_{\mathbb{R}}$-measurable, and hence by definition it is a random or stochastic variable. Since $\mathbb{P}_x[A] = \mathbb{E}_x[\mathbf{1}_A]$, $A \in \mathcal{F}$, and $\mathbb{P}_x[\Omega] = \mathbb{E}_x[\mathbf{1}] = 1$, the measure $\mathbb{P}_x$ is a probability measure on $\mathcal{F}$.

2.25. Remark. Let $F : \Omega \to \mathbb{R}$ be a bounded stochastic variable. The variable $\mathbb{E}_{X_n}[F]$ is a stochastic variable which is measurable with respect to the σ-field $\sigma(X_n)$, i.e. the σ-field generated by X_n. In fact we have

$$\begin{aligned}\mathbb{E}_{X_n(\omega)}\left[F\right] &= \mathbb{E}\left[F \mid X_0 = X_n(\omega)\right] = \frac{\mathbb{E}\left[F,\, X_0 = X_n(\omega)\right]}{\mathbb{P}\left[X_0 = X_n(\omega)\right]} \\ &= \mathbb{E}\left[\omega' \mapsto F(\omega') \times \mathbf{1}_{\{X_0 = X_n(\omega)\}}(\omega')\right].\end{aligned} \tag{2.91}$$

If we fix $\omega \in \Omega$, then in (2.91) everything is determined, and there should be no ambiguity any more.

2.26. Remark. Fix $n \in \mathbb{N}$. The σ-field $\mathcal{G}_n$ is generated by events of the form

$$\{(X_0, X_1, \ldots, X_n) = (s_0, s_1, \ldots, s_n)\}\,.$$

Here $(s_0, s_1, \ldots, s_n)$ varies over S^{n+1}. It follows that

$$\begin{aligned}\mathcal{G}_n &= \sigma\left\{\{(X_0, X_1, \ldots, X_n) = (s_0, s_1, \ldots, x_n)\} : (s_0, s_1, \ldots, s_n) \in S^{n+1}\right\} \\ &= \left\{\{(X_0, X_1, \ldots, X_n) \in B\} : B \in \mathcal{S}^{\otimes(n+1)}\right\} \\ &= \sigma\left(X_0, X_1, \ldots, X_n\right).\end{aligned} \tag{2.92}$$

The σ-field in (2.92) is the smallest σ-field rendering all state variables X_j, $0 \leqslant j \leqslant n$, measurable. It is noticed that $\mathcal{G}_n \subset \mathcal{F}$.

2.27. Remark. Next we discuss conditional expectations. Again let $F : \Omega \to \mathbb{R}$ be a bounded stochastic variable. If we write $Z = \mathbb{E}\left[F \mid \mathcal{G}_n\right]$, then we mean the following:

(1) The stochastic variable Z is $\mathcal{G}_n$-$\mathcal{B}_{\mathbb{R}}$-measurable. This a qualitative aspect of the notion of conditional expectation.
(2) The stochastic variable Z possesses the property that $\mathbb{E}[F, A] = \mathbb{E}[Z, A]$ for all events $A \in \mathcal{G}_n$. This is the quantitative aspect of the notion of conditional expectation.

Notice that the property in (2) is equivalent to the following one: the stochastic variable Z satisfies the equality $\mathbb{E}\left[F \mid A\right] = \mathbb{E}\left[Z \mid A\right]$ for all events $A \in \mathcal{G}_n$.

2.28. Remark. Let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. The mapping $F \mapsto \mathbb{E}\left[F \mid \mathcal{G}\right]$ is an orthogonal projection from $L^2(\Omega, \mathcal{F}, \mathbb{P})$ onto $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Let $F \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, and put $Z = \mathbb{E}\left[F \mid \mathcal{G}\right]$. In fact we have to verify the following conditions:

(1) $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$;
(2) If $G \in L^2(\Omega, \mathcal{G}, \mathbb{P})$, then the following inequality is satisfied:

$$\mathbb{E}\left[|F - Z|^2\right] \leqslant \mathbb{E}\left[|F - G|^2\right].$$

This claim is left as an exercise for the reader. For more details on conditional expectations see Section 1 in Chapter 1.

2.29. Remark. Next we will give an explicit formula for the conditional expectation in the setting of a enumerable discrete state space S. Let $\mathcal{G}_n = \sigma(X_0, X_1, \ldots, X_n)$ where $X_j : \Omega \to S$, $0 \leqslant j \leqslant n$, are state variables with a discrete countable state space S. In addition, let $F : \Omega \to \mathbb{R}$ be a bounded stochastic variable. Then we have

$$\mathbb{E}\left[F \mid \mathcal{G}_n\right] = \sum_{i_0, \ldots, i_n \in S} \mathbb{E}\left[F \mid X_0 = i_0, \ldots, X_n = i_n\right] \mathbf{1}_{\{X_0 = i_0\}} \cdots \mathbf{1}_{\{X_n = i_n\}}. \tag{2.93}$$

Writing the conditional expectation in (2.93) in an explicit manner as a function of ω yields

$$\begin{aligned} &\mathbb{E}\left[F \mid \mathcal{G}_n\right](\omega) \\ &= \sum_{i_0, \ldots, i_n \in S} \mathbb{E}\left[F \mid X_0 = i_0, \ldots, X_n = i_n\right] \mathbf{1}_{\{X_0 = i_0\}}(\omega) \cdots \mathbf{1}_{\{X_n = i_n\}}(\omega). \end{aligned} \tag{2.94}$$

From (2.93) and also (2.94) it is clear that the conditional expectation

$$\mathbb{E}\left[F \mid \mathcal{G}_n\right] \quad \text{is } \mathcal{G}_n\text{-measurable.}$$

2.30. Remark. Put $\mathcal{G} = \mathcal{G}_\infty = \sigma(X_0, \ldots, X_n, \ldots) = \sigma\left(\overline{X}\right)$ where $\overline{X} : \Omega \to S^{\mathbb{N}}$ is the variable defined by $\overline{X}(\omega) = (X_0(\omega), \ldots, X_n(\omega), \ldots)$, $\omega \in \Omega$. Then

$$\begin{aligned} &\mathcal{G} = \mathcal{G}_\infty \\ &= \left\{\left\{\overline{X} \in B\right\} : B \text{ is measurable with respect to the product } \sigma\text{-field on } S^{\mathbb{N}}\right\}. \end{aligned} \tag{2.95}$$

2.31. Remark. Let $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ be a random variable. This random variable is called a stopping time relative the filtration $(\mathcal{G}_n)_{n \in \mathbb{N}}$, or, more briefly,

τ is called a $(\mathcal{G}_n)_{n\in\mathbb{N}}$-stopping time, provided that for every $k \in \mathbb{N}$ an event of the form $\{\tau \leqslant k\}$ is $\mathcal{G}_k$-measurable. The latter property is equivalent to the following one. For every $k \in \mathbb{N}$ the event $\{\tau = k\}$ is $\mathcal{G}_k$-measurable. Note that $\{\tau = k\} = \{\tau \leqslant k\} \backslash \{\tau \leqslant k-1\}$, $k \in \mathbb{N}$, $k \geqslant 1$, and $\{\tau \leqslant k\} = \cup_{j=0}^{k} \{\tau = j\}$. From these equalities it follows that τ is a $(\mathcal{G}_n)_{n\in\mathbb{N}}$-stopping time if and only if for every $k \in \mathbb{N}$ the event $\{\tau = k\}$ is $\mathcal{G}_k$-measurable.

2.32. Remark. Let B be a subset of S. Important examples of stopping times are

$$\begin{aligned} \tau_B^0 &= \inf\{k \geqslant 0 : X_k \in B\} \quad \text{on} \quad \cup_{k=0}^{\infty}\{X_k \in B\} \quad \text{and } \infty \text{ elsewhere;} \\ \tau_B^1 &= \inf\{k \geqslant 1 : X_k \in B\} \quad \text{on} \quad \cup_{k=1}^{\infty}\{X_k \in B\} \quad \text{and } \infty \text{ elsewhere.} \end{aligned} \tag{2.96}$$

Similarly we also write $\tau_B^s = \inf\{k \geqslant s : X_k \in B\}$ on the event $\cup_{k=s}^{\infty}\{X_k \in B\}$, and $\tau_B^s = \infty$ elsewhere. The time τ_B^0 is called the first income time, and τ_B^1 is called the first hitting time, or the first income time after 0.

We also notice that $\tau_B^1 = \min\{k \geqslant 1 : X_k \in B\}$ on $\cup_{k=1}^{\infty}\{X_k \in B\}$ and $\tau_B^1 = \infty$ on $\cap_{\ell=1}^{\infty}\{X_\ell \in S \backslash B\}$. In addition: $1 + \tau_B^0 \circ \vartheta_1 = \vartheta_B^1$.

2.33. Remark. Again let $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ be a $(\mathcal{G}_n)_{n\in\mathbb{N}}$-stopping time. The σ-field $\mathcal{G}_\tau$ containing the information from the past of τ is defined by

$$\begin{aligned} \mathcal{G}_\tau &= \cap_{k\in\mathbb{N}}\{A \in \mathcal{F} : A \cap \{\tau \leqslant k\} \in \mathcal{G}_k\} \\ &= \sigma(X_{j\wedge\tau} : j \in \mathbb{N}) \end{aligned} \tag{2.97}$$

where $X_{j\wedge\tau}(\omega) = X_{j\wedge\tau(\omega)}(\omega)$, $\omega \in \Omega$.

2.34. Remark. Let F be a stochastic variable. What is meant by $F \circ \vartheta_k$ and $F \circ \vartheta_\tau$ on the event $\{\tau < \infty\}$? Here $k \in \mathbb{N}$, and τ is a $(\mathcal{G}_n)_{n\in\mathbb{N}}$-stopping time. For $F = \prod_{j=0}^n f_j(X_j)$ we write:

$$F \circ \vartheta_k = \prod_{j=0}^n f_j(X_j) \circ \vartheta_k = \prod_{j=0}^n f_j(X_{j+k}),$$

and on the event $\{\tau < \infty\}$

$$F \circ \vartheta_\tau = \prod_{j=0}^n f_j(X_j) \circ \vartheta_\tau = \prod_{j=0}^n f_j(X_{j+\tau}). \tag{2.98}$$

2.35. Proposition. *Let* $\left(T_{x,y}^{(n)}\right)_{(x,y)\in S\times S}$ *be a sequence of square matrices with positive entries, possibly with infinite countably many entries (when S is countable, not finite). Put*

$$T_{x,y}^0 = I,\ T_{x,y}^1 = T_{x,y}^{(1)},\quad T_{x,y}^n = \sum_{s_1,\dots,s_n\in S}\prod_{j=1}^{n+1} T_{s_{j-1},s_j},\quad s_0 = x,\quad s_{n+1} = y. \tag{2.99}$$

The equalities in (2.99) are to be considered as matrix multiplications. Fix $1 \leqslant n_1 < \cdots < n_k \leqslant n$, and let the measure space $\left(S^k, \otimes_{j=1}^k \mathcal{S}, \mu^{0,x}_{n_1,\dots,n_k,n+1,y}\right)$, $(x,y) \in S\times S$, $n \in \mathbb{N}$ be determined by the equalities:

$$\begin{aligned}&\int_{S^k}\prod_{j=1}^n f_j(s_j)\, d\mu^{0,s_0}_{n_1,\dots,n_k,n+1,s_{n+1}}(s_1,\dots,s_k)\\ &= \sum_{(s_1,\dots,s_k)\in S^k}\prod_{j=1}^{k+1} T^{(n_j-n_{j-1})}_{s_{j-1},s_j} f_j(s_j),\quad f_j \in L^\infty(S,\mathcal{S}),\end{aligned} \tag{2.100}$$

where $(s_0, s_{n+1}) = (x,y) \in S\times S$, $n_0 = 0$, and $n_{k+1} = n+1$. Then for every $1 \leqslant j_0 \leqslant n$, and $f_j \in L^\infty(E,\mathcal{E})$, $1 \leqslant j \leqslant n$, this family satisfies the following equality:

$$\begin{aligned}&\int_{S^n}\prod_{j=1}^n f_j(s_j)\, d\mu^{0,s_0}_{1,\dots,n,n+1,s_{n+1}}(s_1,\dots,s_n) \qquad (2.101)\\ &= \int_{S^{n-1}}\prod_{j=1,\,j\neq j_0}^n f_j(s_j)\, d\mu^{0,s_0}_{n,s_{n+1}}(s_1,\dots,s_{j_0-1},s_{j_0+1},\dots,s_n)\, T^2_{s_{j_0-1},s_{j_0+1}}\end{aligned}$$

where $T^2_{x,y} = \sum_{z\in S} T^{(1)}_{x,z}T^{(1)}_{z,y}$ for all $(x,y) \in S\times S$ (matrix multiplication). Let $1 \leqslant n_1 < \cdots < n_k \leqslant n$, and put

$$\mu^{0,x}_{n_1,\dots,n_k,n+1}(B_0\times B) = \sum_{y\in S}\mathbf{1}_{B_0}(x)\mu^{0,x}_{n_1,\dots,n_k,n+1,y}(B),\quad B_0 \in \mathcal{S},\ B \in \otimes^k\mathcal{S}. \tag{2.102}$$

Suppose that the matrices $\left(T_{x,y}^{(n)}\right)_{(x,y)\in S\times S}$, $n \in \mathbb{N}$, are stochastic. Then the measures in (2.102) do not depend on $n+1$. Moreover, the following assertions are equivalent:

(a) *The family of measure spaces*

$$\left\{\left(S^{k+1}, \otimes^{k+1}\mathcal{S}, \mu^{0,x}_{n_1,\dots,n_k,n+1}\right) : 1 \leqslant n_1 < \cdots < n_k \leqslant n,\, n \in \mathbb{N}\right\} \tag{2.103}$$

is a consistent family of probability measure spaces.

(b) *The family of measure spaces defined in (2.100) is consistent.*

(c) *For every $n \in \mathbb{N}$ and $(x,y) \in S \times S$ the equality $T^{(n)}_{x,y} = T^n_{x,y}$ holds.*

Suppose that the family in (2.103) is a consistent family of probability spaces. Then the corresponding process

$$\left\{(\Omega, \mathcal{F}, \mathbb{P}_x)_{x\in S}, (X_n : n \in \mathbb{N}), (\vartheta_n, n \in \mathbb{N}), (S, \mathcal{S})\right\}. \tag{2.104}$$

is a Markov chain if and only if for $(x,y) \in S \times S$ and n, $m \in \mathbb{N}$ the following matrix multiplication equality holds:

$$T^{(n+m)}_{x,y} = \sum_{z\in S}^{+} T^{(n)}_{x,z} T^{(m)}_{z,y} = \sum_{z\in S} T^n_{x,z} T^m_{z,y}. \tag{2.105}$$

This means that

$$\mathbb{P}_x\left[X_0 \in B_0, \dots, X_n \in B_n\right] = \mathbf{1}_{B_0}(x)\mu^{0,x}_{1,\dots,n,n+1}\left(B_1 \times \cdots \times B_n\right),$$

for $B_j \in \mathcal{S}$, $0 \leqslant j \leqslant n$, and that the family in (2.104) possesses the Markov property if and only (2.105) holds.

In addition, we have $T^{(n)}_{x,y} = \mathbb{P}_x\left[X_n = y\right]$, x, $y \in S$; i.e. the quantities $T^{(n)}_{x,y}$ represent the n time step transition probabilities from the state x to the state y.

2.36. Theorem. *Let the notation be as in Theorem 2.22. The following assertions are equivalent:*

(1) *For every $s \in S$, for every bounded function $f : S \to \mathbb{R}$, and for all $n \in \mathbb{N}$ the following equality holds $\mathbb{P}_s$-almost surely:*

$$\mathbb{E}_s\left[f\left(X_{n+1}\right) \mid \mathcal{G}_n\right] = \mathbb{E}_{X_n}\left[f\left(X_1\right)\right]. \tag{2.106}$$

(2) *For every bounded function $f : S \to \mathbb{R}$, and for all $n \in \mathbb{N}$ the following equality holds $\mathbb{P}$-almost surely:*

$$\mathbb{E}\left[f\left(X_{n+1}\right) \mid \mathcal{G}_n\right] = \mathbb{E}_{X_n}\left[f\left(X_1\right)\right]. \tag{2.107}$$

2.37. Remark. From the proof it follows that in Theorem 2.36 we may replace the stochastic variable $f(X_1)$ by any bounded stochastic variable $Y : \Omega \to \mathbb{R}$. At the same $f(X_{n+1}) = f(X_1) \circ \vartheta_n$ has to be replaced by $Y \circ \vartheta_n$.

2.38. Remark. Theorem 2.36 together with Remark 2.37 shows that throughout in Theorem 2.22 we may replace the probability $\mathbb{P}$ with $\mathbb{P}_s$ for any $s \in S$. Consequently, we could have defined a time-homogeneous Markov chain as a quadruple

$$\left\{(\Omega, \mathcal{F}, \mathbb{P}_s)_{s\in S}, (X_n : n \in \mathbb{N}), (\vartheta_k, k \in \mathbb{N}), (S, \mathcal{S})\right\}$$

satisfying the equivalent conditions in Theorem 2.22 with $\mathbb{P}_s$, for all $s \in E$, instead of $\mathbb{P}$.

Proof of Theorem 2.36. (1) $\Longrightarrow$ (2) Let $s \in S$, $f : S \to \mathbb{R}$ a bounded function, and $n \in \mathbb{N}$. From (2.106) we infer

$$\mathbb{E}_s \left[f\left(X_{n+1}\right), A \right] = \mathbb{E}_s \left[\mathbb{E}_{X_n} \left[f\left(X_1\right) \right], A \right] \quad \text{for all } A \in \mathcal{G}_n. \tag{2.108}$$

From (2.108) we infer

$$\begin{aligned} &\mathbb{E}\left[f\left(X_{n+1}\right), A, X_0 = s \right] \\ &= \mathbb{E}\left[\mathbb{E}_{X_n} \left[f\left(X_{n+1}\right) \right], A, X_0 = s \right] \quad \text{for all } A \in \mathcal{G}_n, \text{ and } s \in S. \end{aligned} \tag{2.109}$$

By summing over $s \in S$ in (2.109) we obtain

$$\mathbb{E}\left[f\left(X_{n+1}\right), A \right] = \mathbb{E}\left[\mathbb{E}_{X_n} \left[f\left(X_{n+1}\right) \right], A \right] \quad \text{for all } A \in \mathcal{G}_n. \tag{2.110}$$

From (2.110) the equality in (2.107) easily follows.

(2) $\Longrightarrow$ (1) Let $f : S \to \mathbb{R}$ and $n \in \mathbb{N}$ be such that (2.107) holds. Then, since all events of the form $\{X_0 = s\}$, $s \in S$, belong to $\mathcal{G}_n$, (2.107) implies that (2.109) holds for f and hence by dividing by $\mathbb{P}\left[X_0 = s\right]$, for $s \in S$, we obtain (2.108). Hence (2.106) follows.

All this completes the proof of Theorem 2.36. $\square$

The following theorem is similar to the formulation of Theorem 2.36 but now with stopping times and having remark 2.37 taken into account:

2.39. THEOREM. *Let the notation be as in Theorem 2.22, and let $\tau : \Omega \to [0, \infty]$ be a $(\mathcal{G}_n)_{n\in\mathbb{N}}$-stopping time. The following assertions are equivalent:*

(1) *For every $s \in S$, and for every bounded stochastic variable $Y : \Omega \to \mathbb{R}$ the following equality holds $\mathbb{P}_s$-almost surely on the event $\{\tau < \infty\}$:*
$$\mathbb{E}_s\left[Y \circ \vartheta_\tau \mid \mathcal{G}_\tau\right] = \mathbb{E}_{X_\tau}\left[Y\right]. \tag{2.111}$$
(2) *For every bounded stochastic variable $Y : \Omega \to \mathbb{R}$ the following equality holds $\mathbb{P}$-almost surely on the event $\{\tau < \infty\}$:*
$$\mathbb{E}\left[Y \circ \vartheta_\tau \mid \mathcal{G}_\tau\right] = \mathbb{E}_{X_\tau}\left[Y\right]. \tag{2.112}$$

2.40. REMARK. Let $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ be a $(\mathcal{G}_n)_{n\in\mathbb{N}}$-stopping time, and let $A \in \mathcal{G}_\tau$. Let $m \in \mathbb{N} \cup \{\infty\}$. Put $\tau_m = m\mathbf{1}_{\Omega\backslash A} + \tau\mathbf{1}_A$. Then τ_m is a $(\mathcal{G}_n)_{n\in\mathbb{N}}$-stopping time. If $\mathbb{P}[A] < 1$ and $m = \infty$, then $\tau_m = \infty$ on the event $\Omega\backslash A$ which is non-negligible.

2.41. DEFINITION. Let Ω be a set and let $\mathcal{S}$ be a collection of subsets of Ω. Then $\mathcal{S}$ is called a *Dynkin system*, if it has the following properties:

(a) $\Omega \in \mathcal{S}$;
(b) if A and B belong to $\mathcal{S}$ and if $A \supseteq B$, then $A\backslash B$ belongs to $\mathcal{S}$;
(c) if $(A_n : n \in \mathbb{N})$ is an increasing sequence of elements of $\mathcal{S}$, then the union $\bigcup_{n=1}^{\infty} A_n$ belongs to $\mathcal{S}$.

In the literature Dynkin systems are also called λ-systems: see *e.g.* [**3**]. A π-system is a collection of subsets which is closed under finite intersections. A Dynkin system which is also a π-system is a σ-field. The following result on Dynkin systems, known as the π-λ theorem, gives a stronger result.

2.42. THEOREM. *Let $\mathcal{M}$ be a collection of subsets of Ω, which is stable under finite intersections, so that $\mathcal{M}$ is a π-system on Ω. The Dynkin system generated by $\mathcal{M}$ coincides with the σ-field generated by $\mathcal{M}$.*

PROOF. Let $\mathcal{D}(\mathcal{M})$ be the smallest Dynkin-system containing $\mathcal{M}$, *i.e.* $\mathcal{D}(\mathcal{M})$ is the Dynkin-system generated by $\mathcal{M}$. For all $A \in \mathcal{D}(\mathcal{M})$, we define:
$$\Gamma(A) := \{B \in \mathcal{D}(\mathcal{M}) : A \cap B \in \mathcal{D}(\mathcal{M})\}.$$
then we have

(1) if A belongs to $\mathcal{M}$, $\mathcal{M} \subset \Gamma(A)$,
(2) for all $A \in \mathcal{M}$, $\Gamma(A)$ is a Dynkin system on Ω.
(3) if A belongs to $\mathcal{M}$, then $\mathcal{D}(\mathcal{M}) \subset \Gamma(A)$,
(4) if B belongs to $\mathcal{D}(\mathcal{M})$, then $\mathcal{M} \subset \Gamma(B)$,
(5) for all $B \in \mathcal{D}(\mathcal{M})$ the inclusion, $\mathcal{D}(\mathcal{M}) \subset \Gamma(B)$ holds.

It follows that $\mathcal{D}(\mathcal{M})$ is also a π-system. It is esay to see that a Dynkin system which is at the same time a π-system is in fact a σ-field (or σ-algebra). This completes the proof of Theorem 2.42. □

2.43. Theorem. *Let Ω be a set and let $\mathcal{M}$ be a collection of subsets of Ω, which is stable (or closed) under finite intersections. Let $\mathcal{H}$ be a vector space of real valued functions on Ω satisfying:*

(i) *The constant function $\mathbf{1}$ belongs to $\mathcal{H}$ and $\mathbf{1}_A$ belongs to $\mathcal{H}$ for all $A \in \mathcal{M}$;*
(ii) *if $(f_n : n \in \mathbb{N})$ is an increasing sequence of non-negative functions in $\mathcal{H}$ such that $f = \sup_{n\in\mathbb{N}} f_n$ is finite (bounded), then f belongs to $\mathcal{H}$.*

Then $\mathcal{H}$ contains all real valued functions (bounded) functions on Ω, that are $\sigma(\mathcal{M})$ measurable.

Proof. Put $\mathcal{D} = \{A \subseteq \Omega : 1_A \in \mathcal{H}\}$. Then by (i) Ω belongs to $\mathcal{D}$ and $\mathcal{D} \supseteq \mathcal{M}$. If A and B are in $\mathcal{D}$ and if $B \supseteq A$, then $B\backslash A$ belongs to $\mathcal{D}$. If $(A_n : n \in \mathbb{N})$ is an increasing sequence in $\mathcal{D}$, then $\mathbf{1}_{\cup A_n} = \sup_n \mathbf{1}_{A_n}$ belongs to $\mathcal{D}$ by (ii). Hence $\mathcal{D}$ is a Dynkin system, that contains $\mathcal{M}$. Since $\mathcal{M}$ is closed under finite intersection, it follows by Theorem 2.42 that $\mathcal{D} \supseteq \sigma(\mathcal{M})$. If $f \geqslant 0$ is measurable with respect to $\sigma(\mathcal{M})$, then $f = \sup\limits_n 2^{-n} \sum_{j=1}^{n2^n} 1_{\{f \geqslant j2^{-n}\}}$. Since the subsets $\{f \geqslant j2^{-n}\}$, j, $n \in \mathbb{N}$, belong to $\sigma(\mathcal{M})$, we see that f is a member of $\mathcal{H}$. Here we employed the fact that $\sigma(\mathcal{M}) \subseteq \mathcal{D}$. If f is $\sigma(\mathcal{M})$-measurable, then we write f as a difference of two non-negative $\sigma(\mathcal{M})$-measurable functions. $\square$

The previous theorems (Theorem 2.42 and Theorem 2.43) are used in the following form. Let Ω be a set and let $(S_i, \mathcal{S}_i)_{i\in I}$ be a family of measurable spaces, indexed by an arbitrary set I. For each $i \in I$, let $\mathcal{M}_i$ denote a collection of subsets of S_i, closed under finite intersection, which generates the σ-field $\mathcal{S}_i$, and let $f_i : \Omega \to S_i$ be a map from Ω to S_i. In this context the following two propositions follow.

2.44. Proposition. *Let $\mathcal{M}$ be the collection of all sets of the form $\bigcap_{i\in J} f_i^{-1}(A_i)$, $A_i \in \mathcal{M}_i$, $i \in J$, $J \subseteq I$, J finite. Then $\mathcal{M}$ is a collection of subsets of Ω which is stable under finite intersection and $\sigma(\mathcal{M}) = \sigma(f_i : i \in I)$.*

2.45. Proposition. *Let $\mathcal{H}$ be a vector space of real-valued functions on Ω such that:*

(i) *the constant function $\mathbf{1}$ belongs to $\mathcal{H}$;*
(ii) *if $(h_n : n \in \mathbb{N})$ is an increasing sequence of non-negative functions in $\mathcal{H}$ such that $h = \sup_n h_n$ is finite (bounded), then h belongs to $\mathcal{H}$;*
(iii) *$\mathcal{H}$ contains all products of the form $\prod_{i\in J} 1_{A_i} \circ f_i$, $J \subseteq I$, J finite, and $A_i \in \mathcal{M}_i$, $i \in J$.*

Under these assumptions $\mathcal{H}$ contains all real-valued functions (bounded) functions in $\sigma(f_i : i \in I)$.

The theorems 2.42 and 2.43, and the propositions 2.44 and 2.45 are called the *monotone class theorem*.

In the propositions 2.44 and 2.45 we may take $S_i = S$, $\mathcal{M}_i$ the collection of finite subsets of S_i, and $f_i = X_i$, $i \in I = \mathbb{N}$.

3. More on Brownian motion

Further on in this book on stochastic processes we will discuss Brownian motion in more detail. In fact we will consider Brownian motion as a Gaussian process, as a Markov process, and as a martingale (which includes a discussion on Itô calculus). In addition Brownian motion can be viewed as weak limit of a scaled symmetric random walk. For this result we need a Functional Central Limit Theorem (FCLT) which is a generalization of the classical central limit theorem.

2.46. Theorem (Multivariate Classical Central Limit Theorem). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $\{Z_n : n \in \mathbb{N}\}$ be a sequence of $\mathbb{P}$-independent and $\mathbb{P}$-identically distributed random variables with values in $\mathbb{R}^d$ in $L^1\left(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d\right)$. Let $\mu = \mathbb{E}\left[Z_1\right]$, and let $\mathbb{D}$ be the dispersion matrix of Z_1 (i.e. the variance-covariance of the random vector Z_1). Then there exists a centered Gaussian (or multivariate normal) random vector X with dispersion matrix $\mathbb{D}$ such that the sequence*

$$X_n := \frac{Z_1 + \cdots + Z_n - n\mu}{\sqrt{n}}$$

converges weakly (or in distribution) to a centered random vector X with dispersion matrix $\mathbb{D}$ as $n \to \infty$. The latter means that $\lim_{n\to\infty} \mathbb{E}\left[f\left(Z_n\right)\right] = \mathbb{E}\left[f\left(Z\right)\right]$ for all bounded continuous functions $f : \mathbb{R}^d \to \mathbb{R}$.

Notice that by a non-trivial density argument we only need to prove the equality

$$\lim_{n\to\infty} \mathbb{E}\left[f\left(\frac{Z_1 + \cdots + Z_n - n\mu}{\sqrt{n}}\right)\right] = \mathbb{E}\left[f\left(X\right)\right]$$

for all functions f of the form $f(x) = e^{-i\langle x, \xi\rangle}$, $x \in \mathbb{R}^d$, $\xi \in \mathbb{R}^d$.

Next let us give a (formal) definition of Brownian motion.

2.47. Definition. A one-dimensional Brownian motion with drift μ and diffusion coefficient σ^2 is a stochastic process $\{X(t) : t \geqslant 0\}$ with continuous sample paths having independent Gaussian increments with mean and variance of an increment $X(t+s) - X(t)$ given by $s\mu = \mathbb{E}\left[X(t+s) - X(t)\right]$ and $s\sigma^2 = \mathbb{E}\left[\left(X(t+s) - X(t)\right)^2\right]$, $s, t \geqslant 0$. If $X_0 = x$, then this Brownian is said to start at x. A Brownian motion with drift $\mu = 0$, and $\sigma^2 = 1$ is called a standard Brownian motion.

One of the problems is whether or not such a process exists. One way of resolving this problem is to put the Functional Central Limit Theorem at work. Let us prepare for this approach. Let $\{Z_j : j \in \mathbb{N}\}$ be a sequence of centered independent identically distributed real valued random variables in $L^1\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ with variance $\sigma^2 = \mathbb{E}\left[Z_1^2\right]$. For example these variables could be Bernoulli variables taking the values $+\sigma$ and $-\sigma$ with the same probability $\frac{1}{2}$. Put $S_0 = Z_0 = 0$,

$S_n = Z_1 + \cdots, +Z_n$, $n \in \mathbb{N}$, $n \geqslant 1$. Define for each scale parameter $n \geqslant 1$ the stochastic process $X^{(n)}(t)$ by

$$X^{(n)}(t) = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} = \frac{\sum_{k=0}^{\lfloor nt \rfloor} Z_k}{\sqrt{n}}, \quad t \geqslant 0. \tag{2.113}$$

Here $\lfloor nt \rfloor$ is the integer part of nt, i.e. the largest integer k for which $k \leqslant nt < k+1$. The it is relatively easy to see that

$$\mathbb{E}\left[X^{(n)}(t)\right] = 0, \quad \text{and} \quad \operatorname{Var}\left(X^{(n)}(t)\right) = \mathbb{E}\left[X^{(n)}(t)^2\right] = t\sigma^2. \tag{2.114}$$

Then the classical CLT (Central Limit Theorem) implies that there exists a process $\{X(t) : t \geqslant 0\}$ with the property that for every $m \in \mathbb{N}$, for every choice $(t_1, \ldots, t_m)$ of m positive real numbers, and every bounded continuous function $f : \mathbb{R}^m \to \mathbb{C}$ the following limit equality holds:

$$\lim_{n\to\infty} \mathbb{E}\left[f\left(X^{(n)}(t)\right)\right] = \mathbb{E}\left[f\left(X(t)\right)\right]. \tag{2.115}$$

The equality in (2.115) says the finite-dimensional distributions of the sequence of processes $\left\{X^{(n)}(t) : t \geqslant 0\right\}_{n\in\mathbb{N}}$ converges weakly to the finite-dimensional distributions of the process $\{X(t) : t \geqslant 0\}$. This limit should then be one-dimensional Brownian motion with drift zero and variance σ^2. A posteriori we know that Brownian motion should be $\mathbb{P}$-almost surely continuous. However the processes $\left\{X^{(n)}(t) : t \geqslant 0\right\}_{n\in\mathbb{N}}$ have jumps. It would be nice if we were able to replace these processes which have jumps by processes without jumps. Therefore we employ linear interpolation. This can be done as follows. We introduce the following interpolating sequence of continuous processes:

$$\widetilde{X}^{(n)}(t) = \frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} + \left(nt - \lfloor nt \rfloor\right) \frac{Z_{\lfloor nt \rfloor + 1}}{\sqrt{n}}, \quad t \geqslant 0. \tag{2.116}$$

Let m and n be positive integers. Then on the half open interval $\left[\frac{m}{n}, \frac{m+1}{n}\right)$ the variable $X^n(t)$ is constant in time t at level $\frac{S_m}{n}$, while $\widetilde{X}^n(t)$ changes linearly from

$$\frac{S_m}{n} \text{ at time } t = \frac{m}{n} \text{ to } \frac{S_{m+1}}{\sqrt{n}} = \frac{S_m}{\sqrt{n}} + \frac{Z_{m+1}}{\sqrt{n}} \text{ at time } t = \frac{m+1}{n}. \tag{2.117}$$

It can be proved that the sequence of stochastic processes $\left\{\widetilde{X}^{(n)}(t) : t \geqslant 0\right\}_{n\in\mathbb{N}}$ converges weakly to Brownian motion with drift μ and variance σ^2. This is the contents of the following FCLT (Functional Central Limit Theorem). The following result also goes under the name "Donsker's invariance principle": see, e.g., [**15**] or [**42**].

2.48. Theorem (Functional Central Limit Theorem). *Let* $\left\{\widetilde{X}^{(n)}(t) : t \in [0,T]\right\}$, $n \in \mathbb{N}$, *and* $\{X(t) : t \in [0,T]\}$ *be stochastic processes possessing sample paths which are* $\mathbb{P}$*-almost surely continuous with the property that the finite-dimensional distributions of the sequence* $\left\{\widetilde{X}^{(n)}(t) : t \in [0,T]\right\}_{n\in\mathbb{N}}$ *converge weakly to* *those of* $\{X(t) : t \in [0,T]\}$. *Then the sequence* $\left\{\widetilde{X}^{(n)}(t) : t \in [0,T]\right\}_{n\in\mathbb{N}}$ *converges weakly to* $\{X(t) : t \in [0,T]\}$ *if and only if for every* $\varepsilon > 0$ *the following equality holds:*

$$\lim_{\delta\downarrow 0} \sup_{n\in\mathbb{N}} \mathbb{P}\left[\sup_{0\leqslant s,t\leqslant T,\,|s-t|\leqslant\delta} \left|\widetilde{X}^{(n)}(s) - \widetilde{X}^{(n)}(s)\right| \geqslant \varepsilon\right] = 0. \tag{2.118}$$

This result is based on Prohorov's tightness theorem and the Arzela-Ascoli characterization of compact subsets of $C[0,T]$.

2.49. Theorem (Prohorov theorem). *Let* $(P_n : n \in \mathbb{N})$ *be a sequence of probability measures on a separable complete metrizable topological space* S *with Borel* σ*-field* $\mathcal{S}$. *Then the following assertions are equivalent:*

(i) *For every* $\varepsilon > 0$ *there exists a compact subset* K_ε *of* S *such that* $P_n[K_\varepsilon] \geqslant 1 - \varepsilon$ *for all* $n \in \mathbb{N}$.
(ii) *Every subsequence of* $(P_n : n \in \mathbb{N})$ *has a subsequence which converges weakly to a probability measure on* $(S, \mathcal{S})$.

A sequence $(P_n)_n$ satisfying (i) (or (ii)) in Theorem 2.49 is called a Prohorov set. Theorem 2.48 can be proved by applying Theorem 2.49 with P_n equal to the $\mathbb{P}$-distribution of the process $\left\{\widetilde{X}^{(n)}(t) : 0 \leqslant t \leqslant T\right\}$.

2.50. Theorem (Arzela-Ascoli). *Endow $C[0,T]$ with the topology of uniform convergence. A subset A of $C[0,T]$ has compact closure if and only if it has the following properties:*

(i) $\sup_{\omega\in A}|\omega(0)|<\infty$;
(ii) *The subset A is equi-continuous in the sense that*

$$\lim_{\delta\downarrow 0}\sup_{0\leqslant s,t\leqslant T,\,|s-t|\leqslant\delta}\sup_{\omega\in A}|\omega(s)-\omega(t)|=0.$$

From (i) and (ii) it follows that $\sup_{\omega\in A}\sup_{s\in[0,T]}|\omega(s)|<\infty$, and hence A is uniformly bounded. The result which is relevant here reads as follows. It is the same as Theorem T.8.4 in Bhattacharaya and Waymire [**15**].

2.51. Theorem. *Let $(P_n)_n$ be a sequence of probability measures on $C[0,T]$. Then $(P_n)_n$ is tight if and only if the following two conditions hold.*

(i) *For each $\eta>0$ there is a number B such that*
$$P_n\left[\omega\in C[0,T]:\,|\omega(0)|>B\right]<\eta,\quad n=1,2,\dots$$
(ii) *For each $\varepsilon>0$, $\eta>0$, there is a $0<\delta<1$ such that*
$$P_n\left[\omega\in C[0,T]:\sup_{0\leqslant s,t\leqslant T,\,|s-t|\leqslant\delta}|\omega(s)-\omega(t)|\geqslant\varepsilon\right]\leqslant\eta,\quad n=1,2,\dots$$

Proof. If the sequence $(P_n)_n$ is tight, then given $\eta>0$ there is a compact subset K of $C([0,T])$ such that $P_n(K)>1-\eta$ for all n. By the Arzela-Ascoli theorem (Theorem 2.50), if $B>\sup_{\omega\in K}|\omega(0)|$, then
$$P_n\left[\omega\in C[0,T]:\,|\omega(0)|\geqslant B\right]\leqslant P_n\left[K^c\right]\leqslant 1-(1-\eta)=\eta.$$
Also given $\varepsilon>0$ select $\delta>0$ such that $\sup_{\omega\in K}\sup_{0\leqslant s,t\leqslant T,\,|s-t|\leqslant\delta}|\omega(s)-\omega(t)|<\varepsilon$. Then
$$P_n\left[\omega\in C[0,T]:\sup_{0\leqslant s,t\leqslant T,\,|s-t|\leqslant\delta}|\omega(s)-\omega(t)|\geqslant\varepsilon\right]\leqslant P_n\left[K^c\right]<\eta\ \text{ for all } n\geqslant 1.$$
The converse goes as follows. Given $\eta>0$, first select B using (i) such that $P_n\left[\omega\in C([0,T]):\,|\omega(0)|\leqslant B\right]\geqslant 1-\frac{1}{2}\eta$, for $n\geqslant 1$. Select $\delta_r>0$ using (ii) such that
$$P_n\left[\omega\in C([0,T]):\sup_{0\leqslant s,t\leqslant T,\,|s-t|\leqslant\delta}|\omega(s)-\omega(t)|<\frac{1}{r}\right]\geqslant 1-2^{-(r+1)}\eta\quad\text{for } n\geqslant 1.$$
Now take K to be the uniform closure of
$$\bigcap_{r=1}^{\infty}\left\{\omega\in C([0,T]):\,|\omega(0)|\leqslant B,\ \sup_{0\leqslant s,t\leqslant T,\,|s-t|\leqslant\delta}|\omega(s)-\omega(t)|<\frac{1}{r}\right\}.$$
Then $P_n(K)>1-\eta$ for $n\geqslant 1$, and K is compact by the Arzela-Ascoli theorem. This completes the proof Theorem 2.51. $\square$

For convenience of the reader we formulate some limit theorems which are relevant in the main text of this book. The formulations are taken from Stirzaker [**126**]. For proofs the reader is also referred to Stirzaker. For convenience we also insert proofs which are based on Birkhoff's ergodic theorem. Define $S_n = \sum_{k=0}^{n-1} X_k$, where the variables $\{X_k\}_{k\in\mathbb{N}} \subset L^1(\Omega, \mathcal{F}, \mathbb{P})$ are independent and identically distributed (i.i.d.). Then we have the following three classic results.

2.52. THEOREM (Central limit theorem, standard version). *If* $\mathbb{E}[X_k] = \mu$ *and* $0 < \operatorname{var}(X_k) = \sigma^2 < \infty$, *then*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - n\mu}{(n\sigma^2)^{1/2}} \leqslant x\right] = \Phi(x), \quad x \in \mathbb{R},$$

where $\Phi(x)$ *is the standard normal distribution,* i.e. $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-\frac{1}{2}x^2}\, dx.$

PROOF. Let $f : \mathbb{R} \to \mathbb{C}$ be a bounded C^2-function with a bounded second derivative. Then by Taylor's formula (or by integration by parts) we have

$$f(y) = f(0) + yf'(0) + \frac{1}{2}y^2 f''(0) + \int_0^1 (1-s)\, y^2 \{f''(sy) - f''(0)\}\, ds. \tag{2.119}$$

Put $Y_{n,k} = \dfrac{X_k - \mu}{\sigma\sqrt{n}}$. Inserting $y = Y_{n,k}$ into (2.119) yields

$$f(Y_{n,k}) = f(0) + Y_{n,k} f'(0) + \frac{1}{2}Y_{n,k}^2 f''(0) + \int_0^1 (1-s)\, Y_{n,k}^2 \{f''(sY_{n,k}) - f''(0)\}\, ds. \tag{2.120}$$

Then we take expectations in (2.120) to obtain:

$$\mathbb{E}[f(Y_{n,k})] = f(0) + \frac{1}{2n} f''(0) + \int_0^1 (1-s)\, \mathbb{E}\left[Y_{n,k}^2 \{f''(sY_{n,k}) - f''(0)\}\right] ds. \tag{2.121}$$

Put

$$\varepsilon_n(t) = n \int_0^1 (1-s) \mathbb{E}\left[Y_{n,1}^2 \left(1 - e^{-istY_{n,1}}\right)\right] ds, \quad t \in \mathbb{R},$$

and choose $f(y) = e^{-ity}$. Observe that, uniformly in t on compact subsets of $\mathbb{R}$, $\lim_{n\to\infty} \varepsilon_n(t) = 0$. Then, since the variables $Y_{n,k}$, $1 \leqslant k \leqslant n$, are i.i.d., from (2.121) we get

$$\mathbb{E}\left[e^{-itY_{n,k}}\right] = 1 - \frac{t^2}{2n} + \frac{t^2 \varepsilon_n(t)}{n}. \tag{2.122}$$

From (2.122) we infer

$$\mathbb{E}\left[e^{-it\sum_{k=1}^n Y_{n,k}}\right] = \left(\mathbb{E}\left[e^{-itY_{n,1}}\right]\right)^n = \left(1 - \frac{t^2}{2n} + \frac{t^2\varepsilon_n(t)}{n}\right)^n. \tag{2.123}$$

Let $Y : \Omega \to \mathbb{R}$ be a standard normally distributed random variable. From the properties of the sequence $\{\varepsilon_n(t)\}_n$ and (2.123) we see that, for every $0 < R < \infty$,

$$\lim_{n\to\infty} \sup_{|t|\leqslant R} \left\{\mathbb{E}\left[e^{-it\sum_{k=1}^n Y_{n,k}} - e^{-itY}\right]\right\}$$

$$= \lim_{n\to\infty} \sup_{|t|\leqslant R} \left\{ \mathbb{E}\left[e^{-it\sum_{k=1}^{n} Y_{n,k}} \right] - e^{-\frac{1}{2}t^2} \right\}$$

$$= \lim_{n\to\infty} \sup_{|t|\leqslant R} \left\{ \mathbb{E}\left[e^{-it\sum_{k=1}^{n} Y_{n,k}} \right] - \left(1 - \frac{t^2}{2n} + \frac{t^2\varepsilon_n(t)}{n} \right)^n \right\} = 0. \tag{2.124}$$

The conclusion in Theorem 2.52 then follows from (2.124) together with Lévy's continuity theorem: see Theorem 5.42, and Theorem 5.43 assertions (9) and (10). □

2.53. Theorem (Weak law of large numbers). *If* $\mathbb{E}[X_k] = \mu < \infty$, *then for all* $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left[\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right] = 0.$$

For a proof of the following theorem see (the proof of) Theorem 5.60. It is proved as a consequence of the (pointwise) ergodic theorem of Birkhoff: see Theorems 5.59 and 5.66, and Corollary 5.67.

2.54. Theorem (Strong law of large numbers). *The equality*

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu, \quad \text{holds } \mathbb{P}\text{-almost surely} \tag{2.125}$$

for some finite constant μ, *if and only if* $\mathbb{E}\left[|X_k|\right] < \infty$, *and then* $\mu = \mathbb{E}\left[X_1\right]$. *Moreover, the limit in (2.125) also exists in* L^1*-sense.*

We will show that Theorem 2.53 is a consequence of Theorem 2.54.

Proof of Theorem 2.53. Let $\varepsilon > 0$ be arbitrary, and let $\{X_k\}_k$ and μ be as in Theorem 2.53. Then

$$\mathbb{P}\left[\left|\frac{S_n - n\mu}{n}\right| \geqslant \varepsilon\right] \leqslant \frac{1}{\varepsilon}\mathbb{E}\left[\left|\frac{S_n}{n} - \mu\right|\right]. \tag{2.126}$$

By the L^1-version of Theorem 2.54 it follows that the right-had side of (2.126) converges to 0. This shows that Theorem 2.53 is a consequence of Theorem 2.54. □

The central limit theorem is the principal reason for the appearance of the normal (or "bell-shaped") distribution in so many statistical and scientific contexts. The first version of this theorem was proved by Abraham de Moivre before 1733. The laws of large numbers supply a solid foundation for our faith in the usefulness and good behavior of averages. In particular, as we have remarked above, they support one of our most appealing interpretations of probability as long-term relative frequency. The first version of the weak law was proved by James Bernoulli around 1700; and the first form of the strong law by Emile Borel in 1909. We include proofs of these results in the form as stated. As noted above a proof of Theorem 2.54 will be based on Birkhoff's ergodic theorem.

2.55. Remark. The following papers and books give information about the central limit theorem in the context of Stein's method which stems from Stein [**124**]: see Barbour and Hall [**9**], Barbour and Chen [**8**], Chen, Goldstein and Shao [**31**], Nourdin and Peccati [**101**], Berckmoes et al [**13**]. This is a very interesting and elegant method to prove convergence and give estimates for partial sums of so-called standard triangular arrays (STA). It yields sharp estimates: see the forthcoming paper [**14**].

4. Gaussian vectors.

The following theorem gives a definition of a Gaussian (or a multivariate normally distributed) vector purely in terms of its characteristic function (Fourier transform of its distribution.

2.56. Theorem. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, and let* $X = (X_1, \dots, X_n)$ *be an* $\mathbb{R}^n$*-valued Gaussian vector in the sense that there exists a vector* $\mu := (\mu_1, \dots, \mu_n) \in \mathbb{R}^n$ *and a symmetric square matrix* $\sigma := (\sigma_{j,k})_{j,k=1}^n$ *such that the characteristic function of the vector* X *is given by*

$$\mathbb{E}\left[e^{-i\langle \xi, X\rangle}\right] = e^{-i\langle \xi,\mu\rangle - \frac{1}{2}\sum_{j,k=1}^n \xi_j\xi_k\sigma_{j,k}} \quad \text{for all } \xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n. \tag{2.127}$$

Then for every $1 \leqslant j \leqslant n$ the variable X_j belongs to $L^2(\Omega, \mathcal{F}, \mathbb{P})$, $\mu_j = \mathbb{E}[X_j]$, and

$$\sigma_{j,k} = \operatorname{cov}(X_j, X_k) = \mathbb{E}\left[(X_j - \mathbb{E}[X_j])(X_k - \mathbb{E}[X_k])\right]. \tag{2.128}$$

Proof. Put $Y = X - \mu$, and fix $\varepsilon > 0$. Then the equality in (2.127) is equivalent to

$$\mathbb{E}\left[e^{-i\langle \xi, Y\rangle}\right] = e^{-\frac{1}{2}\sum_{j,k=1}^n \xi_j \xi_k \sigma_{j,k}} \quad \text{for all } \xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n. \tag{2.129}$$

From Cauchy's theorem and the equality $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} e^{-\frac{1}{2}\eta^2} d\eta = 1$ we obtain

$$e^{-\frac{1}{2}\varepsilon|Y|^2} = \frac{1}{\left(\sqrt{2\pi\varepsilon}\right)^n}\int_{\mathbb{R}^n} e^{-i\langle \eta, Y\rangle} e^{-\frac{1}{2\varepsilon}|\eta|^2}\, d\eta = \frac{1}{\left(\sqrt{2\pi}\right)^n}\int_{\mathbb{R}^n} e^{-i\langle \sqrt{\varepsilon}\eta, Y\rangle} e^{-\frac{1}{2}|\eta|^2}\, d\eta. \tag{2.130}$$

From (2.129) and (2.130) we infer:

$$\mathbb{E}\left[e^{-i\langle \xi, Y\rangle} e^{-\frac{1}{2}\varepsilon|Y|^2}\right] = \frac{1}{\left(\sqrt{2\pi}\right)^n}\int_{\mathbb{R}^n} \mathbb{E}\left[e^{-i\langle \xi + \sqrt{\varepsilon}\eta, Y\rangle}\right] e^{-\frac{1}{2}|\eta|^2}\, d\eta$$

(employ (2.129) with $\xi + \sqrt{\varepsilon}\eta$ instead of ξ)

$$= \frac{1}{\left(\sqrt{2\pi}\right)^n}\int_{\mathbb{R}^n} e^{-\frac{1}{2}\sum_{j,k=1}^n (\xi_j + \sqrt{\varepsilon}\eta_j)(\xi_k + \sqrt{\varepsilon}\eta_k)\sigma_{j,k}} e^{-\frac{1}{2}|\eta|^2}\, d\eta. \tag{2.131}$$

Next we take $1 \leqslant \ell_1, \ell_2 \leqslant n$, and we differentiate the right-hand side and left-hand side of (2.131) with respect to ξ_{ℓ_2} and the result with respect ξ_{ℓ_1}. In addition we write a negative sign in front of this. Then we obtain:

$$\begin{aligned}
&\mathbb{E}\left[e^{-i\langle \xi, Y\rangle} Y_{\ell_1} Y_{\ell_2} e^{-\frac{1}{2}\varepsilon|Y|^2}\right] \\
&= \frac{1}{\left(\sqrt{2\pi}\right)^n}\int_{\mathbb{R}^n} e^{-\frac{1}{2}\sum_{j,k=1}^n (\xi_j + \sqrt{\varepsilon}\eta_j)(\xi_k + \sqrt{\varepsilon}\eta_k)\sigma_{j,k}} e^{-\frac{1}{2}|\eta|^2} \\
&\qquad \left(\frac{\sigma_{\ell_1,\ell_2} + \sigma_{\ell_2,\ell_1}}{2} - \left(\sum_{j=1}^n \left(\xi_j + \sqrt{\varepsilon}\eta_j\right)\frac{\sigma_{j,\ell_2} + \sigma_{\ell_1,j}}{2}\right)^2\right) d\eta.
\end{aligned} \tag{2.132}$$

Inserting $\xi = 0$ into (2.132) yields:

$$\begin{aligned}
\mathbb{E}\left[Y_{\ell_1} Y_{\ell_2} e^{-\frac{1}{2}\varepsilon|Y|^2}\right] &= \frac{1}{\left(\sqrt{2\pi}\right)^n}\int_{\mathbb{R}^n} e^{-\frac{1}{2}\varepsilon\sum_{j,\,k=1}^n \eta_j\eta_k\sigma_{j,k}} e^{-\frac{1}{2}|\eta|^2} \\
&\qquad \left(\frac{\sigma_{\ell_1,\ell_2} + \sigma_{\ell_2,\ell_1}}{2} - \varepsilon\left(\sum_{j=1}^n \eta_j \frac{\sigma_{j,\ell_2} + \sigma_{\ell_1,j}}{2}\right)^2\right) d\eta.
\end{aligned} \tag{2.133}$$

First assume that $\ell_1 = \ell_2 = \ell$. Then the left-hand side of (2.133) increases to $\mathbb{E}\left[Y_\ell^2\right]$, and the right-hand side increases to $\dfrac{1}{\left(\sqrt{2\pi}\right)^n}\displaystyle\int_{\mathbb{R}^n} e^{-\frac{1}{2}|\eta|^2}\, d\eta\, \sigma_{\ell,\ell} = \sigma_{\ell,\ell}$ if ε decreases to zero. Consequently, $Y_\ell \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}\left[Y_\ell^2\right] = \sigma_{\ell,\ell}$, $1 \leqslant \ell \leqslant n$.

It follows that Y_ℓ belongs to $L^1(\Omega, \mathcal{F}, \mathbb{P})$, and that we also have that $Y_{\ell_1} Y_{\ell_2} \in L^1(\Omega, \mathcal{F}, \mathbb{P})$. By applying the same procedure as above we also obtain that

$$\mathbb{E}\left[Y_{\ell_1} Y_{\ell_2}\right] = \frac{\sigma_{\ell_1,\ell_2} + \sigma_{\ell_2,\ell_1}}{2} = \sigma_{\ell_1,\ell_2}. \tag{2.134}$$

In (2.134) we employed the symmetry of the matrix $\left(\sigma_{\ell_1,\ell_2}\right)_{\ell_1,\ell_2=1}^n$. Again we fix $1 \leqslant \ell \leqslant n$, and we differentiate the equality in (2.131) with respect to ξ_ℓ to obtain

$$\begin{aligned} & i\mathbb{E}\left[e^{-i\langle \xi, Y\rangle} Y_\ell e^{-\frac{1}{2}\varepsilon|Y|^2}\right] \\ &= \frac{1}{\left(\sqrt{2\pi}\right)^n} \int_{\mathbb{R}^n} d\eta\, e^{-\frac{1}{2}\sum_{j,k=1}^n (\xi_j + \sqrt{\varepsilon}\eta_j)} \left(\xi_k + \sqrt{\varepsilon}\eta_k\right) e^{-\frac{1}{2}|\eta|^2} \\ &\quad \sum_{j=1}^n \left(\xi_j + \sqrt{\varepsilon}\eta_j\right) \frac{\sigma_{j,\ell} + \sigma_{\ell,j}}{2}. \end{aligned} \tag{2.135}$$

In (2.135) we set $\xi = 0$, and we let $\varepsilon \downarrow 0$ to obtain $\mathbb{E}\left[Y_\ell\right] = 0$, and hence $X_\ell \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}\left[X_\ell\right] = \mu_\ell$. This completes the proof of Theorem 2.56. ☐

5. Radon-Nikodym Theorem

We begin by formulating a convenient version of Radon-Nikodym's theorem. For a proof the reader is referred to Bauer [**10**] or Stroock [**130**].

2.57. Theorem (Radon-Nikodym theorem). *Let $(\Omega, \mathcal{F}, \mu)$ be a σ-finite measure space, and let ν be a finite measure on $\mathcal{F}$. Suppose that ν is absolute continuous relative to μ, i.e. $\mu(B) = 0$ implies $\nu(B) = 0$. Then there exists a function $f \in L^1(\Omega, \mathcal{F}, \mu)$ such that $\nu(B) = \int_B f\, d\mu$ for all $B \in \mathcal{F}$. In particular the function f is $\mathcal{F}$-measurable.*

The following corollary follows from Theorem 2.57 by taking $\mathcal{F} = \mathcal{B}$, μ the measure $\mathbb{P}$ confined to $\mathcal{B}$, and $\nu(B) = \mathbb{E}\left[X, B\right]$, $B \in \mathcal{B}$.

2.58. Corollary. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and let $\mathcal{B}$ be a subfield (i.e. a sub-σ-field) of $\mathcal{A}$. Let X be a stochastic variable in $L^1(\Omega, \mathcal{A}, \mathbb{P})$. Then there exists a $\mathcal{B}$-measurable variable Z on Ω with the following properties:*

(1) *(qualitative property) the variable Z is $\mathcal{B}$-measurable;*
(2) *(quantitative property) for every $B \in \mathcal{B}$ the equality $\mathbb{E}\left[Z, B\right] = \mathbb{E}\left[X, B\right]$ holds.*

The variable Z is called the conditional expectation of X, and is denoted by $Z = \mathbb{E}\left[X \mid \mathcal{B}\right]$. The existence is guaranteed by the Radon-Nikodym theorem.

6. Some martingales

Let E be a locally compact Hausdorff space which is second countable, and let

$$\left\{\left(\Omega, \mathcal{F}, \mathbb{P}_x\right), \left(X(t), t \geqslant 0\right), \left(\vartheta_t, t \geqslant 0\right), \left(E, \mathcal{E}\right)\right\} \tag{2.136}$$

be a time-homogeneous strong Markov process with right-continuous paths, which also have left limits in the state space E on their life time. Put $S(t)f(x) = \mathbb{E}_x\left[f\left(X(t)\right)\right]$, $f \in C_0(E)$, and assume that $S(t)f \in C_0(E)$ whenever $f \in C_0(E)$. Here a real or complex valued function f belongs to $C_0(E)$ provided that it is continuous and that for every $\varepsilon > 0$ the subset $\{x \in E : |f(x)| \geqslant \varepsilon\}$ is compact in E. Let the operator L be the generator of this process. This means that its domain $D(L)$ consists of those functions $f \in C_0(E)$ for which the limit $Lf = \lim_{t\downarrow 0} \dfrac{S(t)f - f}{t}$ exists in $C_0(E)$ equipped with the supremum norm, i.e. $\|f\|_\infty = \sup_{x\in E} |f(x)|$, $f \in C_0(E)$. The Markov property of the process in (2.136) together with the right continuity of paths implies that the family $\{S(t) : t \geqslant 0\}$ is a Feller, or, more properly, a Feller-Dynkin semigroup.

(1) The semigroup property can be expressed as follows:

$$S\left(t_1 + t_2\right) = S\left(t_1\right) S\left(t_2\right), \quad t_1,\, t_2 \geqslant 0, \quad S(0) = I.$$

(2) Moreover, the right-continuity of paths implies

$$\lim_{t\downarrow 0} S(t)f(x) = \lim_{t\downarrow 0} \mathbb{E}_x\left[f\left(X(t)\right)\right] = \mathbb{E}_x\left[f\left(X(0)\right)\right] = f(x), \quad f \in C_0(E).$$

(3) In addition, if $0 \leqslant f \leqslant \mathbf{1}$, then $0 \leqslant S(t)f \leqslant \mathbf{1}$.

A semigroup with the properties (1), (2) and (3) is a called a Feller, or Feller-Dynkin semigroup. In fact, it can be proved that a Feller-Dynkin semigroup $\{S(t) : t \geqslant 0\}$ satisfies

$$\lim_{s\to t,\, s>0} \|S(s)f - S(t)f\|_\infty = 0, \quad t \geqslant 0, \quad f \in C_0(E).$$

Let $\{S(t) : t \geqslant 0\}$ be a Feller-Dynkin semigroup. Then it can be shown that there a exists a Markov process, as in (2.136) with right-continuous paths such that $S(t)f(x) = \mathbb{E}_x\left[f\left(X(t)\right)\right]$, $f \in C_0(E)$, $t \geqslant 0$. For details, see Blumenthal and Getoor [**20**]. Similar results are true for states spaces which are Polish; see, e.g., [**146**].

Let $t \mapsto M(t)$, $t \geqslant 0$, be an adapted right-continuous multiplicative process, i.e. $M(0) = 1$ and $M(s)M(t) \circ \vartheta_s = M(s + t)$, $s,\, t \geqslant 0$. Put $S_M(t)f(x) = \mathbb{E}_x\left[M(t)f\left(X(t)\right)\right]$, $f \in C_0(E)$, $t \geqslant 0$. Assume that the operators $S_M(t)$ leave the space $C_0(E)$ invariant, so that $S_M(t)f$ belongs to $C_0(E)$ whenever $f \in C_0(E)$. Then the family $\{S_M(t) : t \geqslant 0\}$ has the semigroup property $S_M(s+t) = S_M(s)S_M(t)$, $s,\, t \geqslant 0$, and $\lim_{t\downarrow 0} S_M(t)f(x) = f(x)$, $t \geqslant 0$, $f \in C_0(E)$. If, in addition, for every $f \in C_0(E)$ there exists a $\delta > 0$ such that $\sup_{0\leqslant t\leqslant \delta} \|S_M(t)f\|_\infty < \infty$, then

$$\lim_{t\downarrow 0} \|S_M(t)f - f\|_\infty = 0, \quad f \in C_0(E). \tag{2.137}$$

Moreover, there a exists a closed densely defined linear operator L_M such that

$$L_M f = C_0(E)\text{-}\lim_{t\downarrow 0} \frac{S_M(t)f - f}{t} \tag{2.138}$$

for $f \in D\left(L_M\right)$, the domain of L_M. If $M(t) = \mathbf{1}$, then $L_M = L$.

2.59. Proposition. *The following processes are $\mathbb{P}_x$-martingales:*

$$t \mapsto M(t) f\left(X(t)\right) - M(0) f(X(0)) - \int_0^t M(s) L_M f\left(X(s)\, ds\right),$$

$$t \geqslant 0, \ \ f \in D\left(L_M\right), \tag{2.139}$$

$$s \mapsto M(s) \mathbb{E}_{X(s)}\left[M(t-s) f\left(X(t-s)\right)\right], \ \ 0 \leqslant s \leqslant t, \ \ f \in C_0(E), \tag{2.140}$$

$$s \mapsto M(s) \mathbb{E}_{X(s)}\left[M(t-s-u) p\left(u, X(t-s-u), y\right)\right], \ \ 0 \leqslant s \leqslant t-u. \tag{2.141}$$

In (2.141) it is assumed that there exists a "reference" measure m on the Borel field $\mathcal{E}$ together with an density function $p(t,x,y)$, $(t,x,y) \in (0,\infty) \times E \times E$ such that $\mathbb{E}_x\left[f\left(X(t)\right)\right] = \int p\left(t,x,y\right) f(y)\, dm(y)$ for all $f \in C_0(E)$ and for all $x \in E$ and all $t > 0$. From the semigroup property it follows that $p(s+t,x,y) = \int p(s,x,z) p(t,z,y) dm(z)$ for m-almost all $y \in E$. Assuming that $m(O) > 0$ for all non-empty open subsets of E, and that the function $(t,x,y) \mapsto p(t,x,y)$ is continuous on $(0,\infty) \times E \times E$, it follows that the equality $p(s+t,x,y) = \int p(s,x,z) p(t,z,y) dm(z)$ holds for s, $t > 0$ and for all x, $y \in E$.

The following corollary is the same as Proposition 2.59 with $M = \mathbf{1}$.

2.60. Corollary. *The following processes are $\mathbb{P}_x$-martingales:*

$$t \mapsto f(X(t)) - f(X(0)) - \int_0^t Lf(X(s)\,ds), \quad t \geqslant 0, \quad f \in D(L), \tag{2.142}$$

$$s \mapsto \mathbb{E}_{X(s)}[f(X(t-s))], \; 0 \leqslant s \leqslant t, \; f \in C_0(E), \tag{2.143}$$

$$s \mapsto p(t-s, X(s), y), \; 0 \leqslant s < t. \tag{2.144}$$

Like in Proposition 2.59 in (2.144) it is assumed that there exists a "reference" strictly positive Borel measure m such that for a (unique) continuous density function $p(t,x,y)$ the identity $\mathbb{E}_x[f(X(t))] = \int p(t,x,y)\,f(y)\,dm(y)$ holds for all $f \in C_0(E)$ and for all $x \in E$ and all $t > 0$.

2.61. Lemma. *Let the continuous density be as in Proposition 2.59, and let $z \in E$. Then the following equality holds for all $0 \leqslant s < t$ and for all $y \in E$:*

$$\mathbb{E}_z[p(t-s, X(s), y)] = p(t,x,y). \tag{2.145}$$

Proof of Lemma 2.61. Let the notation be as in Lemma 2.61. Then by the identity of Chapman-Kolmogorov we have

$$\mathbb{E}_z[p(t-s, X(s), y)] = \int p(s,z,w)\,p(t-s,w,y)dm(w) = p(t,x,y). \tag{2.146}$$

The equality in (2.146) is the same as the one in (2.145), which completes the proof of Lemma 2.61. $\square$

Proof of Proposition 2.59. First let f belong to the domain of L_M, and let $t_2 > t_1 \geqslant 0$. Then we have

$$\begin{aligned}
&\mathbb{E}_x\left[M(t_2) f(X(t_2)) - M(0)f(X(0)) - \int_0^{t_2} M(s)L_M f(X(s))\,ds \;\Big|\; \mathcal{F}_{t_1}\right] \\
&\quad - M(t_1) f(X(t_2)) + M(0)f(X(0)) + \int_0^{t_1} M(s)L_M f(X(s))\,ds \\
&= \mathbb{E}_x\Bigg[M(t_1)\Bigg(M(t_2-t_1) f(X(t_2-t_1)) - M(0)f(X(0)) \\
&\quad - \int_0^{t_2-t_1} M(s)L_M f(X(s))\,ds\Bigg) \circ \vartheta_{t_1} \;\Big|\; \mathcal{F}_{t_1}\Bigg]
\end{aligned}$$

(Markov property)

$$\begin{aligned}
&= M(t_1)\,\mathbb{E}_{X(t_1)}\Bigg[M(t_2-t_1) f(X(t_2-t_1)) - M(0)f(X(0)) \\
&\qquad - \int_0^{t_2-t_1} M(s)L_M f(X(s))\,ds\Bigg]
\end{aligned}$$

(definition of the operator $S_M(t)$; put $z = X(t_1)$, and $t = t_2 - t_1$)

$$= M(t_1)\left(S_M(t) f(z) - \mathbb{E}_z[M(0)f(X(0))] - \int_0^t S_M(s)L_M f(z)\,ds\right)$$

$$= M(t_1)\left(S_M(t)\,f(z) - \mathbb{E}_z\left[M(0)f(X(0))\right] - \int_0^t \frac{\partial}{\partial s} S_M(s) f(z)\,ds\right)$$
$$= M(t_1)\left(S_M(t)\,f(z) - \mathbb{E}_z\left[M(0)f(X(0))\right] - S_M(t)f(z) + S_M(0)f(z)\right) = 0. \tag{2.147}$$

The equalities in (2.147) show the equality in (2.139).

Let $f \in C_0(E)$ and $t > 0$. In order to show that the process
$$s \mapsto M(s)\mathbb{E}_{X(s)}\left[M(t-s)f(X(t-s))\right]$$
is a $\mathbb{P}_x$-martingale we proceed as follows:
$$M(s)\mathbb{E}_{X(s)}\left[M(t-s)f(X(t-s))\right]$$

(Markov property)
$$= M(s)\mathbb{E}_x\left[M(t-s)\circ\vartheta_s f(X(t-s))\circ\vartheta_s \,\middle|\, \mathcal{F}_s\right]$$
$$= \mathbb{E}_x\left[M(s)M(t-s)\circ\vartheta_s f(X(t)) \,\middle|\, \mathcal{F}_s\right] = \mathbb{E}_x\left[M(t)f(X(t)) \,\middle|\, \mathcal{F}_s\right]. \tag{2.148}$$

It is clear that the process in (2.148) is a martingale. This proves that the process in (2.140) is a martingale. A similar argument shows the equality:
$$M(s)\mathbb{E}_{X(s)}\left[M(t-s-u)p(u, X(t-s-u), y)\right]$$
$$= \mathbb{E}_x\left[M(t-u)p(u, X(t-u), y) \,\middle|\, \mathcal{F}_s\right],\ 0 \leqslant s < t-u. \tag{2.149}$$

Again it is clear that the process in (2.149) as a function of s is a $\mathbb{P}_x$-martingale.

Altogether this proves Proposition 2.59. □

Proof of Corollary 2.60. The fact that the processes in (2.142) and (2.143) are $\mathbb{P}_x$-martingales is an immediate consequence of (2.139) and (2.140) respectively by inserting $M(\rho) = 1$ for all $0 \leqslant \rho \leqslant t$. If $M(\rho) = 1$ for all $0 < \rho < t$, then by Lemma 2.61 we get
$$M(s)\mathbb{E}_{X(s)}\left[M(t-s-u)p(u, X(t-s-u), y)\right]$$
$$= \mathbb{E}_{X(s)}\left[p(u, X(t-s-u), y)\right] = p(t-s, X(s), y). \tag{2.150}$$

On the other hand by the Markov property we also have:
$$\mathbb{E}_{X(s)}\left[p(u, X(t-s-u), y)\right] = \mathbb{E}_x\left[p(u, X(t-u), y) \,\middle|\, \mathcal{F}_s\right]. \tag{2.151}$$

As a consequence of (2.150) and (2.151) we see that the process in (2.144) is a martingale. This completes the proof of Corollary 2.60. □

2.62. Remark. In general the process $s \mapsto p(t-s, X(s), y)$, $0 \leqslant s < t$, is not a closed martingale. In many concrete examples we have $\lim\limits_{s\uparrow t, s<t} p(t-s, X(s), y) = 0$, $\mathbb{P}_x$-almost surely, on the one hand, and $\mathbb{E}_x\left[p(t-s, X(s), y)\right] = p(t, x, y) > 0$ on the other. For an example of this situation take d-dimensional Brownian motion. By Scheffé's theorem it follows that the $\mathbb{P}_x$-martingale $s \mapsto p(t-s, X(s), y)$, $0 \leqslant s < t$, can not be a closed martingale. If it were, then there would exist an $\mathcal{F}_t$-measurable variable $F(t) = \lim_{s\uparrow, s<t} p(t-s, X(s), y)$ with the property that $p(t-s, X(s), y) = \mathbb{E}_x\left[F(t) \,\middle|\, \mathcal{F}_s\right]$. Since $F(t) = 0$, $\mathbb{P}_x$-almost surely, this is a contradiction.

In the following corollary we consider a special multiplicative process: $M(s) = \mathbf{1}_{\{T>s\}}$, where T is a terminal stopping time, i.e. $T = s + T \circ \vartheta_s$, $\mathbb{P}_x$-almost surely, on the event $\{T > s\}$ for all $s > 0$ and for all $x \in E$.

2.63. Corollary. *The following processes are $\mathbb{P}_x$-martingales:*

$$t \mapsto \mathbf{1}_{\{T>t\}} f\left(X(t)\right) - \mathbf{1}_{\{T>0\}} f(X(0)) - \int_0^{t\wedge T} L_M f\left(X(s)\, ds\right),$$

$$t \geqslant 0, \quad f \in D\left(L_M\right), \tag{2.152}$$

$$s \mapsto \mathbf{1}_{\{T>s\}} \mathbb{E}_{X(s)}\left[f\left(X(t-s)\right), T > t-s\right], \; 0 \leqslant s \leqslant t, \; f \in C_0(E), \tag{2.153}$$

$$s \mapsto \mathbf{1}_{\{T>s\}} \left(p\left(t-s, X(s), y\right) - \mathbb{E}_{X(s)}\left[p\left(t-s-T, X(T), y\right), T < t-s\right]\right),$$

$$0 \leqslant s < t. \tag{2.154}$$

In (2.141) it is assumed that there exists a "reference" measure m on the Borel field $\mathcal{E}$ together with an density function $p(t,x,y)$, $(t,x,y) \in (0,\infty) \times E \times E$ such that $\mathbb{E}_x\left[f\left(X(t)\right)\right] = \int p\left(t,x,y\right) f(y)\, dm(y)$ for all $f \in C_0(E)$ and for all $x \in E$ and all $t > 0$.

It is noticed that the definition of $L_M f(x)$ is only defined pointwise, and that for certain points $x \in E$ the limit

$$L_M f(x) := \lim_{t \downarrow 0} \frac{\mathbb{E}_x \left[M(t) f\left(X(t)\right)\right] - \mathbb{E}_x \left[M(0) f\left(X(0)\right)\right]}{t}$$

does not even exist. A good example is obtained by taking for T the exit time from an open subset U: $T = \tau_U = \inf \{s > 0 : X(s) \in E \backslash U\}$. If the $\lim_{t \downarrow 0} \frac{\mathbb{P}_x [T \leqslant t]}{t} = 0$ for all $x \in U$, then $L_M f(x) = Lf(x)$ for $x \in U$.

Proof of Corollary 2.63. It is only (2.154) which needs some explanation; the others are direct consequences of Proposition 2.59. To this end we fix $0 < u < t$. Then by (2.141) the process:

$$s \mapsto \mathbf{1}_{\{T > s\}} \mathbb{E}_{X(s)} \left[p\left(u, X(t-s-u), y\right), T > t-s-u \right]$$

is a martingale on the closed interval $[0, t-u]$. Next we rewrite

$$\begin{aligned} &\mathbb{E}_{X(s)} \left[p\left(u, X(t-s-u), y\right), T > t-s-u \right] \\ &= \mathbb{E}_{X(s)} \left[p\left(u, X(t-s-u), y\right)\right] - \mathbb{E}_{X(s)} \left[p\left(u, X(t-s-u), y\right), T \leqslant t-s-u \right] \end{aligned} \tag{2.155}$$

(the process $\rho \mapsto p\left(t-s-\rho, X(\rho), y\right)$ is $\mathbb{P}_z$-martingale with $z = X(s)$; put $u = t-s$ in the first term, and $u = t-s-T$ in the second term of the right-hand side of (2.155))

$$= \mathbb{E}_{X(s)} \left[p\left(t-s, X(0), y\right)\right] - \mathbb{E}_{X(s)} \left[p\left(t-s-T, X(T), y\right), T \leqslant t-s-u \right]. \tag{2.156}$$

By letting $u \downarrow 0$ in (2.156) and using (2.141) of Proposition 2.59 we obtain that the process in (2.154) is a $\mathbb{P}_x$-martingale. This completes the proof of Corollary 2.63. □

Next let

$$\left\{ (\Omega, \mathcal{F}, \mathbb{P}_x), (B(t), t \geqslant 0), (\vartheta_t, t \geqslant 0), \left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right) \right\}$$

be the Markov process of Brownian motion. Another application of martingale theory is the following example. Let U be an open subset of $\mathbb{R}^d$ with smooth enough boundary ∂U (C^1 will do), and let $f : \partial U \to \mathbb{R}$ be a bounded continuous function on the boundary ∂U of U. Let $u : \overline{U} \to \mathbb{R}$ be a continuous function such that $u(x) = f(x)$ for $x \in \partial U$ and such that $\Delta u(x) = 0$ for $x \in U$. Let τ_U be the first exit time from U: $\tau_U = \inf \left\{ s > 0 : B(s) \in \mathbb{R}^d \backslash U \right\}$. Then

$$u(x) = \mathbb{E}_x \left[f\left(B\left(\tau_U\right)\right) : \tau_U < \infty \right] + \lim_{t \to \infty} \mathbb{E}_x \left[u\left(B\left(t\right)\right) : \tau_U = \infty \right]. \tag{2.157}$$

Notice that the first expression in (2.157) makes sense, because it can be proved that Brownian motion is $\mathbb{P}_x$-almost surely continuous for $x \in \mathbb{R}^d$. The proof uses the following facts: stopped martingales are again martingales, and the processes

$$t \mapsto f\left(B(t)\right) - f\left(B(0)\right) - \frac{1}{2} \int_0^t \Delta f\left(B(s)\right) ds, \quad f \in C_b\left(\mathbb{R}^d\right), \ \Delta f \in C_b\left(\mathbb{R}^d\right), \tag{2.158}$$

are martingales. The fact that a process of the form (2.158) is a martingale follows from (2.142) in Corollary 2.60. It can also be proved using the equality

$$\frac{\partial}{\partial t} p_d(t,x,y) = \frac{1}{2}\Delta_y p_d(t,x,y) \tag{2.159}$$

where

$$p_d(t,x,y) = \frac{1}{\sqrt{(2\pi t)^d}} e^{-\frac{|x-y|^2}{2t}}.$$

A proof of (2.158) runs as follows. Pick $t_2 > t_1 \geqslant 0$ en a function $f \in C_b\left(\mathbb{R}^d\right)$, such that Δf also belongs to $C_b\left(\mathbb{R}^d\right)$. Then we have:

$$\begin{aligned}
&\mathbb{E}_x\left[f(B(t_2)) - f(B(0)) - \frac{1}{2}\int_0^{t_2} \Delta f(B(s))\, ds \;\Big|\; \mathcal{F}_{t_1}\right] \\
&\quad - f(B(t_1)) + f(B(0)) + \frac{1}{2}\int_0^{t_1} \Delta f(B(s))\, ds \\
&= \mathbb{E}_x\left[\left(f(B(t_2-t_1)) - f(B(0)) - \frac{1}{2}\int_0^{t_2-t_1} \Delta f(B(s))\, ds\right)\circ\vartheta_{t_1} \;\Big|\; \mathcal{F}_{t_1}\right]
\end{aligned}$$

(Markov property of Brownian motion)

$$= \mathbb{E}_{B(t_1)}\left[f(B(t_2-t_1)) - f(B(0)) - \frac{1}{2}\int_0^{t_2-t_1} \Delta f(B(s))\, ds\right]$$

(put $z = B(t_1)$, and $t = t_2 - t_1$)

$$\begin{aligned}
&= \mathbb{E}_z\left[f(B(t))\right] - \mathbb{E}_z\left[f(B(0))\right] - \frac{1}{2}\int_0^t \mathbb{E}_z\left[\Delta f(B(s))\right] ds \\
&= \mathbb{E}_z\left[f(B(t))\right] - \mathbb{E}_z\left[f(B(0))\right] - \frac{1}{2}\int_0^t \int_{\mathbb{R}^d} p_d(s,z,y)\,\Delta f(y)\, dy\, ds \\
&= \mathbb{E}_z\left[f(B(t))\right] - \mathbb{E}_z\left[f(B(0))\right] - \lim_{\varepsilon\downarrow 0}\frac{1}{2}\int_\varepsilon^t \int_{\mathbb{R}^d} p_d(s,z,y)\,\Delta f(y)\, dy
\end{aligned}$$

(integration by parts)

$$= \mathbb{E}_z\left[f(B(t))\right] - \mathbb{E}_z\left[f(B(0))\right] - \lim_{\varepsilon\downarrow 0}\int_\varepsilon^t \int_{\mathbb{R}^d} \frac{1}{2}\Delta_y p_d(s,z,y)\, f(y)\, dy\, ds$$

(use the equality in (2.159))

$$= \mathbb{E}_z\left[f(B(t))\right] - \mathbb{E}_z\left[f(B(0))\right] - \lim_{\varepsilon\downarrow 0}\int_\varepsilon^t \int_{\mathbb{R}^d} \frac{\partial}{\partial s} p_d(s,z,y)\, f(y)\, dy\, ds$$

(interchange integration and differentiation)

$$= \mathbb{E}_z\left[f(B(t))\right] - \mathbb{E}_z\left[f(B(0))\right] - \lim_{\varepsilon\downarrow 0}\int_\varepsilon^t \frac{\partial}{\partial s}\int_{\mathbb{R}^d} p_d(s,z,y)\, f(y)\, dy\, ds$$

(fundamental rule of calculus)

$$= \mathbb{E}_z\left[f(B(t))\right] - \mathbb{E}_z\left[f(B(0))\right]$$

$$
\begin{aligned}
&-\lim_{\varepsilon\downarrow 0}\left(\int_{\mathbb{R}^d} p_d\left(t,z,y\right) f\left(y\right)\, dy - \int_{\mathbb{R}^d} p_d\left(\varepsilon,z,y\right) f\left(y\right)\, dy\right)\\
&= \mathbb{E}_z\left[f\left(B(t)\right)\right] - \mathbb{E}_z\left[f\left(B(0)\right)\right] - \mathbb{E}_z\left[f\left(B(t)\right)\right] + \lim_{\varepsilon\downarrow 0}\mathbb{E}_z\left[f\left(B(\varepsilon)\right)\right]\\
&= \lim_{\varepsilon\downarrow 0}\mathbb{E}_z\left[f\left(B(\varepsilon)\right)\right] - \mathbb{E}_z\left[f\left(B(0)\right)\right] = 0. \qquad (2.160)
\end{aligned}
$$

From Doob's optional sampling theorem it follows that processes of the form

$$
t \mapsto f\left(B\left(\tau_U \wedge t\right)\right) - f\left(B(0)\right) - \frac{1}{2}\int_0^{\tau_U\wedge t} \Delta f\left(B(s)\right)\, ds, \quad f \in C_b\left(\mathbb{R}^d\right), \qquad (2.161)
$$

$f \in C_b\left(\mathbb{R}^d\right)$, $\Delta f \in C_b\left(\mathbb{R}^d\right)$, are $\mathbb{P}_x$-martingales for $x \in U$. We can apply this property to our harmonic function u. It follows that the process

$$
t \mapsto u\left(B\left(\tau_U \wedge t\right)\right) - u\left(B(0)\right) - \frac{1}{2}\int_0^{\tau_U\wedge t} \Delta u\left(B(s)\right)\, ds = u\left(B\left(\tau_U \wedge t\right)\right) - u\left(B(0)\right) \qquad (2.162)
$$

is a martingale. Consequently, from (2.162) we get

$$
\begin{aligned}
u(x) = u\left(B(0)\right) &= \mathbb{E}_x\left[u\left(B\left(\tau_U\wedge t\right)\right)\right]\\
&= \mathbb{E}_x\left[u\left(B\left(\tau_U\wedge t\right)\right), \tau_U \leqslant t\right] + \mathbb{E}_x\left[u\left(B\left(\tau_U\wedge t\right)\right), \tau_U > t\right]\\
&= \mathbb{E}_x\left[u\left(B\left(\tau_U\right)\right), \tau_U \leqslant t\right] + \mathbb{E}_x\left[u\left(B(t)\right), \tau_U > t\right] \qquad (2.163)
\end{aligned}
$$

In (2.163) we let $t \to \infty$ to obtain the equality in (2.157).

2.64. Proposition. *Let $t \mapsto M_1(t)$ and $t \mapsto M_2(t)$ be two continuous martingales in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with covariation process $t \mapsto \langle M_1, M_2 \rangle(t)$, so that in particular the process $t \mapsto M_1(t)M_2(t) - \langle M_1, M_2 \rangle(t)$ is a martingale in $L^1(\Omega, \mathcal{F}, \mathbb{P})$. Then the process*

$$t \mapsto (M_1(t) - M_1(s))(M_2(t) - M_2(s)) - \langle M_1, M_2 \rangle(t) + \langle M_1, M_2 \rangle(s), \quad t \geqslant s, \tag{2.164}$$

is a martingale.

In fact by Itô calculus we have the following integration by parts formula:

$$\begin{aligned}
&(M_1(t) - M_1(s))(M_2(t) - M_2(s)) \\
&= \int_s^t (M_1(\rho) - M_1(s))\, dM_2(\rho) + \int_s^t (M_2(\rho) - M_2(s))\, dM_1(\rho) \\
&\quad + \langle M_1, M_2 \rangle(t) - \langle M_1, M_2 \rangle(s), \quad t \geqslant s.
\end{aligned} \tag{2.165}$$

Proof of Proposition 2.64. Fix $t_2 > t_1 \geqslant s$. Then we calculate:

$$\begin{aligned}
&\mathbb{E}\left[(M_1(t_2) - M_1(s))\left(M_2(t_2) - M_2(s) \,\middle|\, \mathcal{F}_{t_1}\right)\right] \\
&\quad - \mathbb{E}\left[\langle M_1, M_2 \rangle(t_2) - \langle M_1, M_2 \rangle(s) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad - (M_1(t_1) - M_1(s))(M_2(t_1) - M_2(s)) + \langle M_1, M_2 \rangle(t_1) - \langle M_1, M_2 \rangle(s) \\
&= \mathbb{E}\left[(M_1(t_2) - M_1(s))(M_2(t_2) - M_2(s)) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad\quad - \mathbb{E}\left[\langle M_1, M_2 \rangle(t_2) - \langle M_1, M_2 \rangle(t_1) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad - (M_1(t_1) - M_1(s))(M_2(t_1) - M_2(s)) \\
&= \mathbb{E}\left[(M_1(t_2) - M_1(t_1) + M_1(t_1) - M_1(s))\right. \\
&\quad\quad (M_2(t_2) - M_2(t_1) + M_2(t_1) - M_2(s)) \\
&\quad\quad\quad \left. - \langle M_1, M_2 \rangle(t_2) + \langle M_1, M_2 \rangle(t_1) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad - (M_1(t_1) - M_1(s))(M_2(t_1) - M_2(s)) \\
&= \mathbb{E}\left[(M_1(t_2) - M_1(t_1))(M_2(t_2) - M_2(t_1)) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad\quad - \mathbb{E}\left[\langle M_1, M_2 \rangle(t_2) + \langle M_1, M_2 \rangle(t_1) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad + \mathbb{E}\left[(M_1(t_1) - M_1(s))(M_2(t_2) - M_2(t_1)) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad + \mathbb{E}\left[(M_1(t_2) - M_1(t_1))(M_2(t_1) - M_2(s)) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad + \mathbb{E}\left[(M_1(t_1) - M_1(s))(M_2(t_1) - M_2(s)) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad - (M_1(t_1) - M_1(s))(M_2(t_1) - M_2(s)) \\
&= \mathbb{E}\left[M_1(t_2)\,M_2(t_2) - \langle M_1, M_2 \rangle(t_2) + \langle M_1, M_2 \rangle(t_1) - M_1(t_1)\,M_2(t_1) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad - \mathbb{E}\left[M_1(s)(M_2(t_2) - M_2(t_1)) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad - \mathbb{E}\left[(M_1(t_2) - M_1(t_1))\,M_2(s) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad + \mathbb{E}\left[(M_1(t_1) - M_1(s))(M_2(t_1) - M_2(s)) \,\middle|\, \mathcal{F}_{t_1}\right] \\
&\quad - (M_1(t_1) - M_1(s))(M_2(t_1) - M_2(s)) = 0.
\end{aligned} \tag{2.166}$$

In the final step of (2.166) we employed the martingale property of the following processes:

$$t \mapsto M_1(t)M_2(t) - \langle M_1, M_2 \rangle (t), \; t \mapsto M_1(t), \quad \text{and} \quad t \mapsto M_2(t).$$

This completes the proof of Proposition 2.64. $\square$

CHAPTER 3

An introduction to stochastic processes: Brownian motion, Gaussian processes and martingales

In this chapter of the book we will study several aspects of Brownian motion: Brownian motion as a Gaussian process, Brownian motion as a Markov process, Brownian motion as a martingale. It also includes a discussion on stochastic integrals and Itô's formula.

1. Gaussian processes

We begin with an important extension theorem of Kolmogorov, which enables us to construct stochastic processes like Gaussian processes, Lévy processes, Poisson processes and others. It is also useful for the construction of Markov processes. In Theorem 3.1 the symbol Ω_J, $J \subseteq I$, stands for the product space $\Omega_J = \prod_{j\in J} \Omega_j$ endowed with the product σ-field $\mathcal{F}_J$. By saying that the system $\{(\Omega_J, \mathcal{F}_J, \mathbb{P}_J) : J \subseteq I,\ J \text{ finite}\}$ is a projective system (or a consistent system, or a cylindrical measure) we mean that

$$\mathbb{P}_{J_1}\left[p^{J_1}_{J_2} \in A\right] = \mathbb{P}_{J_1}\left[\left(p^{J_1}_{J_2}\right)^{-1}(A)\right] = \mathbb{P}_{J_2}[A],$$

where $A \in \mathcal{F}_{J_2}$, and where $J_2 \subseteq J_1 \subseteq I$, J_1 finite. The mapping $p^{J_1}_{J_2}$, $J_2 \subseteq J_1$, is defined by $p^{J_1}_{J_2}(\omega_j)_{j\in J_1} = (\omega_j)_{j\in J_2}$. In practice this means that in order to prove that the system $\{(\Omega_J, \mathcal{F}_J, \mathbb{P}_J) : J \subseteq I,\quad J \quad \text{finite}\}$ is a projective system indeed, we have to show an equality of the form ($j_0 \notin J$):

$$\mathbb{P}_{J\cup\{j_0\}}\left[B \times \Omega_{j_0}\right] = \mathbb{P}_J[B], \quad B \in \mathcal{F}_J.$$

The following proposition says that under certain conditions a cylindrical measure in fact is a genuine measure.

3.1. Theorem (Extension theorem of Kolmogorov). *Let*

$$\{(\Omega_J, \mathcal{F}_J, \mathbb{P}_J) : J \subseteq I,\ J \text{ finite}\}$$

be a projective system of probability spaces (or distributions). Suppose that each Ω_i is a metrizable and σ-compact Hausdorff space endowed with its Borel field $\mathcal{A}_i$. Then there exists a unique probability measure $\mathbb{P}_I$ on $(\Omega_I, \mathcal{A}_I)$, such that

$$\mathbb{P}_I[p_J \in A] = \mathbb{P}_I\left(p_J^{-1}(A)\right) = \mathbb{P}_J(A) \tag{3.1}$$

for every $J \subseteq I$, J finite, and for every $A \in \mathcal{A}_J$.

For an extensive discussion on Kolmogorov's extension theorem see, *e.g.*, the Probability Theory lecture notes of B. Driver [**40**]. These lecture notes include a discussion on standard Borel spaces and on Polish spaces. The Kolmogorov's extension theorem is also valid if the spaces Ω_i are Polish spaces, or Souslin spaces which are continuous images of Polish spaces. For more details see Appendix 17.6 in [**40**]. The reader may also consult [**21**] or [**137**]. In Theorem 7.4.3 of [**21**] the author shows that finite positive measures on Souslin spaces are regular and concentrated on σ-compact subsets. Bogachev's book contains lots of information on Souslin spaces. In fact much material which is presented in this book, can also be found in the lecture notes by Bruce Driver. A proof of Kolmogorov's extension theorem is supplied in Section 4 of Chapter 5: see (the proof of) Theorem 5.81.

Next we recall Bochner's theorem.

3.2. Theorem. *(Bochner) Let $\varphi : \mathbb{R}^n \to \mathbb{C}$ be a continuous complex function, that is positive definite in the sense that for all $r \in \mathbb{N}$*

$$\sum_{k,\ell=1}^{r} \lambda_k \overline{\lambda}_\ell \varphi\left(\xi^k - \xi^\ell\right) \geqslant 0, \tag{3.2}$$

for all $\lambda_1, \dots, \lambda_r \in \mathbb{C}$ and for all $\xi^1, \dots, \xi^r \in \mathbb{R}^n$. Then there exists a unique non-negative Borel measure μ on $\mathbb{R}^n$ such that its Fourier transform

$$\int \exp(-i \langle \xi, x \rangle) d\mu(x)$$

is equal to $\varphi(\xi)$ for $\xi \in \mathbb{R}^n$. In particular $\mu(\mathbb{R}^n) = \varphi(0)$.

3.3. Example. Let, for every $i \in I$, $\mathbb{P}_i$, $i \in I$, be a probability measures on Ω_i and define $\mathbb{P}_J$ on Ω_J, $J \subseteq I$, J finite, by $\mathbb{P}_J(A) = \mathbb{P}_{j_1} \otimes \cdots \otimes \mathbb{P}_{j_n}(A)$, where A belongs to $\mathcal{A}_J$ and where $J = (j_1, \dots, j_n)$. Then the family $\{\mathbb{P}_J : J \subset I, \ J \text{ finite}\}$ is a consistent system or cylindrical measure.

3.4. Example. Let $\sigma : I \times I \to \mathbb{R}$ be a symmetric (*i.e.* $\sigma(i,j) = \sigma(j,i)$ for all i, j in I) function such that for every finite subset $J = (j_1, \dots, j_n)$ of I the matrix $(\sigma(i,j))_{i,j \in J}$ is positive-definite in the sense that

$$\sum_{i,j \in J} \sigma(i,j) \xi_i \xi_j \geqslant 0, \tag{3.3}$$

for all $\xi_{j_1}, \dots, \xi_{j_n} \in \mathbb{R}$. In the non-degenerate case we shall assume that the inequality in (3.3) is strict whenever the vector $(\xi_{j_1}, \dots, \xi_{j_n})$ is non-zero. Define the process $(i, \omega) \mapsto X_i(\omega)$ by $X_i(\omega) = \omega_i$, where $\omega \in \Omega_I = \mathbb{R}^I$ is given by $\omega = (\omega_i)_{i \in I}$. Let $\mu = (\mu_i) \in \mathbb{R}^I$ be a map from I to $\mathbb{R}$. There exists a unique probability measure $\mathbb{P}$ on the σ-field on Ω_I generated by $(X_i)_{i \in I}$ with the following property:

$$\mathbb{E}\left(\exp\left(-i \sum_{j \in J} \xi_j X_j\right)\right) = \exp\left(-i \sum_{j \in J} \xi_j \mu_j\right) \exp\left(-\frac{1}{2} \sum_{i,j \in J} \sigma(i,j) \xi_i \xi_j\right). \tag{3.4}$$

This measure possesses the following additional properties:

$$\mathbb{E}(X_j) = \mu_j, \quad j \in I, \quad \text{and} \quad \operatorname{cov}(X_i, X_j) = \sigma(i,j), \quad i,j \in I. \tag{3.5}$$

Notice that $\sum_{u,v=1}^{n} \xi_u \xi_v \operatorname{cov}\left(X_{j_u}, X_{j_v}\right) \geqslant 0$ whenever $\xi_1, \dots, \xi_n$ belong to $\mathbb{R}$. For a proof of this result we shall employ both Bochner's theorem as well as Kolmogorov's extension theorem. Therefore let $J = (j_1, \dots, j_n)$ be a finite subset of I, let λ_k, $1 \leqslant k \leqslant r$, be complex numbers and let ξ^k, $1 \leqslant k \leqslant r$, be vectors in $\mathbb{R}^n \equiv \mathbb{R}^J$. Put $\lambda'_k = \lambda_k \exp\left(i \sum_{u=1}^{n} \xi^k_{j_u} \mu^k_{j_u}\right)$ and let U be an orthogonal matrix with the property that the matrix $\left(U\sigma U^{-1}(u,v)\right)_{u,v=1}^{n}$ has the diagonal form $\left(U\sigma U^{-1}(u,v)\right) = \begin{pmatrix} s_1^2 & 0 & \dots & 0 \\ 0 & s_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & s_n^2 \end{pmatrix}$. We also write $\left(\eta_1^\ell, \dots, \eta_n^\ell\right) = U \begin{pmatrix} \xi^\ell_{j_1} \\ \vdots \\ \xi^\ell_{j_n} \end{pmatrix}$. We may and do suppose that the eigenvalues $s_1, \dots, s_m$, $m \leqslant n$, are non-zero and the others (if any) are 0. Then we get

$$
\begin{aligned}
&\sum_{k,\ell=1}^{r} \lambda_k \overline{\lambda}_\ell \exp\left(-\frac{1}{2} \sum_{u,v=1}^{n} \sigma(j_u, j_v) \left(\xi^\ell_{j_u} - \xi^k_{j_u}\right) \left(\xi^\ell_{j_v} - \xi^k_{j_v}\right)\right) \\
&\qquad \times \exp\left(-i \sum_{u=1}^{n} \left(\xi^\ell_{j_u} - \xi^k_{j_u}\right) \mu_{j_u}\right) \\
&= \sum_{k,\ell=1}^{r} \lambda'_k \overline{\lambda'_\ell} \exp\left(-\frac{1}{2} \sum_{u,v=1}^{n} \left(U\sigma U^{-1}\right)(u,v) \left(\eta^\ell_u - \eta^k_u\right)\left(\eta^\ell_v - \eta^k_v\right)\right) \\
&= \sum_{k,\ell=1}^{r} \lambda'_k \overline{\lambda'_\ell} \exp\left(-\frac{1}{2} \sum_{u=1}^{n} s_u^2 \left(\eta^\ell_u - \eta^k_u\right)^2\right) \\
&= \sum_{k,\ell=1}^{r} \lambda'_k \overline{\lambda'_\ell} \exp\left(-\frac{1}{2} \sum_{u=1}^{m} s_u^2 \left(\eta^\ell_u - \eta^k_u\right)^2\right) \\
&= \sum_{k,\ell=1}^{r} \lambda'_k \overline{\lambda'_\ell} \frac{1}{(\sqrt{2\pi})^m} \frac{1}{\prod_{u=1}^{m} s_u} \int \dots \int dx_1 \dots dx_m \\
&\qquad \exp\left(-i \sum_{u=1}^{m} \left(\eta^\ell_u - \eta^k_u\right) x_u\right) \exp\left(-\frac{1}{2} \sum_{u=1}^{m} \frac{x_u^2}{s_u^2}\right) \\
&= \frac{1}{(\sqrt{2\pi})^m} \frac{1}{\prod_{u=1}^{m} s_u} \int \dots \int dx_1 \dots dx_m \\
&\qquad \left|\sum_{k=1}^{r} \lambda'_k \exp\left(i \sum_{u=1}^{m} \eta^\ell_u x_u\right)\right|^2 \exp\left(-\frac{1}{2} \sum_{u=1}^{m} \frac{x_u^2}{s_u^2}\right) \geqslant 0. \qquad (3.6)
\end{aligned}
$$

From Bochner's Theorem 3.2 it follows that there exists a probability measure Π_J on $\mathbb{R}^J$ such that, for all $\xi \in \mathbb{R}^n$,

$$
\int_{\mathbb{R}^J} \exp\left(-i \sum_{u=1}^{n} \xi_u x_u\right) d\Pi_J
$$

$$= \exp\left(-i\sum_{u=1}^{n}\xi_u\mu_u\right)\exp\left(-\frac{1}{2}\sum_{u,v=1}^{n}\sigma(j_u,j_v)\xi_u\xi_v\right). \tag{3.7}$$

Define the probability measure $\mathbb{P}_J$ on $\Omega_J = \prod_{j\in J}\Omega_j$ by

$$\mathbb{P}_J\left((X_{j_1},\dots,X_{j_n})\in B\right) = \Pi_J(B),$$

where B is a Borel subset of $\mathbb{R}^J$. The collection $(\Omega_J, \mathcal{A}_J, \mathbb{P}_J)$ is a projective system, because let $J' := \{j_0\}\cup J$ be a subset of I, which is of size $1+$ size $J = 1+n$ and let B be a Borel subset of $\mathbb{R}^J$. The Fourier transform of the measure $B \mapsto \Pi_{J'}[\mathbb{R}\times B]$ is given by the function:

$$\begin{aligned}
(\xi_{j_1},\dots,\xi_{j_n}) \mapsto & \int_{\mathbb{R}^J}\exp\left(-i\sum_{j\in J}\xi_j x_j\right)\Pi_{J'}[\mathbb{R}\times dx] \\
&= \int_{\mathbb{R}^J}\int_{\mathbb{R}}\exp\left(-i\sum_{j\in J'}\xi_j x_j\right)\Pi_{J'}[dy\times dx] \\
&= \exp\left(-i\sum_{j\in J'}\xi_j\mu_j\right)\exp\left(-\frac{1}{2}\sum_{i,j\in J'}\sigma(i,j)\xi_i\xi_j\right) \\
&= \exp\left(-i\sum_{j\in J}\xi_j\mu_j\right)\exp\left(-\frac{1}{2}\sum_{i,j\in J}\sigma(i,j)\xi_i\xi_j\right) \\
&= \int_{\mathbb{R}^J} e^{-i\sum_{j\in J}\xi_j x_j}\,\Pi_J(dx).
\end{aligned}$$

In the previous formula we used the equality $\xi_{j_0} = 0$ several times: $J' = J \cup \{j_0\}$. It follows that $\Pi_J[B] = \Pi_{J'}[\mathbb{R} \times B]$. An application of the extension theorem of Kolmogorov yields the desired result in Example 3.4.

Suppose that the matrix $(\sigma(j_u, j_v))_{u,v=1}^n$ be non-degenerate (i.e. suppose that its determinant is non-zero) and let $(\alpha(u,v))_{u,v=1}^n$ be its inverse. Then

$$\begin{aligned}
&\mathbb{P}\left((X_{j_1}, \dots, X_{j_n}) \in B\right) \qquad (3.8)\\
&= \frac{(\det \alpha)^{1/2}}{(2\pi)^{n/2}} \int \dots \int dx_1 \dots dx_n 1_B(x_1, \dots, x_n)\\
&\qquad \exp\left(-\frac{1}{2}\sum_{u,v=1}^n \alpha(u,v)\,(x_u - \mu_u)\,(x_v - \mu_v)\right).
\end{aligned}$$

Equality (3.8) can be proved by showing that the Fourier transforms of both measures in (3.8) coincide. In the following propositions (Propositions 3.5 and 3.6) we mention some elementary facts on Gaussian vectors. Gaussian vectors are multivariate normally distributed random vectors.

3.5. PROPOSITION. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X^i : \Omega \mapsto \mathbb{R}^{n_i}$, $i = 1,\ 2$, be random vectors with the property that the random vector $X(\omega) := (X^1(\omega), X^2(\omega))$ is Gaussian in the sense that ($n = n_1 + n_2$)*

$$\mathbb{E}\left(\exp\left(-i\sum_{k=1}^n \xi_k X_k\right)\right) = \exp\left(-i\sum_{k=1}^n \xi_k\mu_k - \frac{1}{2}\sum_{k,\ell=1}^n \sigma(k,\ell)\xi_k\xi_\ell\right), \qquad (3.9)$$

where the matrix $\sigma(k,\ell)_{k,\ell=1}^n$ is positive definite and where $(\mu_1, \dots, \mu_n)$ is a vector in $\mathbb{R}^n$. The vectors X^1 and X^2 are $\mathbb{P}$-independent if and only if they are uncorrelated in the sense that

$$\mathbb{E}\left(X_i^1 X_j^2\right) = \mathbb{E}\left(X_i^1\right)\mathbb{E}\left(X_j^2\right) \qquad (3.10)$$

for all $1 \leqslant i \leqslant n_1$ and for all $1 \leqslant j \leqslant n_2$.

PROOF. The necessity is clear. For the sufficiency we proceed as follows. Put

$$\left(X^1, X^2\right) = \left(X_1, \dots, X_{n_1}, X_{n_1+1}, \dots, X_{n_1+n_2}\right).$$

Since the vectors X^1 and X^2 are uncorrelated (see (3.10)), it follows that

$$\sum_{k,\ell=1}^n \sigma(k,\ell)\xi_k\xi_\ell = \sum_{k,\ell=1}^{n_1} \sigma(k,\ell)\xi_k\xi_\ell + \sum_{k,\ell=n_1+1}^{n} \sigma(k,\ell)\xi_k\xi_\ell. \qquad (3.11)$$

From (3.9) it follows that

$$\begin{aligned}
&\mathbb{E}\left(\exp\left(-i\sum_{k=1}^n \xi_k X_k\right)\right)\\
&= \mathbb{E}\left(\exp\left(-i\sum_{k=1}^{n_1} \xi_k X_k\right)\right)\mathbb{E}\left(\exp\left(-i\sum_{k=n_1+1}^{n_1+n_2} \xi_k X_k\right)\right) \qquad (3.12)
\end{aligned}$$

and hence that the random vectors X^1 and X^2 are independent. □

3.6. PROPOSITION. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.*

(a) *Let $Q : \mathbb{R}^m \to \mathbb{R}^n$ be a linear map. If $X : \Omega \to \mathbb{R}^m$ is a Gaussian vector, then so is QX.*
(b) *A random vector $X : \Omega \to \mathbb{R}^n$ is Gaussian if and only if for every $\xi \in \mathbb{R}^n$ the random variable $\omega \mapsto \langle \xi, X(\omega)\rangle$ is Gaussian.*

PROOF. (a) A random vector X is Gaussian if and only if the Fourier transform of the measure $B \mapsto \mathbb{P}(X \in B)$ is of the form

$$\xi \mapsto \exp\left(-i\langle \xi, \mu\rangle - \frac{1}{2}\langle \sigma\xi, \xi\rangle\right).$$

By a standard result on image measures the Fourier transform of the measure $B \mapsto \mathbb{P}(QX \in B)$, where $X : \Omega \to \mathbb{R}^n$ is Gaussian and where B is a Borel subset of $\mathbb{R}^m$, is given by

$$\begin{aligned} \xi \mapsto \mathbb{E}\left[\exp\left(-i\langle \xi, QX\rangle\right)\right] &= \mathbb{E}\left[\exp\left(-i\langle Q^*\xi, X\rangle\right)\right] \\ &= \exp\left(-i\langle Q^*\xi, \mu\rangle\right)\exp\left(-\frac{1}{2}\langle \sigma Q^*\xi, Q^*\xi\rangle\right). \end{aligned} \tag{3.13}$$

This proves (a). It also proves that the dispersion matrix of QX is given by $Q\sigma Q^*$.

(b) For the necessity we apply (a) with the linear map $Qx := \langle \xi, x\rangle$, $x \in \mathbb{R}^n$, where $\xi \in \mathbb{R}^n$ is fixed. For the sufficiency we again fix $\xi \in \mathbb{R}^n$. Since $Y := \langle \xi, X\rangle$ is a Gaussian variable we have

$$\begin{aligned} \mathbb{E}\left(\exp\left(-i\langle \xi, X\rangle\right)\right) &= \mathbb{E}\left(\exp\left(-iY\right)\right) \\ &= \exp\left(-i\mathbb{E}(Y)\right)\exp\left(-\frac{1}{2}\mathbb{E}\left(Y - \mathbb{E}(Y)\right)^2\right) \\ &= \exp\left(-i\langle \xi, \mu\rangle\right)\exp\left(-\frac{1}{2}\langle \sigma\xi, \xi\rangle\right), \end{aligned} \tag{3.14}$$

where $\mu = \mathbb{E}(X)$ and where

$$\sigma(k, \ell) = \operatorname{cov}\left(X_k, X_\ell\right) = \mathbb{E}\left(X_k - \mathbb{E}(X_k)\right)\left(X_\ell - \mathbb{E}\left(X_\ell\right)\right).$$

This completes the proof of (b). □

3.7. THEOREM. *Let $\sigma : I \times I \to \mathbb{R}$ be a positive-definite function and let $\mu : I \to \mathbb{R}$ be a map. There exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ together with a Gaussian process $(t, \omega) \mapsto X_t(\omega) = X(t, \omega)$, $t \in I$, $\omega \in \Omega$, such that $\mathbb{E}(X_t) = \mu_t$ and such that cov$(X_s, X_t) = \sigma(s, t)$ for all s, $t \in I$.*

PROOF. The proof is essentially given in Example 3.4. □

We conclude this section with the introduction of Brownian motion and Brownian bridge as Gaussian processes. First we show that the function $\sigma : [0, \infty) \times [0, \infty) \to \mathbb{R}$, defined by $\sigma(u, v) = \min(u, v)$, u, $v \in [0, \infty)$, and, for $t \geqslant 0$ fixed, the function $\sigma_t : [0, t] \times [0, t] \to \mathbb{R}$, defined by $\sigma_t(u, v) = t\min(u, v) - uv$, u, $v \in [0, t]$, are positive definite.

3.8. PROPOSITION. *The functions* $\sigma(u,v) = \min(u,v)$, $u,\ v \in [0,\infty)$, $\sigma_t(u,v) = t\min(u,v) - uv$, $u,\ v \in [0,t]$, *and* $\sigma_{\mathbb{R}}(u,v) = \dfrac{1}{2}\exp\left(-|u-v|\right)$, $u,\ v \in \mathbb{R}$, *are positive definite. In addition, the function* $\sigma_0(u,v)$ *defined by*

$$\sigma_0(u,v) = \frac{1}{2}\exp\bigl(-(u+v)\bigr)\left(\exp\left(2\min(u,v)\right) - 1\right),\ \ u,\ \ v \geqslant 0,$$

is positive definite.

PROOF. Let $0 = s_0 < s_1 < s_2 < s_3 < \cdots < s_n < t$ and let $\lambda_1, \dots, \lambda_n$ be complex numbers. The following identities are valid:

$$
\begin{aligned}
&\sum_{j=1}^{n} \left| \sum_{k=j}^{n} \lambda_k (t - s_k) \right|^2 \frac{t(s_j - s_{j-1})}{(t - s_j)(t - s_{j-1})} \\
&= \sum_{j=1}^{n} \sum_{k_1=j}^{n} \sum_{k_2=j}^{n} \lambda_{k_1} \overline{\lambda}_{k_2} (t - s_{k_1})(t - s_{k_2}) \left\{ \frac{s_j}{t - s_j} - \frac{s_{j-1}}{t - s_{j-1}} \right\} \\
&= \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \sum_{j=1}^{\min(k_1,k_2)} \lambda_{k_1} \overline{\lambda}_{k_2} \left\{ \frac{s_j}{t - s_j} - \frac{s_{j-1}}{t - s_{j-1}} \right\} (t - s_{k_1})(t - s_{k_2}) \\
&= \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \lambda_{k_1} \overline{\lambda}_{k_2} \frac{s_{\min(k_1,k_2)}}{t - s_{\min(k_1,k_2)}} (t - s_{k_1})(t - s_{k_2}) \\
&= \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \lambda_{k_1} \overline{\lambda}_{k_2} s_{\min(k_1,k_2)} \left(t - s_{\max(k_1,k_2)}\right) = \sum_{k_1,k_2=1}^{n} \lambda_{k_1} \overline{\lambda}_{k_2} \sigma_t (s_{k_1}, s_{k_2})
\end{aligned}
\tag{3.15}
$$

and hence the function σ_t is positive definite. Since

$$
\sum_{j,k=1}^{n} \lambda_j \overline{\lambda}_k t \min(s_j, s_k) = \sum_{j,k=1}^{n} \lambda_j \overline{\lambda}_k \sigma_t(s_j, s_k) + \left| \sum_{j=1}^{n} \lambda_j s_j \right|^2,
$$

it follows that the function σ is positive definite as well.

In order to prove that the function $\sigma_{\mathbb{R}}$ is positive definite we first notice that the Fourier transform of the function $t \mapsto \exp(-|t|)$ is given by

$$
\begin{aligned}
\int_{-\infty}^{\infty} e^{-i\xi t} e^{-|t|} dt &= 2 \int_0^{\infty} \cos(\xi t)\, e^{-t} dt \\
&= 2\mathrm{Re} \int_0^{\infty} e^{-t(1-i\xi)} dt = 2\mathrm{Re} \frac{1}{1 - i\xi} = \frac{2}{1 + \xi^2}.
\end{aligned}
\tag{3.16}
$$

Hence upon taking the inverse Fourier transform we obtain:

$$
\frac{1}{2} e^{-|t-s|} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(i\xi(t - s))}{\xi^2 + 1} d\xi. \tag{3.17}
$$

Let $\lambda_1, \dots, \lambda_n$ be complex numbers and let $s_1, \dots, s_n$ be real numbers. From (3.16) and (3.17) it follows that

$$
\sum_{k,\ell=1}^{n} \lambda_k \overline{\lambda}_\ell \frac{1}{2} \exp(-|s_k - s_\ell|) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{\xi^2 + 1} \left| \sum_{k=1}^{n} \lambda_k \exp(i\xi s_k) \right|^2 d\xi. \tag{3.18}
$$

An easier way to establish the positive-definiteness of $\sigma_{\mathbb{R}}(u, v)$ is the following. For $\lambda_1, \dots, \lambda_n$ in $\mathbb{C}$ and for real numbers $s_1, \dots, s_n$ we write

$$
\begin{aligned}
&\sum_{k,\ell=1}^{n} \lambda_k \overline{\lambda}_\ell \exp(-|s_k - s_\ell|) \\
&= \sum_{k,\ell=1}^{n} \lambda_k \overline{\lambda}_\ell \min\left(\exp\bigl(-(s_k - s_\ell)\bigr), \exp\bigl(-(s_\ell - s_k)\bigr)\right)
\end{aligned}
$$

$$= \sum_{k,\ell=1}^{n} \exp(-s_k)\lambda_k \exp(-s_\ell)\overline{\lambda}_\ell \min\left(\exp(2s_k), \exp(2s_\ell)\right)$$

$$= \int_0^\infty \left| \sum_{k=1}^{n} \exp\left(-s_k\right) \lambda_k 1_{[0,\exp(2s_k)]}(\xi) \right|^2 d\xi \geqslant 0.$$

A similar argument can be used to prove that the function $\sigma_0(u,v)$ is positive definite. $\square$

We now give existence theorems for the Wiener process (or Brownian motion), for Brownian bridge and for the oscillator process.

3.9. THEOREM. *The following assertions are true.*

(a) *There exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *together with a real-valued Gaussian process* $\{b(s) : s \geqslant 0\}$*, called* Wiener process *or* Brownian motion*, such that* $\mathbb{E}(b(s)) = 0$ *and such that* $\mathbb{E}(b(s_1)b(s_2)) = \min(s_1, s_2)$ *for all* $s_1, s_2 \geqslant 0$.

(b) *Fix* $t > 0$. *There exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *together with a real-valued Gaussian process* $\{X_t(s) : t \geqslant s \geqslant 0\}$*, called* Brownian bridge, *such that* $\mathbb{E}(X_t(s)) = 0$ *and such that*

$$\mathbb{E}(X_t(s_1)X_t(s_2)) = \min(s_1, s_2) - \frac{s_1 s_2}{t}$$

for all $s_1, s_2 \in [0, t]$.

(c) *There exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *together with a real-valued Gaussian process* $\{q(s) : s \in \mathbb{R}\}$*, called* oscillator process*, which is centered,* i.e. $\mathbb{E}(q(s)) = 0$ *and which is such that*

$$\mathbb{E}(q(s_1)q(s_2)) = \frac{1}{2}\exp(-|s_1 - s_2|)$$

for all $s_1, s_2 \in \mathbb{R}$.

(d) *There exists a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *together with a real-valued Gaussian process* $\{X(s) : s \geqslant 0\}$*, called* Ornstein-Uhlenbeck process, *such that* $\mathbb{E}(X(s)) = 0$ *and such that*

$$\begin{aligned}\mathbb{E}(X(s_1)X(s_2)) &= \frac{1}{2}\exp\big(-(s_1+s_2)\big)\big(\exp\big(2\min(s_1,s_2)\big) - 1\big) \qquad (3.19)\\ &= \frac{1}{2}\big(\exp(-|s_1 - s_2|) - \exp\big(-(s_1+s_2)\big)\big) \quad \textit{for all } s_1,\ s_2 \geqslant 0.\end{aligned}$$

2. Brownian motion and related processes

In what follows x and y are real numbers and so is μ. Let $\{b(s) : s \geqslant 0\}$ be Brownian motion (starting in 0) on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ (*i.e.* $\mathbb{E}[b(s)] = 0$ and $\mathbb{E}[b(s_1)\, b(s_2)] = \min(s_1, s_2)$). Then the process $\{x + b(s) + \mu s : s \geqslant 0\}$ is a *Brownian motion with drift* μ *starting at* x. Let $\{X_t(s) : 0 \leqslant s \leqslant t\}$ be a Brownian bridge on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then the process $s \mapsto \left(1 - \frac{s}{t}\right)x + \frac{s}{t}y + X_t(s)$ $0 \leqslant s \leqslant t$ is called *pinned* Brownian motion, namely pinned at x at time 0 and pinned at y at time t. Let $\{b_j(s) : s \geqslant 0\}$, $1 \leqslant j \leqslant d$, be d independent Brownian motions on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The process $\{(b_1(s), \dots, b_d(s)) : s \geqslant 0\}$ is called d-dimensional Brownian motion. The characteristic function for d-dimensional Brownian motion starting at $x \in \mathbb{R}^d$ is given by:

$$\begin{aligned}\mathbb{E}_x\left[e^{-i\sum_{j=1}^n \langle X(s_j), \xi^j\rangle}\right] &= e^{-i\sum_{j=1}^n \langle \xi^j, x\rangle} e^{-\frac{1}{2}\sum_{j,k=1}^n \langle \xi^j, \xi^k\rangle \min(s_j, s_k)}\\ &= e^{-i\sum_{j=1}^n \langle \xi^j, x\rangle} e^{-\frac{1}{2}\sum_{\ell=1}^n (s_\ell - s_{\ell-1})\left|\sum_{j=\ell}^n \xi^j\right|^2}, \qquad (3.20)\end{aligned}$$

where $x_0 = x$ and where $0 = s_0 < s_1 < \cdots < s_n$. A similar definition can be given for d-dimensional Brownian bridge and for the d-dimensional oscillator process. Notice that a d-dimensional process $\{b(s) = (b_1(s), \ldots, b_d(s)) : s \geqslant 0\}$ is a d-dimensional Brownian motion, starting at 0, on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if and only if $\mathbb{E}\left(b_j(s_1), b_k(s_2)\right) = \delta_{j,k} \min(s_1, s_2)$. Let us prove the above equalities.

3.10. Theorem. *Let $0 = s_0 < s_1 < \cdots < s_n < \infty$. Fix the vectors x and $\xi_1, \ldots, \xi_n$ in $\mathbb{R}^d$. Put $s_0 = 0$ and $x_0 = x$. The following equalities are valid:*

$$\sum_{\ell=1}^{n} (s_\ell - s_{\ell-1}) \left|\sum_{j=\ell}^{n} \xi_j\right|^2 = \sum_{j,k=1}^{n} \langle \xi_j, \xi_k \rangle \min(s_j, s_k); \tag{3.21}$$

$$\int_{\mathbb{R}^d} dx_1 \ldots \int_{\mathbb{R}^d} dx_n \exp\left(-i \sum_{j=1}^{n} \langle \xi_j, x_j \rangle\right) \tag{3.22}$$

$$\prod_{j=1}^{n} \frac{1}{\left(\sqrt{2\pi (s_j - s_{j-1})}\right)^d} \exp\left(-\frac{|x_j - x_{j-1}|^2}{2(s_j - s_{j-1})}\right)$$

$$= \exp\left(-i \left\langle \sum_{j=1}^{n} \xi_j, x \right\rangle\right) \exp\left(-\frac{1}{2} \sum_{j,k=1}^{n} \langle \xi_j, \xi_k \rangle \min(s_j, s_k)\right). \tag{3.23}$$

For a and $b \in \mathbb{R}^d$ we write $(a + bi)^2 = |a|^2 + 2i \langle a, b \rangle - |b|^2$.

Proof. In order to see the first equality we write

$$\sum_{\ell=1}^{n} (s_\ell - s_{\ell-1}) \left|\sum_{j=\ell}^{n} \xi_j\right|^2 = \sum_{\ell=1}^{n} (s_\ell - s_{\ell-1}) \sum_{j_1, j_2 = \ell}^{n} \langle \xi_{j_1}, \xi_{j_2} \rangle$$

$$= \sum_{j_1, j_2 = 1}^{n} \sum_{\ell=1}^{\min(j_1, j_2)} (s_\ell - s_{\ell-1}) \langle \xi_{j_1}, \xi_{j_2} \rangle = \sum_{j_1, j_2 = 1}^{n} \left(s_{\min(j_1, j_2)} - s_0\right) \langle \xi_{j_1}, \xi_{j_2} \rangle$$

$$= \sum_{j_1, j_2 = 1}^{n} s_{\min(j_1, j_2)} \langle \xi_{j_1}, \xi_{j_2} \rangle . = \sum_{j_1, j_2 = 1}^{n} \min(s_{j_1}, s_{j_2}) \langle \xi_{j_1}, \xi_{j_2} \rangle . \tag{3.24}$$

For the second equality we proceed as follows:

$$\int_{\mathbb{R}^d} dx_1 \ldots \int_{\mathbb{R}^d} dx_n \exp\left(-i \sum_{j=1}^{n} \langle \xi_j, x_j \rangle\right)$$

$$\prod_{j=1}^{n} \frac{1}{\left(\sqrt{2\pi (s_j - s_{j-1})}\right)^d} \exp\left(-\frac{|x_j - x_{j-1}|^2}{2(s_j - s_{j-1})}\right) \tag{3.25}$$

(substitute $x_j = x + y_1 + \cdots + y_j$)

$$= \exp\left(-i \sum_{j=1}^{n} \langle \xi_j, x \rangle\right) \int_{\mathbb{R}^d} dy_1 \ldots \int_{\mathbb{R}^d} dy_n \exp\left(-i \sum_{\ell=1}^{n} \left\langle \sum_{j=\ell}^{n} \xi_j, y_\ell \right\rangle\right)$$

$$\prod_{j=1}^{n} \frac{1}{\left(\sqrt{2\pi\left(s_j - s_{j-1}\right)}\right)^d} \exp\left(-\frac{\left|y_j\right|^2}{2\left(s_j - s_{j-1}\right)}\right)$$

(substitute $y_\ell = \left(s_\ell - s_{\ell-1}\right)^{1/2} z_\ell$)

$$= \exp\left(-i\sum_{j=1}^{n}\langle \xi_j, x\rangle\right) \int_{\mathbb{R}^d} dz_1 \dots \int_{\mathbb{R}^d} dz_n$$

$$\exp\left(-i\sum_{\ell=1}^{n}\left(s_\ell - s_{\ell-1}\right)^{1/2}\sum_{j=\ell}^{n}\langle \xi_j, z_\ell\rangle\right) \prod_{j=1}^{n}\frac{1}{\left(\sqrt{2\pi}\right)^d} \exp\left(-\frac{\left|z_j\right|^2}{2}\right)$$

$$= \exp\left(-i\sum_{j=1}^{n}\langle \xi_j, x\rangle\right) \exp\left(-\frac{1}{2}\sum_{\ell=1}^{n}\left(s_\ell - s_{\ell-1}\right)\left|\sum_{j=\ell}^{n}\xi_j\right|^2\right)$$

$$\prod_{\ell=1}^{n}\frac{1}{\left(\sqrt{2\pi}\right)^d}\int_{\mathbb{R}^d} dz_\ell \exp\left(-\frac{1}{2}\left(z_\ell + i\sqrt{s_\ell - s_{\ell-1}}\sum_{j=\ell}^{n}\xi_j\right)^2\right),$$

From Cauchy's theorem, it then follows that

$$\int_{\mathbb{R}^d} dx_1 \dots \int_{\mathbb{R}^d} dx_n \exp\left(-i\sum_{j=1}^n \langle \xi_j, x_j\rangle\right) \tag{3.26}$$

$$\prod_{j=1}^n \frac{1}{\left(\sqrt{2\pi\left(s_j - s_{j-1}\right)}\right)^d} \exp\left(-\frac{\left|x_j - x_{j-1}\right|^2}{2\left(s_j - s_{j-1}\right)}\right) \tag{3.27}$$

$$= \exp\left(-i\sum_{j=1}^n \langle \xi_j, x\rangle\right) \exp\left(-\frac{1}{2}\sum_{\ell=1}^n \left(s_\ell - s_{\ell-1}\right)\left|\sum_{j=\ell}^n \xi_j\right|^2\right) \tag{3.28}$$

$$\prod_{\ell=1}^n \frac{1}{\left(\sqrt{2\pi}\right)^d} \int_{\mathbb{R}^d} dz_\ell \exp\left(-\frac{1}{2}\left|z_\ell\right|^2\right) \tag{3.29}$$

$$= \exp\left(-i\sum_{j=1}^n \langle \xi_j, x\rangle\right) \exp\left(-\frac{1}{2}\sum_{\ell=1}^n \left(s_\ell - s_{\ell-1}\right)\left|\sum_{j=\ell}^n \xi_j\right|^2\right) \tag{3.30}$$

(first equality)

$$= \exp\left(-i\sum_{j=1}^n \langle \xi_j, x\rangle\right) \exp\left(-\frac{1}{2}\sum_{j,k=1}^n \langle \xi_j, \xi_k\rangle \min\left(s_j, s_k\right)\right).$$

This completes the proof of Theorem 3.10. □

In the following proposition we collect a number of interesting properties of the (finite dimensional) joint distributions of some of the Gaussian processes we introduced so far.

3.11. Proposition. *Let $\{b(s) : s \geqslant 0\}$ be d-dimensional Brownian motion and let*

$$\{X_t(s) : 0 \leqslant s \leqslant t\}$$

be d-dimensional Brownian bridge. In addition let x and y be vectors in $\mathbb{R}^d$ and let $Q : \mathbb{R}^d \to \mathbb{R}^d$ be an orthogonal linear map. Also fix a strictly positive number a.

(a) *The joint distributions of the processes*

$$\{b(s) : s > 0\} \quad \text{and} \quad \left\{sb\left(\frac{1}{s}\right) : s > 0\right\}$$

coincide.

(b) *The joint distributions of the processes*

$$\{b(as) : s \geqslant 0\} \quad \text{and} \quad \left\{\sqrt{a}b\left(s\right) : s \geqslant 0\right\}$$

coincide.

(c) *The joint distributions of the processes*

$$\{q(s) : s \in \mathbb{R}\} \quad \text{and} \quad \left\{e^{-s}b\left(\frac{e^{2s}}{2}\right) : s \in \mathbb{R}\right\}$$

coincide.

(d) *The joint distributions of the processes*

$$\{X(t) : t \geqslant 0\} \quad \text{and} \quad \left\{e^{-t} b\left(\frac{e^{2t}-1}{2}\right) : t \geqslant 0\right\}$$

coincide. The process $\{X(t) : t \geqslant 0\}$ *also possesses the same joint distribution as* $\left\{\int_0^t \exp\left(-(t-s)\right) db(s) : t \geqslant 0\right\}$.

(e) *The joint distributions of the following processes also coincide:*

$$\left(1-\frac{s}{t}\right) x + \frac{s}{t} y + X_t(s), \quad 0 < s < t, \tag{3.31}$$

$$\left(1-\frac{s}{t}\right) x + \frac{s}{t} y + \left(1-\frac{s}{t}\right) b\left(\frac{st}{t-s}\right), \quad 0 < s < t, \tag{3.32}$$

$$\left(1-\frac{s}{t}\right) x + \frac{s}{t} y + b(s) - \frac{s}{t} b(t), \quad 0 < s < t. \tag{3.33}$$

(f) *The process* $\{Qb(s) : s \geqslant 0\}$ *is d-dimensional Brownian motion and so its joint distribution coincides with that of* $\{b(s) : s \geqslant 0\}$.

Notice that instead of the "distribution" of a random variable or a stochastic process, the name "law" is in vogue.

3.12. Remark. Put $b^x(t) = x + b(t)$. Then $\{b^x(t) : t \geqslant 0\}$ is *Brownian motion that starts in* x. Put $X^x(t) = \exp(-t)x + X(t)$. Then the process $\{X^x(t) : t \geqslant 0\}$ is the *Ornstein-Uhlenbeck process of initial velocity* x.

3.13. Remark. The stochastic integral $\int_0^t \exp(-(t-s))db(s)$ can be defined as the L^2-limit of $\sum_{j=1}^n e^{-(t-s_{j-1})}\left(b(s_j) - b(s_{j-1})\right)$, whenever $\max\limits_{1 \leqslant j \leqslant n} (s_j - s_{j-1})$ tends to zero. Here $0 = s_0 < s_1 < \cdots < s_n = t$ is a subdivision of the interval $[0, t]$.

3.14. Remark. Let $f : \mathbb{R}^d \to \mathbb{C}$ be a bounded Borel measurable function. Then $\mathbb{E}\left[f(X^x(t))\right]$ is given by

$$\mathbb{E}\left[f\left(X^x(t)\right)\right] = \int f\left(e^{-t}x + \sqrt{1-e^{-2t}}\, y\right) \frac{\exp\left(-|y|^2\right)}{\left(\sqrt{\pi}\right)^d} dy.$$

Moreover the Ornstein-Uhlenbeck process is a strong Markov process.

3.15. Remark. Let $\{b^x(t) : t \geqslant 0\}$ be Brownian motion that starts at x (and has drift zero). Fix $s > 0$. The processes

$$\{b^x(s+t) - b^x(s) : t \geqslant 0\} \quad \text{and} \quad \{b^x(t) - x : t \geqslant 0\}$$

possess the same (joint) distribution. In order to see this one may calculate the Fourier transforms, or characteristic functions, of their distributions.

3.16. Remark. Suppose that the Markov process

$$\left\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t), t \geqslant 0), (\vartheta_t, t \geqslant 0), (\mathbb{R}^n, \mathcal{B})\right\} \tag{3.34}$$

is Brownian motion in $\mathbb{R}^n$, and put $p_0(t,x,y) = \dfrac{1}{(2\pi t)^{n/2}} \exp\left(-\dfrac{|x-y|^2}{2t}\right)$, $t > 0$, x, $y \in \mathbb{R}^n$. Define the measure $\mu_{0,t}^{x,y}$ by

$$\mu_{0,x}^{t,y}(A) = \mathbb{E}_x\left[1_A p_0(t-s, X(s), y)\right], \tag{3.35}$$

where the event A belongs to $\mathcal{F}_s = \sigma(X(u) : u \leqslant s)$, for $s < t$. Since the process $s \mapsto p_0(t-s, X(s), y)$ is a $\mathbb{P}_x$-martingale on the half-open interval $0 \leqslant s < t$, it follows that the quantity $\mu_{0,x}^{t,y}(A)$ is well-defined: its value does not depend on s, as long as A belongs to $\mathcal{F}_s$ and $s < t$. From the monotone class theorem it follows that $\mu_{0,t}^{x,y}$ can be considered as a positive measure on the σ-field $\mathcal{F}_{t-}$ given by $\mathcal{F}_{t-} = \sigma(X(s) : 0 \leqslant s < t)$. Then the measure $\mu_{0,x}^{t,y}$ defined in (3.35) is called the conditional Brownian bridge measure. It can be normalized upon dividing it by the density $p_0(t, x, y)$.

Proof of Proposition 3.11. Since all the indicated processes are d-dimensional Gaussian (the definition of a d-dimensional Gaussian process should be obvious: in fact in the discussion of 3.4 and in Theorem 3.7. The expected value μ should be map from I to $\mathbb{R}^d$ and the entries of the diffusion matrix σ should be $d \times d$-matrices), it suffices to show that the corresponding expectations and covariance matrices are the same for the indicated processes. In most cases this is a simple exercise. For example let us prove (f). Let $q(k, \ell)$ be the entries of the matrix Q. Then

$$\begin{aligned}
&\mathbb{E}\left(\left(Qb(s_1)\right)_j \left(Qb(s_2)\right)_k\right) = \sum_{m=1}^{d} q(j,m) \sum_{n=1}^{d} q(k,n) \mathbb{E}\left(b_m(s_1) b_n(s_2)\right) \\
&\quad = \sum_{m=1}^{d} q(j,m) \sum_{n=1}^{d} q(k,n) \delta_{m,n} \min(s_1, s_2) = \sum_{m=1}^{d} q(j,m) q(k,m) \min(s_1, s_2) \\
&\quad = \left(QQ^*\right)(j,k) \min(s_1, s_2) = \delta_{j,k} \min(s_1, s_2). \qquad (3.36)
\end{aligned}$$

This proves that $\{Qb(s) : s \geqslant 0\}$ is again d-dimensional Brownian motion. This completes the brief outline of the proof of Proposition 3.11. $\square$

In the proof of the existence of a continuous version of Brownian motion, we shall employ the following maximal inequality of Lévy.

3.17. Theorem. *(Lévy) Let $X_1, \dots, X_n$ be random variables with values in $\mathbb{R}^d$. Suppose that the joint distribution of $X_1, \dots, X_n$ is invariant under any change of sign $(x_1, \dots, x_n) \mapsto (\epsilon_1 x_1, \dots, \epsilon_n x_n)$, where $\epsilon_j = \pm 1$. Put $S_k = \sum_{j=1}^k X_j$. Then for any $\lambda > 0$*

$$\mathbb{P}\left(\max_{1\leqslant k\leqslant n} |S_k| \geqslant \lambda\right) \leqslant 2\mathbb{P}\left(|S_n| \geqslant \lambda\right). \tag{3.37}$$

If $d = 1$, then

$$\mathbb{P}\left(\max_{1\leqslant k\leqslant n} S_k \geqslant \lambda\right) \leqslant 2\mathbb{P}\left(S_n \geqslant \lambda\right). \tag{3.38}$$

Proof. We prove (3.37). Put

$$A_k = \bigcap_{j=1}^{k-1} \{|S_j| < \lambda\} \cap \{|S_k| \geqslant \lambda\}$$

and put $A = \bigcup_{k=1}^n A_k$. Write $T_k = \sum_{j=1}^k X_j - \sum_{j=k+1}^n X_j$. Then $S_k = \frac{1}{2}S_n + \frac{1}{2}T_k$ and so

$$\{|S_k| \geqslant \lambda\} \subset \{|S_n| \geqslant \lambda\} \cup \{|T_k| \geqslant \lambda\}.$$

Hence, from the invariance of the joint distribution of $(X_1, \dots, X_n)$ under sign changes we see

$$\begin{aligned} \mathbb{P}\left(A_k\right) &= \mathbb{P}\left(A_k, |S_k| \geqslant \lambda\right) \\ &\leqslant \mathbb{P}\left(A_k, |S_n| \geqslant \lambda\right) + \mathbb{P}\left(A_k, |T_k| \geqslant \lambda\right) = 2\mathbb{P}\left(A_k, |S_n| \geqslant \lambda\right). \end{aligned}$$

Since the events A_k, $1 \leqslant k \leqslant n$, are mutually disjoint, we infer

$$\begin{aligned} &\mathbb{P}\left(\max_{1\leqslant k\leqslant n} |S_k| \geqslant \lambda\right) \\ &= \mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A_k) \leqslant 2\sum_{k=1}^n \mathbb{P}\left(A_k, |S_n| \geqslant \lambda\right) \leqslant 2\mathbb{P}\left(|S_n| \geqslant \lambda\right). \end{aligned}$$

This proves (3.37). The proof of (3.38) is similar and will be left to the reader. Altogether this completes the proof Theorem 3.17. □

Let $\{X(t) : t \geqslant 0\}$ be Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We shall prove that there exists a continuous process $\{b(t) : t \geqslant 0\}$, that is indistinguishable from the process $\{X(t) : t \geqslant 0\}$. This means that $\mathbb{P}(X(t) = b(t)) = 1$ for all $t \geqslant 0$.

3.18. Theorem. *Let $\{X(t) : t \geqslant 0\}$ be Brownian motion on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then there exists stochastic process $\{b(t) : t \geqslant 0\}$ which is $\mathbb{P}$-almost surely continuous, and that is also a Brownian motion on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and that is indistinguishable from the process $\{X(s) : s \geqslant 0\}$. Here we suppose that $\mathcal{F}$ contains the $\mathbb{P}$-zero sets.*

PROOF. Without loss of generality we may and do assume that the Brownian motion $\{X(s) : s \geqslant 0\}$ has drift 0 and diffusion matrix identity. For the proof we shall rely on Theorem 3.17 and on the Borel-Cantelli lemma, which reads as follows. Let $(A_n : n \in \mathbb{N})$ be a sequence of events with $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. Then $\mathbb{P}\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n\right) = 0$. In Theorem 3.18 we choose the sequence $(A_n : n \in \mathbb{N})$ as follows. Let D be the set of non-negative dyadic rational numbers and put

$$A_n = \left\{ \max_{0 \leqslant k < n2^n} \sup_{q \in D \cap [k2^{-n},(k+1)2^{-n}]} \left|X(q) - X(k2^{-n})\right| > \frac{1}{n} \right\}.$$

An application of Theorem 3.17, with $X(t+j\delta 2^{-m}) - X(t+(j-1)\delta 2^{-m})$ replacing X_j yields

$$\begin{aligned}
&\mathbb{P}\left(\max_{1 \leqslant j \leqslant 2^m} \left|X\left(t + j\delta 2^{-m}\right) - X(t)\right| \geqslant \alpha \right) \leqslant 2\mathbb{P}\left(\left|X(t+\delta) - X(t)\right| \geqslant \alpha\right) \\
&\leqslant \frac{2}{\alpha^4} \mathbb{E}\left|X(t+\delta) - X(t)\right|^4 = \frac{2\delta^2}{\alpha^4} \frac{1}{\left(\sqrt{2\pi}\right)^d} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}|y|^2\right) |y|^4 \, dy \\
&= \frac{2\delta^2\left(2d + d^2\right)}{\alpha^2}.
\end{aligned} \tag{3.39}$$

In (3.39) we let m tend to infinity to obtain:

$$\mathbb{P}\left(\sup_{0 \leqslant q \leqslant 1, q \in D} \left|X(t + q\delta) - X(t)\right| > \alpha \right) \leqslant \frac{2\delta^2\left(2d + d^2\right)}{\alpha^4}. \tag{3.40}$$

Hence, with $J_{n,k} = \left[k2^{-n}, (k+1)2^{-n}\right]$ (see also (3.46) below), and with $t = k2^{-n}$ and $\delta = 2^{-n}$,

$$\begin{aligned}
&\mathbb{P}\left(\max_{0 \leqslant k < n2^n} \sup_{q \in D \cap J_{n,k}} \left|X(q) - X(k2^{-n})\right| > \frac{1}{n} \right) \\
&\leqslant \sum_{k=0}^{n2^n - 1} \mathbb{P}\left(\sup_{q \in D \cap J_{n,k}} \left|X(q) - X(k2^{-n})\right| > \frac{1}{n} \right) \\
&\leqslant n2^n \frac{2\left(2d2^{-2n} + d^2 2^{-2n}\right)}{\left(n^{-1}\right)^4} = \frac{2\left(2d + d^2\right)n^5}{2^n} \leqslant \frac{6d^2 n^5}{2^n}.
\end{aligned} \tag{3.41}$$

Since the sequence in (3.41) is summable, we may apply Borel-Cantelli's lemma to conclude that $\mathbb{P}$-almost surely, for all $t > 0$, the path $q \mapsto X(q)$ is uniformly continuous on $D \cap [0,t]$. So it makes sense to define the $\mathbb{P}$-almost surely continuous function $s \mapsto b(s)$ by $b(s) = \lim_{q \to s, q \in D} X(s)$. It is not so difficult to see that the process $\{b(s) : s \geqslant 0\}$ is also a Brownian motion. In fact let $\xi_1, \ldots, \xi_n$ be n vectors in $\mathbb{R}^d$ and suppose $0 = s_0 < s_1 < \cdots < s_n$. Then we choose sequences $0 = q_0(m) < s_1 < q_1(m) < s_2 < q_2(m) < s_{n-1} < \cdots < q_{n-1}(m) < s_n < q_n(m)$, $m \in \mathbb{N}$, in D, such that $q_k(m) \downarrow s_k$, if m tends to infinity and this for $1 \leqslant k \leqslant n$. Since $\{X(s) : s \geqslant 0\}$ is d-dimensional Brownian motion we have

$$\mathbb{E}\left(\exp\left(-i \sum_{k=1}^{n} \left\langle \xi_k, X\left(q_k(m)\right)\right\rangle \right) \right)$$

$$= \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\left|\sum_{k=j}^{n}\xi_k\right|^2 \left(q_j(m)-q_{j-1}(m)\right)\right). \tag{3.42}$$

In (3.42) we let m tend to ∞ to obtain

$$\mathbb{E}\left(\exp\left(-i\sum_{k=1}^{n}\left\langle \xi_k, b\left(s_k\right)\right\rangle\right)\right) = \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\left|\sum_{k=j}^{n}\xi_k\right|^2 \left(s_j - s_{j-1}\right)\right). \tag{3.43}$$

This equality shows that $\{b(s) : s \geqslant 0\}$ is a Brownian motion. In order to prove that it cannot be distinguished from the process $\{X(s) : s \geqslant 0\}$, we notice first that

$$\mathbb{E}\left(\exp\left(-i\left\langle \xi, X(t+s) - X(t)\right\rangle\right)\right) = \exp\left(-\frac{1}{2}\left|\xi\right|^2 s\right), \quad \xi \in \mathbb{R}^d. \tag{3.44}$$

Hence, for $\xi \in \mathbb{R}^d$,

$$\begin{aligned}
&\mathbb{E}\left|\exp\left(-i\left\langle \xi, X(t)\right\rangle\right) - \exp\left(-i\left\langle \xi, b(t)\right\rangle\right)\right|^2 \\
&\quad = \mathbb{E}\left(2 - \exp\left(i\left\langle \xi, X(t) - b(t)\right\rangle\right) - \exp\left(-i\left\langle \xi, X(t) - b(t)\right\rangle\right)\right) \\
&\quad = \lim_{q\downarrow t, q\in D}\left(2 - \mathbb{E}\left(\exp\left(-i\left\langle \xi, X(q) - X(t)\right\rangle\right)\right) - \mathbb{E}\left(\exp\left(-i\left\langle \xi, X(t) - X(q)\right\rangle\right)\right)\right) \\
&\quad = 2 - 2\lim_{q\downarrow t, q\in D}\exp\left(-\frac{1}{2}\left|\xi\right|^2 (q-t)\right) = 0. \qquad (3.45)
\end{aligned}$$

From (3.45) it readily follows that the processes $\{X(s) : s \geqslant 0\}$ and $\{b(s) : s \geqslant 0\}$ cannot be distinguished. $\square$

In this proof of Theorem 3.18 we have also used the fourth moment

$$\mathbb{E}\left|X(t+s)-X(t)\right|^4.$$

From (3.44) it follows that this moment does not depend on t and hence

$$\mathbb{E}\left|X(t+s)-X(t)\right|^4 = \mathbb{E}\left|X(s)-X(0)\right|^4 = \mathbb{E}\left|X(s)\right|^4.$$

A way of computing $\mathbb{E}\left|X(s)\right|^4$ is the following:

$$\begin{aligned}
\mathbb{E}\left|X(s)\right|^4 &= \left(\sum_{j=1}^d \frac{\partial^2}{\partial \xi_j^2}\right)^2 \mathbb{E}\left(\exp\left(-i\left\langle \xi, X(s)\right\rangle\right)\right)\big|_{\xi=0} \\
&= \left(\sum_{j=1}^d \frac{\partial^2}{\partial \xi_j^2}\right)^2 \exp\left(-\frac{1}{2}\left|\xi\right|^2 s\right)\big|_{\xi=0} \\
&= \left(2ds^2 - 2s^3\left|\xi\right|^2 + s^4\left|\xi\right|^4 - 2ds^3\left|\xi\right|^2 + d^2s^2\right)\exp\left(-\frac{1}{2}\left|\xi\right|^2\right)\big|_{\xi=0} \\
&= \left(2ds^2 - 2(d+1)s^2\left|\xi\right|^2 + s^4\left|\xi\right|^4 + d^2s^2\right)\exp\left(-\frac{1}{2}\left|\xi\right|^2 s\right)\big|_{\xi=0} \\
&= 2ds^2 + d^2s^2. \qquad (3.46)
\end{aligned}$$

In the following theorem we compute the finite dimensional distributions of d-dimensional Brownian motion starting at 0 and possessing drift μ. Therefore we define the Gaussian kernel $p(t,x,y)$ by $p(t,x,y) = \dfrac{1}{(2\pi t)^{\frac{1}{2}d}}\exp\left(-\dfrac{\left|x-y\right|^2}{2t}\right)$. Notice the Chapman-Kolmogorov identity

$$p(s,x,z)p(t,z,y) = p(s+t,x,y)p\left(\frac{st}{s+t}, \frac{sx+ty}{s+t}, z\right).$$

3.19. Theorem. *Let $\{b(s) : s \geqslant 0\}$ be d-dimensional Brownian motion with diffusion matrix identity, with drift 0 and which starts in 0. Let $f_1, \ldots, f_n$ be bounded Borel measurable functions on $\mathbb{R}^d$ and let $0 = s_0 < s_1 < \cdots < s_n$. Then*

$$\begin{aligned}
&\mathbb{E}\left(\prod_{j=1}^n f_j(x + b(s_j) + \mu s_j)\right) \qquad (3.47) \\
&= \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d} dx_1\ldots dx_n \prod_{j=1}^n f_j(x_j)\prod_{j=1}^n p\left(s_j - s_{j-1}, x_{j-1} - \mu s_{j-1}, x_j - \mu s_j\right),
\end{aligned}$$

where $x_0 = x$.

3.20. Remark. Equality (3.47) determines the joint distribution of the process

$$\{X(s) := x + b(s) + \mu s : s \geqslant 0\}.$$

This will follow from the monotone class theorem. The vector μ is the so-called drift vector and the process $X = \{X(s) : s \geqslant 0\}$ starts at x in $\mathbb{R}^d$.

3.21. Remark. Another consequence of equality (3.47) is the fact that the random vector $b(t)-b(s)$, $t>s$ fixed, is independent of the σ-field generated by the process $\{b(\sigma):0\leqslant\sigma\leqslant s\}$. This fact also follows from (3.48) below together with the monotone class theorem. For $\xi\in\mathbb{R}^d$, $\xi_j\in\mathbb{R}^d$, $1\leqslant j\leqslant n$, $t>s\geqslant s_n>\cdots s_1>s_0=0$ the following identity is valid and relevant:

$$\begin{aligned}
&\mathbb{E}\left(\exp\left(-i\left\langle \xi, b(t)-b(s)\right\rangle - i\sum_{j=1}^{n}\left\langle \xi_j, b(s_j)\right\rangle\right)\right)\\
&=\exp\left(-\frac{1}{2}\left|\xi\right|^2(t-s)-\frac{1}{2}\sum_{j,k=1}^{n}\min(s_j,s_k)\left\langle \xi_j,\xi_k\right\rangle\right)\\
&=\mathbb{E}\left(\exp\left(-i\left\langle \xi, b(t)-b(s)\right\rangle\right)\right)\mathbb{E}\left(\exp\left(-i\sum_{j=1}^{n}\left\langle \xi_j, b(s_j)\right\rangle\right)\right). \qquad (3.48)
\end{aligned}$$

In other words a Brownian motion (diffusion matrix identity) is a Gaussian process $\{b(s):s\geqslant 0\}$ with independent increments $b(t)-b(s)$, $t>s$, with mean $\mu(t-s)$ and covariance matrix $\operatorname{cov}(b_k(t)-b_k(s), b_\ell(t)-b_\ell(s))=\delta_{k,\ell}(t-s)$.

Proof. Theorem 3.10 shows that the equality in (3.47) holds for functions f_j, $1\leqslant j\leqslant n$, of the form

$$f_j(x)=\int\exp\left(-i\left\langle \xi,x\right\rangle\right)d\mu_j(\xi), \qquad (3.49)$$

where $\mu_j=\delta_{\xi_j}$ is the Dirac measure ξ_j. Fubini's theorem then implies that (3.47) also holds for functions f_j, $1\leqslant j\leqslant n$, of the form (3.49) with $\mu_j(B)=\int_B g_j(\xi)\,d\xi$, with $g_j\in L^1\left(\mathbb{R}^d\right)$, $1\leqslant j\leqslant n$. Since, by the Stone-Weierstrass theorem functions of the form (3.49) with $\mu_j(B)=\int_B g_j(\xi)\,d\xi$ where $g_j\in L^1\left(\mathbb{R}^d\right)$, are dense in the space $C_0\left(\mathbb{R}^d\right)$, it follows that (3.47) holds for functions $f_j\in C_0\left(\mathbb{R}^d\right)$, $1\leqslant j\leqslant d$. By approximating indicator functions of open subsets from below by functions in $C_0\left(\mathbb{R}^d\right)$ it follows that the equality in (3.47) holds for functions f_j which are indicator functions of open subsets. A Dynkin argument (or the monotone class theorem) then shows that (3.47) is also true if the functions f_j are indicator functions of Borel subsets B_j, $1\leqslant j\leqslant n$. But then this equality also holds for bounded Borel functions f_j, $1\leqslant j\leqslant n$.

This completes the proof of Theorem 3.19. $\square$

Next we want to define standard Brownian motion, with drift vector μ, that starts at $x\in\mathbb{R}^d$.

3.22. Definition. The standard Brownian motion, starting at $x\in\mathbb{R}^d$ and with drift μ is defined as the canonical Gaussian process $\{X(s):s\geqslant 0\}$ defined on $(\Omega,\mathcal{F},\mathbb{P}_x)$ with the property that the increments $X(t+h)-X(t)$ are mutually independent and have $\mathbb{P}_x$-expectation μh. Moreover it starts $\mathbb{P}_x$-almost surely at x, i.e. $\mathbb{P}_x(X(0)=x)=1$ and $\operatorname{cov}\left(X_k(t+h)-X_k(t), X_\ell(t+h)-X_\ell(t+h)\right)=\delta_{k,\ell}h$. The covariance is of course also taken with respect to $\mathbb{P}_x$. The process is canonical because for Ω we take $\Omega=C\left([0,\infty),\mathbb{R}^d\right)$, for $X(t)$ we take $X(t)(\omega)=$

$\omega(t)$, $\omega \in \Omega$. For $\mathcal{F}$ we take the σ-field in Ω, generated by the *state variables* $\{X(s) : s \geqslant 0\}$. For all this we often write

$$\left\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), \left(\mathbb{R}^d, \mathcal{B}\right)\right\}.$$

Here the *shift* or *translation* operators ϑ_t, $t \geqslant 0$, are defined by $\vartheta_t(\omega)(s) = \omega(s+t)$, $\omega \in \Omega$. We also introduce the filtration $(\mathcal{F}_t : t \geqslant 0)$ defined as the full history: $\mathcal{F}_t$ is the σ-field generated by the variables $X(s)$, $0 \leqslant s \leqslant t$. We also shall need the right closure $\mathcal{F}_{t+}$ defined by $\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$.

In the following result we give some interesting martingale properties for Brownian motion.

3.23. Proposition. *Let*

$$\left\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), \left(\mathbb{R}^d, \mathcal{B}\right)\right\}$$

be standard Brownian motion that starts at $x \in \mathbb{R}^d$ and that has drift μ. For $t > s$ the variable $X(t) - X(s)$ does not depend on the σ-field $\mathcal{F}_s$. The following processes are $\mathbb{P}_x$-martingales with respect to the filtration $\mathcal{F}_t$, $t \geqslant 0$:

$$t \mapsto X(t) - t\mu, \qquad t \mapsto |X(t) - t\mu|^2 - dt.$$

Proof. The fact that the increment $X(t) - X(s)$ does not depend on the *past* $\mathcal{F}_s$ is explained in Remark 3.21 following Theorem 3.19. The other assertions are consequences of this. Let s and t be positive real numbers. Then we have

$$\begin{aligned}&\mathbb{E}_x\left(X(s+t) - (s+t)\mu \,\middle|\, \mathcal{F}_s\right) - (X(s) - s\mu)\\ &\quad = \mathbb{E}_x\left(X(s+t) - X(s) \,\middle|\, \mathcal{F}_s\right) - t\mu\end{aligned}$$

(increments are independent of the past)

$$= \mathbb{E}_x\left(X(s+t) - X(s)\right) - t\mu = t\mu - t\mu = 0. \tag{3.50}$$

Similarly, but more complicated, we also see

$$\begin{aligned}&\mathbb{E}_x\left[|X(s+t) - (s+t)\mu|^2 - d(s+t) - |X(s) - s\mu|^2 + ds \,\middle|\, \mathcal{F}_s\right]\\ &= \mathbb{E}_x\left[|X(s+t) - X(s) - t\mu + X(s) - s\mu|^2 - dt - |X(s) - s\mu|^2 \,\middle|\, \mathcal{F}_s\right]\\ &= \mathbb{E}_x\left[|X(s+t) - X(s) - t\mu|^2 - dt + \langle X(s+t) - X(s) - t\mu, X(s) - s\mu\rangle \,\middle|\, \mathcal{F}_s\right]\end{aligned}$$

(use (3.50))

$$= \mathbb{E}_x\left[|X(s+t) - X(s) - t\mu|^2 \,\middle|\, \mathcal{F}_s\right] - dt$$

(again an application of (3.50))

$$= \mathbb{E}_x\left[|X(s+t) - X(s) - t\mu|^2\right] - dt$$

$$= \sum_{k=1}^{d} \operatorname{cov}\left(X_k(s+t) - X_k(s), X_k(s+t) - X_k(s)\right) - dt = dt - dt = 0. \tag{3.51}$$

This proves Proposition 3.23. $\square$

So far we have looked at Brownian motion as a Gaussian process. On the other hand it is also a Markov process. We would like to discuss that now. In fact mathematically speaking equality (3.47) in Theorem 3.19 is an equivalent form of the Markov property. As already indicated in Remark 3.20 following Theorem 3.19 the monotone class theorem is important for the proofs of the several versions of the Markov property.

3.24. Definition. Let Ω be a set and let $\mathcal{S}$ be a collection of subsets of Ω. Then $\mathcal{S}$ is a *Dynkin* system if it has the following properties:

(a) $\Omega \in \mathcal{S}$;

(b) if A and B belong to $\mathcal{S}$ and if $A \supseteq B$, then $A \backslash B$ belongs to $\mathcal{S}$;

(c) if $(A_n : n \in \mathbb{N})$ is an increasing sequence of elements of $\mathcal{S}$, then the union $\bigcup_{n=1}^{\infty} A_n$ belongs to $\mathcal{S}$.

The following result on Dynkin systems is well-known.

3.25. Theorem. *Let $\mathcal{M}$ be a collection of subsets of Ω, which is stable under finite intersections. The Dynkin system generated by $\mathcal{M}$ coincides with the σ-field generated by $\mathcal{M}$.*

3.26. Theorem. *Let Ω be a set and let $\mathcal{M}$ be a collection of subsets of Ω, which is stable (or closed) under finite intersections. Let $\mathcal{H}$ be a vector space of real valued functions on Ω satisfying:*

(i) The constant function 1 belongs to $\mathcal{H}$ and 1_A belongs to $\mathcal{H}$ for all $A \in \mathcal{M}$;

(ii) if $(f_n : n \in \mathbb{N})$ is an increasing sequence of non-negative functions in $\mathcal{H}$ such that $f = \sup_{n\in\mathbb{N}} f_n$ is finite (bounded), then f belongs to $\mathcal{H}$.

Then $\mathcal{H}$ contains all real valued functions (bounded) functions on Ω, that are $\sigma(\mathcal{M})$ measurable.

Proof. Put $\mathcal{D} = \{A \subseteq \Omega : 1_A \in \mathcal{H}\}$. Then by (i) Ω belongs to $\mathcal{D}$ and $\mathcal{D} \supseteq \mathcal{M}$. If A and B are in $\mathcal{D}$ and if $B \supseteq A$, then $B\backslash A$ belongs to $\mathcal{D}$. If $(A_n : n \in \mathbb{N})$ is an increasing sequence in $\mathcal{D}$, then $1_{\cup A_n} = \sup_n 1_{A_n}$ belongs to $\mathcal{D}$ by (ii). Hence $\mathcal{D}$ is a Dynkin system, that contains $\mathcal{M}$. Since $\mathcal{M}$ is closed under finite intersection, it follows by Theorem 3.25 that $\mathcal{D} \supseteq \sigma(\mathcal{M})$. If $f \geqslant 0$ is measurable with respect to $\sigma(\mathcal{M})$, then

$$f = \sup_n 2^{-n} \sum_{j=1}^{n2^n} 1_{\{f \geqslant j2^{-n}\}}. \tag{3.52}$$

Since $1_{\{f\geqslant j2^{-n}\}}$, j, $n \in \mathbb{N}$, belong to $\sigma(\mathcal{M})$, we see that f belongs to $\mathcal{H}$. Here we employed the fact that $\sigma(\mathcal{M}) \subseteq \mathcal{D}$. If f is $\sigma(\mathcal{M})$-measurable, then we write f as a difference of two non-negative $\sigma(\mathcal{M})$-measurable functions. $\square$

The previous theorems (Theorems 3.25 and 3.26) are used in the following form. Let Ω be a set and let $(E_i, \mathcal{E}_i)_{i\in I}$ be a family of measurable spaces, indexed by an arbitrary set I. For each $i \in I$, let $\mathcal{S}_i$ denote a collection of subsets of E_i, closed under finite intersection, which generates the σ-field $\mathcal{E}_i$, and let $f_i : \Omega \to E_i$ be a map from Ω to E_i. In this context the following two propositions follow.

3.27. Proposition. *Let $\mathcal{M}$ be the collection of all sets of the form*

$$\bigcap_{i\in J} f_i^{-1}(A_i), \quad A_i \in \mathcal{S}_i,$$

$i \in J$, $J \subseteq I$, J finite. Then $\mathcal{M}$ is a collection of subsets of Ω which is stable under finite intersection and $\sigma(\mathcal{M}) = \sigma(f_i : i \in I)$.

3.28. Proposition. *Let $\mathcal{H}$ be a vector space of real-valued functions on Ω such that:*

(i) *the constant function 1 belongs to $\mathcal{H}$;*
(ii) *if $(h_n : n \in \mathbb{N})$ is an increasing sequence of non-negative functions in $\mathcal{H}$ such that $h = \sup_n h_n$ is finite (bounded), then h belongs to $\mathcal{H}$;*
(iii) *$\mathcal{H}$ contains all products of the form $\prod_{i\in J} 1_{A_i} \circ f_i$, $J \subseteq I$, J finite, and $A_i \in \mathcal{S}_i$, $i \in J$.*

Under these assumptions $\mathcal{H}$ contains all real-valued functions (bounded) functions in $\sigma(f_i : i \in I)$.

The Theorems 3.25 and 3.26 and the Propositions 3.27 and 3.28 are called the *monotone class theorem*.

In the following theorem $\mathcal{F}$ is the σ-field generated by $\{X(s) : s \geqslant 0\}$ and $\mathcal{F}_t$ is the σ-field generated by the *past* or *full history*, i.e. $\mathcal{F}_t = \sigma\{X(s) : 0 \leqslant s \leqslant t\}$. If T is an $(\mathcal{F}_{t+})$-stopping time we write

$$\mathcal{F}_{T+} = \bigcap_{t\geqslant 0} \{A \in \mathcal{F} : A \cap \{T \leqslant t\} \in \mathcal{F}_{t+}\}.$$

An $(\mathcal{F}_{t+})$-stopping is an $\mathcal{F}$-measurable map T from Ω to $[0,\infty]$ with the property that $\{T \leqslant t\}$ belongs to $\mathcal{F}_{t+}$ for all $t \geqslant 0$.

Notice that stopping times may take infinite values. Often this is very interesting.

3.29. THEOREM. *Let $\left\{(\Omega,\mathcal{F},\mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), \left(\mathbb{R}^d,\mathcal{B}\right)\right\}$, $x \in \mathbb{R}^d$, be d-dimensional Brownian motions. Then the following conditions are verified:*

- (a$_1$) *For every $\alpha > 0$, for every $t \geqslant 0$ and for every open subset U of $\mathbb{R}^d$, the set $\left\{x \in \mathbb{R}^d : \mathbb{P}_x(X(t) \in U) > \alpha\right\}$ is open;*
- (a$_2$) *For every $\alpha > 0$, for every $t \geqslant 0$ and for every compact subset K of $\mathbb{R}^d$, the set $\left\{x \in \mathbb{R}^d : \mathbb{P}_x(X(t) \in K) \geqslant \alpha\right\}$ is compact;*
- (b) *For every open subset U of $\mathbb{R}^d$ and for every $x \in U$, the equality $\lim_{t\downarrow 0} \mathbb{P}_x(X(t) \in U) = 1$ is valid.*

Moreover d-dimensional Brownian motion has the following properties:

(i) For all $t \geqslant 0$ and for all bounded random variables $Y : \Omega \to \mathbb{C}$ the equality

$$\mathbb{E}_x\left(Y \circ \vartheta_t \mid \mathcal{F}_t\right) = \mathbb{E}_{X(t)}(Y) \tag{3.53}$$

holds $\mathbb{P}_x$-almost surely for all $x \in \mathbb{R}^d$;

(ii) For all finite tuples $0 \leqslant t_1 < t_2 < \ldots < t_n < \infty$ together with Borel subsets $B_1, \ldots, B_n$ of $\mathbb{R}^d$ the equality

$$\begin{aligned}
&\mathbb{P}_x\left(X(t_1) \in B_1, \ldots, X(t_n) \in B_n\right) \\
&= \int_{B_1} \cdots \int_{B_{n-1}} \int_{B_n} P(t_n - t_{n-1}, x_{n-1}, dx_n) P(t_{n-1} - t_{n-2}, x_{n-2}, dx_{n-1}) \\
&\qquad \ldots P(t_2 - t_1, x_1, dx_2) P(t_1, x, dx_1)
\end{aligned} \tag{3.54}$$

is valid for all $x \in \mathbb{R}^d$ (here $\mathbb{P}_x(X(t) \in B) = P(t,x,B)$);

(iii) For every $(\mathcal{F}_{t+})$-stopping time T and for every bounded random variable $Y : \Omega \to \mathbb{C}$ the equality

$$\mathbb{E}_x\left(Y \circ \vartheta_T \mid \mathcal{F}_{T+}\right) = \mathbb{E}_{X(T)}(Y), \tag{3.55}$$

holds $\mathbb{P}_x$-almost surely on $\{T < \infty\}$ for all $x \in \mathbb{R}^d$;

(iv) Let $\mathcal{B}_1$ be the Borel field of $[0,\infty)$. For every bounded function $F:[0,\infty)\times\Omega\to\mathbb{C}$, which is measurable with respect to $\mathcal{B}\otimes\mathcal{F}$, and for every $(\mathcal{F}_{t+})$-stopping time T the equality

$$\mathbb{E}_x\left(\{\omega\mapsto F\left(T(\omega),\vartheta_T(\omega)\right)\}\mid\mathcal{F}_{T+}\right)=\left\{\omega'\mapsto\mathbb{E}_{X(T(\omega'))}\left\{\omega\mapsto F\left(T(\omega'),\omega\right)\right\}\right\} \tag{3.56}$$

holds $\mathbb{P}_x$-almost surely $\{T<\infty\}$ for all $x\in\mathbb{R}^d$.

Since d-dimensional Brownian motion verifies (a_1), (a_2) and (b), the properties in (i), (ii), (iii) and (iv) are all equivalent. Properties (i) and (ii) are always equivalent and also (iii) and (iv). The implication (iii) $\Rightarrow$ (ii) is also clear. For the reverse implication the full strength of (a_1), (a_2) and (b) is employed. The fact that Brownian motion possesses property (ii) is a consequence of Theorem 3.19. In fact the right continuity of paths is very important. Since we have proved that Brownian motion possesses continuous paths $\mathbb{P}_x$-almost surely this condition is verified. Property (i) is called the *Markov property* and property (iii) is called the *strong Markov property*. Equality (3.56) is called the *strong time-dependent Markov property*. We shall not prove this result. It is part of the general theory of Markov processes and their sample path properties. It is also closely connected to the theory of Feller semigroups. As in Theorem 3.29 let $\left\{(\Omega,\mathcal{F},\mathbb{P}_x),(X(t):t\geqslant 0),(\vartheta_t:t\geqslant 0),\left(\mathbb{R}^d,\mathcal{B}\right)\right\}$ be Brownian motion starting in x. In fact the family of operators $\{P(t):t\geqslant 0\}$ defined by $[P(t)f](x)=\mathbb{E}_x\left(f(X(t)\right)$, $f\in L^\infty(\mathbb{R}^d)$, $t\geqslant 0$, is a Feller semigroup, because it possesses the properties mentioned in the following definition.

In what follows E is a second countable locally compact Hausdorff space, *e.g.* $E=\mathbb{R}^d$. We define a Feller semigroup as follows.

3.30. Definition. A family $\{P(t):t\geqslant 0\}$ of operators defined on $L^\infty(E)$ is a *Feller semigroup*, or, more precisely, a *Feller-Dynkin semigroup* on $C_0(E)$ if it possesses the following properties:

(i) It leaves $C_0(E)$ invariant: $P(t)C_0(E)\subseteq C_0(E)$ for $t\geqslant 0$;
(ii) It is a semigroup: $P(s+t)=P(s)\circ P(t)$ for all s, $t\geqslant 0$, and $P(0)=I$;
(iii) It consists of contraction operators: $\|P(t)f\|_\infty\leqslant\|f\|_\infty$ for all $t\geqslant 0$ and for all $f\in C_0(E)$;
(iv) It is positivity preserving: $f\geqslant 0$, $f\in C_0(E)$, implies $P(t)f\geqslant 0$;
(v) It is continuous for $t=0$: $\lim_{t\downarrow 0}[P(t)f](x)=f(x)$, for all $f\in C_0(E)$ and for all $x\in E$.

In the presence of (iii) and (ii), property (v) is equivalent to:

(v$'$) $\lim_{t\downarrow 0}\|P(t)f-f\|_\infty=0$ for all $f\in C_0(E)$.

So that a Feller semigroup is in fact *strongly continuous* in the sense that, for every $f\in C_0(E)$,

$$\lim_{s\to t}\|P(s)f-P(t)f\|_\infty=0.$$

It is perhaps useful to observe that $C_0(E)$, equipped with the supremum-norm $\|\cdot\|_\infty$ is a Banach space (in fact it is a Banach algebra). A function $f : E \to \mathbb{C}$ belongs to $C_0(E)$ if it is continuous and if for every $\epsilon > 0$, there exists a compact subset K of E such that $|f(x)| < \epsilon$ for $x \notin K$. We need one more definition. Let $\{P(t) : t \geqslant 0\}$ be a Feller semigroup. Define for U an open subset of E, the *transition probability* $P(t, x, U)$, $t \geqslant 0$, $x \in E$, by

$$P(t, x, U) = \sup\{[P(t)u](x) : 0 \leqslant u \leqslant 1_U,\ u \in C_0(E)\}.$$

This transition function can be extended to all Borel subsets by writing

$$P(t, x, K) = \inf\{P(t, x, U) : U \ \text{ open } \ U \supseteq K\},$$

for K a compact subset of E. If B is a Borel subset of E, then we write

$$\begin{aligned} P(t, x, B) &= \inf\{P(t, x, U) : U \supseteq B, U \text{ open }\} \\ &= \sup\{P(t, x, K) : K \subseteq B, K \text{ compact }\}. \end{aligned}$$

It then follows that the mapping $B \mapsto P(t, x, B)$ is a Borel measure on $\mathcal{E}$, the Borel field of E. The Feller semigroup is said to be *conservative* if, for all $t \geqslant 0$ and for all $x \in E$, $P(t, x, E) = 1$.

We want to conclude this section with a convergence result for Gaussian processes.

3.31. Proposition. *Let* $\left(X_s^{(n)} : s \in I\right)$, $n \in \mathbb{N}$, *be a sequence of Gaussian processes. Let* $(X_s : s \in I)$ *be a process with property that*

$$\mathbb{E}\left[X_u X_v\right] = \lim_{n\to\infty} \mathbb{E}\left[X_u^{(n)} X_v^{(n)}\right], \quad \text{for all } u \text{ and } v \text{ in } I \text{ and}$$
$$\mathbb{E}\left[X_u\right] = \lim_{n\to\infty} \mathbb{E}\left[X_u^{(n)}\right], \quad \text{for all } u \in I.$$

Also suppose that, in weak sense,

$$\lim_{n\to\infty} \left(X_{u_1}^{(n)}, \dots, X_{u_m}^{(n)}\right) = \left(X_{u_1}, \dots, X_{u_m}\right)$$

for all finite subsets $(u_1, \dots, u_m)$ *of* I. *Then the process* $(X_s : s \in I)$ *is Gaussian as well.*

Proof. Let $\xi_1, \dots, \xi_m$ be real numbers, and let $u_1, \dots, u_m$ be members of I. Then

$$\begin{aligned}&\mathbb{E}\left(\exp\left(-i\sum_{k=1}^m \xi_k X_{u_k}^{(n)}\right)\right)\\ &= \exp\left(-i\sum_{k=1}^m \xi_k \mathbb{E}\left(X_{u_k}^{(n)}\right) - \frac{1}{2}\sum_{k,\ell=1}^m \xi_k\xi_\ell \operatorname{cov}\left(X_{u_k}^{(n)}, X_{u_\ell}^{(n)}\right)\right).\end{aligned}$$

Next let n tend to infinity to obtain (here we employ Lévy's theorem on weak convergence):

$$\begin{aligned}&\mathbb{E}\left(\exp\left(-i\sum_{k=1}^m \xi_k X_{u_k}\right)\right)\\ &= \exp\left\{-i\sum_{k=1}^m \xi_k \mathbb{E}\left[X_{u_k}\right] - \frac{1}{2}\sum_{k,\ell=1}^m \xi_k\xi_\ell \mathbb{E}\left[\left(X_{u_k} - \mathbb{E}\left(X_{u_k}\right)\right)\left(X_{u_\ell} - \mathbb{E}\left(X_{u_\ell}\right)\right)\right]\right\}.\end{aligned}$$

So the result in Proposition 3.31 follows. □

For Lévy's weak convergence theorem see Theorem 5.42.

3.32. Theorem. *Brownian motion is a Markov process. More precisely, (3.53) is satisfied.*

Proof. Let F be a bounded stochastic variable. We have to show the following identity:

$$\mathbb{E}_x\left[F \circ \vartheta_t \,\middle|\, \mathcal{F}_t\right] = \mathbb{E}_{X(t)}\left[F\right], \quad \mathbb{P}_x\text{-almost surely.}$$

It suffices to show that

$$\mathbb{E}_x\left[F \circ \vartheta_t \times G\right] = \mathbb{E}_x\left[\mathbb{E}_{X(t)}\left[F\right] G\right] \tag{3.57}$$

for all bounded stochastic variables F and for all bounded $\mathcal{F}_t$-measurable functions G. By an application of the monotone class theorem twice (see Proposition 3.28) it suffices to take F of the form $F = \prod_{j=1}^m f_j\left(X\left(s_j\right)\right)$, $0 \leqslant s_1 < s_2 <$

$\cdots s_m < \infty$, and G of the form $G = \prod_{j=1}^{n} g_j\left(X\left(t_j\right)\right)$, $0 \leqslant t_1 < t_2 < \cdots < t_n < t$. Here $f_1, \ldots, f_m$ and $g_1, \ldots, g_m$ are bounded continuous functions from $\mathbb{R}^d$ to $\mathbb{R}$ or $\mathbb{C}$. Once the monotone class theorem is applied to the vector space

$$\left\{G \in L^\infty\left(\Omega, \mathcal{F}_t\right) : \mathbb{E}_x\left[\mathbb{E}_{X(t)}\left[F\right] \times G\right] = \mathbb{E}_x\left[F \circ \vartheta_t \times G\right]\right\},$$

where F is as above, and once to the vector space

$$\left\{F \in L^\infty\left(\Omega, \mathcal{F}\right) : \mathbb{E}_{X(t)}\left[F\right] = \mathbb{E}_x\left[F \circ \vartheta_t \mid \mathcal{F}_t\right], \quad \mathbb{P}_x\text{-almost surely}\right\}.$$

Then (3.57) may be rewritten as

$$\begin{aligned}
&\mathbb{E}_x\left[f_1\left(X\left(s_1+t\right)\right)\cdots f_m\left(X\left(s_m+t\right)\right) g_1\left(X\left(t_1\right)\right)\cdots g_n\left(X\left(t_n\right)\right)\right] \\
&= \mathbb{E}_x\left[\mathbb{E}_{X(t)}\left[f_1\left(X\left(s_1\right)\right)\cdots f_m\left(X\left(s_m\right)\right)\right] g_1\left(X\left(t_1\right)\right)\cdots g_n\left(X\left(t_n\right)\right)\right]. \qquad (3.58)
\end{aligned}$$

Put $\tau_j = t_j$, $1 \leqslant j \leqslant n$, $\tau_{n+k} = s_k + t$, $1 \leqslant k \leqslant m$; $h_j = g_j$, $1 \leqslant j \leqslant n$, $h_{n+k} = f_k$, $1 \leqslant k \leqslant m$. By definition we have

$$\begin{aligned}
&\mathbb{E}_x\left[f_1\left(X\left(s_1+t\right)\right)\cdots f_m\left(X\left(s_m+t\right)\right) g_1\left(X\left(t_1\right)\right)\cdots g_n\left(X\left(t_n\right)\right)\right] \\
&= \mathbb{E}_x\left[h_j\left(x\left(\tau_1\right)\right)\cdots h_{n+m}\left(X\left(\tau_{n+m}\right)\right)\right] \\
&= \int \ldots \int dx_1 \ldots dx_{n+m} h_1\left(x_1\right)\cdots h_{n+m}\left(x_{n+m}\right) \\
&\qquad p\left(\tau_1, x, x_1\right)\cdots p\left(\tau_{n+m} - \tau_{n+m-1}, x_{n+m-1}, x_{n+m}\right). \qquad (3.59)
\end{aligned}$$

Next we rewrite the right-hand side of (3.58):

$$\begin{aligned}
&\mathbb{E}_x\left[\mathbb{E}_{X(t)}\left[f_1\left(X\left(s_1\right)\right)\cdots f_m\left(X\left(s_m\right)\right)\right] g_1\left(X\left(t_1\right)\right)\cdots g_n\left(X\left(t_n\right)\right)\right] \\
&= \mathbb{E}_x\Bigg[g_1\left(X\left(t_1\right)\right)\cdots g_n\left(X\left(t_n\right)\right)\int \ldots \int dy_1 \ldots dy_m f_1\left(y_1\right)\cdots f_m\left(y_m\right) \\
&\qquad p\left(s_1, X(t), y_1\right)\cdots p\left(s_m - s_{m-1}, y_{m-1}, y_m\right)\Bigg] \\
&= \int \ldots \int dz_1 \ldots dz_n g_1\left(z_1\right)\cdots g_n\left(z_n\right) \\
&\qquad p\left(t_1, x, z_1\right)\cdots p\left(t_n - t_{n-1}, z_{n-1}, z_n\right)\int dz p\left(t - t_n, z_n, z\right) \\
&\qquad \int \ldots \int dy_1 \ldots dy_m f_1\left(y_1\right)\cdots f_m\left(y_m\right) \\
&\qquad p\left(s_1, z, y_1\right)\cdots p\left(s_m - s_{m-1}, y_{m-1}, y_m\right)
\end{aligned}$$

(Chapman-Kolmogorov: $\int p\left(t - t_n, z_n, z\right) p\left(s_1, z, y\right)\, dz = p\left(s_1 + t - t_n, z_n, y\right)$)

$$\begin{aligned}
&= \int \ldots \int dz_1 \ldots dz_n g_1\left(z_1\right)\cdots g_n\left(z_n\right)\int \ldots \int dy_1 \cdots dy_m f\left(y_1\right)\cdots f_m\left(y_m\right) \\
&\qquad p\left(t_1, x, z_1\right)\cdots p\left(t_n - t_{n-1}, z_{n-1}, z_n\right) \\
&\qquad p\left(s_1 + t - t_n, z_n, y_1\right)\cdots p\left(s_m - s_{m-1}, y_{m-1}, y_m\right) \\
&= \mathbb{E}_x\left[f_1\left(X\left(s_1+t\right)\right)\cdots f_m\left(X\left(s_m+t\right)\right) g_1\left(X\left(t_1\right)\right)\cdots g_n\left(X\left(t_n\right)\right)\right]. \qquad (3.60)
\end{aligned}$$

Since the expressions in (3.59) and (3.60) are the same, this proves the Markov property of Brownian motion. The proof of Theorem 3.32 is now complete. □

3. Some results on Markov processes, on Feller semigroups and on the martingale problem

Let E be a second countable locally compact Hausdorff space, let $E^{\triangle}$ be its one-point compactification or, if E is compact, let $\triangle$ be an isolated point of $E^{\triangle} = E \bigcup \triangle$. Define the path space Ω as follows. The path space Ω is a subset of $\left(E^{\triangle}\right)^{[0,\infty)}$ with the following properties:

(i) If ω belongs to Ω, if $t \geqslant 0$ is such that $\omega(t) = \triangle$ and if $s \geqslant t$, then $\omega(s) = \triangle$;
(ii) Put $\zeta(\omega) = \inf\{s > 0 : \omega(s) = \triangle\}$ for $\omega \in \Omega$. If ω belongs to Ω, then ω possesses left limits in $E^{\triangle}$ on the interval $[0, \zeta]$ and it is right-continuous on $[0, \infty)$;
(iii) If ω belongs to Ω, if $t \geqslant 0$ is such that $\omega(t)$ belongs to E, then the closure of the set $\{\omega(s) : 0 \leqslant s \leqslant t\}$ is a compact subset of E or, equivalently, if $t > 0$ is such that $\omega(t-) = \triangle$ and if $s \geqslant t$, then $\omega(s) = \triangle$.

3.33. Definition. The random variable ζ, defined in (iii.) is called the *life time* of ω. A path $\omega \in \Omega$ is said to be *cadlag* on its life time. We also define the *state variables* $X(t) : \Omega \to E^{\triangle}$ by $X(t)(\omega) = X(t, \omega) = \omega(t)$, $t \geqslant 0$, $\omega \in \Omega$. The *translation* or *shift operators* are defined in the following way: $[\vartheta_t(\omega)](s) = \omega(s + t)$, s, $t > 0$ and $\omega \in \Omega$. The largest subset of $\left(E^{\triangle}\right)^{[0,\infty)}$ with the properties (i), (ii) and (iii) is sometimes written as $D\left([0, \infty), E^{\triangle}\right)$ or as $D_{E^{\triangle}}([0, \infty))$. Let $\mathcal{F}$ be a σ-field on Ω. A function $Y : \Omega \to \mathbb{C}$ is called a *random variable* if it is measurable with respect to $\mathcal{F}$. Of course $\mathbb{C}$ is supplied with its Borel field. The so-called *state space* E is also equipped with its Borel field $\mathcal{E}$ and $E^{\triangle}$ is also equipped with its Borel field $\mathcal{E}^{\triangle}$. The path $\omega_{\triangle}$ is given by $\omega_{\triangle}(s) = \triangle$, $s \geqslant 0$. Unless specified otherwise we write $\Omega = D([0, \infty), E^{\triangle})$. The space $D([0, \infty), E^{\triangle})$ is also called *Skorohod space*. In addition let $\mathcal{F}$ be a σ-field on Ω and let $\{\mathcal{F}_t : t \geqslant 0\}$ be a filtration on Ω. Suppose $\mathcal{F}_t \subseteq \mathcal{F}$, $t \geqslant 0$, and suppose that every state variable $X(t)$, $t \geqslant 0$, is measurable with respect to $\mathcal{F}_t$. (This is the case where e.g. $\mathcal{F}_t$ is the σ-field generated by $\{X(s) : s \leqslant t\}$.)

We also want to make a digression to operator theory. Let L be a linear operator with domain $D(L)$ and range $R(L)$ contained in $C_0(E)$. The operator L is said to be *closable* if the closure of its graph is again the graph of an operator. Here the *graph* of L, $G(L)$, is defined by $G(L) = \{(f, Lf) : f \in D(L)\}$. Its *closure* is the closure of $G(L)$ in the cartesian product $C_0(E) \times C_0(E)$. If the closure of $G(L)$ is the graph of an operator, then this operator is, by definition, the *closure* of L. It is written as $\overline{L}$. Sometimes $\overline{L}$ is called the *smallest closure* of L.

3.34. Definition. Let L be a linear operator with domain and range in $C_0(E)$.

(i) The operator L is said to be *dissipative* if, for all $\lambda > 0$ and for all $f \in D(L)$,
$$\|\lambda f - Lf\|_\infty \geqslant \lambda \|f\|_\infty . \tag{3.61}$$
(ii) The operator L is said to verify the *maximum principle* if for every $f \in D(L)$ with $\sup\{\mathrm{Re}\, f(x) : x \in E\}$ strictly positive, there exists $x_0 \in E$ with the property that
$$\mathrm{Re}\, f(x_0) = \sup\{\mathrm{Re}\, f(x) : x \in E\} \quad \text{and} \quad \mathrm{Re}\, Lf(x_0) \leqslant 0.$$
(iii) The martingale problem is said to be *uniquely solvable*, or *well-posed*, for the operator L, if for every $x \in E$ there exists a unique probability $\mathbb{P} = \mathbb{P}_x$ which satisfies:
 (a) For every $f \in D(L)$ the process
$$f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds, \quad t \geqslant 0,$$
 is a $\mathbb{P}$-martingale;
 (b) $\mathbb{P}(X(0) = x) = 1$.

(iv) The operator L is said to solve the *martingale problem maximally* if for L the martingale problem is uniquely solvable and if its closure $\overline{L}$ is maximal for this property. This means that, if L_1 is any linear operator with domain and range in $C_0(E)$, which extends $\overline{L}$ and for which the martingale problem is uniquely solvable, then $\overline{L}$ coincides with L_1.
(v) The operator L is said to be the (infinitesimal) *generator* of a Feller semigroup
$$\{P(t) : t \geqslant 0\},$$
if $L = \text{s-}\lim_{t\downarrow 0} \dfrac{P(t) - I}{t}$. This means that a function f belongs to $D(L)$ whenever $Lf := \lim_{t\downarrow 0} \dfrac{P(t)f - f}{t}$ exists in $C_0(E)$.

An operator which verifies the maximum principle is dissipative (see e.g. [**141**], p. 14) and can be considered as kind of a generalized second order derivative operator. A prototype of such an operator is the Laplace operator. An operator for which the martingale problem is uniquely solvable is closable. This follows from (3.125) below. Our main result says that linear operators in $C_0(E)$ which maximally solve the martingale problem are generators of Feller semigroups and conversely.

3.35. Definition. Next suppose that, for every $x \in E$, a probability measure $\mathbb{P}_x$ on $\mathcal{F}$ is given. Suppose that for every bounded random variable $Y : \Omega \to \mathbb{R}$ the equality $\mathbb{E}_x \left(Y \circ \vartheta_t \mid \mathcal{F}_t \right) = \mathbb{E}_{X(t)}(Y)$ holds P_x-almost surely for all $x \in E$ and for all $t \geqslant 0$. Then the process

$$\left\{ (\Omega, \mathcal{F}, \mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E}) \right\}$$

is called a *Markov process*. If the fixed time t may be replaced with a stopping time T, the process $\left\{ (\Omega, \mathcal{F}, \mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E}) \right\}$ is called a *strong Markov process*. By definition $\mathbb{P}_\triangle(A) = 1_A(\omega_\triangle) = \delta_{\omega_\triangle}(A)$. Here A belongs to $\mathcal{F}$. If the process $\left\{ (\Omega, \mathcal{F}, \mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E}) \right\}$ is a Markov process, then we write

$$P(t, x, B) = \mathbb{P}_x(X(t) \in B), \quad t \geqslant 0, \quad B \in \mathcal{E}, \quad x \in E, \tag{3.62}$$

for the corresponding *transition function*. The operator family $\{P(t) : t \geqslant 0\}$ is defined by $[P(t)f](x) = \mathbb{E}_x \left(f(X(t)) \right)$, $f \in C_0(E)$.

An relevant book on Markov processes is Ethier and Kurtz [**54**]. An elementary theory of diffusions is given in Durrett [**45**]. In this aspect the books of Stroock and Varadhan [**133**], Stroock [**132**] [**131**], and Ikeda and Watanabe [**61**] are of interest as well.

We shall mainly be interested in the case that the function $P(t)f$ is a member of $C_0(E)$ whenever f is so. In the following theorem $\mathcal{F}$ is the σ-field generated by $\{X(s) : s \geqslant 0\}$ and $\mathcal{F}_t$ is the σ-field generated by the *past* or *full history*, i.e. $\mathcal{F}_t = \sigma \{X(s) : 0 \leqslant s \leqslant t\}$. If T is a stopping time we write $\mathcal{F}_{T+} = \bigcap_{t \geqslant 0} \{A \in \mathcal{F} : A \cap \{T \leqslant t\} \in \mathcal{F}_{t+}\}$.

3.36. Theorem. *Let $(\Omega, \mathcal{F}, \mathbb{P}_x)$, $x \in E$, be probability spaces with the following properties:*

- (a_1) *For every $\alpha > 0$, for every $t \geqslant 0$ and for every open subset U of E, the set $\{x \in E : \mathbb{P}_x(X(t) \in U) > \alpha\}$ is open;*
- (a_2) *For every $\alpha > 0$, for every $t \geqslant 0$ and for every compact subset K of E, the set $\{x \in E : \mathbb{P}_x(X(t) \in K) \geqslant \alpha\}$ is compact;*
- (c) *For every open subset U of E and for every $x \in U$, the equality $\lim_{t \downarrow 0} \mathbb{P}_x \left(X(t) \in U \right) = 1$ is valid.*

The following assertions are equivalent:

(i) For all $t \geqslant 0$ and for all bounded random variables $Y : \Omega \to \mathbb{C}$ the equality

$$\mathbb{E}_x \left(Y \circ \vartheta_t \mid \mathcal{F}_t \right) = \mathbb{E}_{X(t)}(Y) \tag{3.63}$$

holds $\mathbb{P}_x$-almost surely for all $x \in E$;

(ii) For all finite tuples $0 \leqslant t_1 < t_2 < \ldots < t_n < \infty$ together with Borel subsets $B_1, \ldots, B_n$ of E the equality

$$\begin{aligned} &\mathbb{P}_x \left(X(t_1) \in B_1, \ldots, X(t_n) \in B_n \right) \\ &= \int_{B_1} \ldots \int_{B_{n-1}} \int_{B_n} P(t_n - t_{n-1}, x_{n-1}, dx_n) P(t_{n-1} - t_{n-2}, x_{n-2}, dx_{n-1}) \end{aligned}$$

$$\ldots P(t_2 - t_1, x_1, dx_2) P(t_1, x, dx_1) \tag{3.64}$$

is valid for all $x \in E$ *(here* $\mathbb{P}_x(X(t) \in B) = P(t, x, B)$*);*

(iii) For every $(\mathcal{F}_{t+})$*-stopping time* T *and for every bounded random variable* $Y : \Omega \to \mathbb{C}$ *the equality*

$$\mathbb{E}_x \left(Y \circ \vartheta_T \mid \mathcal{F}_{T+} \right) = \mathbb{E}_{X(T)} \left(Y \right), \tag{3.65}$$

holds $\mathbb{P}_x$*-almost surely on* $\{T < \infty\}$ *for all* $x \in E$*;*

(iv) Let $\mathcal{B}$ *be the Borel field of* $[0, \infty)$*. For every bounded function* $F : [0, \infty) \times \Omega \to \mathbb{C}$*, which is measurable with respect to* $\mathcal{B} \otimes \mathcal{F}$*, and for every* $(\mathcal{F}_{t+})$*-stopping time* T *the equality*

$$\mathbb{E}_x \left(\{\omega \mapsto F\left(T(\omega), \vartheta_T(\omega)\right)\} \mid \mathcal{F}_{T+} \right) = \left\{\omega' \mapsto \mathbb{E}_{X(T(\omega'))} \left\{\omega \mapsto F\left(T(\omega'), \omega\right)\right\}\right\} \tag{3.66}$$

holds $\mathbb{P}_x$*-almost surely* $\{T < \infty\}$ *for all* $x \in E$*.*

Equality (3.66) is called the *strong time-dependent Markov property*. We shall not prove this result.

3.37. Theorem. *Let $\{P(t) : t \geqslant 0\}$ be a Feller semigroup. There exists a collection of probabilities $(\mathbb{P}_x)_{x\in E}$ on the σ-field $\mathcal{F}$ generated by the state variables $\{X(t) : t \geqslant 0\}$ defined on $\Omega := D\left([0,\infty), E^{\triangle}\right)$ in such a way that*

$$\begin{aligned}
&\mathbb{E}_x\left[f\left(X(t_1),\dots,X(t_n)\right)\right] = \int f\left(X(t_1),\dots,X(t_n)\right) d\mathbb{P}_x \\
&= \int \dots \int f(x_1,x_2,\dots,x_n) P(t_n - t_{n-1}, x_{n-1}, dx_n) P(t_{n-1} - t_{n-2}, x_{n-2}, dx_{n-1}) \\
&\qquad \dots P(t_2 - t_1, x_1, dx_2) P(t_1, x, dx_1), \qquad (3.67)
\end{aligned}$$

where f is any bounded complex or non-negative Borel measurable function defined on $E^{\triangle} \times \dots \times E^{\triangle}$, that vanishes outside of $E \times \dots \times E$. Let the measure spaces $(\Omega, \mathcal{F}, \mathbb{P}_x)_{x\in E}$ be as in (3.67). The process

$$\left\{(\Omega,\mathcal{F},\mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E})\right\}$$

is a strong Markov process.

The proof of this result is quite technical. The first part follows from a well-known theorem of Kolmogorov on projective systems of measures: see Theorems 1.14, 3.1, 5.81. In the second part we must show that the indicated path space has full measure, so that no information is lost. Proofs are omitted. They can be found in for example Blumenthal and Getoor [**20**], Theorem 9.4. p. 46. For a discussion in the context of Polish spaces see, *e.g.*, Sharpe [**120**] or Van Casteren [**146**]. For the convenience of the reader we include an outline of the proof of Theorem 3.37. The following lemma is needed in the proof.

3.38. Lemma. *Let $(\Omega, \mathcal{F}_t, \mathbb{P})$ be a probability space, and let $t \mapsto Y(t)$, be a supermartingale, which attains positive values. Fix $t > 0$ and let D be a dense countable subset of $[0,\infty)$. Then*

$$\mathbb{P}\left[Y(t) > 0, \inf_{0<s<t,\, s\in D} Y(s) = 0\right] = 0. \qquad (3.68)$$

Proof. Let $(s_j)_j$ be an enumeration of the set $D \cap [0,\infty)$. Fix n, $N \in \mathbb{N}$, and define the stopping time $S_{n,N}$ by

$$S_{n,N} = \min\left\{s_j : \min_{1\leqslant j\leqslant N} Y(s_j) < 2^{-n}\right\}.$$

Then we have $Y(S_{n,N}) < 2^{-n}$ on the event $\{S_{n,N} < \infty\}$. In addition, by the supermartingale property (for discrete stopping times) we infer

$$\begin{aligned}
\mathbb{E}\left[Y(t), \min_{1\leqslant j\leqslant N} Y(s_j) < 2^{-n}\right] &\leqslant \mathbb{E}\left[Y(t),\ S_{n,N} < t\right] \\
&= \mathbb{E}\left[Y(t),\ \min(S_{n,N}, t) < t\right] \\
&\leqslant \mathbb{E}\left[Y(\min(S_{n,N}, t)),\ \min(S_{n,N}, t) < t\right] \\
&= \mathbb{E}\left[Y(S_{n,N}),\ S_{n,N} < t\right] \\
&\leqslant \mathbb{E}\left[2^{-n},\ S_{n,N} < t\right] \leqslant 2^{-n}. \qquad (3.69)
\end{aligned}$$

In (3.69) we let $N \to \infty$ to obtain:

$$\mathbb{E}\left[Y(t),\ \inf_{0<s<t,\, s\in D} Y(s) < 2^{-n}\right] \leqslant 2^{-n}. \tag{3.70}$$

In (3.70) we let $n \to \infty$ to get:

$$\mathbb{E}\left[Y(t),\ \inf_{0<s<t,\, s\in D} Y(s) = 0\right] = 0. \tag{3.71}$$

Let $\alpha > 0$ be arbitrary. From (3.71) it follows that

$$\mathbb{P}\left[Y(t) > \alpha,\ \inf_{0<s<t,\, s\in D} Y(s) = 0\right] \leqslant \frac{1}{\alpha}\mathbb{E}\left[Y(t),\ \inf_{0<s<t,\, s\in D} Y(s) = 0\right] = 0. \tag{3.72}$$

Then in (3.72) we let $\alpha \downarrow 0$ to complete the proof of Lemma 3.38. $\square$

Let $\{P(t) : t \geqslant 0\}$ be a Feller-Dynkin semigroup acting on $C_0(E)$ where E is a locally compact Hausdorff space. In the proof of Theorem 3.37 we will also use the resolvent operators $\{R(\alpha) : \alpha > 0\}$: $R(\alpha)f(x) = \int_0^\infty e^{-\alpha t}P(t)f(x)\,dt$, $\alpha > 0$, $f \in C_0(E)$. An important property is the resolvent equation:

$$R(\beta) - R(\alpha) = (\alpha - \beta)\, R(\alpha)R(\beta), \quad \alpha,\, \beta > 0.$$

The latter property is a consequence of the semigroup property.

3.39. Remark. The space E is supposed to be a second countable (*i.e.* it is a topological space with a countable base for its topology) locally compact Hausdorff space (in particular it is a Polish space). A second-countable locally-compact Hausdorff space is Polish. Let $(U_i)_i$ be a countable basis of open subsets with compact closures, choose for each $i \in \mathbb{N}$, $y_i \in U_i$, together with a continuous function $f_i : E \to [0,1]$ such that $f_i(y_i) = 1$ and such that $f_i(y) = 0$ for $y \notin U_i$. Since a locally compact Hausdorff space is completely regular this choice is possible. Put

$$d(x,y) = \sum_{i=1}^{\infty} 2^{-i}\left|f_i(x) - f_i(y)\right| + \left|\frac{1}{\sum_{i=1}^{\infty} 2^{-i} f_i(x)} - \frac{1}{\sum_{i=1}^{\infty} 2^{-i} f_i(y)}\right|, \quad x,\, y \in E.$$

This metric gives the same topology, and it is not too difficult to verify its completeness. For this notice that the sequence $(f_i)_i$ separates the points of E, and therefore the algebraic span (*i.e.* the linear span of the finite products of the functions f_i) is dense in $C_0(E)$ for the topology of uniform convergence. A proof of the fact that a locally compact space is completely regular can be found in Willard [**152**] Theorem 19.3. The connection with Urysohn's metrization theorem is also explained. A related construction can be found in Garrett [**57**]: see Dixmier [**39**] Appendix V as well.

3.40. Remark. Next we present the notion of Skorohod space. Let $D([0,1],\mathbb{R})$ be the space of real-valued functions ω defined on the interval $[0,1]$ that are right-continuous and have left-hand limits, *i.e.*, $\omega(t) = \omega(t+) = \lim_{s\downarrow t}\omega(s)$ for all $0 \leqslant t < 1$, and $\omega(t-) = \lim_{s\uparrow t}\omega(s)$ exists for all $0 < t \leqslant 1$. (In probabilistic literature, such a function is also said to be a cadlag function, "cadlag" being an acronym for the French "continu à droite, limites à gauche".) The supremum norm on $D([0,1],\mathbb{R})$, given by

$$\|\omega\|_\infty = \sup_{t\in[0,1]} |\omega(t)|, \quad \omega \in D([0,1],\mathbb{R}),$$

turns the space $D([0,1],\mathbb{R})$ into a Banach space which is non-separable. This non-separability causes well-known problems of measurability in the theory of weak convergence of measures on the space. To overcome this inconvenience, A.V. Skorohod introduced a metric (and topology) under which the space becomes a separable metric space. Although the original metric introduced by Skorohod has a drawback in the sense that the metric space obtained is not complete, it turned out (see Kolmogorov [**70**]) that it is possible to construct an equivalent metric (*i.e.*, giving the same topology) under which the space $D([0,1],\mathbb{R})$ becomes a separable and complete metric space. Such metric space the term Polish space is often used. This metric is defined as follows, and taken from Paulauskas in [**110**]. Let Λ denote the class of strictly increasing continuous mappings of $[0,1]$ onto itself. For $\lambda \in \Lambda$, let

$$\|\lambda\| = \sup_{0\leqslant s<t\leqslant 1} \left|\log \frac{\lambda(t)-\lambda(s)}{t-s}\right|.$$

Then for ω_1 and $\omega_2 \in D([0,1],\mathbb{R})$ we define

$$d(\omega_1,\omega_2) = \inf_{\lambda\in\Lambda} \max\left(\|\lambda\|, \|\omega_1-\omega_2\circ\gamma\|_\infty\right).$$

The topology generated by this metric is called the Skorohod topology and the complete separable metric space $D([0,1],\mathbb{R})$ is called the Skorohod space. This space is very important in the theory of stochastic processes. The general theory of weak convergence of probability measures on metric spaces and, in particular, on the space $D([0,1],\mathbb{R})$ is well developed. This theory was started in the fundamental papers like Chentsov [**33**], Kolmogorov [**70**], Prohorov [**112**], Skorohod [**122**]. A well-known reference on these topics is Billingsley [**17**]. Generalizations of the Skorohod space are worth mentioning. Instead of real-valued functions on $[0,1]$ it is possible to consider functions defined on $[0,\infty)$ and taking values in a metric space E. The space of cadlag functions obtained in this way is denoted by $D([0,\infty),E)$ and if E is a Polish space, then $D([0,\infty),E)$, with the appropriate topology, is also a Polish space, see Ethier and Kurtz [**54**] and Pollard [**111**], where these spaces are treated systematically.

Outline of a proof of Theorem 3.37. Firstly, the Riesz representation theorem, applied to the functionals $f \mapsto P(t)f(x)$, $f \in C_0(E)$, $(t,x) \in [0,\infty)\times E$, provides a family of sub-probability measures $B \mapsto P(t,x,B)$, $B \in \mathcal{E}$, $(t,x)\in[0,\infty)\times E$, with $P(0,x,B) = \delta_x(B) = 1_B(x)$. From the semigroup property, *i.e.* $P(s+t) = P(s)P(t)$, s, $t \geqslant 0$, it follows that the family $\{P(t,x,\cdot) : (t,x)\in[0,\infty)\}$ obeys the Chapman-Kolmogorov identity:

$$P(s+t,x,B) = \int P(t,y,B)\,P(s,x,dy), \quad B\in\mathcal{E},\ s\geqslant 0,\ t\geqslant 0,\ x\in E. \tag{3.73}$$

The measures $B \mapsto P(t,x,B)$, $B \in \mathcal{E}$, are inner and outer regular in the sense that, for all Borel subsets B (*i.e.* $B \in \mathcal{E}$),

$$\begin{aligned} P(t,x,B) &= \sup \{P(t,x,K) : K \subset B,\ K \text{ compact}\} \\ &= \inf \{P(t,x,O) : O \supset B,\ O \text{ open}\}. \end{aligned} \tag{3.74}$$

In general we have $0 \leqslant P(t,x,B) \leqslant 1$, $(t,x,B) \in [0,\infty) \times E \times \mathcal{E}$. In order to apply Kolmogorov's extension theorem we need that, for every $x \in E$, the function $t \mapsto P(t,x,E)$ is constant. Since $P(0,x,E) = 1$ this constant must be 1. This can be achieved by adding an absorption point $\triangle$ to E. So instead of E we consider the state space $E^{\triangle} = E \cup \{\triangle\}$, which, topologically speaking, can be considered as the one-point compactification of E, if E is not compact. If E is compact, $\triangle$ is an isolated point of $E^{\triangle}$. Let $\mathcal{E}^{\triangle}$ be the Borel field of $E^{\triangle}$. Then the new family of probability measures $\{N(t,x,\cdot) : (t,x) \in [0,\infty) \times E\}$ is defined as follows:

$$\begin{aligned} &N(t,x,B) = P(t,x,B \cap E) + (1 - P(t,x,E))\, 1_B(\triangle), \quad (t,x) \in [0,\infty) \times E, \\ &N(t,\triangle,B) = 1_B(\triangle), \quad t \geqslant 0, \ B \in \mathcal{E}^{\triangle}. \end{aligned} \tag{3.75}$$

Compare this construction with the one in (1.1). The family

$$\mathcal{N} := \{N(t,x,\cdot) : (t,x) \in [0,\infty) \times E\}$$

again satisfies the Chapman-Kolmogorov identity with state space $E^{\triangle}$ instead of E. Notice that $N(t,x,B) = P(t,x,B)$ whenever (t,x,B) belongs to $[0,\infty)\times E\times \mathcal{E}$. Employing the family $\mathcal{N}$ we define a family of probability spaces as follows. For every $x_0 \in E$, and every increasing n-tuple $0 \leqslant t_1 < \cdots < t_n < \infty$ in $[0,)^n$ we consider the probability space $\left(\left(E^{\triangle}\right)^n, \otimes^n \mathcal{E}^{\triangle}, P_{x_0,t_1,\dots,t_n}\right)$: the probability measure $P_{x_0,t_1,\dots,t_n}$ is defined by

$$P_{x_0,t_1,\dots,t_n}(B) = \int\limits_B\cdots\int N(t_1-t_0,x_0,dx_1)\cdots N(t_n-t_{n-1},x_{n-1},dx_n), \tag{3.76}$$

where $B \in \otimes^n\mathcal{E}^{\triangle}$. By an appeal to the Chapman-Kolmogorov identity it follows that, for $x_0 \in E$ fixed, the family of probability spaces:

$$\left\{\left(\left(E^{\triangle}\right)^n, \otimes^n \mathcal{E}^{\triangle}, P_{x_0,t_1,\dots,t_n}\right) : 0 \leqslant t_1 < \cdots < t_n < \infty,\, n \in \mathbb{N}\right\} \tag{3.77}$$

is a projective system of probability spaces. Put $\widetilde{\Omega}^{\triangle} = \left(E^{\triangle}\right)^{[0,\infty)}$, and equip this space with the product σ-field $\widetilde{\mathcal{F}}^{\triangle} := \otimes^{[0,\infty)}\mathcal{E}^{\triangle}$. In addition, write $\widetilde{X}(t)(\omega) = \omega(t)$, $\vartheta_t\omega(s) = \omega(s+t)$, s, $t \geqslant 0$, $\omega \in \widetilde{\Omega}^{\triangle}$. The variables $\widetilde{X}(t)$, $t \geqslant 0$, are called the state variables, and the mappings ϑ_t, $t \geqslant 0$, are called the (time) translation or shift operators. By Kolmogorov's extension theorem there exists, for every $x \in E^{\triangle}$, a probability measure $\widetilde{\mathbb{P}}_x$ on the σ-field $\widetilde{\mathcal{F}}^{\triangle}$ such that

$$\widetilde{\mathbb{P}}_x\left[\left(\widetilde{X}(t_1),\cdots,\widetilde{X}(t_n)\right) \in B\right] = P_{x,t_1,\dots,t_n}[B]. \tag{3.78}$$

In (3.78) B belongs to $\otimes^n\mathcal{E}^{\triangle}$, and $0 \leqslant t_1 < \cdots < t_n < \infty$ is an arbitrary increasing n-tuple in $[0,\infty)$. Another appeal to the Chapman-Kolmogorov identity and the monotone class theorem shows that the quadruple

$$\left\{\left(\widetilde{\Omega}^{\triangle}, \widetilde{\mathcal{F}}^{\triangle}, \widetilde{\mathbb{P}}_x\right), \left(\widetilde{X}(t), t \geqslant 0\right), (\vartheta_t, t \geqslant 0), \left(E^{\triangle}, \mathcal{E}^{\triangle}\right)\right\} \tag{3.79}$$

is a Markov process relative to the internal history, *i.e.* relative to the filtration $\widetilde{\mathcal{F}}_t^{\triangle} = \sigma\left(\widetilde{X}(s) : 0 \leqslant s \leqslant t\right)$, $t \geqslant 0$. Moreover, we have $\widetilde{\mathbb{P}}_x\left[\widetilde{X}(0) = x\right] = 1$, and, by the Markov property, we also have, for $x \in E^{\triangle}$, $t > s \geqslant 0$,

$$\widetilde{\mathbb{P}}_x\left[\widetilde{X}(t) = \triangle,\, \widetilde{X}(s) = \triangle\right] = \widetilde{\mathbb{E}}_x\left[\widetilde{\mathbb{P}}_x\left[\widetilde{X}(t-s)\circ\vartheta_s = \triangle \mid \widetilde{\mathcal{F}}^{\triangle}\right], \widetilde{X}(s) = \triangle\right]$$

(Markov property)

$$\begin{aligned}
&= \widetilde{\mathbb{E}}_x\left[\widetilde{\mathbb{P}}_{\widetilde{X}(s)}\left[X(t-s) = \triangle\right], \widetilde{X}(s) = \triangle\right] \\
&= \widetilde{\mathbb{E}}_x\left[\widetilde{\mathbb{P}}_{\triangle}\left[\widetilde{X}(t-s) = \triangle\right], \widetilde{X}(s) = \triangle\right] \\
&= N(t-s,\triangle,\{\triangle\})\cdot N(s,x,\{\triangle\}) \\
&= N(s,x,\{\triangle\}) = \widetilde{\mathbb{P}}_x\left[\widetilde{X}(s) = \triangle\right].
\end{aligned} \tag{3.80}$$

The equality in (3.80) says that once the process $t \mapsto \widetilde{X}(t)$ enters $\triangle$ it stays there. In other words $\triangle$ is an absorption point for the process $\widetilde{X}$. Define, for $t >$

0, the mapping $\widetilde{P}(t) : C\left(E^{\triangle}\right) \to C\left(E^{\triangle}\right)$ by $\widetilde{P}(t)f(x) = \widetilde{\mathbb{E}}_x\left[f\left(\widetilde{X}(t)\right)\right]$, $f \in C\left(E^{\triangle}\right)$. From the Markov property of the process $\widetilde{X}$, and since the semigroup $t \mapsto P(t)$ is a Feller-Dynkin semigroup, it follows that the mappings $\widetilde{P}(t)$, $t \geqslant 0$, constitute a Feller (or Feller-Dynkin semigroup) on $C\left(E^{\triangle}\right)$. Consequently, for any $f \in C\left(E^{\triangle}\right)$, and any $t_0 \geqslant 0$, we have

$$\lim_{s,t \to t_0,\, s,\, t \geqslant 0} \sup_{x \in E^{\triangle}} \widetilde{\mathbb{E}}_x\left[\left|f\left(\widetilde{X}(t)\right) - f\left(\widetilde{X}(s)\right)\right|\right] = 0. \tag{3.81}$$

Let D be the collection of non-negative dyadic rational numbers. Since the space $E^{\triangle}$ is compact-metrizable, it follows from (3.81) that, for all $x \in E^{\triangle}$, the following limits

$$\lim_{s \uparrow t,\, s \in D} \widetilde{X}(s), \quad \text{and} \quad \lim_{s \downarrow t,\, s \in D} \widetilde{X}(s), \tag{3.82}$$

exist in $E^{\triangle}$ $\widetilde{\mathbb{P}}_x$-almost surely. Define the mapping $\pi : \widetilde{\Omega}^{\triangle} \to \Omega$ by

$$\pi(\omega)(t) = \lim_{s \downarrow t,\, s \in D} \widetilde{X}(s)\,(\omega) =: X(t)\,(\pi(\omega))\,, \quad t \geqslant 0, \quad \omega \in \widetilde{\Omega}^{\triangle}. \tag{3.83}$$

Then we have that, for every $x \in E^{\triangle}$ fixed, the processes $t \mapsto X(t) \circ \pi$ and $t \mapsto \widetilde{X}(t)$ are $\widetilde{\mathbb{P}}_x$-indistinguishable in the sense that there exists an event $\widetilde{\Omega}^{\triangle,\prime} \subset \widetilde{\Omega}^{\triangle}$ such that $\widetilde{P}_x\left[\widetilde{\Omega}^{\triangle,\prime}\right] = 1$ and such that for all $t \in [0,\infty)$ the equality $\widetilde{X}(t) = X(t) \circ \pi$ holds on the event $\widetilde{\Omega}^{\triangle,\prime}$. This assertion is a consequence of the following argument. For every $(t,x) \in [0,\infty) \times E^{\triangle}$ and for every $f \in C\left(E^{\triangle}\right)$ we see

$$\begin{aligned}\widetilde{\mathbb{E}}_x\left[f\left(X(t)\circ\pi\right) - f\left(\widetilde{X}(t)\right)\right] &= \lim_{s \downarrow t,\, s \in D} \widetilde{\mathbb{E}}_x\left[f\left(\widetilde{X}(s)\right) - f\left(\widetilde{X}(t)\right)\right] \\ &= \widetilde{\mathbb{E}}_x\left[f\left(\widetilde{X}(t)\right) - f\left(\widetilde{X}(t)\right)\right] = 0. \end{aligned}\tag{3.84}$$

Since the space $E^{\triangle}$ is second countable, the space $C\left(E^{\triangle}\right)$ is separable, the equalities in (3.82) and (3.84) imply that, up to an event which is $\widetilde{\mathbb{P}}_x$-negligible $\widetilde{X}(t) = X(t) \circ \pi$ for all $t \geqslant 0$. See Definition 5.88 as well. In addition, we have that, for $\omega \in \widetilde{\Omega}$ the realization $t \mapsto \pi(\omega)(t)$ belongs to the Skorohod space Ω, *i.e.* it is continuous from the right and possesses left limits. We still need to show that $\widetilde{X}(t) \in E$ implies that the closure of the orbit $\left\{\widetilde{X}(s) : 0 \leqslant s \leqslant t,\, s \in D\right\}$ is a closed, and so, compact subset of E. For this purpose we choose a strictly positive function $f \in C_0(E)$ which we extend to a function, again called f, such that $f(\triangle) = 0$. It is convenient to employ the resolvent operators $R(\alpha)$, $\alpha > 0$, here. We will prove that, for $\alpha > 0$ fixed, the process $t \mapsto e^{-\alpha t} R(\alpha) f\left(\widetilde{X}(t)\right)$ is $\widetilde{\mathbb{P}}_x$-supermartingale relative to the filtration $\left\{\widetilde{\mathfrak{F}}_t^{\triangle};\, t \geqslant 0\right\}$. Therefore, let $t_2 > t_1 \geqslant 0$. Then we write:

$$\begin{aligned}&\widetilde{\mathbb{E}}_x\left[e^{-\alpha t_2} R(\alpha) f\left(\widetilde{X}\left(t_2\right)\right) \mid \widetilde{\mathfrak{F}}_{t_1}^{\triangle}\right] \\ &= \widetilde{\mathbb{E}}_x\left[e^{-\alpha t_2}\int_0^\infty e^{-\alpha s}\widetilde{\mathbb{E}}_{\widetilde{X}(t_2)}\left[f\left(\widetilde{X}(s)\right)\right] ds \mid \widetilde{\mathfrak{F}}_{t_1}^{\triangle}\right]\end{aligned}$$

(Markov property)

$$= \widetilde{\mathbb{E}}_x \left[e^{-\alpha t_2} \int_0^\infty e^{-\alpha s} \widetilde{\mathbb{E}}_x \left[f\left(\widetilde{X}(s+t_2)\right) \mid \widetilde{\mathfrak{F}}_{t_2}^{\triangle} \right] ds \mid \widetilde{\mathfrak{F}}_{t_1}^{\triangle} \right]$$

(Fubini's theorem and tower property of condiitonal expectation)

$$= \widetilde{\mathbb{E}}_x \left[e^{-\alpha t_2} \int_0^\infty e^{-\alpha s} f\left(\widetilde{X}(s+t_2)\right) ds \mid \widetilde{\mathfrak{F}}_{t_1}^{\triangle} \right]$$

$$= \widetilde{\mathbb{E}}_x \left[\int_{t_2}^\infty e^{-\alpha s} f\left(\widetilde{X}(s)\right) ds \mid \widetilde{\mathfrak{F}}_{t_1}^{\triangle} \right]$$

(the function f is non-negative and $t_2 > t_1$)

$$\leqslant \widetilde{\mathbb{E}}_x \left[\int_{t_1}^\infty e^{-\alpha s} f\left(\widetilde{X}(s)\right) ds \mid \widetilde{\mathfrak{F}}_{t_1}^{\triangle} \right]$$

$$= \widetilde{\mathbb{E}}_x \left[e^{-\alpha t_1} \int_0^\infty e^{-\alpha s} f\left(\widetilde{X}(s+t_1)\right) ds \mid \widetilde{\mathfrak{F}}_{t_1}^{\triangle} \right]$$

(Fubini's theorem in combinaton with the Markov property)

$$= e^{-\alpha t_1} \int_0^\infty e^{-\alpha s} \widetilde{\mathbb{E}}_{\widetilde{X}(t_1)} \left[f\left(\widetilde{X}(s)\right) \right] ds = e^{-\alpha t_1} R(\alpha) f\left(\widetilde{X}(t_1)\right). \tag{3.85}$$

Put $\widetilde{Y}(t) = e^{-\alpha t}R(\alpha)f\left(\widetilde{X}(t)\right)$, and fix $x \in E$. From (3.85) we see that the process $t \mapsto \widetilde{Y}(t)$ is a $\widetilde{\mathbb{P}}_x$-supermartingale relative to the filtration $\left(\widetilde{\mathcal{F}}_t^{\triangle}\right)_{t\geqslant 0}$. From Lemma 3.38 with $\widetilde{Y}(t)$ instead of $Y(t)$ and $\widetilde{\mathbb{P}}_x$ in place of $\mathbb{P}$ we infer

$$\widetilde{P}_x\left[\widetilde{X}(t) \in E, \inf_{s\in D\cap(0,t)} \widetilde{Y}(s) = 0\right] = \widetilde{P}_x\left[\widetilde{Y}(t) > 0, \inf_{s\in D\cap(0,t)} \widetilde{Y}(s) = 0\right] = 0. \tag{3.86}$$

From (3.86) we see that, $\widetilde{P}_x$-almost surely, $\widetilde{X}(t) \in E$ implies $\inf_{s\in D\cap(0,t)} \widetilde{Y}(s) > 0$. Consequently, for every $x \in E$, the equality

$$\widetilde{P}_x\left[\widetilde{X}(t) \in E\right] = \widetilde{P}_x\left[\widetilde{X}(t) \in E, \text{ closure}\left\{\widetilde{X}(s) : s \in D \cap (0,t)\right\} \subset E\right], \tag{3.87}$$

holds. In other words: the closure of the orbit $\left\{\widetilde{X}(s) : s \in D \cap (0,t)\right\}$ is contained in E whenever $\widetilde{X}(t)$ belongs to E. We are almost at the end of the proof. We still have to carry over the Markov process in (3.79) to a process of the form

$$\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t), t \geqslant 0), (\vartheta_t, t \geqslant 0), (E, \mathcal{E})\} \tag{3.88}$$

with $\Omega = D\left([0,\infty), E^{\triangle}\right)$ the Skorohod space of paths with values in $E^{\triangle}$. This can be done as follows. Define the state variables $X(t) : \Omega \to E^{\triangle}$ by $X(t)(\omega) = \omega(t)$, $\omega \in \Omega$, and let $\vartheta_t : \Omega \to \Omega$ be defined as above, *i.e.* $\vartheta_t(\omega)(s) = \omega(s+t)$, $\omega \in \Omega$. Let the mapping $\pi : \widetilde{\Omega} \to \Omega$ be defined as in (3.83). Then, as shown above, for every $x \in E$, the processes $\widetilde{X}(t)$ and $X(t)\circ\pi$ are $\widetilde{P}_x$-indistinguishable. The probability measures $\mathbb{P}_x$, $x \in E$, are defined by $\mathbb{P}_x[A] = \widetilde{\mathbb{P}}_x[\pi \in A]$ where A is a Borel subset of Ω. Then all ingredients of (3.88) are defined. It is clear that the quadruple in (3.88) is a Markov process. Since the paths, or realizations, are right-continuous, it represents a strong Markov process.

This completes an outline of the proof of Theorem 3.37. □

As above L is a linear operator with domain $D(L)$ and range $R(L)$ in $C_0(E)$. Suppose that the domain $D(L)$ of L is dense in $C_0(E)$. The problem we want to address is the following. Give necessary and sufficient conditions on the operator L in order that for every $x \in E$ there exists a unique probability measure $\mathbb{P}_x$ on $\mathcal{F}$ with the following properties:

(i) For every $f \in D(L)$ the process $f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds$, $t \geqslant 0$, is a $\mathbb{P}_x$-martingale;
(ii) $\mathbb{P}_x(X(0) = x) = 1$.

Here we suppose $\Omega = D\left([0,\infty), E^{\triangle}\right)$ (Skorohod space) and $\mathcal{F}$ is the σ-field generated by the state variables $X(t)$, $t \geqslant 0$. Let $P(\Omega)$ be the set of all probability measures on $\mathcal{F}$ and define the subset $P'(\Omega)$ of $P(\Omega)$ by

$$P'(\Omega) = \bigcup_{x\in E^{\triangle}} \Big\{\mathbb{P} \in P(\Omega) : \mathbb{P}[X(0) = x] = 1 \text{ and for every } f \in D(L) \text{ the process}$$

$$f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds, t \geqslant 0, \text{ is a } \mathbb{P}\text{-martingale} \Big\}. \tag{3.89}$$

Let $(v_j : j \in \mathbb{N})$ be a sequence of continuous functions defined on E with the following properties:

(i) $v_0 \equiv 1$;
(ii) $\|v_j\|_\infty \leqslant 1$ and v_j belongs to $D(L)$ for $j \geqslant 1$;
(iii) The linear span of v_j, $j \geqslant 0$, is dense in $C(E^\triangle)$.

In addition let $(f_k : k \in \mathbb{N})$ be a sequence in $D(L)$ such that the linear span of $\{(f_k, Lf_k) : k \in \mathbb{N}\}$ is dense in the graph $G(L) := \{(f, Lf) : f \in D(L)\}$ of the operator L. Moreover let $(s_j : j \in \mathbb{N})$ be an enumeration of the set $\mathbb{Q} \cap [0, \infty)$. The subset $P'(\Omega)$ may be described as follows:

$$\begin{aligned} P'(\Omega) = &\bigcap_{n=1}^\infty \bigcap_{k=1}^\infty \bigcap_{m=1}^\infty \bigcap_{(j_1,\dots,j_{m+1})\in\mathbb{N}^{m+1}} \bigcap_{0\leqslant s_{j_1}<\dots<s_{j_{m+1}}} \\ &\Big\{\mathbb{P} \in P(\Omega) : \inf_{x\in E} \max_{1\leqslant j\leqslant n} \left| \int v_j(X(0))d\mathbb{P} - v_j(x) \right| = 0, \\ &\int \left(f_k(X(s_{j_{m+1}})) - \int_0^{s_{j_{m+1}}} Lf_k(X(s))ds \right) \prod\nolimits_{k=1}^m v_{j_k}\left(X(s_{j_k})\right) d\mathbb{P} \\ &\quad = \int \left(f_k(X(s_{j_m})) - \int_0^{s_{j_m}} Lf_k(X(s))ds \right) \prod\nolimits_{k=1}^m v_{j_k}\left(X(s_{j_k})\right) d\mathbb{P} \Big\}. \end{aligned} \tag{3.90}$$

It follows that $P'(\Omega)$ is a weakly closed subset of $P(\Omega)$. In fact we shall prove that, if for the operator L, the martingale problem is uniquely solvable, then the set $P'(\Omega)$ is compact metrizable for the metric $d(\mathbb{P}_1, \mathbb{P}_2)$ given by

$$d(\mathbb{P}_1, \mathbb{P}_2) = \sum\nolimits_{\Lambda\subset\mathbb{N},|\Lambda|<\infty} 2^{-|\Lambda|} \sum_{(\ell_j)_{j\in\Lambda}} \left| \int \prod_{j\in\Lambda} 2^{-j-\ell_j} v_j\left(X\left(s_{\ell_j}\right)\right) d\left(\mathbb{P}_2 - \mathbb{P}_1\right) \right|. \tag{3.91}$$

The following result should be compared to the comments in 6.7.4. of [**133**]. It is noticed that in Proposition 3.41 below the uniqueness of the solutions to the martingale problem is not used.

3.41. Proposition. *The set $P'(\Omega)$ supplied with the metric d defined in (3.91) is a compact Hausdorff space.*

Proof. Let $(\mathbb{P}_n : n \in \mathbb{N})$ be any sequence in $P'(\Omega)$. Let $(\mathbb{P}_{n_\ell} : \ell \in \mathbb{N})$ be a subsequence with the property that for every $m \in \mathbb{N}$, for every m-tuple $(j_1, \dots, j_m)$ in $\mathbb{N}^m$ and for every m-tuple $(s_{j_1}, \dots, s_{j_m}) \in \mathbb{Q}^m$ the limit

$$\lim_{\ell\to\infty} \int \prod\nolimits_{k=1}^m v_{j_k}\left(X\left(s_{j_k}\right)\right) d\mathbb{P}_{n_\ell}$$

exists. We shall prove that for every $m \in \mathbb{N}$, for every m-tuple $(j_1, \ldots, j_m)$ in $\mathbb{N}^m$ and for every m-tuple $(t_{j_1}, \ldots, t_{j_m}) \in [0, \infty)^m$ the limit

$$\lim_{\ell \to \infty} \int \prod_{k=1}^{m} u_{j_k} \left(X \left(t_{j_k} \right) \right) d\mathbb{P}_{n_\ell} \tag{3.92}$$

exists for all sequences $(u_j : j \in \mathbb{N})$ in $C_0(E)$. But then there exists, by Kolmogorov's extension theorem, a probability measure $\mathbb{P}$ such that

$$\lim_{\ell \to \infty} \int \prod_{k=1}^{m} u_{j_k} \left(X \left(t_{j_k} \right) \right) d\mathbb{P}_{n_\ell} = \int \prod_{k=1}^{m} u_{j_k} \left(X \left(t_{j_k} \right) \right) d\mathbb{P}, \tag{3.93}$$

for all $m \in \mathbb{N}$, for all $(j_1, \ldots, j_m) \in \mathbb{N}^m$ and for all $(t_{j_1}, \ldots, t_{j_m}) \in [0, \infty)^m$. From the description (3.89) of $P'(\Omega)$ it then readily follows that $\mathbb{P}$ is a member of $P'(\Omega)$. So the existence of the limit in (3.92) remains to be verified together with the fact that $D([0, \infty), E^{\triangle})$ has full $\mathbb{P}$-measure. Let t be in $\mathbb{Q}$. Since, for every $j \in \mathbb{N}$, the process $v_j(X(s)) - v_j(X(0)) - \int_0^s Lv_j(X(\sigma))d\sigma$, $s \geqslant 0$, is a martingale for the measures $\mathbb{P}_{n_\ell}$, we infer

$$\int \int_0^t Lv_j(X(s))ds d\mathbb{P}_{n_\ell} = \int v_j(X(t))d\mathbb{P}_{n_\ell} - \int v_j(X(0))d\mathbb{P}_{n_\ell}.$$

and hence the limit $\lim_{\ell \to \infty} \int \int_0^t Lv_j(X(s))ds d\mathbb{P}_{n_\ell}$ exists.

Next let t_0 be in $[0,\infty)$. Again using the martingale property we see

$$\begin{aligned}
&\int v_j\left(X(t_0)\right) d\left(\mathbb{P}_{n_\ell}-\mathbb{P}_{n_k}\right)\\
&=\int\left(\int_0^t Lv_j\left(X(s)\right) ds\right) d\left(\mathbb{P}_{n_\ell}-\mathbb{P}_{n_k}\right)+\int v_j(X(0)) d\left(\mathbb{P}_{n_\ell}-\mathbb{P}_{n_k}\right)\\
&\quad-\int\left(\int_{t_0}^t Lv_j\left(X(s)\right) ds\right) d\left(\mathbb{P}_{n_\ell}-\mathbb{P}_{n_k}\right),
\end{aligned}\tag{3.94}$$

where t is any number in $\mathbb{Q}\cap[0,\infty)$. From (3.94) we infer

$$\begin{aligned}
&\left|\int v_j\left(X(t_0)\right) d\left(\mathbb{P}_{n_\ell}-\mathbb{P}_{n_k}\right)\right|\\
&\leqslant\left|\int\left(\int_0^t Lv_j\left(X(s)\right) ds\right) d\left(\mathbb{P}_{n_\ell}-\mathbb{P}_{n_k}\right)\right|+\left|\int v_j(X(0)) d\left(\mathbb{P}_{n_\ell}-\mathbb{P}_{n_k}\right)\right|\\
&\quad+2\left|t-t_0\right|\left\|Lv_j\right\|_\infty.
\end{aligned}\tag{3.95}$$

If we let ℓ and k tend to infinity, we obtain

$$\limsup_{\ell,k\to\infty}\left|\int v_j\left(X(t_0)\right) d\left(\mathbb{P}_{n_\ell}-\mathbb{P}_{n_k}\right)\right|\leqslant 2\left|t-t_0\right|\left\|Lv_j\right\|_\infty.\tag{3.96}$$

Consequently for every $s\geqslant 0$ the limit $\lim_{\ell\to\infty}\int v_j\left(X(s)\right) d\mathbb{P}_{n_\ell}$ exists. The inequality

$$\begin{aligned}
\left|\int v_j\left(X(t)\right) d\mathbb{P}_{n_\ell}-\int v_j\left(X(t_0)\right) d\mathbb{P}_{n_\ell}\right|&=\left|\int\int_{t_0}^t Lv_j\left(X(s)\right) ds d\mathbb{P}_{n_\ell}\right|\\
&\leqslant\left|t-t_0\right|\left\|Lv_j\right\|_\infty
\end{aligned}$$

shows that the functions $t\mapsto\lim_{\ell\to\infty}\int v_j\left(X(t)\right) d\mathbb{P}_{n_\ell}$, $j\in\mathbb{N}$, are continuous. Since the linear span of $(v_j : j\geqslant 1)$ is dense in $C_0(E)$, it follows that for $v\in C_0(E)$ and for every $t\geqslant 0$ the limit

$$t\mapsto\lim_{\ell\to\infty}\int v\left(X(t)\right) d\mathbb{P}_{n_\ell}\tag{3.97}$$

exists and that this limit, as a function of t, is continuous. The following step consists in proving that for every $t_0\in[0,\infty)$ the equality

$$\lim_{t\to t_0}\limsup_{\ell\to\infty}\int\left|v_j\left(X(t)\right)-v_j\left(X(t_0)\right)\right| d\mathbb{P}_{n_\ell}=0\tag{3.98}$$

holds. For $t>s$ the following (in-)equalities are valid:

$$\begin{aligned}
&\left(\int\left|v_j\left(X(t)\right)-v_j\left(X(s)\right)\right| d\mathbb{P}_{n_\ell}\right)^2\leqslant\int\left|v_j\left(X(t)\right)-v_j\left(X(s)\right)\right|^2 d\mathbb{P}_{n_\ell}\\
&=\int\left|v_j(X(t))\right|^2 d\mathbb{P}_{n_\ell}-\int\left|v_j(X(s))\right|^2 d\mathbb{P}_{n_\ell}\\
&\quad-2\mathrm{Re}\int\left(v_j(X(t))-v_j(X(s))\right)\overline{v}_j(X(s)) d\mathbb{P}_{n_\ell}\\
&=\int\left|v_j(X(t))\right|^2 d\mathbb{P}_{n_\ell}-\int\left|v_j(X(s))\right|^2 d\mathbb{P}_{n_\ell}
\end{aligned}$$

$$\begin{aligned}&- 2\mathrm{Re} \int \left(\int_s^t Lv_j(X(\sigma))d\sigma \right) \overline{v}_j(X(s))d\mathbb{P}_{n_\ell} \\ &\leqslant \int |v_j(X(t))|^2 \, d\mathbb{P}_{n_\ell} - \int |v_j(X(s))|^2 \, d\mathbb{P}_{n_\ell} + 2(t-s) \left\| Lv_j \right\|_\infty . \end{aligned} \tag{3.99}$$

Hence (3.97) together with (3.99) implies (3.93). By (3.93), we may apply Kolmogorov's theorem to prove that there exists a probability measure $\mathbb{P}$ on $\Omega' := (E^\triangle)^{[0,\infty)}$ with the property that

$$\int \prod_{k=1}^m v_{j_k}(X(s_{j_k}))d\mathbb{P} = \lim_{\ell\to\infty} \int \prod_{k=1}^m v_{j_k}(X(s_{j_k}))d\mathbb{P}_{n_\ell}, \tag{3.100}$$

holds for all $m \in \mathbb{N}$ and for all $(s_{j_1}, \ldots, s_{j_m}) \in [0,\infty)^m$. It then also follows that the equality in (3.100) is also valid for all m-tuples $f_1, \ldots, f_m$ in $C(E^\triangle)$ instead of $v_{j_1}, \ldots, v_{j_m}$. This is true because the linear span of the sequence $(v_j : j \in \mathbb{N})$ is dense in $C(E^\triangle)$. In addition we conclude that the processes $f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds$, $t \geqslant 0$, $f \in D(L)$ are $\mathbb{P}$-martingales. We still have to show that $D([0,\infty), E^\triangle)$ has $\mathbb{P}$-measure 1. From (3.98) it essentially follows that set of $\omega \in (E^\triangle)^{[0,\infty)}$ for which the left and right hand limits exist in $E^\triangle$ has "full" $\mathbb{P}$-measure. First let $f \geqslant 0$ be in $C_0(E)$. Then the process $[G_\lambda f](t) := \mathbb{E}\left(\int_t^\infty e^{-\lambda\sigma} f(X(\sigma))d\sigma \mid \mathcal{F}_t\right)$ is a $\mathbb{P}$-supermartingale with respect to the filtration $\{\mathcal{F}_t : t \geqslant 0\}$. It follows that the limits $\lim_{t\uparrow t_0} [G_\lambda f](t)$ and $\lim_{t\downarrow t_0} [G_\lambda f](t)$ both exist $\mathbb{P}$-almost surely for all $t_0 \geqslant 0$ and for all $f \in C_0(E)$. In particular these limits exist $\mathbb{P}$-almost surely for all $f \in D(L)$. By the martingale property it follows that, for $f \in D(L)$,

$$\begin{aligned}\left| f(X(t)) - \lambda e^{\lambda t} [G_\lambda f](t) \right| &= \left| \lambda e^{\lambda t} \mathbb{E}\left(\int_t^\infty e^{-\lambda\sigma} \left(f(X(\sigma)) - f(X(t)) \right) d\sigma \mid \mathcal{F}_t \right) \right| \\ &= \left| \lambda e^{\lambda t} \mathbb{E}\left(\int_t^\infty e^{-\lambda\sigma} \left(\int_t^\sigma Lf(X(s))ds \right) d\sigma \mid \mathcal{F}_t \right) \right| \\ &\leqslant \lambda e^{\lambda t} \int_t^\infty e^{-\lambda\sigma}(\sigma - t) \left\| Lf \right\|_\infty d\sigma = \lambda^{-1} \left\| Lf \right\|_\infty .\end{aligned}$$

Consequently, we may conclude that, for all s, $t \geqslant 0$,

$$|f(X(t)) - f(X(s))| \leqslant 2\lambda^{-1} \left\| Lf \right\|_\infty + \left| \lambda e^{\lambda t} [G_\lambda f](t) - \lambda e^{\lambda s} [G_\lambda f](s) \right|$$

and hence that the limits $\lim_{t\downarrow s} f(X(t))$ and $\lim_{t\uparrow s} f(X(t))$ exist $\mathbb{P}$-almost surely for all $f \in D(L)$. By separability and density of $D(L)$ it follows that the limits $\lim_{t\downarrow s} X(t)$ and $\lim_{t\uparrow s} X(t)$ exist $\mathbb{P}$-almost surely for all $s \geqslant 0$. Put $Z(s)(\omega) = \lim_{t\downarrow s, t\in\mathbb{Q}} X(t)(\omega)$, $t \geqslant 0$. Then, for $\mathbb{P}$-almost all ω and for all $s \geqslant 0$, $Z(s)(\omega)$ is well-defined, possesses left limits and is right continuous. In addition we have

$$\begin{aligned}\mathbb{E}\left(f(Z(s))g(s)\right) &= \mathbb{E}\left(f(X(s+))g(X(s))\right) = \lim_{t\downarrow s} \mathbb{E}\left(f(X(t))g(X(s))\right) \\ &= \mathbb{E}\left(f(X(s))g(X(s))\right), \text{for all } f,\ g \in C_0(E)\end{aligned}$$

and for all $s \geqslant 0$: see (3.98). But then we may conclude that $X(s) = Z(s)$ $\mathbb{P}$-almost surely for all $s \geqslant 0$. Hence we may replace X with Z and consequently (see the arguments in the proof of Theorem 9.4. of Blumenthal and Getoor [[**20**], p. 49])

$$\mathbb{P}\left(\omega \in \Omega' : \omega \text{ is right continuous and has left limits in } E^{\triangle}\right) = 1.$$

Fix $s > t$. We are going to show that the set of paths $\omega \in (E^{\triangle})^{[0,\infty)}$ for which $\omega(s) = X(s)(\omega)$ belongs to E and for which, $\omega(t-) = \lim_{\tau\uparrow t} X(\tau)(\omega) = \triangle$ possesses $\mathbb{P}$-measure 0. It suffices to prove that, for $f \in C_0(E)$, $1 \geqslant f(x) > 0$ for all $x \in E$ fixed, the following integral equalities hold:

$$\mathbb{E}\left[f(X(s)), f(X(t-)) = 0\right] = \int f(X(s)) 1_{\{f=0\}}(X(t-))\, d\mathbb{P} = 0.$$

This can be achieved as follows. From the (in-)equalities

$$\begin{aligned}
&\mathbb{E}\left(f(X(s)), f(X(t)) = 0\right) \\
&= \lim_{n\to\infty} \mathbb{E}\left(f(X(s))\left(1 - (f(X(t)))^{1/n}\right)\right) \\
&= \lim_{n\to\infty} \mathbb{E}\left(f(X(s))\left(1 - f(X(t))^{1/n}\right)\right) \\
&= \lim_{n\to\infty} \mathbb{E}\left(f(X(s))\int_0^{1/n} f(X(t))^{\sigma} \log\frac{1}{f(X(t))} d\sigma\right) \\
&= \lim_{n\to\infty} \int_0^{1/n} \mathbb{E}\left(f(X(s))f(X(t))^{\sigma} \log\frac{1}{f(X(t))}\right) d\sigma \\
&= \lim_{n\to\infty} \int_0^{1/n} \left(\mathbb{E} - \mathbb{E}_{n_\ell}\right)\left(f(X(s))f(X(t))^{\sigma} \log\frac{1}{f(X(t))}\right) d\sigma \\
&\quad + \lim_{n\to\infty} \int_0^{1/n} \mathbb{E}_{n_\ell}\left(f(X(s))f(X(t))^{\sigma} \log\frac{1}{f(X(t))}\right) d\sigma \\
&\leqslant \int_0^1 \left|\mathbb{E}\left(f(X(s))f(X(t))^{\sigma} \log\frac{1}{f(X(t))}\right)\right. \\
&\qquad \left. -\mathbb{E}_{n_\ell}\left(f(X(s))f(X(t))^{\sigma} \log\frac{1}{f(X(t))}\right)\right| d\sigma \\
&\qquad + \mathbb{E}_{n_\ell}\left(f(X(s)), f(X(t)) = 0\right),
\end{aligned}$$

we conclude that

$$\begin{aligned}
\mathbb{E}\left(f(X(s)), f(X(t)) = 0\right) &= \lim_{n\to\infty} \mathbb{E}\left(f(X(s))\left(1 - f(X(t))^{1/n}\right)\right) \\
\leqslant \int_0^1 &\left|\mathbb{E}\left(f(X(s))f(X(t))^{\sigma} \log\frac{1}{f(X(t))}\right)\right. \\
&\left. -\mathbb{E}_{n_\ell}\left(f(X(s))f(X(t))^{\sigma} \log\frac{1}{f(X(t))}\right)\right| d\sigma.
\end{aligned}$$

Since the function $x \mapsto f(x)^{\sigma} \log\dfrac{1}{f(x)}$ belongs to $C_0(E)$ for every $\sigma > 0$, we obtain upon letting ℓ tend to ∞, that $\mathbb{E}\left(f(X(s)), f(X(t)) = 0\right) = 0$, where $s > t$. To see this apply Scheffé's theorem (see e.g. Bauer [[**10**], Corollary 2.12.5. p. 105]) to the sequence $\sigma \mapsto \mathbb{E}_{n_\ell}\left(f(X(s))f(X(t))^{\sigma} \log\dfrac{1}{f(X(t))}\right)$. From description (3.90), it then follows that $\mathbb{P}$ belongs to $P'(\Omega)$: it is also clear that the limits in (3.92) exist. □

3.42. Proposition. *Suppose that for every $x \in E$ the martingale problem is uniquely solvable. Define the map $F : P'(\Omega) \to E^{\triangle}$ by $F(\mathbb{P}) = x$, where $\mathbb{P} \in P'(\Omega)$ is such that $\mathbb{P}(X(0) = x) = 1$. Also notice that $F(P_{\triangle}) = \triangle$. Then F is a homeomorphism from $P'(\Omega)$ onto $E^{\triangle}$. In fact it follows that for every $u \in C_0(E)$ and for every $s \geqslant 0$, the function $x \mapsto \mathbb{E}_x(u(X(s))$ belongs to $C_0(E)$.*

Proof. Since the martingale problem is uniquely solvable for every $x \in E$ the map F is a one-to-one map from the compact metric Hausdorff space $P'(\Omega)$

onto $E^{\triangle}$ (see Proposition 3.41). Let for $x \in E$ the probability $\mathbb{P}_x$ be the unique solution to the martingale problem:

(i) For every $f \in D(L)$ the process $f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds$, $t \geqslant 0$, is a $\mathbb{P}_x$-martingale;
(ii) $\mathbb{P}_x(X(0) = x) = 1$.

Then, by definition $F(\mathbb{P}_x) = x$, $x \in E$, and $F(\mathbb{P}_{\triangle}) = \triangle$. Moreover, since for every $x \in E$ the martingale problem is uniquely solvable we see $P'(\Omega) = \left\{\mathbb{P}_x : x \in E^{\triangle}\right\}$. Let $(x_\ell : \ell \in \mathbb{N})$ be a sequence in $E^{\triangle}$ with the property that $\lim_{\ell\to\infty} d\left(\mathbb{P}_{x_\ell}, \mathbb{P}_x\right) = 0$ for some $x \in E^{\triangle}$. Then $\lim_{\ell\to\infty} \left|v_j\left(x_\ell\right) - v_j(x)\right| = 0$, for all $j \in \mathbb{N}$, where, as above, the span of the sequence $(v_j : j \in \mathbb{N})$ is dense in $C\left(E^{\triangle}\right)$. It follows that $\lim_{\ell\to\infty} x_\ell = x$ in $E^{\triangle}$. Consequently the mapping F is continuous. Since F is a continuous bijective map from one compact metric Hausdorff space $P'(\Omega)$ onto another such space $E^{\triangle}$, its inverse is continuous as well. Among others this implies that, for every $s \in \mathbb{Q} \cap [0, \infty)$ and for every $j \geqslant 1$, the function $x \mapsto \int v_j\left(X(s)\right) d\mathbb{P}_x$ belongs to $C_0(E)$. Since the linear span of the sequence $(v_j : j \geqslant 1)$ is dense in $C_0(E)$ it also follows that for every $v \in C_0(E)$, the function $x \mapsto \int v\left(X(s)\right) d\mathbb{P}_x$ belongs to $C_0(E)$. Next let $s_0 \geqslant 0$ be arbitrary. For every $j \geqslant 1$ and every $s \in \mathbb{Q} \cap [0, \infty)$, $s > s_0$, we have by the martingale property:

$$\begin{aligned}\sup_{x\in E} \left|\mathbb{E}_x\left(v_j(X(s))\right) - \mathbb{E}_x\left(v_j(X(s_0))\right)\right| &= \sup_{x\in E} \left|\int_{s_0}^{s} \mathbb{E}_x\left(Lv_j\left(X(\sigma)\right)\right) d\sigma\right| \\ &\leqslant (s - s_0) \left\|Lv_j\right\|_\infty . \end{aligned} \tag{3.101}$$

Consequently, for every $s \in [0, \infty)$, the function $x \mapsto \mathbb{E}_x\left(v_j\left(X(s)\right)\right)$, $j \geqslant 1$, belongs to $C_0(E)$. It follows that, for every $v \in C_0(E)$ and every $s \geqslant 0$, the function $x \mapsto \mathbb{E}_x\left(v(X(s))\right)$ belongs to $C_0(E)$. This proves Proposition 3.42. ☐

The proof of the following proposition may be copied from Ikeda and Watanabe [**61**], Theorem 5.1. p. 205.

3.43. Proposition. *Suppose that for every $x \in E^{\triangle}$ the martingale problem:*

(i) *For every $f \in D(L)$ the process $f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds$, $t \geqslant 0$, is a $\mathbb{P}$-martingale;*
(ii) $\mathbb{P}(X(0) = x) = 1$,

has a unique solution $\mathbb{P} = \mathbb{P}_x$. Then the process

$$\left\{\left(\Omega, \mathcal{F}, \mathbb{P}_x\right), \left(X(t) : t \geqslant 0\right), \left(\vartheta_t : t \geqslant 0\right), (E, \mathcal{E})\right\}$$

is a strong Markov process.

Proof. Fix $x \in E$ and let T be a stopping time and choose a realization of

$$A \mapsto \mathbb{E}_x\left[1_A \circ \vartheta_T \mid \mathcal{F}_T\right], \quad A \in \mathcal{F}.$$

Fix any $\omega \in \Omega$ for which

$$A \mapsto Q_y(A) := \mathbb{E}_x\left[1_A \circ \vartheta_T \mid \mathcal{F}_T\right](\omega),$$

is defined for all $A \in \mathcal{F}$. Here, by definition, $y = \omega(T(\omega))$. Notice that, sice the space E is a topological Hausdorff space that satisfies the second countability axiom, this construction can be performed for $\mathbb{P}_x$-almost all ω. Let f be in $D(L)$ and fix $t_2 > t_1 \geqslant 0$. Moreover fix $C \in \mathcal{F}_{t_1}$. Then $\vartheta_T^{-1}(C)$ is a member of $\mathcal{F}_{t_1+T}$. Put $M_f(t) = f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds$, $t \geqslant 0$. We have

$$\mathbb{E}_y\left(M_f(t_2)1_C\right) = \mathbb{E}_y\left(M_f(t_1)1_C\right). \tag{3.102}$$

We also have

$$\int \left(f(X(t_2)) - f(X(0)) - \int_0^{t_2} Lf(X(s))ds \right) 1_C dQ_y \tag{3.103}$$

$$= \mathbb{E}_x\left[\left(f\left(X(t_2+T)\right) - f(X(T)) - \int_0^{t_2} Lf\left(X(s+T)\right)ds \right) 1_C \circ \vartheta_T \mid \mathcal{F}_T \right](\omega)$$

$$= \mathbb{E}_x\left[\left(f\left(X(t_2+T)\right) - f(X(T)) - \int_T^{t_2+T} Lf\left(X(s)\right)ds \right) \left(1_C \circ \vartheta_T\right) \mid \mathcal{F}_T \right](\omega)$$

$$= \mathbb{E}_x\left[\mathbb{E}_x\left[\left(f\left(X(t_2+T)\right) - f(X(T)) - \int_T^{t_2+T} Lf\left(X(s)\right)ds \right) \mid \mathcal{F}_{t_1+T} \right] \cdot 1_C \circ \vartheta_T \mid \mathcal{F}_T \right](\omega).$$

By Doob's optional sampling theorem, the process

$$f\left(X(t+T)\right) - f(X(T)) - \int_T^{t+T} Lf\left(X(s)\right) ds$$

is a $\mathbb{P}_x$-martingale with respect to the fields $\mathcal{F}_{t+T}$, $t \geqslant 0$. So from (3.103) we obtain:

$$\begin{aligned}
&\int \left(f(X(t_2)) - f(X(0)) - \int_0^{t_2} Lf(X(s))ds \right) 1_C dQ_y \\
&= \mathbb{E}_x \left[\left(f\left(X(t_1+T)\right) - f(X(T)) - \int_T^{t_1+T} Lf\left(X(s)\right) ds \right) 1_C \circ \vartheta_T \mid \mathcal{F}_T \right] (\omega) \\
&= \int \left(f(X(t_1)) - f(X(0)) - \int_0^{t_1} Lf(X(s))ds \right) 1_C dQ_y. \qquad (3.104)
\end{aligned}$$

It follows that, for $f \in D(L)$, the process $M_f(t)$ is a $\mathbb{P}_y$- as well as a Q_y-martingale. Since $\mathbb{P}_y[X(0) = y] = 1$ and since

$$\begin{aligned}
Q_y(X(0) = y) &= \mathbb{E}_x \left[1_{\{X(0)=y\}} \circ \vartheta_T \mid \mathcal{F}_T \right] (\omega) \\
&= \mathbb{E}_x \left[1_{\{X(T)=y\}} \mid \mathcal{F}_T \right] (\omega) = 1_{\{X(T)=y\}}(\omega) = 1, \qquad (3.105)
\end{aligned}$$

we conclude that the probabilities $\mathbb{P}_y$ and Q_y are the same. Equality (3.105) follows, because, by definition, $y = X(T)(\omega) = \omega(T(\omega))$. Since $\mathbb{P}_y = Q_y$, it then follows that

$$\mathbb{P}_{X(T)(\omega)}(A) = \mathbb{E}_x \left[1_A \circ \vartheta_T \mid \mathcal{F}_T \right] (\omega), \quad A \in \mathcal{F}.$$

Or putting it differently:

$$\mathbb{P}_{X(T)}(A) = \mathbb{E}_x \left[1_A \circ \vartheta_T \mid \mathcal{F}_T \right], \quad A \in \mathcal{F}. \qquad (3.106)$$

However, this is exactly the strong Markov property and completes the proof of Proposition 3.43. □

The following proposition can be proved in the same manner as Theorem 5.1 Corollary in Ikeda and Watanabe [**61**], p. 206.

3.44. Proposition. *If an operator L generates a Feller semigroup, then the martingale problem is uniquely solvable for L.*

Proof. Let $\{P(t) : t \geqslant 0\}$ be the Feller semigroup generated by L and let

$$\left\{ (\Omega, \mathcal{F}, \mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E}) \right\}$$

be the associated strong Markov process (see Theorem 3.37). If f belongs to $D(L)$, then the process $M_f(t) := f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds$, $t \geqslant 0$, is a $\mathbb{P}_x$-martingale for all $x \in E$. This can be seen as follows. Fix $t_2 > t_1 \geqslant 0$. Then

$$\begin{aligned}
&\mathbb{E}_x \left[M_f(t_2) \mid \mathcal{F}_{t_1} \right] \\
&= M_f(t_1) + \mathbb{E}_x \left(\left(f(X(t_2)) - \int_{t_1}^{t_2} Lf(X(s))ds \right) \mid \mathcal{F}_{t_1} \right) - f(X(t_1)) \\
&= M_f(t_1) + \mathbb{E}_x \left[\left(f(X(t_2 - t_1 + t_1)) - \int_0^{t_2-t_1} Lf(X(s+t_1))ds \right) \mid \mathcal{F}_{t_1} \right]
\end{aligned}$$

$$- f(X(t_1))$$

(Markov property)

$$= M_f(t_1) + \mathbb{E}_{X(t_1)}\left(f(X(t_2-t_1)) - \int_0^{t_2-t_1} Lf(X(s))ds\right) - f(X(t_1)). \quad (3.107)$$

Next we compute, for $y \in E$ and $s > 0$, the quantity:

$$\begin{aligned}
&\mathbb{E}_y\left(f(X(s)) - \int_0^s Lf(X(\sigma))d\sigma\right) - f(y) \\
&= [P(s)f](y) - \int_0^s [P(\sigma)(Lf)](y)d\sigma - f(y) \\
&= [P(s)f](y) - \int_0^s \frac{d}{d\sigma}[P(\sigma)f](y)d\sigma - f(y) \\
&= [P(s)f]](y) - ([P(s)f](y) - [P(0)f](y)) - f(y) = 0. \qquad (3.108)
\end{aligned}$$

Hence from (3.107) and (3.108) it follows that the process $M_f(t)$, $t \geqslant 0$, is a $\mathbb{P}_x$-martingale. Next we shall prove the uniqueness of the solutions of the martingale problem associated to the operator L. Let $\mathbb{P}_x^1$ and $\mathbb{P}_x^2$ be solutions "starting" in $x \in E$. We have to show that these probabilities coincide. Let f belong to $D(L)$ and let T be a stopping time. Then, via partial integration, we infer

$$\begin{aligned}
&\lambda\int_0^\infty e^{-\lambda t}\left\{f(X(t+T)) - \int_T^{t+T} Lf(X(\tau))d\tau - f(X(T))\right\}dt + f(X(T)) \\
&= \lambda\int_0^\infty e^{-\lambda t}\left\{f(X(t+T)) - \int_T^{t+T} Lf(X(\tau))d\tau\right\}dt \\
&= \lambda\int_0^\infty e^{-\lambda t} f(X(t+T))dt - \lambda\int_0^\infty e^{-\lambda t}\int_0^t Lf(X(\tau+T))d\tau dt \\
&= \lambda\int_0^\infty e^{-\lambda t} f(X(t+T))dt - \lambda\int_0^\infty \left(\int_\tau^\infty e^{-\lambda t}dt\right) Lf(X(\tau+T))d\tau \\
&= \int_0^\infty e^{-\lambda t}[(\lambda I - L)f](X(t+T))dt. \qquad (3.109)
\end{aligned}$$

From Doob's optional sampling theorem together with (3.109) we obtain:

$$\begin{aligned}
&\int_0^\infty e^{-\lambda t}\mathbb{E}_x^1((\lambda I - L)f(X(t+T)) \mid \mathcal{F}_T)dt - f(X(T)) \\
&= \lambda\int_0^\infty e^{-\lambda t}\mathbb{E}_x^1\left\{\left(f(X(t+T)) - \int_T^{t+T} Lf(X(\tau))d\tau - f(X(T))\right) \mid \mathcal{F}_T\right\}dt \\
&= 0 \\
&= \lambda\int_0^\infty e^{-\lambda t}\mathbb{E}_x^2\left\{\left(f(X(t+T)) - \int_T^{t+T} Lf(X(\tau))d\tau - f(X(T))\right) \mid \mathcal{F}_T\right\}dt \\
&= \int_0^\infty e^{-\lambda t}\mathbb{E}_x^2((\lambda I - L)f(X(t+T)) \mid \mathcal{F}_T)dt - f(X(T)). \qquad (3.110)
\end{aligned}$$

Next we set

$$[R(\lambda)f](x) = \int_0^\infty e^{-\lambda t} [P(t)f](x)dt, \quad x \in E, \ \lambda > 0, \ f \in C_0(E). \tag{3.111}$$

Then

$$(\lambda I - L)R(\lambda)f = f, \ f \in C_0(E), \ R(\lambda)(\lambda I - L)f = f, \ f \in D(L). \tag{3.112}$$

Among other things we see that $R(\lambda I - L) = C_0(E)$, $\lambda > 0$. From (3.110) it then follows that, for $g \in C_0(E)$,

$$\begin{aligned}\int_0^\infty e^{-\lambda t}\mathbb{E}_x^1 \left(g(X(t+T)) \mid \mathfrak{F}_T\right) dt &= \int_0^\infty e^{-\lambda t}[P(t)g](X(T))dt \\ &= \int_0^\infty e^{-\lambda t}\mathbb{E}_x^2 \left(g(X(t+T)) \mid \mathfrak{F}_T\right) dt. \end{aligned}\tag{3.113}$$

Since Laplace transforms are unique, since g belongs to $C_0(E)$ and since paths are right continuous, we conclude

$$\mathbb{E}_x^1 \left(g(X(t+T)) \mid \mathfrak{F}_T\right) = [P(t)g](X(T)) = \mathbb{E}_x^2 \left(g(X(t+T)) \mid \mathfrak{F}_T\right), \tag{3.114}$$

whenever g belongs to $C_0(E)$, whenever $t \geqslant 0$ and whenever T is a stopping time. The first equality in (3.114) holds $\mathbb{P}_x^1$-almost surely and the second $\mathbb{P}_x^2$-almost surely. As in Theorem 3.36 it then follows that

$$\mathbb{E}_x^1 \left(\prod_{j=1}^n f_j(X(t_j))\right) = \mathbb{E}_x^2 \left(\prod_{j=1}^n f_j(X(t_j))\right) \tag{3.115}$$

for $n = 1, 2, \ldots$ and for $f_1, \ldots, f_n$ in $C_0(E)$. But then the probabilities $\mathbb{P}_x^1$ and $\mathbb{P}_x^2$ are the same. This proves Proposition 3.44. $\square$

The theorem we want to prove reads as follows.

3.45. THEOREM. *Let L be a linear operator with domain $D(L)$ and range $R(L)$ in $C_0(E)$. Let Ω be the path space $\Omega = D\left([0,\infty), E^{\triangle}\right)$. The following assertions are equivalent:*

(i) *The operator L is closable and its closure generates a Feller semigroup;*
(ii) *The operator L solves the martingale problem maximally and its domain $D(L)$ is dense in $C_0(E)$;*
(iii) *The operator L verifies the maximum principle, its domain $D(L)$ is dense in $C_0(E)$ and there exists $\lambda_0 > 0$ such that the range $R(\lambda_0 I - L)$ is dense in $C_0(E)$.*

3.46. REMARK. The hard part in (iii) is usually the range property: there exists $\lambda_0 > 0$ such that the range $R(\lambda_0 I - L)$ is dense in $C_0(E)$. The theorem also shows, in conjunction with the results on Feller semigroups and Markov processes, the relations which exist between the unique solvability of the martingale problem, the strong Markov property and densely defined operators verifying the maximum principle together with the range property. However if L is in fact a second order differential operator, then we want to read of the range property from the coefficients. There do exist results in this direction. The interested reader is referred to the literature: Stroock and Varadhan [**133**] and also Ikeda and Watanabe [**61**].

In what follows we shall assume that the equivalence of (i) and (iii) already has been established. A proof can be found in [**141**], Theorem 2.2., p.14. In the proof of (ii) $\Rightarrow$ (i) we shall use this result. We shall also show the implication (i) $\Rightarrow$ (ii).

PROOF. (ii) $\Rightarrow$ (i). Let, for $x \in E$, the probability $\mathbb{P}_x$ be the unique solution of the martingale problem associated to the operator L. From Proposition 3.43 it follows that the process $\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E})\}$ is a strong Markov process. Define the operators $\{P(t) : t \geqslant 0\}$ as follows:

$$[P(t)f](x) = \mathbb{E}_x\left(f(X(t))\right), \quad f \in C_0(E), \quad t \geqslant 0. \tag{3.116}$$

We also define the operators $\{R(\lambda) : \lambda > 0\}$ as follows:

$$[R(\lambda)f](x) = \int_0^\infty e^{-\lambda t}[P(t)f](x)dt, \quad f \in C_0(E), \quad \lambda > 0. \tag{3.117}$$

From Proposition 3.42 it follows that the operators $P(t)$ leave $C_0(E)$ invariant and hence we also have $R(\lambda)C_0(E) \subseteq C_0(E)$. From the Markov property it follows that $\{P(t) : t \geqslant 0\}$ is a Feller semigroup and that the family $\{R(\lambda) : \lambda > 0\}$ is a resolvent family in the sense that

$$P(s+t) = P(s) \circ P(t), \quad s,\ t \geqslant 0, \tag{3.118}$$

$$R(\lambda_2) - R(\lambda_1) = (\lambda_1 - \lambda_2)\, R(\lambda_1) \circ R(\lambda_2), \quad \lambda_1,\ \lambda_2 > 0. \tag{3.119}$$

For $\lambda > 0$ fixed the operator $\widetilde{L}$ is defined in $C_0(E)$ as follows:

$$\widetilde{L} : R(\lambda)f \mapsto \lambda R(\lambda)f - f, \quad f \in C_0(E). \tag{3.120}$$

Here the domain $D(\widetilde{L})$ is given by $D(\widetilde{L}) = \{R(\lambda)f : f \in C_0(E)\}$. The operator $\widetilde{L}$ is well-defined. For, if f_1 and f_2 in $C_0(E)$ are such that $R(\lambda)f_1 = R(\lambda)f_2$, then by the resolvent property (3.119) we see $\mu R(\mu)f_1 = \mu R(\mu)f_2$, $\mu > 0$. Let μ tend ∞, to obtain $f_1 = f_2$. Since the operator $R(\lambda)$ is continuous, the operator $\widetilde{L}$ is closed. Next we shall prove that $\widetilde{L}$ is an extension of L. By partial integration, it follows that, for $f \in D(L)$,

$$\begin{aligned} e^{-\lambda t}&\left\{f(X(t)) - f(X(0)) - \int_0^t Lf(X(\tau))d\tau\right\} \\ &\quad + \lambda\int_0^t e^{-\lambda s}\left\{f(X(s)) - f(X(0)) - \int_0^s Lf(X(\tau))d\tau\right\} ds \\ &= e^{-\lambda t}f(X(t)) - f(X(0)) + \int_0^t e^{-\lambda s}(\lambda I - L)f(X(s))ds. \end{aligned} \tag{3.121}$$

As a consequence upon applying (3.121) once more, the processes

$$\left\{e^{-\lambda t}f(X(t)) - f(X(0)) + \int_0^t e^{-\lambda s}(\lambda I - L)f(X(s))ds : t \geqslant 0\right\}, \quad f \in D(L), \tag{3.122}$$

are $\mathbb{P}_x$-martingales for all $x \in E$. Here we employ the fact that the processes

$$\left\{f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds : t \geqslant 0\right\}, \quad f \in D(L),$$

are $\mathbb{P}_x$-martingales. This is part of assertion (ii). From assertion (3.122) it follows that

$$0 = \mathbb{E}_x\left(e^{-\lambda t}f(X(t)) - f(X(0)) + \int_0^t e^{-\lambda s}(\lambda I - L)f(X(s))ds\right), \quad f \in D(L). \tag{3.123}$$

Let t tend to infinity in (3.123) to obtain

$$0 = -\mathbb{E}_x\left(f(X(0))\right) + \int_0^\infty e^{-\lambda s}\mathbb{E}_x\left((\lambda I - L)f(X(s))\right)ds, \quad f \in D(L). \tag{3.124}$$

From (3.124) we obtain $f(x) = \int_0^\infty e^{-\lambda s}\left[P(s)(\lambda I - L)f\right](x)ds$, $f \in D(L)$. Or writing this differently $f = R(\lambda)(\lambda I - L)f$, $f \in D(L)$. Let f belong to $D(L)$. Then $f = R(\lambda)g$, with $g = (\lambda I - L)f$ and hence f belongs to $D(\widetilde{L})$. Moreover we see

$$\widetilde{L}f = \widetilde{L}(R(\lambda)g) = \lambda R(\lambda)g - g = \lambda f - (\lambda f - Lf) = Lf. \tag{3.125}$$

It follows that $\widetilde{L}$ is a closed linear extension of L. In addition we have $R(\lambda I - \widetilde{L}) = C_0(E)$. We shall show that the operator $\widetilde{L}$ verifies the maximum principle. This can be achieved as follows. Let f in $C_0(E)$ be such that, for some $x_0 \in E$,

$$\text{Re }(R(\lambda)f)(x_0) = \sup\{\text{Re } R(\lambda)f(x) : x \in E\} > 0. \tag{3.126}$$

Then Re $(R(\lambda)f)(x_0) \geqslant$ Re $R(\lambda)f(X(t))$, $t \geqslant 0$, and hence

$$\text{Re }(R(\lambda)f)(x_0) \geqslant \text{Re }\int_0^\infty e^{-\lambda s}\mathbb{E}_{X(t)}\left(f(X(s))\right)ds, \quad t \geqslant 0. \tag{3.127}$$

So that, upon employing the Markov property, we obtain for $t \geqslant 0$:

$$\begin{aligned}\operatorname{Re}\,(R(\lambda)f)(x_0) &\geqslant \operatorname{Re}\int_0^\infty e^{-\lambda s}\mathbb{E}_{x_0}\left(\mathbb{E}_{X(t)}\left(f(X(s))\right)\right)ds \\ &= \operatorname{Re}\,\mathbb{E}_{x_0}\left(\left[R(\lambda)f\right](X(t))\right). \end{aligned}\tag{3.128}$$

Hence, for $\mu > 0$, we obtain

$$\begin{aligned}\frac{1}{\mu}\operatorname{Re}\,(R(\lambda)f)(x_0) &= \int_0^\infty e^{-\mu t}\operatorname{Re}\,R(\lambda)f(x_0)dt \\ &\geqslant \operatorname{Re}\int_0^\infty e^{-\mu t}\mathbb{E}_{x_0}\left(\left[R(\lambda)f\right](X(t))\right)dt \\ &= \operatorname{Re}\,R(\mu)R(\lambda)f(x_0)\end{aligned}$$

(resolvent equation (3.119)

$$= \operatorname{Re}\,\frac{R(\lambda)f(x_0) - R(\mu)f(x_0)}{\mu - \lambda}. \tag{3.129}$$

Consequently: $[\lambda \operatorname{Re}\,R(\lambda)f](x_0) \leqslant \operatorname{Re}\,[\mu R(\mu)f](x_0)$, $\mu > \lambda$. Let μ tend to infinity, use right continuity of paths and the continuity of f to infer

$$\operatorname{Re}\,\widetilde{L}R(\lambda)f(x_0) = \operatorname{Re}\,\{\lambda R(\lambda)f(x_0) - f(x_0)\} \leqslant 0. \tag{3.130}$$

This proves that $\widetilde{L}$ verifies the maximum principle. Employing the implication (iii) $\Rightarrow$ (i) in Theorem 3.45 yields that $\widetilde{L}$ is the generator of a Feller semigroup. From Proposition 3.44 it then follows that for $\widetilde{L}$ the martingale problem is uniquely solvable. Since L solves the martingale problem maximally and since $\widetilde{L}$ extends L, it follows that $\widetilde{L} = \overline{L}$, the closure of L. Consequently the operator L is closable and its closure generates a Feller semigroup.

(i) $\Rightarrow$ (ii) Let the closure of L, $\overline{L}$, be the generator of a Feller semigroup. From Proposition 3.44 it follows that for $\overline{L}$ the martingale problem is uniquely solvable. Hence this is true for L. We still have to prove that L is maximal with respect to this property. Let L_1 be any closed linear extension of L for which the martingale problem is uniquely solvable. Define $\widetilde{L_1}$ in the same fashion as in the proof of the implication (ii) $\Rightarrow$ (i), with L_1 replacing L. Then $\widetilde{L}_1$ is a closed linear operator, which extends L_1. So that $\widetilde{L}_1$ extends $\overline{L}$. As in the proof of the implication (ii) $\Rightarrow$ (i) it also follows that $\widetilde{L}_1$ generates a Feller semigroup. Since, by (i), the closure of L, also generates a Feller semigroup, we conclude by uniqueness of generators, that $\overline{L} = \widetilde{L}_1$. Since $\widetilde{L}_1 \supseteq L_1 = \overline{L_1} \supseteq \overline{L} \supseteq L$, it follows that the closure of L coincides with L_1. This proves the maximality property of L, and so the proof of Theorem 3.45 is complete. □

In fact a careful analysis of the proof of Theorem 3.45 shows the following result.

3.47. Proposition. *Let L be a densely defined operator for which the martingale problem is uniquely solvable, and which is maximal for this property. Then there exists a unique closed linear extension L_0 of L, which is the generator of a Feller semigroup.*

Proof. *Existence*. Let $\{\mathbb{P}_x : x \in E\}$ be the solution for L, and assume that for all $f \in C_0(E)$ the function $x \mapsto [P(t)f](x)$ belongs to $C_0(E)$ for all $t \geqslant 0$. Here $P(t)f(x)$ is defined by

$$[P(t)f](x) = \mathbb{E}_x\left(f(X(t))\right), \quad [R(\lambda)f](x) = \int_0^\infty e^{-\lambda s}[P(s)f](x)ds,$$
$$L_0(R(\lambda)f) := \lambda R(\lambda)f - f, \quad f \in C_0(E).$$

Here $t \geqslant 0$ and $\lambda > 0$ is fixed. Then, as follows from the proof of Theorem 3.45, the operator L_0 generates a Feller semigroup.

Uniqueness. Let L_1 and L_2 be closed linear extensions of L, which both generate Feller semigroups. Let

$$\left\{(\Omega, \mathcal{F}, \mathbb{P}_x^1), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E})\right\}$$

respectively

$$\left\{(\Omega, \mathcal{F}, \mathbb{P}_x^2), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E})\right\}$$

be the corresponding Markov processes. For every $f \in D(L)$, the process $f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds$, $t \geqslant 0$, is a martingale with respect to $\mathbb{P}_x^1$ as well as with respect to $\mathbb{P}_x^2$. Uniqueness implies $\mathbb{P}_x^1 = \mathbb{P}_x^2$ and hence $L_1 = L_2$.

The proof of Proposition 3.47 is complete now. □

3.48. Corollary. *Let L be a densely defined linear operator with domain $D(L)$ and range $R(L)$ in $C_0(E)$. The following assertions are equivalent:*

(i) Some extension of L generates a Feller semigroup.

(ii) For some extension of L the martingale problem is uniquely solvable for every $x \in E$.

Proof. (i) $\Longrightarrow$ (ii). Let L_0 be an extension of L that generates a Feller semigroup. Let $\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E})\}$ be the corresponding Markov process. For $x \in E$ the probability $\mathbb{P}_x$ is the unique solution for the martingale problem starting in x.

(ii) $\Longrightarrow$ (i). Let L_0 be an extension of L for which the martingale problem is uniquely solvable for every $x \in E$. Also suppose that L_0 is maximal for this property. Let $\{\mathbb{P}_x : x \in E\}$ be the unique solution of the corresponding martingale problem. Define the operators $P(t)$, $t \geqslant 0$, by $[P(t)f](x) = \mathbb{E}_x(f(X(t)))$, $f \in C_0(E)$. From the proof of Theorem 3.45 it follows that $\{P(t) : t \geqslant 0\}$ is a Feller semigroup with generator L_0.

This completes the proof of Corollary 3.48. $\square$

3.49. Example. Let L_0 be an unbounded generator of a Feller semigroup in $C_0(E)$ and let μ_k and ν_k, $1 \leqslant k \leqslant n$, be finite (signed) Borel measures on E. Define the operator $L_{\vec{\mu},\vec{\nu}}$ as follows:

$$\begin{aligned} D\left(L_{\vec{\mu},\vec{\nu}}\right) &= \bigcap_{k=1}^{n} \left\{ f \in D(L_0) : \int L_0 f \, d\mu_k = \int f \, d\nu_k \right\}, \\ L_{\vec{\mu},\vec{\nu}} f &= L_0 f, \quad f \in D\left(L_{\vec{\mu},\vec{\nu}}\right). \end{aligned}$$

Then the martingale problem is uniquely solvable for $L_{\vec{\mu},\vec{\nu}}$. In fact let

$$\{(\Omega, \mathcal{F}, \mathbb{P}_x), (X(t) : t \geqslant 0), (\vartheta_t : t \geqslant 0), (E, \mathcal{E})\}$$

be the strong Markov process associated to the Feller semigroup generated by L_0. Then $\mathbb{P} = \mathbb{P}_x$ solves the martingale problem

(a) For every $f \in D(L_{\vec{\mu},\vec{\nu}})$ the process
$$f(X(t)) - f(X(0)) - \int_0^t L_{\vec{\mu},\vec{\nu}} f(X(s)) ds, \quad t \geqslant 0,$$
is a $\mathbb{P}$-martingale;

(b) $\mathbb{P}(X(0) = x) = 1$,

uniquely. This can be seen as follows. We may and do suppose that the functionals $f \mapsto \int L_0 f d\mu_k - \int f d\nu_k$, $f \in D(L_0)$, $1 \leqslant k \leqslant n$, are linearly independent. If some μ_k belongs to $D(L_0^*)$, then $D(L_{\vec{\mu},\vec{\nu}})$ is not dense and if none of the measures μ_k belongs to $D(L_0^*)$, then $D(L_{\vec{\mu},\vec{\nu}})$ is dense in $C_0(E)$. In either case there exists a unique extension, in fact L_0, of $L_{\vec{\mu},\vec{\nu}}$ which generates a Feller semigroup. Therefore we choose functions $u_k \in D(L_0)$, $1 \leqslant k \leqslant n$, in such a way that $\int L_0 u_k d\mu_\ell - \int u_k d\nu_\ell = \delta_{k,\ell}$, $1 \leqslant k, \ell \leqslant n$. Suppose that $\mathbb{P}_x^1$ and $\mathbb{P}_x^2$ are probabilities, that start in x, with the property that for all $f \in D(L_{\vec{\mu},\vec{\nu}})$ the process

$t \mapsto f(X(t)) - f(X(0)) - \int_0^t L_0 f(X(s))ds$ is a $\mathbb{P}_x^1$- as well as a $\mathbb{P}_x^2$-martingale. As in (3.110) we see that for all $f \in D(L_0)$ ($v_k = (\lambda I - L_0)\, u_k$, $1 \leqslant k \leqslant n$):

$$f(x) - \sum_{k=1}^n \left(\int L_0 f d\mu_k - \int f d\nu_k \right) u_k(x) \tag{3.131}$$
$$= \int_0^\infty e^{-\lambda s} \mathbb{E}_x^1 \left((\lambda I - L_0) \left(f - \sum_{k=1}^n \left(\int L_0 f d\mu_k - \int f d\nu_k \right) u_k \right) (X(s)) \right) ds$$
$$= \int_0^\infty e^{-\lambda s} \mathbb{E}_x^2 \left((\lambda I - L_0) \left(f - \sum_{k=1}^n \left(\int L_0 f d\mu_k - \int f d\nu_k \right) u_k \right) (X(s)) \right) ds.$$

Write $f = (\lambda I - L_0)^{-1} g = R(\lambda) g$. From (3.131) we obtain

$$R(\lambda)g(x) - \sum_{k=1}^n \left(\int (\lambda R(\lambda) g - g)\, d\mu_k - \int R(\lambda) g d\nu_k \right) u_k(x) \tag{3.132}$$
$$= \int_0^\infty e^{-\lambda s} \mathbb{E}_x^1 \left[g(X(s)) - \sum_{k=1}^n \left(\int (\lambda R(\lambda) g - g)\, d\mu_k - \int R(\lambda) g d\nu_k \right) v_k(X(s)) \right] ds$$
$$= \int_0^\infty e^{-\lambda s} \mathbb{E}_x^2 \left[g(X(s)) - \sum_{k=1}^n \left(\int (\lambda R(\lambda) g - g)\, d\mu_k - \int R(\lambda) g d\nu_k \right) v_k(X(s)) \right] ds.$$

Put $\overrightarrow{F}(\lambda) = (F_1(\lambda), \dots, F_n(\lambda))$ and put $U(\lambda) = (u_{k,\ell}(\lambda))$, where, for $1 \leqslant k \leqslant n$,

$$F_k(\lambda) = \int_0^\infty e^{-\lambda s} \left(\mathbb{E}_x^1 \left[(\lambda I - L_0) u_k(X(s)) \right] - \mathbb{E}_x^2 \left[(\lambda I - L_0) u_k(X(s)) \right] \right) ds,$$

and where $u_{k,\ell}$, $1 \leqslant k, \ell \leqslant n$, is given by

$$u_{k,\ell}(\lambda) = \int \lambda R(\lambda) u_\ell \, d\mu_k - \int u_\ell d\mu_k - \int R(\lambda) u_\ell \, d\nu_k.$$

Since (3.132) is valid for all $g \in C_0(E)$, it follows that $\overrightarrow{F}(\lambda) = U(\lambda)\overrightarrow{F}(\lambda)$. Since, in addition $\lim_{\lambda \to \infty} U(\lambda) = 0$, we see $F_k(\lambda) = 0$ for all $\lambda > 0$ and for $1 \leqslant k \leqslant n$. So that $\int_0^\infty e^{-\lambda s} \mathbb{E}_x^1 \left(u_k(X(s)) \right) ds = \int_0^\infty e^{-\lambda s} \mathbb{E}_x^2 \left[u_k(X(s)) \right] ds$ for all $\lambda > 0$ and for all $1 \leqslant k \leqslant n$. Again an application of (3.132) yields $\mathbb{E}_x^1 \left[g(X(s)) \right] = \mathbb{E}_x^2 \left[g(X(s)) \right]$ for all $g \in C_0(E)$. Since these arguments are valid for any $x \in E$, we conclude just as in Proposition 2.9 and its Corollary on page 206 of Ikeda and Watanabe [**61**]), that $\mathbb{P}_x^1 = \mathbb{P}_x^2 = \mathbb{P}_x$, $x \in E$, In particular we may take $E = [0,1]$, $L_0 f = \frac{1}{2} f''$, $D(L_0) = \{ f \in C^2[0,1] : f'(0) = f'(1) = 0 \}$, $\mu_k(I) = \int_{\alpha_k}^{\beta_k} 1_I(s) ds$, $\nu_k = 0$, $0 \leqslant \alpha_k < \beta_k \leqslant 1$, $1 \leqslant k \leqslant n$. Then L_0 generates the Feller semigroup of *reflected Brownian motion*: see Liggett [**86**], Example 5.8., p. 45. For the operator $L_{\vec{\mu},\vec{\nu}}$ the martingale problem is uniquely (but not maximally uniquely) solvable. However it does not generate a Feller semigroup. The previous arguments do not seem to be entirely correct. It ought to be replaced with some results in Section 10 (*e.g.* Theorem 3.110).

Problem. We want to close this section with the following question. Suppose that the operator L possesses a unique extension L_0, that generates a Feller semigroup. Is it true that for L the martingale problem is uniquely solvable?

In general the answer is no, but if we require that L solves the martingale problem maximally, then the answer is yes, provided as sample space we take the Skorohod space. This result is proved in Theorem 3.45.

For the time being we will not pursue the Markov property. However, we will continue with Brownian motion and stochastic integrals. First we give the definition of some interesting processes.

4. Martingales, submartingales, supermartingales and semimartingales

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{\mathcal{F}_t : t \geqslant 0\}$ be an increasing family of σ-fields in $\mathcal{F}$. If necessary we suppose that the filtration $\{\mathcal{F}_t : t \geqslant 0\}$ is right continuous, i.e. $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$, or complete in the sense that, for every $t > 0$, the σ-field $\mathcal{F}_t$ contains all $A \in \mathcal{F}$, with $\mathbb{P}(A) = 0$.

Let $\{H(t) : t \geqslant 0\}$ be a collection of $\mathbb{R}$-valued functions defined on Ω. Such a family is called a (real-valued) process.

3.50. Definition. The following processes and σ-fields will play a role in the sequel.

(a) The process $\{H(t) : t \geqslant 0\}$ is said to be *adapted* or *non-anticipating* if, for every $t \geqslant 0$, the variable $H(t)$ is measurable with respect to $\mathcal{F}_t$.
(b1 The symbol Λ denotes the σ-field ($= \sigma$-algebra) of subsets of $[0, \infty) \times \Omega$, which is generated by the adapted processes which are right-continuous and which possess left limits. These are the so-called *cadlag processes*.
(b2) The process $\{H(t) : t \geqslant 0\}$ is said to be *optional* if the function $(t, \omega) \to H(t, \omega)$ is measurable with respect to Λ.
(c1) The symbol Π denotes the σ-field of subsets of $[0, \infty) \times \Omega$, which is generated by the adapted processes which are left-continuous adapted processes.
(c2) The process $\{H(t) : t \geqslant 0\}$ is said to be *predictable* if the function $(t, \omega) \to H(t, \omega)$ is measurable with respect to Π.

3.51. Proposition. *The collection* $\{(a, b] \times A \colon 0 \leqslant a < b, A \in \mathcal{F}_a\}$ *generates the* σ*-field* Π.

Proof. Let A belong to $\mathcal{F}_a$. The variable $\omega \mapsto 1_{(a,b]}(s) 1_A(\omega)$ is measurable with respect to $\mathcal{F}_s$ and the function $s \mapsto 1_{(a,b]}(s) 1_A(\omega)$ is left continuous. This proves that $(a, b] \times A$ belongs to Π.

Conversely let F be adapted and left continuous. Put

$$F_n(s, \omega) = \sum\nolimits_{k=0}^{\infty} F\left(k2^{-n}, \omega\right) 1_{(k2^{-n},(k+1)2^{-n}]}(s) = F\left(\left(\lceil 2^n s\rceil - 1\right) 2^{-n}, \omega\right).$$

Then, by left continuity, $\lim_{n\to\infty} F_n(s, \omega) = F(s, \omega)$, $\mathbb{P}$-almost surely. Moreover the processes $\{F_n(t) : t \geqslant 0\}$ are adapted and are measurable with respect to the σ-field generated by $\{(a, b] \times A : 0 \leqslant a < b, A \in \mathcal{F}_a\}$. All this completes the proof of Proposition 3.51. □

3.52. Remark. Since $1_{(a,b]}(s) \times 1_A(\omega) = \lim\limits_{n\to\infty} 1_{[a_n,b_n)}(s) 1_A(\omega)$, where $a_n \downarrow a$ and where $b_n \downarrow b$, it follows that $\Pi \subseteq \Lambda$. Here we employ Proposition 3.51.

3.53. Definition. Let $\{X(t) : t \geqslant 0\}$ be an adapted process.

(a) The family $\{X(t) : t \geqslant 0\}$ is a martingale if $\mathbb{E}\left(|X(t)|\right) < \infty$, $t \geqslant 0$, and if, for every $t > s \geqslant 0$, $\mathbb{E}\left(X(t) \mid \mathcal{F}_s\right) = X(s)$, $\mathbb{P}$-almost surely.

(b) The family $\{X(t) : t \geqslant 0\}$ is a submartingale if $\mathbb{E}(|X(t)|) < \infty$, $t \geqslant 0$, and if, for every $t > s \geqslant 0$, $\mathbb{E}(X(t) \mid \mathcal{F}_s) \geqslant X(s)$, $\mathbb{P}$-almost surely.
(c) The family $\{X(t) : t \geqslant 0\}$ is a supermartingale if $\mathbb{E}(|X(t)|) < \infty$, $t \geqslant 0$, and if, for every $t > s \geqslant 0$, $\mathbb{E}(X(t) \mid \mathcal{F}_s) \leqslant X(s)$, $\mathbb{P}$-almost surely.
(d) It is $\mathbb{P}$-almost surely of finite variation (on $[0,t]$) if

$$\sup\left\{\sum\nolimits_{j=1}^{n} |X(t_j) - X(t_{j-1})| : 0 \leqslant t_0 < t_1 < \ldots < t_n < \infty\right\} < \infty,$$
$$\left(\sup\left\{\sum\nolimits_{j=1}^{n} |X(t_j) - X(t_{j-1})| : 0 \leqslant t_0 < t_1 < \ldots < t_n \leqslant t\right\} < \infty,\right)$$

$\mathbb{P}$-almost surely.
(e) It is a local martingale if there exists an increasing sequence of stopping times $(T_n : n \in \mathbb{N})$ for which $\lim_{n\to\infty} T_n = \infty$, $\mathbb{P}$-almost surely, and for which the processes

$$\{X(T_n \wedge t) : t \geqslant 0\}, \quad n = 1, 2, \ldots$$

are martingales with respect to the filtration $\{\mathcal{F}_{T_n \wedge t} : t \geqslant 0\}$.
(f) Let T be a stopping time. The process $\{X(t) : t \geqslant 0\}$ is a local martingale on $[0,T)$ if there exists a sequence of stopping times $(T_n : n \in \mathbb{N})$ which is increasing for which $\lim_{n\to\infty} T_n = T$, $\mathbb{P}$-almost surely, and for which the processes $\{X(T_n \wedge t) : t \geqslant 0\}$, $n = 1, 2, \ldots$ are martingales with respect to the filtration $\{\mathcal{F}_{T_n \wedge t} : t \geqslant 0\}$.
(g) The definition of "local submartingale", "local supermartingale" and "being locally $\mathbb{P}$-almost surely of finite variation" are now self-explanatory.
(h) The process $\{X(t) : t \geqslant 0\}$ is called a semi-martingale if it can be written in the form $X(t) = M(t) + A(t)$, where $\{M(t) : t \geqslant 0\}$ is a martingale and where $\{A(t) : t \geqslant 0\}$ is an adapted process which is finite variation, $\mathbb{P}$-almost surely, on $[0,t]$ for every $t > 0$, and for which $\mathbb{E}|A(t)| < \infty$, $t \geqslant 0$.
(i) The process $\{X(t) : t \geqslant 0\}$ is of class (DL) if for every $t > 0$ the family

$$\{X(\tau) : 0 \leqslant \tau \leqslant t, \tau \text{ is a } (\mathcal{F}_t)\text{-stopping time}\}$$

is uniformly integrable.

3.54. Remark. Let $\{X(t) : t \geqslant 0\}$ be a semi-martingale. The decomposition $X(t) = M(t) + A(t)$, where $\{M(t) : t \geqslant 0\}$ is a martingale and where for every $t > 0$ the process $\{A(t) : t \geqslant 0\}$ is $\mathbb{P}$-almost surely of finite variation and where $\{A(t) : t \geqslant 0\}$ is predictable and right continuous $\mathbb{P}$-almost surely is unique, provided $A(0) = 0$, $\mathbb{P}$-almost surely. This follows from the fact that a right-continuous martingale which is predictable and of finite variation is necessarily constant: this is a consequence of the uniqueness part of the Doob-Meyer decomposition: see Theorem 1.24. A proof of the Doob-Meyer decomposition theorem may start as follows. Put

$$X_j(t) = \mathbb{E}\left[X\left(\frac{\lceil 2^j t\rceil}{2^j}\right) \Bigm| \mathcal{F}_t\right] \quad \text{and} \tag{3.133}$$
$$A_j(t) = A_j(0) + \sum_{0\leqslant k<2^j t} \mathbb{E}\left[X\left(\frac{k+1}{2^j}\right) - X\left(\frac{k}{2^j}\right) \Bigm| \mathcal{F}_{k2^{-j}}\right],$$

and prove $M_j(t) := X_j(t) - A_j(t)$ is a martingale. Then let $j \to \infty$ to obtain: $X(t) = M(t) + A(t)$, where $M(t) = \lim_{j\to\infty} M_j(t)$ and $A(t) = \lim_{j\to\infty} A_j(t)$.

3.55. Remark. An $\mathcal{F}_t$-martingale $\{M(t) : t \geqslant 0\}$ is of class (DL), an increasing adapted process $\{A(t) : t \geqslant 0\}$ in $L^1(\Omega, \mathcal{F}, \mathbb{P})$ is of class (DL) and hence the sum

$$\{M(t) + A(t) : t \geqslant 0\}$$

is of class (DL). If $\{X(t) : t \geqslant 0\}$ is a submartingale and if μ is a real number, then the process $\{\max(X(t), \mu) : t \geqslant 0\}$ is a submartingale of class (DL). Processes of class (DL) are important in the Doob-Meyer decomposition theorem.

We continue with some examples of martingales, submartingales and the like.

3.56. Example. Let $T : \Omega \to [0, \infty]$ be a stopping time. Since T is a stopping time and since the process $\left\{1_{\{T<t\}} : t \geqslant 0\right\}$ is left continuous, it is predictable. It follows that the process $\left\{1_{\{T\geqslant t\}} : t \geqslant 0\right\}$ is predictable as well.

3.57. Example. Let I be an open interval in $\mathbb{R}$ and let $\varphi : I \to (-\infty, \infty)$ be an increasing convex function. If $\{X(t) : t \geqslant 0\}$ is a submartingale with values in I, then the process $\{\varphi(X(t)) : t \geqslant 0\}$ is also a submartingale. For let $t > s \geqslant 0$. Then by the Jensen inequality and the monotonicity of φ it follows that

$$\mathbb{E}\left[\varphi(X(t)) \mid \mathcal{F}_s\right] \geqslant \varphi\left[\mathbb{E}\left(X(t) \mid \mathcal{F}_s\right)\right] \geqslant \varphi\left(X(s)\right).$$

3.58. Example. Let $(B(t), \mathbb{P}_0)$ be one-dimensional Brownian motion starting in 0. Then $\{B(t) : t \geqslant 0\}$ is a martingale. Since the definition of martingale also makes sense for vector valued processes, we also see that an $\mathbb{R}^\nu$-valued Brownian motion is a martingale.

3.59. Example. Let $(B(t), \mathbb{P}_0)$ be $\mathbb{R}^\nu$-valued Brownian motion starting in 0. The process $\left\{|B(t)|^2 - \nu t : t \geqslant 0\right\}$ is a martingale.

3.60. Example. Let $\{X(t), \mathbb{P}_x\}$ be a (strong) Markov process such that

$$\mathbb{E}_x\left[f(X(t))\right] = \int p(t, x, y) f(y) dm(y), \quad f \geqslant 0,$$

where the density $p(t, x, y)$ verifies the Chapman-Kolmogorov identity:

$$p(s + t, x, y) = \int p(s, x, z) p(t, z, y) dm(z).$$

The process $\{p(t - s, X(s), y) : 0 \leqslant s < t\}$ is a martingale on $[0, t)$. For example for $X(t)$ we may take $B(t)$, d-dimensional Brownian motion. Then

$$p(t, x, y) = p_d(t, x, y) = \frac{1}{\left(\sqrt{2\pi t}\right)^d} \exp\left(-\frac{|x - y|^2}{2t}\right).$$

3.61. Example. Let $\{X(t) : t \geqslant 0\}$ be a right-continuous martingale and let T be a stopping time. The process $\{X(T \wedge t) : t \geqslant 0\}$ is a martingale with respect to $\{\mathcal{F}_t : t \geqslant 0\}$ and also with respect to the filtration $\{\mathcal{F}_{T\wedge t} : t \geqslant 0\}$.

3.62. Example. This is a standard example of a closed martingale, *i.e.* a martingale which is written as conditional expectations on σ-fields taken from a filtration. Let Y be an random variable in $L^1(\Omega, \mathcal{F}, \mathbb{P})$. The process $s \mapsto \mathbb{E}[Y \mid \mathcal{F}_s]$, $s \geqslant 0$, is a martingale.

We want to insert an inequality on the second moment of a martingale. This is a special case of the Burkholder-Davis-Gundy inequality.

3.63. Proposition. *Let $\{M(t) : t \geqslant 0\}$ be a continuous martingale with $M(0) = 0$. Then*

$$\mathbb{E}\left(M(t)^2\right) \leqslant \mathbb{E}\left(M^*(t)^2\right) \leqslant 4\mathbb{E}\left(M(t)^2\right).$$

Here $M^*(t) = \sup_{0 \leqslant s \leqslant t} |M(s)|$.

Proof. Define for $\xi > 0$ the stopping time T_ξ by

$$T_\xi = \inf\{t > 0 : M^*(t) \geqslant \xi\}.$$

Then $\{M^*(t) > \xi\} \subseteq \{T_\xi < t\}$ and $\{T_\xi < t\} \subseteq \{M^*(t) \geqslant \xi\}$ and hence, since $|M(t)|$ is a submartingale we obtain upon using Doob's optional sampling

$$\mathbb{E}\left(M^*(t)^2\right) = \int_0^\infty \mathbb{P}\left(M^*(t)^2 > \lambda\right) d\lambda$$

(make the substitution $\lambda = \xi^2$)

$$\begin{aligned} &= 2\int_0^\infty \xi \mathbb{P}\left(M^*(t) > \xi\right) d\xi \leqslant 2\int_0^\infty \xi \mathbb{P}\left(T_\xi < t\right) d\xi \\ &= 2\int_0^\infty \mathbb{E}\left(|M(T_\xi)| : T_\xi < t\right) d\xi \end{aligned}$$

(Doob's optional sampling)

$$\begin{aligned} &\leqslant 2\int_0^\infty \mathbb{E}\left(|M(t)| : T_\xi < t\right) d\xi \\ &= 2\int_0^\infty \mathbb{E}\left(|M(t)| : M^*(t) \geqslant \xi\right) d\xi \\ &= 2\mathbb{E}\left(|M(t)|\, M^*(t)\right) \end{aligned}$$

(Cauchy-Schwarz' inequality)

$$\leqslant 2\left(\mathbb{E}\left(M(t)^2\right)\right)^{1/2}\left(\mathbb{E}\left(M^*(t)^2\right)\right)^{1/2}.$$

Consequently $\mathbb{E}\left(M^*(t)^2\right) \leqslant 4\mathbb{E}\left(M(t)^2\right)$. This completes the proof of Proposition 3.63. □

3.64. Remark. The method of works very well if $\mathbb{E}\left(M^*(t)^2\right)$ is finite. If this is not the case we may use a localization technique. The reader should provide the details. Perhaps truncating is also possible.

5. Regularity properties of stochastic processes

In Theorem 3.18 we proved that Brownian motion possesses a continuous version. We want to amplify this result. In fact we shall prove that Brownian motion has Hölder continuous paths of any order $\alpha < \frac{1}{2}$. This means that for every $\alpha < \frac{1}{2}$ and for every $a > 0$, $a \in \mathbb{R}$, there exists a random variable $C(b)$, depending on Brownian motion such that for all $0 \leqslant s < t \leqslant a$, the inequality

$$|b(t) - b(s)| \leqslant C(b)\,|t-s|^\alpha$$

holds $\mathbb{P}$-almost surely. This will be the content of Theorem 3.67 below. We begin with a rather general result, due to Kolmogorov, for arbitrary stochastic processes.

3.65. THEOREM. *Fix a finite interval $[a, b]$. Let $\{X(s) : a \leqslant s \leqslant b\}$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that there exist constants K, r and p, such that $0 < r < p < \infty$ and such that*

$$\mathbb{E}\left(|X(t) - X(s)|^p\right) \leqslant K\,|t-s|^{1+r} \tag{3.134}$$

for all $a \leqslant s, t \leqslant b$. Fix $0 < \alpha < r/p$. Then there exists a random variable $C(X)$, which is finite $\mathbb{P}$-almost surely, such that

$$|X(t) - X(s)| \leqslant C(X)\,|t-s|^{\alpha} \tag{3.135}$$

for all dyadic rational numbers s and t in the interval $[a, b]$. In particular it follows that a process $X = \{X(s) : a \leqslant s \leqslant b\}$ verifying (3.134) has a Hölder continuous version of order α, $\alpha < r/p$.

PROOF. It suffices to prove (3.135), because the version problem can be taken care of as in Theorem 3.18. Without loss of generality we may and do suppose that $a = 0$ and that $b = 1$. Otherwise we consider the process Y defined by $Y(s) = X\left((a_0 + s(b_0 - a_0)\right)$, $0 \leqslant s \leqslant 1$, where a_0 and b_0 are dyadic rational with $a_0 \leqslant a$ and with $b \leqslant b_0$ and where outside of the interval $[a, b]$ the process X is defined by $X(t) = X(a)$, if $a_0 \leqslant t \leqslant a$, and $X(t) = X(b)$, if $b \leqslant t \leqslant b_0$. Put $\epsilon = r - \alpha p$. Then

$$\mathbb{P}\left(|X(t) - X(s)| \geqslant |t-s|^{\alpha}\right) \leqslant |t-s|^{-\alpha p}\,\mathbb{E}\left(|X(t) - X(s)|^p\right) \leqslant K\,|t-s|^{1+\epsilon}, \tag{3.136}$$

so that

$$\mathbb{P}\left(\left|X\left(\frac{k+1}{2^n}\right) - X\left(\frac{k}{2^n}\right)\right| \geqslant 2^{-n\alpha}\right) \leqslant K 2^{-n} 2^{-n\epsilon}. \tag{3.137}$$

Hence

$$\begin{aligned}\sum_{n=1}^{\infty}\sum_{k=0}^{2^n-1}&\mathbb{P}\left(\left|X\left(\frac{k+1}{2^n}\right) - X\left(\frac{k}{2^n}\right)\right| \geqslant 2^{-n\alpha}\right)\\ &\leqslant K\sum_{n=1}^{\infty} 2^n 2^{-n} 2^{-n\epsilon} = \frac{K}{2^{\epsilon}-1}.\end{aligned} \tag{3.138}$$

By the Borel-Cantelli lemma it follows that

$$\mathbb{P}\left(\bigcup_{m=1}^{\infty}\bigcap_{n\geqslant m}\left\{\max_{0\leqslant k\leqslant 2^n-1}\left|X\left(\frac{k+1}{2^n}\right) - X\left(\frac{k}{2^n}\right)\right| \leqslant 2^{-n\alpha}\right\}\right) = 1. \tag{3.139}$$

Hence there exists a random integer $\nu(X)$ with the following property: For $\mathbb{P}$-almost all ω the inequality

$$\max_{0\leqslant k\leqslant 2^n-1}\left|X\left(\frac{k+1}{2^n}\right) - X\left(\frac{k}{2^n}\right)\right| \leqslant 2^{-n\alpha} \tag{3.140}$$

is valid for $n \geqslant \nu(X)$. Next let $n \geqslant \nu(X)$ and let t be a dyadic rational in the interval $[k2^{-n}, (k+1)2^{-n}]$. Write $t = k2^{-n} + \sum_{j=1}^{m}\gamma_j 2^{-n-j}$, each γ_j equals 0 or 1. Then

$$\left|X(t) - X\left(\frac{k}{2^n}\right)\right| \leqslant \sum_{j=1}^{m}\frac{\gamma_j}{2^{\alpha(n+j)}} \leqslant \frac{1}{2^{\alpha}-1}\frac{1}{2^{n\alpha}}. \tag{3.141}$$

Similarly we have, with $t = \ell 2^{-N}$, $N \geqslant n$, $(k+1)2^{-n} = \ell 2^{-N} + \sum_{j=1}^{m'} \gamma'_j 2^{-N-j}$, γ'_j equals 0 or 1,

$$\left| X(t) - X\left(\frac{k+1}{2^n}\right)\right| \leqslant \sum_{j=1}^{m'} \frac{\gamma'_j}{2^{\alpha(N+j)}} \leqslant \frac{1}{2^\alpha - 1}\frac{1}{2^{N\alpha}} \leqslant \frac{1}{2^\alpha - 1}\frac{1}{2^{n\alpha}}. \tag{3.142}$$

Next let s and t be dyadic rationale numbers with $0 < t - s \leqslant 2^{-\nu(X)}$. Take $n \in \mathbb{N}$ with $2^{-n-1} \leqslant t-s < 2^{-n}$ and pick k in such a way that $k2^{-n-1} \leqslant s < (k+1)2^{-n-1}$. Then $(k+1)2^{-n-1} \leqslant t = t-s+s < 2^{-n} + (k+1)2^{-n-1} = (k+3)2^{-n-1}$. It follows that, since t belongs to $[(k+1)2^{-n-1}, (k+2)2^{-n-1}]$ or to the interval $[(k+2)2^{-n-1}, (k+3)2^{-n-1}]$,

$$\begin{aligned} &|X(t) - X(s)| \\ &\leqslant \left| X(t) - X\left(\frac{k+2}{2^{n+1}}\right)\right| + \left| X\left(\frac{k+2}{2^{n+1}}\right) - X\left(\frac{k+1}{2^{n+1}}\right)\right| + \left| X\left(\frac{k+1}{2^{n+1}}\right) - X(s)\right| \\ &\leqslant \frac{3}{2^\alpha - 1} 2^{-(n+1)\alpha} \leqslant \frac{3}{2^\alpha - 1}|t-s|^\alpha. \end{aligned} \tag{3.143}$$

If $1 \geqslant t-s > 2^{-\nu(X)}$, we choose k and $\ell \in \mathbb{N}$ in such a way that $2^{\nu(X)} > \ell > k \geqslant 0$ and that $\ell 2^{-\nu(X)} < t \leqslant (\ell+1)2^{-\nu(X)}$ and $k2^{-\nu(X)} < s \leqslant (k+1)2^{-\nu(X)}$. Then we get

$$\begin{aligned} |X(t) - X(s)| &\leqslant \left| X(t) - X\left(\frac{\ell+1}{2^{\nu(X)}}\right)\right| + \sum_{j=k}^{j=\ell} \left| X\left(\frac{j+1}{2^{\nu(X)}}\right) - X\left(\frac{j}{2^{\nu(X)}}\right)\right| \\ &\quad + \left| X\left(\frac{k}{2^{\nu(X)}}\right) - X(s)\right| \\ &\leqslant \frac{2 + 2^{\nu(X)}}{2^\alpha - 1} 2^{-\alpha\nu(X)} \leqslant \frac{2 + 2^{\nu(X)}}{2^\alpha - 1}|t-s|^\alpha. \end{aligned} \tag{3.144}$$

From (3.143) and (3.144) the result in Theorem 3.65 follows. □

In order to apply the previous result to Brownian motion, we insert a general equality for a Gaussian variable X.

3.66. Proposition. *Let $X : \Omega \to \mathbb{R}$ be a non-constant Gaussian variable. Then its distribution is given by*

$$\mathbb{P}(X \in B) = \frac{1}{\left(2\pi \mathbb{E}\left(X^2 - (\mathbb{E}(X))^2\right)\right)^{1/2}} \int_B \exp\left(-\frac{1}{2}\frac{|x - \mathbb{E}(X)|^2}{\mathbb{E}(X^2) - (\mathbb{E}(X))^2}\right) dx) \tag{3.145}$$

and its moments $\mathbb{E}(|X - \mathbb{E}(X)|^p)$, $p > -1$, are given by

$$\mathbb{E}(|X - \mathbb{E}(X)|^p) = \frac{2^{\frac{1}{2}p}\Gamma\left(\frac{1}{2}p + \frac{1}{2}\right)}{\sqrt{\pi}} \left(\sqrt{\mathbb{E}\left(X^2 - (\mathbb{E}(X))^2\right)}\right)^p. \tag{3.146}$$

PROOF. Equality (3.145) follows from formula (3.8) and formula (3.146) is proved by using (3.145). The formal arguments read (we write $Y = X - \mathbb{E}(X)$):

$$\begin{aligned}
\mathbb{E}\left(|Y|^p\right) &= \frac{1}{\left(2\pi\mathbb{E}\left(Y^2\right)\right)^{1/2}} \int |y|^p \exp\left(-\frac{1}{2}\frac{y^2}{\mathbb{E}(Y^2)}\right) dy \\
&= \left(\sqrt{\mathbb{E}(Y^2)}\right)^p \frac{2}{\sqrt{2\pi}} \int_0^\infty y^p \exp\left(-\frac{1}{2}y^2\right) dy \\
&= \left(\sqrt{\mathbb{E}(Y^2)}\right)^p \frac{2^{\frac{1}{2}p}\Gamma\left(\frac{1}{2}p + \frac{1}{2}\right)}{\sqrt{\pi}}.
\end{aligned}$$

The latter is the same as (3.146). $\square$

3.67. THEOREM. *Let $\{b(s) : s \geqslant 0\}$ be d-dimensional Brownian motion. This process is $\mathbb{P}$-almost surely Hölder continuous of order α for any $\alpha < \frac{1}{2}$.*

PROOF. It suffices to prove Theorem 3.67 for 1-dimensional Brownian motion. So suppose $d = 1$ and let $\alpha < 1/2$. Choose $p > 1$ so large that $\alpha < \dfrac{1}{2} - \dfrac{1}{p}$. From inequality (3.146) in Proposition 3.66 with $X = b(t) - b(s)$ we obtain

$$\begin{aligned}\mathbb{E}\left(|b(t)-b(s)|^p\right) &= \mathbb{E}\left(|b(t-s)|^p\right) = C_p\left(\mathbb{E}\left(|b(t-s)|^2\right)\right)^{p/2}\\ &= C_p\,|t-s|^{p/2} = C_p\,|t-s|^{1+r}, \end{aligned}\tag{3.147}$$

where $r = p/2 - 1 > p\alpha$. An application of Theorem 3.65 yields the desired result. □

The following theorem says that Brownian motion is nowhere differentiable.

3.68. THEOREM. *Fix $\alpha > \frac{1}{2}$. Then with probability one, $t \mapsto b(t)$ is nowhere Hölder continuous of order α. More precisely*

$$\mathbb{P}\left(\inf_{0\leqslant t\leqslant 1}\left[\limsup_{h\to 0}\frac{|b(t+h)-b(t)|}{|h|^{\alpha}}\right] = \infty\right) = 1.$$

PROOF. For a proof we refer the reader to the literature; e.g. Simon [[**121**], Theorem 5.4. p. 46]. □

In the theory of stochastic integration we will have a need for the following lemma. The following lemma can also be proved by the strong law of large numbers: see *e.g.* Smythe [**123**].

3.69. LEMMA. *Let $\{b(s) : s \geqslant 0\}$ be one-dimensional Brownian motion. Then, $\mathbb{P}$-almost surely, $\lim_{n\to\infty}\sum_{k=0}^{2^n-1}\left|b\left((k+1)2^{-n}t\right) - b\left(k2^{-n}t\right)\right|^2 = t$.*

PROOF. Put $\triangle_{k,n} = \left|b\left((k+1)2^{-n}t\right) - b\left(k2^{-n}t\right)\right|^2 - 2^{-n}t$. Then the variables $\triangle_{k,n}$, $0 \leqslant k \leqslant 2^n - 1$, are independent and have expectation 0. So that

$$\begin{aligned}\mathbb{E}\left(\sum_{k=0}^{2^n-1}\triangle_{k,n}\right)^2 &= \sum_{k=0}^{2^n-1}\mathbb{E}\left(\triangle_{k,n}\right)^2 = \sum_{k=0}^{2^n-1}\mathbb{E}\left(\left|b\left(2^{-n}t\right)\right|^2 - 2^{-n}t\right)^2\\ &= 2^n\left(\mathbb{E}\left|b\left(2^{-n}t\right)\right|^4 - 2\mathbb{E}\left|b\left(2^{-n}t\right)\right|^2 2^{-n}t + 2^{-2n}t^2\right) = 2\times 2^{-n}t^2. \end{aligned}\tag{3.148}$$

Tchebychev's inequality gives

$$\mathbb{P}\left(\left(\sum_{k=0}^{2^n-1}\triangle_{k,n}\right)^2 > \epsilon\right) \leqslant \frac{2}{\epsilon}t^2 2^{-n}.$$

Hence $\displaystyle\sum_{n=1}^{\infty}\mathbb{P}\left(\left(\sum_{k=0}^{2^n-1}\triangle_{k,n}\right)^2 > \epsilon\right) \leqslant \frac{2t^2}{\epsilon}$. Thus we may apply the Borel-Cantelli lemma to prove the claim in Lemma 3.69. □

3.70. PROPOSITION. *Brownian motion is nowhere of bounded variation.*

PROOF. Just as in the previous lemma we have that, for $t > s \geqslant 0$,

$$\lim_{n\to\infty} \sum_{k=0}^{2^n-1} \left|b\left(s+(k+1)2^{-n}(t-s)\right) - b\left(s+k2^{-n}(t-s)\right)\right|^2 - (t-s) = 0,$$

$\mathbb{P}$-almost surely. Since Brownian paths are almost surely continuous it follows that (for $\delta > 0$)

$$\begin{aligned}
0 < t-s &\leqslant \lim_{n\to\infty} \sum_{k=0}^{2^n-1} \left|b\left(s+(k+1)2^{-n}(t-s)\right) - b(\left(s+k2^{-n}(t-s)\right)\right|^2 \\
&\leqslant \liminf_{n\to\infty} \max_{0\leqslant \ell\leqslant 2^n} \left|b\left(s+(\ell+1)2^{-n}(t-s)\right) - b(\left(s+\ell 2^{-n}(t-s)\right)\right| \\
&\quad \times \sum_{k=0}^{2^n-1} \left|b\left(s+(k+1)2^{-n}(t-s)\right) - b(\left(s+k2^{-n}(t-s)\right)\right| \\
&\leqslant \sup_{s\leqslant\sigma_1,\sigma_2\leqslant t, |\sigma_2-\sigma_1|\leqslant\delta} |b(\sigma_2)-b(\sigma_1)| \times \text{variation of } b \text{ on the interval } [s,t].
\end{aligned}$$

The statement in the Proposition 3.70 now follows from the continuity of paths. □

Next we will see how to transfer properties of discrete time semi-martingales to continuous time semi-martingales. Most of the results in the remainder of this section are taken from Bhattacharya and Waymire [**15**]. We begin with an upcrossing inequality for a discrete time sub-martingale. Consider a sequence $\{Z_n : n \in \mathbb{N}\}$ of real-valued random variables and sigma-fields $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$, such that, for every $n \in \mathbb{N}$, the variable Z_n is $\mathcal{F}_n$-measurable. An *upcrossing* of an interval (a,b) by $\{Z_n\}$ is a passage to a value equal to or exceeding b from an value equal to or below a at an earlier time. Define the random variables X_n, $n \in \mathbb{N}$, by $X_n = \max(Z_n - a, 0)$. If the process $\{Z_n\}$ is a sub-martingale, then so is the process $\{X_n\}$. The upcrossings of $(0, b-a)$ by $\{X_n\}$ are the upcrossings of the interval (a,b) by $\{Z_n\}$. We define the successive upcrossing times η_{2k}, $k \in \mathbb{N}$, of $\{X_n\}$ as follows:

$$\begin{aligned}
\eta_1 &= \inf\{n \geqslant 1 : X_n = 0\}; \\
\eta_2 &= \inf\{n \geqslant \eta_1 : X_n \geqslant b-a\}; \\
\eta_{2k+1} &= \inf\{n \geqslant \eta_{2k} : X_n = 0\}; \\
\eta_{2k+2} &= \inf\{n \geqslant \eta_{2k+1} : X_n \geqslant b-a\}.
\end{aligned}$$

Then every η_k is an $\{\mathcal{F}_n\}$-stopping time. Fix $N \in \mathbb{N}$ and put $\tau_k = \min(\eta_k, N)$. Then every τ_k is also a stopping time and $\tau_k = N$ for $k > \lfloor N/2 \rfloor$, the largest integer smaller than or equal to $N/2$. It follows that $X_{\tau_{2k}} = X_N$ for $k > \lfloor n/2 \rfloor$ and we also have $\eta_k \geqslant k$ and so $k \leqslant \tau_k \leqslant N$. Let $U_N(a,b)$ be the number of upcrossings of (a,b) by the process $\{Z_n\}$ at time N. That means

$$U_N(a,b) = \sup\{k \geqslant 1 : \eta_{2k} \leqslant N\} \tag{3.149}$$

with the convention that the supremum over the empty set is 0. Notice that $U_N(a,b)$ is also the number of upcrossings of the interval $(0, b-a)$ by $\{X_n\}$ in time N.

3.71. Proposition (Upcrossing inequality). *Let $\{Z_n\}$ be an $\{\mathcal{F}_n\}$-submartingale. For each pair (a,b), $a<b$, the expected number of upcrossings of (a,b) by $Z_1,\dots,Z_N$ satisfies the inequality:*

$$\begin{aligned}\mathbb{E}\left(U_N(a,b)\right) &\leqslant \frac{\mathbb{E}\left(\max\left(Z_N-a,0\right)-\max\left(Z_1-a,0\right)\right)}{b-a}\\ &\leqslant \frac{\mathbb{E}\left(\max\left(Z_N-Z_1,0\right)\right)}{b-a}.\end{aligned} \tag{3.150}$$

Proof. Since $X_{\tau_{2k}} = X_N$ for $k > \lfloor N/2\rfloor$, we may write (setting $\tau_0 = 1$):

$$X_N - X_1 = \sum_{k=1}^{\lfloor N/2\rfloor+1}\left(X_{\tau_{2k-1}} - X_{\tau_{2k-2}}\right) + \sum_{k=1}^{\lfloor N/2\rfloor+1}\left(X_{\tau_{2k}} - X_{\tau_{2k-1}}\right). \tag{3.151}$$

Next let ν be the largest integer k with the property that $\eta_k \leqslant N$, i.e. ν is the last time $\leqslant N$ of an upcrossing or a downcrossing. It readily follows that $U_N(a,b) = \lfloor \nu/2\rfloor$. If ν is *even*, then

$$\begin{aligned}&X_{\tau_{2k}} - X_{\tau_{2k-1}} \geqslant b-a \quad \text{provided} \quad 2k-1<\nu;\\ &X_{\tau_{2k}} - X_{\tau_{2k-1}} = X_N - X_N = 0 \quad \text{provided} \quad 2k-1>\nu.\end{aligned} \tag{3.152}$$

Now suppose that ν is *odd*. Then we have

$$\begin{aligned}&X_{\tau_{2k}} - X_{\tau_{2k-1}} \geqslant b-a \quad \text{provided} \quad 2k-1<\nu;\\ &X_{\tau_{2k}} - X_{\tau_{2k-1}} = X_{\tau_{2k}} - X_{\nu} \geqslant X_{\tau_{2k}} - 0 = X_{\tau_{2k}} \quad \text{provided} \quad 2k-1=\nu;\\ &X_{\tau_{2k}} - X_{\tau_{2k-1}} = X_N - X_N = 0 \quad \text{provided} \quad 2k-1>\nu.\end{aligned} \tag{3.153}$$

From (3.152) and (3.153) it follows that

$$\begin{aligned}\sum_{k=1}^{\lfloor N/2\rfloor+1}\left(X_{\tau_{2k}} - X_{\tau_{2k-1}}\right) &\geqslant \sum_{k=1}^{\lfloor \nu/2\rfloor}\left(X_{\tau_{2k}} - X_{\tau_{2k-1}}\right)\\ &\geqslant \lfloor \nu/2\rfloor(b-a) = (b-a)U_N(a,b).\end{aligned} \tag{3.154}$$

Consequently

$$X_N - X_1 \geqslant \sum_{k=1}^{\lfloor N/2\rfloor+1}\left(X_{\tau_{2k-1}} - X_{\tau_{2k-2}}\right) + (b-a)U_N(a,b). \tag{3.155}$$

So far we did not make use of the fact that the process $\{X_n\}$ is a sub-martingale. It then follows that the process $\{X_{\tau_k} : k\in\mathbb{N}\}$ is a $\{\mathcal{F}_{\tau_n}\}$-martingale and hence $k \mapsto \mathbb{E}\left(X_{\tau_k}\right)$ is an increasing sequence of non-negative real numbers. So that (3.155) yields

$$\begin{aligned}\mathbb{E}\left(X_N - X_1\right) &\geqslant \sum_{k=1}^{\lfloor N/2\rfloor+1}\mathbb{E}\left(X_{\tau_{2k}} - X_{\tau_{2k-1}}\right) + (b-a)\mathbb{E}\left(U_N(a,b)\right)\\ &\geqslant (b-a)\mathbb{E}\left(U_N(a,b)\right).\end{aligned} \tag{3.156}$$

The desired result in Proposition 3.71 follows from (3.156). $\square$

3.72. Theorem. *(Sub-martingale convergence theorem) Let* $\{Z_n\}$ *be a sub-martingale with the property that* $\sup_{n\in\mathbb{N}} \mathbb{E}(|Z_n|) < \infty$. *Then the sequence* $\{Z_n\}$ *converges almost surely to an integrable random variable* Z_∞. *Moreover we have* $\mathbb{E}(|Z_\infty|) \leqslant \liminf_{n\to\infty} \mathbb{E}(|Z_n|)$.

3.73. Remark. In general we do not have $\mathbb{E}(Z_\infty) = \lim_{n\to\infty} \mathbb{E}(Z_n)$. In fact there exist martingales $\{M_n : n \in \mathbb{N}\}$ such that $M_n \geqslant 0$, such that $\mathbb{E}(M_n) = 1$, $n \in \mathbb{N}$, and such that $M_\infty = \lim_{n\to\infty} M_n = 0$, $\mathbb{P}$-almost surely. To be specific, let $\{b(s) : s \geqslant 0\}$ be ν-dimensional Brownian motion starting at $x \in \mathbb{R}^\nu$ and let $p(t, x, y)$ be the corresponding transition density. Fix $t > 0$ and $y \neq x$ and put

$$M_n = \frac{p\left(t/n, b(t - t/n), y\right)}{p(t, x, y)}. \tag{3.157}$$

The process $\{M_n : n \in \mathbb{N}\}$ defined in (3.157) is $\mathbb{P}_x$-martingale with respect to the sigma-fields $\mathcal{F}_n$ generated by $b(s)$, $0 \leqslant s \leqslant t - t/n$.

PROOF. Let $U(a,b)$ be the total number of upcrossings of (a,b) by the process $\{Z_n : n \in \mathbb{N}\}$. Then $U_N(a,b) \uparrow U(a,b)$ as $N \uparrow \infty$. Therefore, by monotone convergence,

$$\mathbb{E}\big((U(a,b)\big) = \lim_{N\to\infty} \mathbb{E}\left(U_N(a,b)\right) \leqslant \sup_{N\in\mathbb{N}} \frac{\mathbb{E}\left(|Z_N|\right) + |a|}{b-a} < \infty. \tag{3.158}$$

In particular it follows that $U(a,b) < \infty$ $\mathbb{P}$-almost surely. Hence

$$\mathbb{P}\left(\liminf Z_n < a < b < \limsup Z_n\right) \leqslant \mathbb{P}\left(U(a,b) = \infty\right) = 0. \tag{3.159}$$

Since

$$\{\liminf Z_n < \limsup Z_n\} = \bigcup_{a<b,a,b\in\mathbb{Q}} \{\liminf Z_n < a < b < \limsup Z_n\}$$

it follows from (3.159) that $\mathbb{P}\left(\liminf Z_n < \limsup Z_n\right) = 0$. By Fatou's lemma it follows that $\mathbb{E}\left(|Z_\infty|\right) = \mathbb{E}\left(\liminf_n |Z_n|\right) \leqslant \liminf \mathbb{E}\left(|Z_n|\right) < \infty$.

This completes the proof of Theorem 3.72. □

3.74. COROLLARY. *A non-negative martingale $\{Z_n\}$ converges almost surely to a finite limit Z_∞. Also $\mathbb{E}\left(Z_\infty\right) \leqslant \mathbb{E}\left(Z_1\right)$.*

Remark. Convergence properties for supermartingales $\{Z_n\}$ are obtained from the sub-martingale results applied to $\{-Z_n\}$. Since a semi-martingale is a difference of two sub-martingales, we also have convergence results for semi-martingales.

3.75. DEFINITION. A continuous time process $\{X(t) : t \in \mathbb{R}\}$ is called *stochastically continuous* at t_0 if for every $\varepsilon > 0$

$$\lim_{t\to t_0} \mathbb{P}\left(|X(t) - X(t_0)| > \varepsilon\right) = 0. \tag{3.160}$$

3.76. REMARK. Brownian motion possesses almost surely continuous sample paths and is stochastically continuous for every $t \geqslant 0$. On the other hand a Poisson process is stochastically continuous, but its sample paths are step functions with unit jumps. In fact, for $t > s$, $X(t) \geqslant X(s)$ $\mathbb{P}$-almost surely and, again for $t > s$, $\mathbb{P}\left(X(t) - X(s) = n\right) = e^{-\lambda(t-s)}\dfrac{\big(\lambda(t-s)\big)^n}{n!}$ and hence, always for $t > s$ and for $\epsilon > 0$,

$$\mathbb{P}\left(X(t) - X(s) \geqslant \epsilon\right) \leqslant e^{-\lambda(t-s)} \sum_{n=1}^{\infty} \frac{\big(\lambda(t-s)\big)^n}{n!} = 1 - e^{-\lambda(t-s)}.$$

3.77. THEOREM. *Let $\{X(t) : t \geqslant 0\}$ be a sub-martingale or a super-martingale that is stochastically continuous at each $t \geqslant 0$. Then there exists a process $\left\{\widetilde{X}(t) : t \geqslant 0\right\}$ with the following properties:*

(i) *(stochastic equivalence) $\left\{\widetilde{X}(t)\right\}$ is equivalent to $\{X(t)\}$ in the sense that*

$$\mathbb{P}\left(\widetilde{X}(t) = X(t)\right) = 1 \quad \textit{for every} \quad t \geqslant 0;$$

(ii) *(sample path regularity) with probability 1 the sample paths of the process* $\left\{\widetilde{X}(t) : t \geqslant 0\right\}$ *are bounded on compact intervals* $[a,b]$, $a < b < \infty$, *are right-continuous and possess left-hand limits at each* $t > 0$ *(in other words* $\left\{\widetilde{X}(t) : t \geqslant 0\right\}$ *is cadlag).*

PROOF. Fix $T > 0$ and let $\mathbb{Q}_T$ denote the set of rational numbers in $[0,T]$. Write $\mathbb{Q}_T = \bigcup_{n=1}^{\infty} R_n$, where R_n is a finite subset of $[0,T]$ and where $T \in R_1 \subset R_2 \subset R_2 \subset \cdots$. By Doob's maximal inequality for sub-martingales we have

$$\mathbb{P}\left(\max_{t\in R_n} |X(t)| > \lambda\right) \leqslant \frac{\mathbb{E}\,|X(T)|}{\lambda}, \quad n = 1,2,\dots$$

and hence

$$\mathbb{P}\left(\sup_{t\in \mathbb{Q}_T} |X(t)| > \lambda\right) \leqslant \lim_{n\to\infty} \mathbb{P}\left(\max_{t\in R_n} |X(t)| > \lambda\right) \leqslant \frac{\mathbb{E}\,|X(T)|}{\lambda}, \quad n = 1,2,\dots$$

For Doob's maximal inequality see *e.g.* Proposition 3.107 or Theorem 5.110. In particular, the paths of $\{X(t) : t \in \mathbb{Q}_T\}$ are bounded with probability 1. Let (c,d) be any interval in $\mathbb{R}$ and let $U^{\{T\}}(c,d)$ denote the number of upcrossings of (c,d) by the process $\{X(t) : t \in \mathbb{Q}_T\}$. Then $U^{\{T\}}(c,d)$ is the limit of the number $U^{\{n\}}(c,d)$ of upcrossings of (c,d) by $\{X(t) : t \in R_n\}$ as n tends to ∞. By the upcrossing inequality we have

$$\mathbb{E}\left(U^{\{n\}}(c,d)\right) \leqslant \frac{\mathbb{E}\left(|X(T)|\right) + |c|}{d-c}. \tag{3.161}$$

Since $U^{\{n\}}(c,d)$ increases with n it follows from (3.161) that

$$\mathbb{E}\left(U^{\{T\}}(c,d)\right) \leqslant \frac{\mathbb{E}\left(|X(T)|\right) + |c|}{d-c}, \tag{3.162}$$

and hence that $U^{\{T\}}(c,d)$ is almost surely finite. Taking unions over all intervals (c,d), with c, $d \in \mathbb{Q}$, and $c < d$, it follows with probability 1 that the process $\{X(t) : t \in \mathbb{Q}_T\}$ has only finitely many upcrossings of any interval. In particular, therefore, left- and right-hand limits must exist at each $t < T$ $\mathbb{P}$-almost surely. To construct a right-continuous version of $\{X(t)\}$ we define $\left\{\widetilde{X}(t) : t \geqslant 0\right\}$ as follows: $\widetilde{X}(t) = \lim_{s\downarrow t, s\in\mathbb{Q}} X(s)$ for $t < T$. That this process $\left\{\widetilde{X}(t)\right\}$ is stochastically equivalent to $\{X(t)\}$ follows from the stochastic continuity of the process $\{X(t)\}$. Further details are left to the reader. This completes the proof of Theorem 3.77. □

Next we prove Doob's optional sampling for continuous time sub-martingales (that are right-continuous) and a similar result holds for martingales (where the inequality sign in (3.163) is replaced with an equality) and for super-martingales (where the inequality is reversed). For discrete sub-martingales the result will be taken for granted: see Theorems 5.104 and 5.114.

3.78. Theorem. *Let $\{X(t) : t \geqslant 0\}$ be a right-continuous sub-martingale of class (DL) and let T be a stopping time. Suppose $t \geqslant s$. Then*

$$\mathbb{E}\left[X\left(\min(t,T)\right) \mid \mathcal{F}_s\right] \geqslant X\left(\min(s,T)\right), \quad \mathbb{P}\text{-almost surely.} \tag{3.163}$$

Proof. Put $s_n = 2^{-n}[2^n s]$, $t_n = 2^{-n}[2^n t]$ and $T_n = 2^{-n}[2^n T]$. If A belongs to $\mathcal{F}_s$, then A also belongs to $\mathcal{F}_{s_n}$ for all $n \in \mathbb{N}$. From Doob's optional sampling for discrete time sub-martingales we infer, upon using the (DL)-property,

$$\mathbb{E}\left[X\left(\min(t_n,T_n)\right) 1_A\right] \geqslant \mathbb{E}\left[X\left(\min(s_n,T_n)\right) 1_A\right]. \tag{3.164}$$

Upon letting n tend to ∞ and using the right-continuity of the process $t \mapsto X(t)$, $t \geqslant 0$, we infer

$$\mathbb{E}\left[X\left(\min(t,T)\right) 1_A\right] \geqslant \mathbb{E}\left[X\left(\min(s,T)\right) 1_A\right], \tag{3.165}$$

where $A \in \mathcal{F}_s$ is arbitrary. Consequently the result in (3.163) follows from (3.165), and so the proof of Theorem is complete 3.78. $\square$

6. Stochastic integrals, Itô's formula

The assumptions are as in Section 4. The process $\{b(t) : t \geqslant 0\}$ is assumed to be one-dimensional Brownian motion and hence the process $t \mapsto b(t)^2 - t$ is a martingale: see Proposition 3.23. The following proposition contains the basic ingredients of (the definition of) a stochastic integral.

3.79. PROPOSITION. *Let $s_1, \dots, s_n$ and $t_1, \dots, t_n$ be non-negative numbers for which $s_{j-1} < t_{j-1} \leqslant s_j < t_j$, $2 \leqslant j \leqslant n$. Let $f_1, \dots, f_n$ be bounded random variables which are measurable with respect to $\mathcal{F}_{s_1}, \dots, \mathcal{F}_{s_n}$ respectively. Put*

$$Y(s,\omega) = \sum\nolimits_{j=1}^{n} f_j(\omega) 1_{(s_j,t_j]}(s)$$

and write

$$\int_0^t Y(s,\cdot)db(s) = \sum\nolimits_{j=1}^{n} f_j \left\{ b(\min(t,t_j)) - b(\min(t,s_j)) \right\}.$$

The following assertions hold true:

(a) *The process $\left\{\int_0^t Y(s)db(s) : t \geqslant 0\right\}$ is a martingale and the process $\left\{\left(\int_0^t Y(s)db(s)\right)^2 : t \geqslant 0\right\}$ is a submartingale.*
(b) *The process $\left\{\left(\int_0^t Y(s)db(s)\right)^2 - \int_0^t Y(s)^2 ds : t \geqslant 0\right\}$ is a martingale.*
(c) *(Itô isometry) The following equality is valid:*

$$\mathbb{E}\left[\left(\int_0^t Y(s)db(s)\right)^2\right] = \mathbb{E}\left[\int_0^t Y(s)^2 ds\right]. \tag{3.166}$$

The equality in assertion (c) is called the Itô isometry. It is an extremely important equality: the entire Itô calculus is justified by the use of the equality in (3.166).

PROOF. The more or less straightforward calculations are left as an exercise to the reader. We insert some ways to simplify the computations. Let F and G be predictable processes of the form $F(s) = f1_{(u,\infty)}(s)$ and $G(s) = g1_{(v,\infty)}(s)$, where f is measurable for the σ-field $\mathcal{F}_u$ and g for $\mathcal{F}_v$. Put

$$I_t(F) = \int_0^t F(s)db(s) := f\left(b(t) - b(\min(u,t))\right)$$

and similarly write

$$I_t(G) = \int_0^t G(s)db(s) = g\left(b(t) - b(\min(v,t))\right).$$

Without loss of generality we assume $v \geqslant u$ (otherwise we interchange the role of F and G). We begin with a proof of (a). Upon employing linearity it suffices to show that the process $t \mapsto I_t(F)$ is a martingale. (Also notice that

$\int_0^t f1_{(u,v]}(s)db(s) = I_t(F) - I_t(F_1)$, where $F_1(s) = f1_{(v,\infty)}(s)$.) Fix $t > s \geqslant 0$ and consider

$$\begin{aligned}
&\mathbb{E}\left(I_t(F) \mid \mathcal{F}_s\right) - I_s(F) = \mathbb{E}\left(I_t(F) - I_s(F) \mid \mathcal{F}_s\right)\\
&= \mathbb{E}\left(f\left(b(t) - b(\min(u,t)) - b(s) + b(\min(u,s))\right) \mid \mathcal{F}_s\right)\\
&= \mathbb{E}\left(\mathbb{E}\left(f\left(b(t) - b(\min(u,t)) - b(s) + b(\min(u,s))\right) \mid \mathcal{F}_{\min(\max(u,s),t)}\right) \mid \mathcal{F}_s\right)\\
&= \mathbb{E}\left(f\mathbb{E}\left(\left(b(t) - b(\min(u,t)) - b(s) + b(\min(u,s))\right) \mid \mathcal{F}_{\min(\max(u,s),t)}\right) \mid \mathcal{F}_s\right)
\end{aligned}$$

(Brownian motion is a martingale)

$$\begin{aligned}
&= \mathbb{E}\left(f\left((b(\min(\max(u,s),t)) - b(\min(u,t)) - b(s) + b(\min(u,s))\right)\right) \mid \mathcal{F}_s\right)\\
&= 0,
\end{aligned}$$

proving that the process $t \mapsto I_t(F)$ is a martingale indeed. Next we shall prove that the process $t \mapsto I_t(F)I_t(G) - \int_0^t F(\tau)G(\tau)d\tau$ is a martingale. Using bilinearity in F and G yields a proof of (b) and hence also of (c). Again we fix $t > s$ and consider

$$\begin{aligned}
&\mathbb{E}\left(I_t(F)I_t(G) - \int_0^t F(\tau)G(\tau)d\tau \mid \mathcal{F}_s\right) - \left(I_s(F)I_s(G) - \int_0^s F(\tau)G(\tau)d\tau\right)\\
&= \mathbb{E}\left(I_t(F)I_t(G) - \int_0^t F(\tau)G(\tau)d\tau - \left(I_s(F)I_s(G) - \int_0^s F(\tau)G(\tau)d\tau\right) \mid \mathcal{F}_s\right)\\
&= \mathbb{E}\left(\left(I_t(F) - I_s(F)\right)\left(I_t(G) - I_s(G)\right) - \int_s^t F(\tau)G(\tau)d\tau \mid \mathcal{F}_s\right)\\
&\quad + \mathbb{E}\left(I_s(F)\left(I_t(G) - I_s(G)\right) + \left(I_t(F) - I_s(F)\right)I_s(G) \mid \mathcal{F}_s\right)\\
&= \mathbb{E}\left(\left(I_t(F) - I_s(F)\right)\left(I_t(G) - I_s(G)\right) - \int_s^t F(\tau)G(\tau)d\tau \mid \mathcal{F}_s\right)\\
&\quad + I_s(F)\mathbb{E}\left(I_t(G) - I_s(G) \mid \mathcal{F}_s\right) + \mathbb{E}\left(I_t(F) - I_s(F) \mid \mathcal{F}_s\right)I_s(G)
\end{aligned}$$

(use the martingale property of $I_t(F)$ and $I_t(G)$)

$$\begin{aligned}
&= \mathbb{E}\left[\left(I_t(F) - I_s(F)\right)\left(I_t(G) - I_s(G)\right) - \int_s^t F(\tau)G(\tau)\,d\tau \mid \mathcal{F}_s\right]\\
&= \mathbb{E}\left[fg\left(b(t) - b\left(\min(\max(u,s),t)\right)\right)\left(b(t) - b\big(\min(\max(v,s),t)\big)\right)\right.\\
&\qquad \left. -fg\left(t - \min(\max(u,v,s),t)\right) \mid \mathcal{F}_s\right]
\end{aligned}$$

(use $v \geqslant u$ and put $u_{s,t} = \min(\max(u,s),t)$, $v_{s,t} = \min(\max(v,s),t)$)

$$\begin{aligned}
&= \mathbb{E}\left[fg\left(\left(b(t) - b\big(v_{s,t}\big)\right)^2\right.\right.\\
&\qquad \left.\left. +fg\left(b\big(v_{s,t}\big) - b\big(u_{s,t}\big)\right)\left(b(t) - b\big(v_{s,t}\big)\right) - fg\left(t - \min(\max(u,v,s),t)\right) \mid \mathcal{F}_s\right)\right]\\
&= \mathbb{E}\left[fg\left(\left(b(t) - b\big(v_{s,t}\big)\right)^2 - \left(t - v_{s,t}\right)\right) \mid \mathcal{F}_s\right]\\
&\quad + \mathbb{E}\left[\left(fg\left(b\big(v_{s,t}\big) - b\big(u_{s,t}\big)\right)\left(b(t) - b\big(v_{s,t}\big)\right)\right) \mid \mathcal{F}_s\right]\\
&= \mathbb{E}\left[fg\mathbb{E}\left[\left(\left(b(t) - b\big(v_{s,t}\big)\right)^2 - \left(t - v_{s,t}\right)\right) \mid \mathcal{F}_{v_{s,t}}\right] \mid \mathcal{F}_s\right]
\end{aligned}$$

$$+\,\mathbb{E}\left[fg\left(b(v_{s,t})-b(u_{s,t})\right)\times\mathbb{E}\left[\left(b(t)-b(v_{s,t})\right)\mid\mathcal{F}_{v_{s,t}}\right]\mid\mathcal{F}_s\right]$$

(the processes $\{b(s)\}$ and $\{b(s)^2-s\}$ are martingales)

$$=\mathbb{E}\left(fg.0\mid\mathcal{F}_s\right)+\mathbb{E}\left(fg\left(b(v_{s,t})-b(u_{s,t})\right).0\mid\mathcal{F}_s\right)=0.$$

The latter yields a proof of (b) (via bilinearity). Altogether this finishes the proof of Proposition 3.79. □

3.80. Definition. A process of the form

$$F(s,\omega)=\sum\nolimits_{j=1}^{n}f_j(\omega)1_{(s_j,t_j]}(s),$$

where $0\leqslant s_{j-1}<t_{j-1}\leqslant s_j<t_j$, $2\leqslant j\leqslant n$, and where the functions $f_1,\dots,f_n$ are bounded and measurable with respect to $\mathcal{F}_{s_1},\dots,\mathcal{F}_{s_n}$ respectively is called a simple predictable process.

3.81. Definition. Again let b be Brownian motion with drift zero and let $\Pi_2(b)$ be the vector space of all predictable processes F with the property that

$$\|F\|_b^2 := \mathbb{E}\left(\int_0^\infty |F(s)|^2\, ds\right) < \infty.$$

Let Q be the σ-additive measure, defined on the predictable field Π, determined by

$$Q\left(A \times (s,t]\right) = \mathbb{E}\left[1_A\right](t-s) = \mathbb{P}(A)(t-s), \quad A \in \mathcal{F}_s. \tag{3.167}$$

The measure Q is called the Doléans measure for Brownian motion.

Then it follows that $\Pi_2(b) = L^2\left([0,\infty) \times \Omega, \Pi, Q\right)$. Moreover we have

$$\|F\|_b^2 = \int |F|^2\, dQ, \quad F \in \Pi_2(b).$$

It also follows that, for given $F \in \Pi_2(b)$, there exists a sequence of simple processes $(F_n : n \in \mathbb{N})$, which are predictable, such that $\lim_{n\to\infty} \|F_n - F\|_b = 0$. Hence in view of Proposition 3.79 it is obvious how to define $\int_0^t F(s)\, db(s)$, $t \geqslant 0$, for $F \in \Pi_2(b)$. In fact

$$\int_0^t F(s) db(s) = L^2\text{-}\lim_{n\to\infty} \int_0^t F_n(s) db(s),$$

where the sequence $(F_n : n \in \mathbb{N})$ verifies $\lim_{n\to\infty} \|F_n - F\|_b = 0$ and where F_n belongs to $\Pi_2(b)$. Let $\Pi_3(b)$ be the vector space of all predictable processes F for which the integrals $\int_0^t |F(s)|^2\, ds$ are finite $\mathbb{P}$-almost surely for all $t > 0$. In order to extend the definition of stochastic integral to processes $F \in \Pi_3(b)$ we proceed as follows. Define the stopping times T_n, $n \in \mathbb{N}$, in the following fashion:

$$T_n = \inf\left\{t > 0 : \int_0^t |F(s)|^2\, ds > n\right\}. \tag{3.168}$$

We also write $F_n(s) = F(s)1_{\{T_n > s\}}$ and we observe that F_n is a predictable process with $\int |F_n(s)|^2\, ds \leqslant n$. Moreover it follows that for $n > m$ the expression

$$\int_0^t F_n(s) db(s) - \int_0^t F_m(s) db(s) = \int F(s) 1_{(T_m, \min(T_n,t)]}(s) db(s) \tag{3.169}$$

vanishes almost everywhere on the event $\{T_m > t\}$. So it makes sense to write

$$\int_0^t F(s) db(s) = \int_0^t F_m(s) db(s), \quad \text{on} \quad \{T_m > t\}.$$

Since $\lim_{n\to\infty} T_n = \infty$, $\mathbb{P}$-almost surely, the quantity $\int_0^t F(s) db(s)$ is unambiguously defined. Hence the integral $\int F(s) db(s)$ is well defined for processes F belonging to $\Pi_3(b)$.

3.82. Corollary. *Let b be Brownian motion and let F and G be processes in $\Pi_3(b)$. The following processes are local martingales:*

$$\left\{\int_0^t F(s) db(s) : t \geqslant 0\right\}, \quad \left\{\int_0^t G(s) db(s) : t \geqslant 0\right\}; \tag{3.170}$$

$$\left\{\left(\int_0^t F(s)db(s)\right)^2 - \int_0^t |F(s)|^2 \, ds : t \geqslant 0\right\}; \tag{3.171}$$

$$\left\{\int_0^t F(s)db(s) \int_0^t G(s)db(s) - \int_0^t F(s)G(s)ds : t \geqslant 0\right\}. \tag{3.172}$$

Put $X(t) = \int_0^t F(s)db(s)$ and $Y(t) = \int_0^t G(s)db(s)$. The following identity is valid:

$$X(t)Y(t) - \int_0^t F(s)G(s)ds = \int_0^t F(s)Y(s)db(s) + \int_0^t X(s)G(s)db(s). \tag{3.173}$$

Proof. The assertions (3.170), (3.171) and (3.172) follow from Proposition 3.79 together with taking appropriate limits. For the proof of (3.173) we first take $F \equiv G \equiv 1$. Then (3.173) reduces to showing that

$$b(t)^2 - 2\int_0^t b(s)db(s) - t = 0. \tag{3.174}$$

Notice that (3.174) is equivalent to $2\int_0^t b(s)db(s) = b(t)^2 - t, \quad t \geqslant 0$. For the proof of (3.174) we use Lemma 3.69. to conclude:

$$\begin{aligned}
&b(t)^2 - 2\int_0^t b(s)db(s) - t \\
&= \lim_{n\to\infty}\left(b(t)^2 - 2\sum_{k=0}^{2^n-1} b\left(k2^{-n}t\right)\left(b\left((k+1)2^{-n}t\right) - b\left(k2^{-n}t\right) - t\right)\right) \\
&= \lim_{n\to\infty}\left(\sum_{k=0}^{2^n-1}\left(b\left((k+1)2^{-n}t\right)^2 - b\left(k2^{-n}t\right)^2\right)\right. \\
&\qquad \left. -2\sum_{k=0}^{2^n-1} b\left(k2^{-n}t\right)\left(b\left((k+1)2^{-n}t\right) - b\left(k2^{-n}\right)\right) - t\right) \\
&= \lim_{n\to\infty}\left(\sum_{k=0}^{2^n-1}\left(b\left((k+1)2^{-n}t\right) - b\left(k2^{-n}t\right)\right)^2 - t\right) = 0.
\end{aligned}$$

For the proof of (3.173) we then take $F(s) = f1_{(s_1,\infty)}(s)$ and $G(s) = g1_{(s_2,\infty)}$, where f is bounded and measurable with respect to $\mathcal{F}_{s_1}$ and g is measurable with respect to $\mathcal{F}_{s_2}$. Formula (3.174) will then yield the desired result. Then we pass over to linear combinations and finally to limits. This completes the proof of Corollary 3.82. □

3.83. Proposition. *Stochastic integrals with integrands in $\Pi_3(b)$ are continuous $\mathbb{P}$-almost surely.*

Proof. It suffices to prove the result for integrands in $\Pi_2(b)$. Since Brownian motion is almost surely continuous, it follows that stochastic integrals of simple predictable processes are continuous. Let F be in $\Pi_2(b)$ and choose a sequence $(F_n)_{n\in\mathbb{N}}$ of simple predictable processes with the property that

$$\begin{aligned}
&\left(\mathbb{E}\left(\int_0^{t_0}|F(s)-F_n(s)|^2\,ds\right)\right)^{1/2} = \lim_{\ell\to\infty}\left(\mathbb{E}\left(\int_0^{t_0}|F_{n+\ell}(s)-F_n(s)|^2\,ds\right)\right)^{1/2}\\
&\leqslant \sum_{\ell=1}^{\infty}\left(\mathbb{E}\left(\int_0^{t_0}|F_{n+\ell}(s)-F_{n+\ell-1}(s)|^2\,ds\right)\right)^{1/2} \leqslant \sum_{\ell=1}^{\infty}2^{-n-\ell-2} = 2^{-n-1}.
\end{aligned}$$

From Proposition 3.63 it follows that, for $k\in\mathbb{N}$,

$$\begin{aligned}
&\left(\mathbb{E}\left[\sup_{0\leqslant t\leqslant t_0}\left|\int_0^t\left(F_{n+k}(s)-F_n(s)\right)db(s)\right|^2\right]\right)^{\frac12}\\
&\leqslant \sum_{\ell=1}^{k}\left(\mathbb{E}\left[\sup_{0\leqslant t\leqslant t_0}\left|\int_0^t\left(F_{n+\ell}(s)-F_{n+\ell-1}(s)\right)db(s)\right|^2\right]\right)^{\frac12}\\
&\leqslant 2\sum_{\ell=1}^{\infty}\left(\mathbb{E}\left[\left|\int_0^{t_0}\left(F_{n+\ell}(s)-F_{n+\ell-1}(s)\right)db(s)\right|^2\right]\right)^{\frac12}\\
&= 2\sum_{\ell=1}^{\infty}\left(\mathbb{E}\left[\int_0^{t_0}\left|F_{n+\ell}(s)-F_{n+\ell-1}(s)\right|^2ds\right]\right)^{\frac12} \leqslant 2^{-n}. \qquad (3.175)
\end{aligned}$$

From (3.175) the sample path continuity of stochastic integrals immediately follows. This completes the proof of Proposition 3.83. $\square$

3.84. Remark. The theory in this section can be extended to (continuous) martingales instead of Brownian motion. To be precise, let $t \mapsto M(t)$ be a continuous martingale with quadratic variation process $t \mapsto \langle M, M\rangle(t)$. Then the process $t \mapsto M(t)^2 - \langle M, M\rangle(t)$ is a martingale, and the space $\Pi_2(b)$ should be replaced with $\Pi_2(M)$, the space of all predictable processes $t \mapsto F(t)$ with the property that

$$\|F\|_M^2 = \mathbb{E}\left[\int_0^\infty |F(s)|^2 \, d\langle M, M\rangle(s)\right] < \infty. \tag{3.176}$$

The corresponding Doléans measure Q_M is given by

$$Q_M(A \times (s,t]) = \mathbb{E}\left[1_A\left(\langle M, M\rangle(t) - \langle M, M\rangle(s)\right)\right], \quad A \in \mathcal{F}_s, \quad s < t. \tag{3.177}$$

It follows that

$$\Pi_2(M) = L^2\left(\Omega \times [0,\infty), \Pi, Q_M\right).$$

The space $\Pi_3(M)$ consists of those predictable processes $t \mapsto F(t)$ which have the property that $\int_0^t |F(s)|^2 \, d\langle M, M\rangle s$ are finite $\mathbb{P}$-almost surely for all $t > 0$. The definition of stochastic integral to processes $F \in \Pi_3(M)$ we proceed as follows. Define the stopping times T_n, $n \in \mathbb{N}$, in the following fashion:

$$T_n = \inf\left\{t > 0 : \int_0^t |F(s)|^2 \, d\langle M, M\rangle(s) > n\right\}. \tag{3.178}$$

As in the case of Brownian motion these stopping times can be used to define stochastic integrals of the form $\int_0^t F(s)\, dM(s)$, $F \in \Pi_3(M)$. These integrals are then local martingales.

Next we extend the equality in (3.173) to the multi-dimensional situation.

3.85. Proposition. *Let $s \mapsto \sigma(s) = (\sigma_{jk}(s))_{1 \leqslant j,k \leqslant \nu}$ be a matrix with predictable entries and with the property that the expression*

$$\sum_{j,k=1}^{\nu} \int_0^t \mathbb{E}\left|\sigma_{jk}(s)\right|^2 ds \tag{3.179}$$

is finite for every $t > 0$. Put $a_{ij}(s) = \sum_{k=1}^{\nu} \sigma_{ik}(s)\sigma_{jk}(s)$, $1 \leqslant i,\, j \leqslant \nu$. Furthermore let $\{b(s) = (b_1(s), \dots, b_\nu(s)) : s \geqslant 0\}$ be ν-dimensional Brownian motion. Put $M_j(t) = \sum_{k=1}^{\nu} \int_0^t \sigma_{jk}(s) db_k(s)$, $1 \leqslant j \leqslant \nu$. Then the following identity is valid:

$$\begin{aligned} &M_i(t)M_j(t) \\ &= \sum_{k=1}^{\nu} \int_0^t M_i(s)\sigma_{jk}(s) db_k(s) + \sum_{k=1}^{\nu} \int_0^t \sigma_{ik}(s) M_j(s) db_k(s) + \int_0^t a_{ij}(s) ds. \end{aligned} \tag{3.180}$$

Proof. First we suppose $\nu = 2$, $M_1(t) = b_1(t)$ and $M_2(t) = b_2(t)$. Then (3.180) reads as follows:

$$b_1(t)b_2(t) = \int_0^t b_1(s) db_2(s) + \int_0^t b_2(s) db_1(s). \tag{3.181}$$

In order to prove (3.181) we write

$$\begin{aligned}
& b_1(t)b_2(t) - \int_0^t b_1(s)db_2(s) - \int_0^t b_2(s)db_1(s) \\
&= \lim_{n\to\infty} \sum_{k=0}^{2^n-1} \Big\{ b_1\left((k+1)2^{-n}t\right) b_2\left((k+1)2^{-n}t\right) - b_1\left(k2^{-n}t\right) b_2\left(k2^{-n}t\right) \\
&\quad - b_1\left(k2^{-n}t\right)\left(b_2\left((k+1)2^{-n}t\right) - b_2\left(k2^{-n}t\right)\right) \\
&\qquad - b_2\left(k2^{-n}t\right)\left(b_1\left((k+1)2^{-n}t\right) - b_1\left(k2^{-n}t\right)\right)\Big\} \\
&= \lim_{n\to\infty} \sum_{k=0}^{2^n-1} \left(b_1\left((k+1)2^{-n}t\right) - b_1\left(k2^{-n}t\right)\right)\left(b_2\left((k+1)2^{-n}t\right) - b_2\left(k2^{-n}t\right)\right).
\end{aligned} \tag{3.182}$$

The limit in (3.182) vanishes, because by independence and martingale properties of the processes b_1 and b_2, we infer

$$\begin{aligned}
& \mathbb{E}\left[\sum_{k=0}^{2^n-1} \left(b_1\left((k+1)2^{-n}t\right) - b_1\left(k2^{-n}t\right)\right)\left(b_2\left((k+1)2^{-n}t\right) - b_2\left(k2^{-n}t\right)\right)\right]^2 \\
&= \mathbb{E}\left(\sum_{k=0}^{2^n-1} \left(b_1\left((k+1)2^{-n}t\right) - b_1\left(k2^{-n}t\right)\right)^2\left(b_2\left((k+1)2^{-n}t\right) - b_2\left(k2^{-n}t\right)\right)^2\right) \\
&= \sum_{k=0}^{2^n-1} \mathbb{E}\left[b_1\left((k+1)2^{-n}t\right) - b_1\left(k2^{-n}t\right)\right]^2 \mathbb{E}\left(b_2\left((k+1)2^{-n}t\right) - b_2\left(k2^{-n}t\right)\right)^2 \\
&= \sum_{k=0}^{2^n-1} \left(2^{-n}t\right)^2 = 2^{-n}t^2.
\end{aligned} \tag{3.183}$$

From Borel-Cantelli's lemma it then easily follows that the limit in (3.182) vanishes and hence that equality (3.181) is true. The validity of (3.180) is then checked for the special case that $\sigma_{jk}(s) = f_{jk}1_{(s_{jk},\infty)}(s)$, where f_{jk} is measurable with respect to $\mathcal{F}_{s_{jk}}$. The general statement follows via bi-linearity and a limiting procedure together with equality (3.173) in Corollary 5.142. The proof of Proposition 3.85 is now complete. $\square$

Next let $M(t) = (M_1(t), \dots, M_\nu(t))$ be a ν-dimensional martingale as in Proposition 3.85 and let $A(t) = (A_1(t), \dots, A_\nu(t))$ be an adapted ν-dimensional process that $\mathbb{P}$-almost surely is of bounded variation on $[0,t]$ for every $t > 0$. This means that

$$\sup_{n\in\mathbb{N}} \sup_{0\leqslant s_0<s_1<\cdots s_n\leqslant t} |A(s_j) - A(s_{j-1})| \quad \text{is finite } \mathbb{P}\text{-almost surely for all } t>0.$$

It follows that the random set function $\mu^A : (a,b] \mapsto A(b) - A(a)$ extends to an $\mathbb{R}^\nu$-valued measure on $[0,t]$ for every $t > 0$. Stieltjes integrals of the form $\int_0^t F(s)dA(s)$ may be interpreted as $\int_0^t F(s)dA(s) = \int_0^t F(s)d\mu^A(s)$. The process A may have jumps. This is not the case for the process M. The latter follows from Proposition 3.83. The process $X := A + M$ is a ν-dimensional semi-martingale with the property that $\mathbb{E}\left(|M(t)|^2\right) < \infty$, $t > 0$. Put

$$\begin{aligned} J_X(t) &= \sum\nolimits_{s\leqslant t} \left(X(s) - X(s-)\right) \\ &= \sum_{s\leqslant t} \left(M(s) - M(s-)\right) + \sum_{s\leqslant t} \left(A(s) - A(s-)\right) = J_A(t). \end{aligned}$$

The definition of $J_A(t)$ does not pose to much of a problem. In fact for $\mathbb{P}$-almost all ω the sum $\sum\limits_{s\leqslant t} |A(s,\omega) - A(s-,\omega)| < \infty$. The process $\{X(t) - J_X(t) : t \geqslant 0\}$ is $\mathbb{P}$-almost surely continuous.

The following result is the fundamental theorem in stochastic calculus.

3.86. Theorem (Itô's formula). *Let $X = (X_1, \dots, X_\nu) = A + M$ be a ν-dimensional local semi-martingale as described above, and let $f : \mathbb{R}^\nu \to \mathbb{R}$ be a twice continuously differentiable function. Put $a_{ij}(t) = \sum_{k=1}^{\nu} \sigma_{ik}(t)\sigma_{jk}(t)$, $1 \leqslant i$, $j \leqslant \nu$. Then, $\mathbb{P}$-almost surely,*

$$\begin{aligned} & f(X(t)) \\ & = f(X(0)) + \sum_{s \leqslant t} \Big(f(X(s)) - f\left(X(s-)\right) - \nabla f\left(X(s-)\right).\left(X(s) - X(s-)\right) \Big) \\ & \qquad + \int_0^t \nabla f\left(X(s-)\right) \cdot dX(s) + \frac{1}{2} \sum_{i,j=1}^{\nu} \int_0^t D_i D_j f(X(s)) a_{ij}(s) ds. \end{aligned} \tag{3.184}$$

Before we prove Theorem 3.86 we want to make some comments and we want to give a reformulation of Itô's formula. Moreover, we shall not prove Itô's formula in its full generality. We shall content ourselves with a proof with $A \equiv 0$.

Remark. The integral $\int_0^t \nabla f\left(X(s-)\right) dX(s)$ has the interpretation:

$$\begin{aligned} & \int_0^t \nabla f\left(X(s-)\right) dX(s) \\ & = \sum_{i=1}^{\nu} \left(\int_0^t D_i f\left(X(s-)\right) dM_i(s) + \int_0^t D_i f\left(X(s-)\right) dA_i(s) \right) \qquad (3.185) \\ & = \sum_{i=1}^{\nu} \left(\int_0^t D_i f\left(X(s-)\right) dM_i(s) + \int_0^t D_i f\left(X(s-)\right) dA_i(s) \right) \\ & = \sum_{i=1}^{\nu} \left(\sum_{k=1}^{\nu} \int_0^t D_i f\left(X(s-)\right) \sigma_{ik}(s) db_k(s) + \int_0^t D_i f\left(X(s-)\right) dA_i(s) \right). \end{aligned}$$

Here $X = M + A$ is the decomposition of the semi-martingale in a martingale part M and a process A which is locally of bounded variation.

For ν-dimensional Brownian motion we have the following corollary.

3.87. Corollary. *Let $b(t) = (b_1(t), \dots, b_\nu(t))$ be ν-dimensional Brownian motion. Let $f : \mathbb{R}^\nu \to \mathbb{R}$ be a twice continuously differentiable function. Then, $\mathbb{P}$-almost surely,*

$$f(b(t)) = f(b(0)) + \int_0^t \nabla f(b(s)) db(s) + \frac{1}{2} \int_0^t \triangle f(b(s)) ds. \tag{3.186}$$

In fact it suffices to suppose that the functions $D_1 f, \dots, D_\nu f$ and $D_1^2 f, \dots, D_\nu^2 f$ are continuous. Next we reformulate Itô's formula.

3.88. Theorem. *Let $X = (X_1, \dots, X_\nu)$ be a ν-dimensional right continuous semi-martingale as in Theorem 3.86 and let $f : \mathbb{R}^\nu \to \mathbb{R}$ be a twice continuously differentiable function. Then, $\mathbb{P}$-almost surely,*

$$f(X(t)) = f(X(0)) + \int_0^t \nabla f\left(X(s-)\right) \cdot dX(s) \tag{3.187}$$

$$+\sum_{i,j=1}^{\nu}\int_0^t\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)d\sigma d\left[X_i,X_j\right](s),$$

where

$$\left[X_i,X_j\right](t)=\int_0^t a_{ij}(s)ds+\sum_{s\leqslant t}\left(X_i(s)-X_i(s-)\right)\left(X_j(s)-X_j(s-)\right). \quad (3.188)$$

Remark. In the proof below we employ the following notation. Let M_i be martingale of the form

$$M_i(t):=\sum_{k=1}^{\nu}\int_0^t\sigma_{ik}(s)db_k(s).$$

Then quadratic covariation process $\left\langle M_i,M_j\right\rangle(t)$ satisfies

$$\left\langle M_i,M_j\right\rangle(t)=\sum_{k=1}^{\nu}\int_0^t\sigma_{ik}(s)\sigma_{jk}(s)\,ds.$$

PROOF OF THEOREMS 3.88 AND 3.86. Since, for a and b in $\mathbb{R}^\nu$

$$\begin{aligned}f(b)-f(a)&=\nabla f(a).(b-a) &(3.189)\\&\quad+\sum_{i,j=1}^{\nu}\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)a+\sigma b\right)d\sigma\times(b_i-a_i)(b_j-a_j)\end{aligned}$$

and since

$$\left[X_i,X_j\right](t)=\int_0^t a_{ij}(s)ds+\sum_{s\leqslant t}\left(X_i(s)-X_i(s-)\right)\left(X_j(s)-X_j(s-)\right),$$

it follows that

$$\begin{aligned}&\sum_{i,j=1}^{\nu}\int_0^t\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)d\sigma d\left[X_i,X_j\right](s)\\
&=\sum_{i,j=1}^{\nu}\int_0^t\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)d\sigma a_{ij}(s)ds\\
&\quad+\sum_{i,j=1}^{\nu}\sum_{s\leqslant t}\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)d\sigma\\
&\qquad\times\left(X_i(s)-X_i(s-)\right)\left(X_j(s)-X_j(s-)\right)\\
&=\sum_{i,j=1}^{\nu}\int_0^t\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)d\sigma a_{ij}(s)ds\\
&\quad+\sum_{s\leqslant t}\left\{f(X(s))-f\left(X(s-)\right)-\nabla f\left(X(s-)\right).\left(X(s)-X(s-)\right)\right\}. &(3.190)\end{aligned}$$

So the formulas in Theorem 3.88 and Theorem 3.86 are equivalent. Also notice that, since $\int_0^t a_{ij}(s)ds$ is a continuous process of finite variation (locally), we have

$$\begin{aligned}
&\int_0^t \int_0^1 (1-\sigma) D_i D_j f\left((1-\sigma)X(s-) + \sigma X(s)\right) d\sigma a_{ij}(s) ds \\
&\quad = \int_0^t \int_0^1 (1-\sigma) D_i D_j f\left((1-\sigma)X(s-) + \sigma X(s-)\right) d\sigma a_{ij}(s) ds \\
&\quad = \frac{1}{2} \int_0^t D_i D_j f\left(X(s-)\right) a_{ij}(s) ds.
\end{aligned}$$

Hence it suffices to prove equality (3.187) in Theorem 3.88. Assume $X(0) = M(0)$ and hence $A(0) = 0$. Upon stopping we may and do assume that in $X = M + A$, $|X(t-)| \leqslant L$ and $\mathrm{var}A(t-) \leqslant L$. This can be achieved by replacing $X(t)$ with $X(\min(t,\tau))$, where τ is the stopping time defined by

$$\tau = \inf\left\{s > 0 : \max\left(|M(s)|,\ \mathrm{var}A(s)\right) > L\right\}.$$

Here $\mathrm{var}A(s)$ is defined by

$$\mathrm{var}A(s) = \sup\left\{\sum_{j=1}^n |A(s_j) - A(s_{j-1})| : 0 \leqslant s_0 < s_1 < \ldots < s_n = s\right\}.$$

Next we define, for every $n \in \mathbb{N}$, the sequence of stopping times $\{T_{n,k} : k \in \mathbb{N}\}$ as follows:

$$\begin{aligned} T_{n,0} &= 0; \\ T_{n,k+1} &= \inf\left\{ s > T_{n,k} : \max\left(s - T_{n,k}, |X(s) - X(T_{n,k})|\right) > \frac{1}{n} \right\}. \end{aligned} \tag{3.191}$$

Since

$$\max\left(T_{n,k+1} - T_{n,k}, |X(T_{n,k+1}) - X(T_{n,k})|\right) \geqslant \frac{1}{n},$$

it follows that $\lim_{k\to\infty} T_{n,k} = \infty$, $\mathbb{P}$-almost surely. Moreover, since

$$\max\left(T_{n,k+1} - T_{n,k}, |X(T_{n,k+1}-) - X(T_{n,k})|\right) \leqslant \frac{1}{n},$$

we have $T_{n,k+1} - T_{n,k} \leqslant \dfrac{1}{n}$. Next we write:

$$\begin{aligned} &f(X(t)) - f(X(0)) \\ &= \sum_{k=0}^{\infty} \{ f\left(X\left(T_{n,k+1} \wedge t-\right)\right) - f\left(X\left(T_{n,k} \wedge t\right)\right) \\ &\quad + f\left(X\left(T_{n,k+1} \wedge t\right)\right) - f\left(X\left(T_{n,k+1} \wedge t-\right)\right) \} \\ &= \sum_{k=0}^{\infty} \Bigg\{ \int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-} \nabla f\left(X\left(T_{n,k} \wedge t\right)\right) \cdot dX(s) \\ &\quad + \sum_{i,j=1}^{\nu} \int_0^1 (1-\sigma) D_i D_j f\left((1-\sigma) X\left(T_{n,k} \wedge t\right) + \sigma X\left(T_{n,k+1} \wedge t-\right)\right) d\sigma \\ &\qquad \times \left(X_i\left(T_{n,k+1} \wedge t-\right) - X_i\left(T_{n,k} \wedge t\right)\right)\left(X_j\left(T_{n,k+1} \wedge t-\right) - X_j\left(T_{n,k} \wedge t\right)\right) \\ &\quad + f\left(X\left(T_{n,k+1} \wedge t\right)\right) - f\left(X\left(T_{n,k+1} \wedge t-\right)\right) \Bigg\}. \end{aligned} \tag{3.192}$$

On the other hand we also have:

$$\begin{aligned} &\int_0^t \nabla f\left(X(s-)\right) dX(s) \\ &\quad + \sum_{i,j=1}^{\nu} \int_0^t \int_0^1 (1-\sigma) D_i D_j f\left((1-\sigma) X(s-) + \sigma X(s)\right) d\sigma d\left[X_i, X_j\right](s) \\ &= \sum_{k=0}^{\infty} \Bigg\{ \int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-} \nabla f\left(X(s-)\right) \cdot dX(s) \\ &\quad + \sum_{i,j=0}^{\nu} \int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-} \int_0^1 (1-\sigma) D_i D_j f\left((1-\sigma) X(s-) + \sigma X(s)\right) d\sigma d\left[X_i, X_j\right](s) \\ &\quad + \nabla f\left(X\left(T_{n,k+1} \wedge t-\right)\right) . \left(X\left(T_{n,k+1} \wedge t\right) - X\left(T_{n,k+1} \wedge t-\right)\right) \\ &\quad + \sum_{i,j=1}^{\nu} \int_0^1 (1-\sigma) D_i D_j f\left((1-\sigma) X\left(T_{n,k+1} \wedge t-\right) + \sigma X\left(T_{n,k+1} \wedge t\right)\right) d\sigma \end{aligned}$$

$$\times \left(X_i\left(T_{n,k+1}\wedge t\right)-X_i\left(T_{n,k+1}\wedge t-\right)\right)\left(X_j\left(T_{n,k+1}\wedge t\right)-X_j\left(T_{n,k+1}\wedge t-\right)\right)\Bigg\}$$

$$=\sum_{k=0}^{\infty}\Bigg\{\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\nabla f\left(X(s-)\right)\cdot dX(s)$$

$$+\sum_{i,j=0}^{\nu}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)d\sigma d\left[X_i,X_j\right](s)$$

$$+f\left(X\left(T_{n,k+1}\wedge t\right)\right)-f\left(X\left(T_{n,k+1}\wedge t-\right)\right)\Bigg\}. \tag{3.193}$$

Upon subtracting (3.193) from (3.192) we infer by employing Proposition 3.85:

$$f(X(t))-f(X(0))-\int_0^t\nabla f\left(X(s-)\right)dX(s)$$

$$-\sum_{i,j=1}^{\nu}\int_0^t\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)d\sigma d\left[X_i,X_j\right](s)$$

$$=\sum_{k=0}^{\infty}\Bigg\{-\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(\nabla f\left(X(s-)\right)-\nabla f\left(X\left(T_{n,k}\right)\right)\right)\cdot dX(s)$$

$$+\sum_{i,j=1}^{\nu}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\Bigg\{\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X\left(T_{n,k}\wedge t\right)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)d\sigma$$

$$-\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)d\sigma\Bigg\}d\left[X_i,X_j\right](s)$$

$$+\sum_{i,j=1}^{\nu}\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X\left(T_{n,k}\wedge t\right)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)d\sigma$$

$$\times\Bigg\{\left(X_i\left(T_{n,k+1}\wedge t-\right)-X_i\left(T_{n,k}\wedge t\right)\right)\left(X_j\left(T_{n,k+1}\wedge t-\right)-X_j\left(T_{n,k}\wedge t\right)\right)$$

$$-\left[X_i,X_j\right]\left(T_{n,k+1}\wedge t-\right)+\left[X_i,X_j\right]\left(T_{n,k}\wedge t\right)\Bigg\}\Bigg\}$$

$$=\sum_{k=0}^{\infty}\Bigg\{-\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(\nabla f\left(X(s-)\right)-\nabla f\left(X\left(T_{n,k}\right)\right)\right)\cdot dX(s)$$

$$+\sum_{i,j=1}^{\nu}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\int_0^1(1-\sigma)\left\{D_iD_jf\left((1-\sigma)X(T_{n,k}\wedge t)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)\right.$$

$$\left.-D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)\right\}d\sigma d\left[X_i,X_j\right](s)$$

$$+\sum_{i,j=1}^{\nu}\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(T_{n,k}\wedge t)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)d\sigma$$

$$\times\Bigg\{\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(X_i(s-)-X_i\left(T_{n,k}\wedge t\right)\right)dX_j(s)$$

$$+\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(X_j(s-)-X_j\left(T_{n,k}\wedge t\right)\right)dX_i(s)\Bigg\}\Bigg\}. \tag{3.194}$$

We shall estimate the following quantities:

$$\mathbb{E}\left(\left(\sum_{k=0}^{\infty}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(D_if\left(X(s-)\right)-D_if\left(X\left(T_{n,k}\wedge t\right)\right)\right)\cdot dM_i(s)\right)^2\right); \tag{3.195}$$

$$\mathbb{E}\left(\left|\sum_{k=0}^{\infty}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(D_if\left(X(s-)\right)-D_if\left(X\left(T_{n,k}\wedge t\right)\right)\right)\cdot dA_i(s)\right|\right); \tag{3.196}$$

$$\begin{aligned}\mathbb{E}\Bigg(\Bigg|\sum_{k=0}^{\infty}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\int_0^1(1-\sigma)\left\{D_iD_jf\left((1-\sigma)X(T_{n,k}\wedge t)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)\right.\\ \left.-D_iD_jf\left((1-\sigma)X(s-)+\sigma X(s)\right)\right\}d\sigma d\left[X_i,X_j\right](s)\Bigg|\Bigg);\end{aligned} \tag{3.197}$$

$$\begin{aligned}\mathbb{E}\Bigg(\Bigg(\sum_{k=0}^{\infty}\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(T_{n,k}\wedge t)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)d\sigma\\ \times\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(X_i(s-)-X_i\left(T_{n,k}\wedge t\right)\right)dM_j(s)\Bigg)^2\Bigg);\end{aligned} \tag{3.198}$$

$$\begin{aligned}\mathbb{E}\Bigg(\Bigg|\sum_{k=0}^{\infty}\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(T_{n,k}\wedge t)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)d\sigma\\ \times\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(X_i(s-)-X_i\left(T_{n,k}\wedge t\right)\right)dA_j(s)\Bigg|\Bigg).\end{aligned} \tag{3.199}$$

Since the process $\int_0^u\left(D_if\left(X(s-)\right)-D_if\left(X\left(T_{n,k}\wedge t\right)\right)\right)dM_i(s)$, $u\geqslant 0$, is a martingale, the quantity in (3.195) verifies

$$\begin{aligned}&\mathbb{E}\left(\left(\sum_{k=0}^{\infty}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(D_if\left(X(s-)\right)-D_if\left(X\left(T_{n,k}\wedge t\right)\right)\right)\cdot dM_i(s)\right)^2\right)\\ &=\sum_{k=0}^{\infty}\mathbb{E}\left(\left(\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(D_if\left(X(s-)\right)-D_if\left(X\left(T_{n,k}\wedge t\right)\right)\right)\cdot dM_i(s)\right)^2\right)\\ &=\sum_{k=0}^{\infty}\mathbb{E}\left(\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(D_if\left(X(s-)\right)-D_if\left(X\left(T_{n,k}\wedge t\right)\right)\right)^2\cdot d\left\langle M_i\right\rangle(s)\right)\\ &\leqslant\sup_{x,y\in\mathbb{R}^\nu:|y-x|\leqslant 1/n,\max(|x|,|y|)\leqslant 2L}\left|D_if(y)-D_if(x)\right|^2.\mathbb{E}\left(\left\langle M_i\right\rangle(t)\right).\end{aligned} \tag{3.200}$$

Similarly we obtain an estimate for the quantity in (3.198):

$$\mathbb{E}\Bigg(\Bigg(\sum_{k=0}^{\infty}\int_0^1(1-\sigma)D_iD_jf\left((1-\sigma)X(T_{n,k}\wedge t)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)d\sigma$$

$$
\begin{aligned}
&\times \int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-} \left(X_i(s-) - X_i\left(T_{n,k}\wedge t\right)\right) dM_j(s)\Bigg)^2\Bigg)\\
&= \mathbb{E}\left(\sum_{k=0}^{\infty}\left(\int_0^1 (1-\sigma) D_i D_j f\left((1-\sigma) X(T_{n,k}\wedge t) + \sigma X\left(T_{n,k+1}\wedge t-\right)\right) d\sigma\right)^2\right.\\
&\qquad \left.\times \left(\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-} \left(X_i(s-) - X_i\left(T_{n,k}\wedge t\right)\right) dM_j(s)\right)^2\right)\\
&\leqslant \frac{1}{2}\sup_{|y|\leqslant 2L}\left|D_i D_j f(y)\right| \mathbb{E}\left(\sum_{k=0}^{\infty}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-} \left(X_i(s-) - X_i\left(T_{n,k}\wedge t\right)\right)^2 d\left\langle M_j\right\rangle(s)\right)\\
&\leqslant \frac{1}{2}\sup_{|y|\leqslant 2L}\left|D_i D_j f(y)\right| \times \frac{1}{n^2}\mathbb{E}\left(\left\langle M_j\right\rangle(t)\right). \qquad (3.201)
\end{aligned}
$$

The other estimates are even easier:

$$
\begin{aligned}
&= \mathbb{E}\left(\left|\sum_{k=0}^{\infty}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(D_i f\left(X(s-)\right) - D_i f\left(X\left(T_{n,k}\right)\right)\right)\cdot dA_i(s)\right|\right) \\
&\leqslant \sup_{x,y\in\mathbb{R}^\nu:|y-x|\leqslant 1/n,\max(|x|,|y|)\leqslant 2L}\left|D_i f(y) - D_i f(x)\right|.\mathbb{E}\left(\int_0^t |dA_i(s)|\right), \qquad (3.202)
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\left(\left|\sum_{k=0}^{\infty}\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\int_0^1(1-\sigma)\left\{D_iD_j f\left((1-\sigma)X(T_{n,k}\wedge t)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)\right.\right.\right. \\
&\qquad \left.\left.\left. -D_iD_j f\left((1-\sigma)X(s-)+\sigma X(s)\right)\right\} d\sigma d\left[X_i,X_j\right](s)\right|\right) \\
&\leqslant \frac{1}{2}\sup_{x,y\in\mathbb{R}^\nu:|y-x|\leqslant 1/n,\max(|x|,|y|)\leqslant 2L}\left|D_iD_j f(y) - D_iD_j f(x)\right| \\
&\qquad \times \mathbb{E}\left(\int_0^t \left|d\left[X_i,X_j\right](s)\right|\right) \\
&\leqslant \frac{1}{2}\sup_{x,y\in\mathbb{R}^\nu:|y-x|\leqslant 1/n,\max(|x|,|y|)\leqslant 2L}\left|D_iD_j f(y) - D_iD_j f(x)\right| \\
&\quad \times \sqrt{\mathbb{E}\left(\left[X_i,X_i\right](t)\right)}\sqrt{\mathbb{E}\left(\left[X_j,X_j\right](t)\right)}, \qquad \text{and} \qquad (3.203)
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\left(\left|\sum_{k=0}^{\infty}\int_0^1(1-\sigma)D_iD_j f\left((1-\sigma)X(T_{n,k}\wedge t)+\sigma X\left(T_{n,k+1}\wedge t-\right)\right)d\sigma\right.\right. \\
&\quad \left.\left.\times\int_{T_{n,k}\wedge t}^{T_{n,k+1}\wedge t-}\left(X_i(s-)-X_i\left(T_{n,k}\wedge t\right)\right)dA_j(s)\right|\right) \\
&\leqslant \frac{1}{2n}\sup_{y\in\mathbb{R}^\nu,|y|\leqslant 2L}\left|D_iD_j f(y)\right|\mathbb{E}\left(\int_0^t |dA_j(s)|\right). \qquad (3.204)
\end{aligned}
$$

The inequality (3.203) will be established shortly. The quantities (3.200), (3.201), (3.202), (3.203) and (3.204) tend to zero if n tends to infinity. Consequently, from (3.194) it then follows that, $\mathbb{P}$-almost surely,

$$
\begin{aligned}
f(X(t)) &= f(X(0)) + \int_0^t \nabla f(X(s))dX(s) \qquad (3.205)\\
&\quad + \sum_{i,j=1}^{\nu}\int_0^t\int_0^1(1-\sigma)D_iD_j f\left((1-\sigma)X(s-)+\sigma X(s)\right)d\sigma d\left[X_i,X_j\right](s).
\end{aligned}
$$

So that the formula of Itô has been established now. For completeness we prove the inequality

$$
\mathbb{E}\left(\int_0^t \left|d\left[X_i,X_j\right](s)\right|\right) \leqslant \sqrt{\mathbb{E}\left(\left[X_i,X_i\right](t)\right)}\sqrt{\mathbb{E}\left(\left[X_j,X_j\right](t)\right)}. \qquad (3.206)
$$

A proof of (3.206) will establish (3.203). For an appropriate sequence of subdivisions $0 = s_0^{(n)} < s_1^{(n)} < \cdots < s_{N_n}^{(n)} = t$ we have with, temporarily, $[X_i] = [X_i, X_i]$,

$$\begin{aligned}
&\int_0^t |d\,[X_i, X_j]\,(s)| = \lim_{n\to\infty} \sum_{k=1}^{N_n} \left| [X_i, X_j]\left(s_k^{(n)}\right) - [X_i, X_j]\left(s_{k-1}^{(n)}\right)\right| \\
&\leqslant \lim_{n\to\infty} \sum_{k=1}^{N_n} \sqrt{[X_i]\left(s_k^{(n)}\right) - [X_i]\left(s_{k-1}^{(n)}\right)}\sqrt{[X_j]\left(s_k^{(n)}\right) - [X_j]\left(s_{k-1}^{(n)}\right)} \\
&\leqslant \lim_{n\to\infty} \left(\sum_{k=1}^{N_n} \left([X_i, X_i]\left(s_k^{(n)}\right) - [X_i, X_i]\left(s_{k-1}^{(n)}\right)\right)\right)^{1/2} \\
&\qquad \left(\sum_{k=1}^{N_n} \left([X_j, X_j]\left(s_k^{(n)}\right) - [X_j, X_j]\left(s_{k-1}^{(n)}\right)\right)\right)^{1/2} \\
&= \left([X_i, X_i]\,(t) - [X_i, X_i]\,(0)\right)^{1/2} \left([X_j, X_j]\,(t) - [X_j, X_j]\,(0)\right)^{1/2} \qquad (3.207)
\end{aligned}$$

Taking expectations and using the inequality of Cauchy-Schwartz once more yields the desired result.

This completes the proofs of Theorems 3.86 and 3.88. □

Remark. In the proof of equality (3.194) there is a gap. It is correct if the process $A \equiv 0$. In order to make the proof complete, Proposition 3.85 has to be supplemented with equalities of the form ($M_i(t) = \sum_{k=1}^{\nu} \int_0^t \sigma_{ik}(s)db_k(s)$):

$$\begin{aligned}
M_i(t)A_j(t) &= \int_0^t M_i(s)dA_j(s) + \sum_{k=1}^{\nu} \int_0^t \sigma_{ik}(s)A_j(s)db_k(s); \\
A_i(t)A_j(t) &= \int_0^t A_i(s)dA_j(s) + \int_0^t A_j(s)dA_i(s).
\end{aligned}$$

This kind of equalities is true for continuous processes. If jumps are present even more care has to be taken. We continue with some examples. We begin with the heat equation.

Example 1. (Heat equation) Let U be an open subset of $\mathbb{R}^\nu$, let $f : U \to \mathbb{R}$ be a function in $C_0(E)$ and let $u : [0, \infty) \times U \to \mathbb{R}$ be a solution to the following problem:

$$\begin{cases} \dfrac{\partial u}{\partial t} & = \frac{1}{2}\Delta u \quad \text{in } [0, \infty) \times U; \\ u & \text{is continuous on } [0, \infty) \times U \text{ and } u(0, x) = f(x). \end{cases}$$

Moreover we assume that $\lim\limits_{x\to b, x\in U} u(t, x) = 0$ if b belongs to ∂U. Then $u(t, x) = \mathbb{E}_x\left[f(b(t)) : \tau > t\right]$, where τ is the exit time of U: $\tau = \inf\{s > 0 : b(s) \in \mathbb{R}^\nu \backslash U\}$. Of course $\{b(s) : s \geqslant 0\}$ stands for ν-dimensional Brownian motion. In order to prove this claim we fix $t > 0$ and we consider the process $\{M(s) : 0 \leqslant s \leqslant t\}$ defined by $M(s) = u(t - s, b(s))1_{\{\tau > s\}}$. An application of Itô's formula yields the following identities:

$$\begin{aligned}
M(s) - M(0) &= -\int_0^s \frac{\partial u}{\partial t}(t-r, b(r))1_{\{\tau > r\}}dr + \int_0^s \nabla u(t-r, b(r))1_{\{\tau > r\}} \cdot db(r) \\
&\quad + \frac{1}{2}\int_0^s \Delta u(t-r, b(r))1_{\{\tau > r\}}dr \\
&= \int_0^s \left\{-\frac{\partial u}{\partial t}(t-r, b(r)) + \frac{1}{2}\Delta u(t-r, b(r))\right\} 1_{\{\tau > r\}}dr \\
&\quad + \int_0^s \nabla u(t-r, b(r))1_{\{\tau > r\}} \cdot db(r) \\
&= \int_0^s \nabla u(t-r, b(r))1_{\{\tau > r\}} \cdot db(r).
\end{aligned}$$

Consequently, the process $\{M(s) : 0 \leqslant s \leqslant t\}$ is a martingale. It follows that

$$\begin{aligned}
u(t,x) &= \mathbb{E}_x\left(u(t, b(0))\right) = \mathbb{E}_x(M(0)) = \mathbb{E}_x(M(t)) \\
&= \mathbb{E}_x\left(u(0, b(t))1_{\{\tau > t\}}\right) = \mathbb{E}_x\left(f(b(t))1_{\{\tau > t\}}\right).
\end{aligned}$$

Example 2. Let U be an open subset of $\mathbb{R}^\nu$, let f belong to $C_0(\partial U)$ and let $g : [0,\infty) \times U \to \mathbb{R}$ be a function in $C_0(E)$ and let $u : [0,\infty) \times U \to \mathbb{R}$ be a solution to the following problem:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{1}{2}\Delta u + g \ \text{ in } [0,\infty) \times U; \\ u \text{ is continuous on } [0,\infty) \times U \text{ and } u(0,x) = f(x). \end{cases}$$

Moreover we assume that $\lim_{x\to b, x\in U} u(t,x) = 0$ if b belongs to ∂U. Then $u(t,x) = \mathbb{E}_x\left(f(b(t)) : \tau > t\right) + \mathbb{E}_x\left(\int_0^{\min(t,\tau)} g(t-r, b(r))dr\right)$, where, as in Example 1, τ is the exit time of U. Also as in Example 1, $\{b(s) : s \geqslant 0\}$ stands for ν-dimensional Brownian motion. A proof can be given following the same lines as in the previous example.

Example 3. (Feynman-Kac formula) Let U be an open subset of $\mathbb{R}^\nu$, let f belong to $C_0(U)$ and let $V : U \to \mathbb{R}$ be an appropriate function and let $u : [0,\infty) \times U \to \mathbb{R}$ be a solution to the following problem:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{1}{2}\Delta u - Vu \text{ in } [0,\infty) \times U; \\ u \ \text{ is continuous on } [0,\infty) \times U \text{ and } u(0,x) = f(x). \end{cases}$$

Moreover we want that $\lim_{x\to b, x\in U} u(t,x) = 0$ if b belongs to ∂U. Then $u(t,x) = \mathbb{E}_x\left(\exp\left(-\int_0^t V(b(r))dr\right) f(b(t)) : \tau > t\right)$.

For the proof we fix $t > 0$ and we consider the process $\{M(s) : 0 \leqslant s \leqslant t\}$ defined by $M(s) = u(t-s, b(s)) \exp\left(-\int_0^s V(b(r))dr\right) 1_{\{\tau > s\}}$ and we apply Itô's formula to obtain:

$$\begin{aligned} &M(s) - M(0) \\ &= \int_0^s \left\{ -\frac{\partial u}{\partial t}(t-r, b(r)) + \frac{1}{2}\Delta u(t-r, b(r)) - V(b(r))u(t-r, b(r)) \right\} \\ &\qquad \exp\left(-\int_0^r V(b(\rho))d\rho\right) \mathbf{1}_{\{\tau > r\}} dr \\ &\quad + \int_0^s \nabla u(t-r, b(r)) \exp\left(-\int_0^r V(b(\rho))d\rho\right) \mathbf{1}_{\{\tau > r\}} \cdot db(r) \\ &= \int_0^s \nabla u(t-r, b(r)) \exp\left(-\int_0^r V(b(\rho))d\rho\right) \mathbf{1}_{\{\tau > r\}} \cdot db(r). \end{aligned}$$

Here we used the fact that u is supposed to be a solution of our initial value problem. It follows that the process $\{M(s) : 0 \leqslant s \leqslant t\}$ is a martingale. Hence we may conclude that

$$\begin{aligned} u(t,x) &= \mathbb{E}_x[M(0)] = \mathbb{E}_x\left[M(t)\right] \\ &= \mathbb{E}_x\left[u(0, b(t)) \exp\left(-\int_0^t V(b(\rho))d\rho\right) : \tau > t\right] \\ &= \mathbb{E}_x\left[f(b(t)) \exp\left(-\int_0^t V(b(\rho))d\rho\right) : \tau > t\right]. \end{aligned}$$

Example 4. (Cameron-Martin or Girsanov transformation). Let U be an open subset of $\mathbb{R}^\nu$, let f belong to $C_0(U)$ and let $c : U \to \mathbb{R}^\nu$ be an appropriate vector field on U and let $u : [0, \infty) \times U \to \mathbb{R}$ be a solution to the following problem:

$$\begin{cases} \dfrac{\partial u}{\partial t} & = \frac{1}{2}\Delta u + c.\nabla u \text{ in } [0,\infty) \times U; \\ u & \text{is continuous on } [0,\infty) \times U \text{ and } u(0,x) = f(x). \end{cases}$$

Moreover we want that $\lim_{x\to b, x\in U} u(t,x) = 0$ if b belongs to ∂U. Then $u(t,x) = \mathbb{E}_x \left(\exp\left(Z(t)\right) f(b(t)) : \tau > t\right)$, where $Z(t) = \int_0^t c(b(r)) \cdot db(r) - \frac{1}{2}\int_0^t \left|c(b(r))\right|^2 dr$.

For a proof we fix $t > 0$ and we consider the process $\{M(s) : 0 \leqslant s \leqslant t\}$ defined by $M(s) = u(t-s, b(s)) \exp\left(Z(s)\right) 1_{\{\tau > s\}}$. An application of Itô's formula to the function $f(s,x,y) = u(t-s,x)\exp(y)$ will yield the following result

$$\begin{aligned}
&M(s) - M(0) \\
&= \int_0^s -\frac{\partial u}{\partial t}(t-r, b(r)) \exp\left(Z(r)\right) \mathbf{1}_{\{\tau > r\}} dr \\
&\quad + \int_0^s \nabla u(t-r, b(r)) \exp\left(Z(r)\right) \mathbf{1}_{\{\tau > r\}} \cdot db(r) \\
&\quad + \int_0^s u(t-r, b(r)) \exp\left(Z(r)\right) dZ(r) \\
&\quad + \frac{1}{2}\int_0^{\min(s,\tau)} \Delta u(t-r, b(r)) \exp\left(Z(r)\right) \mathbf{1}_{\{\tau > r\}} dr \\
&\quad + \sum_{j=1}^{\nu} \int_0^s D_j u(t-r, b(r)) \exp\left(Z(r)\right) \mathbf{1}_{\{\tau > r\}} d\left\langle b_j, Z\right\rangle(r) \\
&\quad + \frac{1}{2}\int_0^s u(t-r, b(r)) \exp\left(Z(r)\right) \mathbf{1}_{\{\tau > r\}} d\left\langle Z, Z\right\rangle(r) \\
&= \int_0^s \left\{ -\frac{\partial u}{\partial t}(t-r, b(r)) + \frac{1}{2}\Delta u(t-r, b(r)) + c(b(r)).\nabla u(t-r, b(r)) \right\} \\
&\qquad \exp\left(Z(r)\right) \mathbf{1}_{\{\tau > r\}} dr \\
&\quad + \int_0^s \left\{\nabla u(t-r, b(r)) + u(t-r, b(r)) c(b(r))\right\} \exp\left(Z(r)\right) \mathbf{1}_{\{\tau > r\}} \cdot db(r) \\
&\quad - \frac{1}{2}\int_0^s u(t-r, b(r)) \exp\left(Z(r)\right) \left|c(b(r))\right|^2 \mathbf{1}_{\{\tau > r\}} dr \\
&\quad + \frac{1}{2}\int_0^s u(t-r, b(r)) \exp\left(Z(r)\right) \left|c(b(r))\right|^2 \mathbf{1}_{\{\tau > r\}} dr \\
&= \int_0^s \left\{\nabla u(t-r, b(r)) + u(t-r, b(r)) c(b(r))\right\} \exp\left(Z(r)\right) \mathbf{1}_{\{\tau > r\}} \cdot db(r).
\end{aligned}$$

As above it will follow that $u(t,x) = \mathbb{E}_x \left(\exp\left(Z(t)\right) f(b(t)) : \tau > t\right)$.

Example 5. (Stochastic differential equation). Let $\left(\sigma(x)\right)_{j,k=1}^{\nu}$, $x \in \mathbb{R}^\nu$, be a continuous square matrix valued function and let $c(x)$ be a so-called drift vector field (see the previous example). Suppose that the process $\{X^x(s) : s \geqslant 0\}$

satisfies the following (stochastic) integral equation:

$$X^x(t) = x + \int_0^t c(X^x(s))ds + \int_0^t \sigma(X^x(s)) \cdot db(s).$$

In other words the process $\{X^x(s) : s \geqslant 0\}$ is a solution of the following stochastic differential equation:

$$dX^x(t) = c(X^x(t))dt + \sigma(X^x(t)) \cdot db(t)$$

together with $X^x(0) = x$. The integral $\int_0^t \sigma(X^x(s))db(s)$ has the interpretation

$$\int_0^t \sigma(X^x(s))db(s) = \left(\sum_{k=1}^{\nu} \int_0^t \sigma_{jk}(X^x(s))db_k(s)\right)_{j=1}^{\nu}.$$

Next let $u : [0, \infty) \times \mathbb{R}^\nu \to \mathbb{R}$ be a twice continuously differentiable function. Then, by Itô's lemma,

$$\begin{aligned}
&u\left(t-s, X^x(s)\right) - u\left(t, X^x(0)\right) \\
&= -\int_0^s \frac{\partial u}{\partial t}\left(t-r, X^x(r)\right) dr \\
&\quad + \int_0^s \nabla u\left(t-r, X^x(r)\right) \cdot dX^x(r) \\
&\quad + \frac{1}{2}\sum_{j,k=1}^{\nu} \int_0^s D_j D_k u\left(t-r, X^x(r)\right) d\left\langle X_j^x, X_k^x\right\rangle(r).
\end{aligned}$$

Next we compute

$$\begin{aligned}
d\left\langle X_j^x, X_k^x\right\rangle(r) &= \sum_{m,n=1}^{\nu} \sigma_{jm}\left(X_j^x m(r)\right)\sigma_{kn}\left(X^x(r)\right) d\left\langle b_m, b_n\right\rangle(r) \\
&= \sum_{m=1}^{\nu} \sigma_{jm}\left(X^x(r)\right)\sigma_{km}\left(X^x(r)\right) dr = \left(\sigma\left(X^x(r)\right)\sigma\left(X^x(r)\right)^\tau\right)_{jk} dr,
\end{aligned}$$

where $\sigma(x)^\tau$ is the transposed matrix of $\sigma(x)$. Next we introduce the differential operator L as follows:

$$[Lf](x) = \frac{1}{2}\sum_{j,k=1}^{\nu} \left(\sigma(x)\sigma(x)^\tau\right)_{jk} D_j D_k f(x) + \sum_{j=1}^{\nu} c_j(x) D_j f(x).$$

For our twice continuously differential function u we obtain:

$$\begin{aligned}
&u\left(t-s, X^x(s)\right) - u\left(t, X^x(0)\right) \\
&= -\int_0^s \frac{\partial u}{\partial t}\left(t-r, X^x(r)\right) dr \\
&\quad + \sum_{j=1}^{\nu}\int_0^s c_j\left(X^x(r)\right) D_j u\left(t-r, X^x(r)\right) dr \\
&\quad + \int_0^s \nabla u\left(t-r, X^x(r)\right)\sigma\left(X^x(r)\right) \cdot db(r)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2}\sum_{j,k=1}^{\nu}\int_0^s \left(\sigma\left(X^x(r)\right)\sigma\left(X^x(r)\right)^{\tau}\right)_{j,k} D_j D_k u\left(t-r, X^x(r)\right) dr \\
&= \int_0^s \nabla u\left(t-r, X^x(r)\right)\sigma\left(X^x(r)\right)\cdot db(r) + \int_0^s \left(L-\frac{\partial}{\partial t}\right) u\left(t-r, X^x(r)\right) dr.
\end{aligned}
$$

So that, if $\left(L-\dfrac{\partial}{\partial t}\right) u \equiv 0$, then, for $0 \leqslant s \leqslant t$,

$$
u\left(t-s, X^x(s)\right) - u\left(t, X^x(0)\right) = \int_0^s \nabla u\left(t-r, X^x(r)\right)\sigma\left(X^x(r)\right) db(r),
$$

and hence, the process $M(s) := u\left(t-s, X^x(s)\right)$ is a martingale on the interval $[0,t]$. It follows that

$$
u(t,x) = \mathbb{E}(M(0)) = \mathbb{E}(M(t)) = \mathbb{E}\left(u\left(0, X^x(t)\right)\right) = \mathbb{E}\left(f\left(X^x(t)\right)\right)
$$

where $u(0,x) = f(x)$. For more details on stochastic differential equations see Chapter 4.

Example 6. (Quantum mechanical magnetic field). Let $\vec{a}$ be an appropriate vector field on $\mathbb{R}^\nu$ and let $H(\vec{a}, V) = \frac{1}{2}\left(i\nabla + \vec{a}\right)^2 + V$ be the (quantum mechanical) Hamiltonian of a particle under the influence of the scalar potential V in a *magnetic field* $\vec{B}(x)$ with vector potential $\vec{a}(x)$: i.e. $\vec{B} = \nabla \times \vec{a}$. Let f be a function in $C_0\left(\mathbb{R}^\nu\right)$ and let $u : [0,\infty) \times \mathbb{R}\nu \to \mathbb{R}$ be a solution to the following problem:

$$\begin{cases} \dfrac{\partial u}{\partial t} & = -H(\vec{a}, V)u \text{ in } [0,\infty) \times \mathbb{R}^\nu; \\ u & \text{is continuous on } [0,\infty) \times \mathbb{R}^\nu \text{ and } u(0,x) = f(x). \end{cases}$$

Moreover we want that $\lim\limits_{x\to\infty} u(t,x) = 0$. Then $u(t,x) = \mathbb{E}_x\left[e^{Z(t)} f(b(t))\right]$, where

$$Z(t) = -i\int_0^t \vec{a}(b(s)) \cdot db(s) - \frac{1}{2}i\int_0^t \nabla \cdot \vec{a}(b(s))ds - \int_0^t V(b(s))ds$$

with $\nabla \cdot \vec{a} = \sum_{j=1}^{\nu} \dfrac{\partial a_j}{\partial x_j}$. Put $M(s) = u(t-s, b(s)) \exp\left(Z(s)\right)$, $0 < s < t$. An application of Itô's formula to the function $f(s,x,y) = u(t-s,x)\exp(y)$ will yield the following result

$$\begin{aligned}
&M(s) - M(0) = f(s, b(s), Z(s)) - f(0, b(0), Z(0)) \\
&= \int_0^s \frac{\partial f}{\partial s}\left(\sigma, b(\sigma), Z(\sigma)\right) d\sigma + \int_0^s \nabla_x f(\sigma, b(\sigma), Z(\sigma)) \cdot db(\sigma) \\
&\quad + \int_0^s \frac{\partial f}{\partial y}(\sigma, b(\sigma), Z(\sigma)) \cdot dZ(\sigma) + \frac{1}{2}\int_0^s \Delta_x f(\sigma, b(\sigma), Z(\sigma)) d\sigma \\
&\quad + \frac{1}{2}\sum_{j=1}^{\nu} \int_0^s \frac{\partial^2}{\partial y \partial x_j} f(\sigma, b(\sigma), Z(\sigma)) d\left\langle Z, b_j\right\rangle(\sigma) \\
&\quad + \frac{1}{2}\sum_{j=1}^{\nu} \int_0^s \frac{\partial^2 f}{\partial x_j \partial y}(\sigma, b(\sigma), Z(\sigma))\, d\left\langle b_j, Z\right\rangle(\sigma) \\
&\quad + \frac{1}{2}\int_0^s \frac{\partial^2 f}{\partial y^2}(\sigma, b(\sigma), Z(\sigma)) d\left\langle Z, Z\right\rangle(\sigma) \\
&= \int_0^s \frac{\partial f}{\partial s}\left(\sigma, b(\sigma), Z(\sigma)\right) d\sigma + \int_0^s \nabla_x f(\sigma, b(\sigma), Z(\sigma)) \cdot db(\sigma) \\
&\quad + \int_0^s f(\sigma, b(\sigma), Z(\sigma)) \cdot dZ(\sigma) + \frac{1}{2}\int_0^s \Delta_x f(\sigma, b(\sigma), Z(\sigma)) d\sigma \\
&\quad - i\sum_{j=1}^{\nu} \int_0^s \frac{\partial f}{\partial x_j}(\sigma, b(\sigma), Z(\sigma)) a_j\big(b(\sigma)\big)\, d\sigma \\
&\quad + \frac{1}{2}\int_0^s f(\sigma, b(\sigma), Z(\sigma)) \sum_{j=1}^{\nu} a_j\left(b(\sigma)\right)^2 d\sigma \\
&= -\int_0^s \frac{\partial u}{\partial t}\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} d\sigma + \int_0^s \nabla_x u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} \cdot db(\sigma)
\end{aligned}$$

$$
\begin{aligned}
&+\int_0^s u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} \cdot dZ(\sigma) + \frac{1}{2}\int_0^s \Delta_x u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} d\sigma \\
&- i \sum_{j=1}^{\nu} \int_0^s \frac{\partial u}{\partial x_j}\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} a_j(b(\sigma) d\sigma \\
&+ \frac{1}{2}\int_0^s u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} \sum_{j=1}^{\nu} a_j\left(b(\sigma)\right)^2 d\sigma \\
=& \int_0^s \left\{ -\frac{\partial u}{\partial t} + \frac{1}{2}\Delta_x u - i\vec{a}(b(\sigma)).\nabla_x u - \frac{1}{2}\left|\vec{a}(b(\sigma))\right|^2 u \right. \\
&\quad \left. -\frac{1}{2} i \nabla \cdot \vec{a}(b(\sigma)) u - V(b(\sigma)) u \right\} \left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} d\sigma \\
&+ \int_0^s \nabla_x u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} \cdot db(\sigma) - i \int_0^s u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} \vec{a}(b(\sigma)) \cdot db(\sigma) \\
=& \int_0^s \left\{ -\frac{\partial}{\partial t} - \frac{1}{2}\left(i\nabla + \vec{a}\right)^2 - V \right\} u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} d\sigma \\
&+ \int_0^s \nabla_x u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} \cdot db(\sigma) - i \int_0^s u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} \vec{a}(b(\sigma)) \cdot db(\sigma) \\
=& \int_0^s \nabla_x u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} \cdot db(\sigma) - i \int_0^s u\left(t-\sigma, b(\sigma)\right) e^{Z(\sigma)} \vec{a}(b(\sigma)) \cdot db(\sigma).
\end{aligned}
$$

Here we used the fact that the function u satisfies the differential equation. The claim in the beginning of the example then follows as in Example 4.

Example 7. A geometric Brownian motion (GBM) (occasionally called exponential Brownian motion) is a continuous-time stochastic process in which the logarithm of the randomly varying quantity follows a Brownian motion, also called a Wiener process: see *e.g.* Ross [**116**] Section 10.3.2. It is applicable to mathematical modelling of some phenomena in financial markets. It is used particularly in the field of option pricing because a quantity that follows a GBM may take any positive value, and only the fractional changes of the random variate are significant. This is a reasonable approximation of stock price dynamics except for rare events.

A stochastic process S_t is said to follow a GBM if it satisfies the following stochastic differential equation:

$$dS(t) = \mu S(t)\, dt + \sigma S(t)\, dW(t)$$

where $W(t)$ is a Wiener process or Brownian motion and μ (“the percentage drift” or “drift rate”) and σ (“the (percentage or ratio) volatility”) are constants.

For an arbitrary initial value $S(0)$ the equation has the analytic solution

$$S(t) = S(0) \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W(t)\right),$$

which is a log-normally distributed random variable with expected value given by $\mathbb{E}[S(t)] = e^{\mu t} S(0)$ and variance by $\mathrm{Var}(S(t)) = e^{2\mu t} S(0)^2 \left(e^{\sigma^2 t} - 1\right)$.

The correctness of the solution can be verified using Itô's lemma. The random variable $\log\left(\frac{S(t)}{S(0)}\right)$ is normally distributed with mean $\left(\mu - \frac{1}{2}\sigma^2\right)t$ and variance $\sigma^2 t$, which reflects the fact that increments of a GBM are normal relative to the current price, which is why the process has the name "geometric".

Example 8. The term Black-Scholes refers to three closely related concepts:

1. The Black-Scholes model is a mathematical model of the market for an equity, in which the equity's price is a stochastic process.
2. The Black-Scholes PDE is a partial differential equation which (in the model) must be satisfied by the price of a derivative on the equity.
3. The Black-Scholes formula is the result obtained by solving the Black-Scholes PDE for a European call option.

Fischer Black and Myron Scholes first articulated the Black-Scholes formula in their 1973 paper, "The Pricing of Options and Corporate Liabilities.": see [**19**]. The foundation for their research relied on work developed by scholars such as Jack L. Treynor, Paul Samuelson, A. James Boness, Sheen T. Kassouf, and Edward O. Thorp. The fundamental insight of Black-Scholes is that the option is implicitly priced if the stock is traded.

Robert C. Merton was the first to publish a paper expanding the mathematical understanding of the options pricing model and coined the term "Black-Scholes" options pricing model.

Merton and Scholes received the 1997 The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel for this and related work. Though ineligible for the prize because of his death in 1995, Black was mentioned as a contributor by the Swedish academy.

7. Black-Scholes model

The text in this section is taken from Wikipedia (English version). The Black-Scholes model of the market for a particular equity makes the following explicit assumptions:

1. It is possible to borrow and lend cash at a known constant risk-free interest rate.
2. The price follows a geometric Brownian motion with constant drift and volatility.
3. There are no transaction costs.
4. The stock does not pay a dividend (see below for extensions to handle dividend payments).
5. All securities are perfectly divisible (i.e. it is possible to buy any fraction of a share).
6. There are no restrictions on short selling.
7. There is no arbitrage opportunity.

From these ideal conditions in the market for an equity (and for an option on the equity), the authors show that it is possible to create a hedged position, consisting of a long position in the stock and a short position in [calls on the same stock], whose value will not depend on the price of the stock.

Notation. We define the following quantities:

- S, the price of the stock (please note as below).
- $V(S,t)$, the price of a financial derivative as a function of time and stock price.
- $C(S,t)$ the price of a European call and $P(S,t)$ the price of a European put option.
- K, the strike of the option.
- r, the annualized risk-free interest rate, continuously compounded.
- μ, the drift rate of S, annualized.
- σ, the volatility of the stock; this is the square root of the quadratic variation of the stock's log price process.
- t a time in years; we generally use now $= 0$, expiry $= T$.
- Π, the value of a portfolio.
- R, the accumulated profit or loss following a delta-hedging trading strategy.
- $N(x)$ denotes the standard normal cumulative distribution function,

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}z^2}\, dz.$$

- $N'(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ denotes the standard normal probability density function.

Black-Scholes PDE. Simulated Geometric Brownian Motions with Parameters from Market Data

In the model as described above, we assume that the underlying asset (typically the stock) follows a geometric Brownian motion. That is,

$$dS(t) = \mu S(t)\,dt + \sigma S(t)\,dW(t),$$

where $W(t)$ is a Brownian motion; the dW term here stands in for any and all sources of uncertainty in the price history of a stock.

The payoff of an option $V(S,T)$ at maturity is known. To find its value at an earlier time we need to know how V evolves as a function of S and T. By Itô's lemma for two variables we have

$$\begin{aligned} dV\left(S(t),t\right) &= \frac{\partial V\left(S(t),t\right)}{\partial S}\,dS(t) + \frac{\partial V\left(S(t),t\right)}{\partial t}\,dt + \frac{1}{2}\frac{\partial^2 V\left(S(t),t\right)}{\partial^2 S}\,d\left\langle S,S\right\rangle(t) \\ &= \left(\mu S(t)\frac{\partial V\left(S(t),t\right)}{\partial S} + \frac{\partial V\left(S(t),t\right)}{\partial t} + \frac{1}{2}\sigma^2 S(t)^2\frac{\partial^2 V\left(S(t),t\right)}{\partial S^2}\right) dt \\ &\quad + \sigma S(t)\frac{\partial V\left(S(t),t\right)}{\partial S}\,dW(t). \end{aligned} \tag{3.208}$$

Now consider a trading strategy under which one holds $a(t)$ units of a single option with value $S(t)$ and $b(t)$ units of a bond with value $\beta(t)$ at time t. The value $V\left(S(t),t\right)$ of the portfolio of the trading strategy $(a(t),b(t))$ is then given by

$$V\left(S(t),t\right) = a(t)S(t) + b(t)\beta(t). \tag{3.209}$$

Observe that (3.209) is equivalent to

$$b(t) = \frac{V\left(S(t),t\right) - a(t)S(t)}{\beta(t)}.$$

In addition, $a(t) = \dfrac{\partial V\left(S(t),t\right)}{\partial s}$, which is called the *delta hedging rule*. Assuming, like in the Black-Sholes model, that the strategy $(a(t),b(t))$ is self-financing, which by definition implies

$$dV\left(S(t),t\right) = a(t)\,dS(t) + b(t)\,d\beta(t), \tag{3.210}$$

we get

$$dV(t) = \mu a(t)S(t)\,dt + b(t)\,d\beta(t) + \sigma a(t)S(t)\,dW(t). \tag{3.211}$$

Assume that the process $t \mapsto \beta(t)$, *i.e.*, the bond price, is of bounded variation. By equating the terms with $dW(t)$ in (3.208) and (3.211) we see $a(t) = \dfrac{\partial V\left(S(t),t\right)}{\partial S}$. From this and again equating the other terms in (3.208) and (3.211) and using (3.209) we also obtain

$$\begin{aligned} &\left(\frac{\partial V\left(S(t),t\right)}{\partial t} + \frac{1}{2}\sigma^2 S(t)^2\frac{\partial^2 V\left(S(t),t\right)}{\partial S^2}\right) dt \\ &= b(t)\,d\beta(t) = \left(V\left(S(t),t\right) - S(t)\frac{\partial V\left(S(t),t\right)}{\partial s}\right)\frac{d\beta(t)}{\beta(t)}. \end{aligned} \tag{3.212}$$

If the interest rate for the bond is constant, *i.e.*, if $d\beta(t) = r\beta(t)\,dt$, or, what amounts to the same, $\beta(t) = \beta(0)e^{rt}$, then from (3.212) it also follows that

$$\begin{aligned}&\frac{\partial V\left(S(t),t\right)}{\partial t} + \frac{1}{2}\sigma^2 S(t)^2 \frac{\partial^2 V\left(S(t),t\right)}{\partial S^2}\\ &= r\left(V\left(S(t),t\right) - S(t)\frac{\partial V\left(S(t),t\right)}{\partial s}\right). \qquad (3.213)\end{aligned}$$

If we trade in a single option continuously trades in the stock in order to hold $-\dfrac{\partial V}{\partial S}$ shares, then at time t, the value of these holdings will be

$$\Pi(t) = V\left(S(t),t\right) - S(t)\frac{\partial V\left(S(t),t\right)}{\partial S}.$$

The composition of this portfolio, called the delta-hedge portfolio, will vary from time-step to time-step. Let $R(t)$ denote the accumulated profit or loss from following this strategy. Then over the time period $[t, t+dt]$, the instantaneous profit or loss is

$$dR(t) = dV\left(S(t),t\right) - \frac{\partial V\left(S(t),t\right)}{\partial S}\,dS(t).$$

By substituting in the equations above we get

$$dR(t) = \left(\frac{\partial V\left(S(t),t\right)}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V\left(S(t),t\right)}{\partial S^2}\right)\,dt.$$

This equation contains no $dW(t)$ term. That is, it is entirely risk free (delta neutral). Black, Scholes and Merton reason that under their ideal conditions, the rate of return on this portfolio must be equal at all times to the rate of return on any other risk free instrument; otherwise, there would be opportunities for arbitrage. Now assuming the risk free rate of return is r we must have over the time period $[t, t+dt]$ (Black-Scholes assumption):

$$r\Pi(t)\,dt = dR(t) = \left(\frac{\partial V\left(S(t),t\right)}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V\left(S(t),t\right)}{\partial S^2}\right)\,dt.$$

Observe that the Black-Sholes assumption comes down to the assumption of self-financing, because the results If we now substitute in for $\Pi(t)$ and divide through by dt we obtain the Black-Scholes PDE:

$$\frac{\partial V\left(S(t),t\right)}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 V\left(S(t),t\right)}{\partial S^2} + rS\frac{\partial V\left(S(t),t\right)}{\partial S} - rV\left(S(t),t\right) = 0. \quad (3.214)$$

Observe that the Black-Sholes assumption comes down to the assumption of self-financing, because the resulting partial differential equation in (3.213) and (3.214) is the same. With the assumptions of the Black-Scholes model, this partial differential equation holds whenever V is twice differentiable with respect to S and once with respect to t. Above we used the method of arbitrage-free pricing ("delta-hedging") to derive some PDE governing option prices given the Black-Scholes model. It is also possible to use a risk-neutrality argument. This latter method gives the price as the expectation of the option payoff under a particular probability measure, called the risk-neutral measure, which differs from the real world measure.

Black-Scholes formula. The Black-Scholes formula is used for obtaining the price of European put and call options. It is obtained by solving the Black-Scholes PDE as discussed - see derivation below.

The value of a call option in terms of the Black-Scholes parameters is given by:

$$C(S,t) = C\left(S(t),t\right) = S(t)N(d_1) - Ke^{-r(T-t)}N(d_2) \quad \text{with} \tag{3.215}$$

$$d_1 = \frac{\log(\frac{S}{K}) + \left(r + \frac{\sigma^2}{2}\right)(T-t)}{\sigma\sqrt{T-t}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T-t}. \tag{3.216}$$

The price of a put option is:

$$P(S,t) = P\left(S(t),t\right) = Ke^{-r(T-t)}N(-d_2) - S(t)N(-d_1). \tag{3.217}$$

For both, as above:

1. $N(\cdot)$ is the standard normal or cumulative distribution function.
2. $T - t$ is the time to maturity.
3. $S = S(t)$ is the spot price of the underlying asset at time t.
4. K is the strike price.
5. r is the risk free interest rate (annual rate, expressed in terms of continuous compounding).
6. σ is the volatility in the log-returns of the underlying asset.

Interpretation. The quantities $N(d_1)$ and $N(d_2)$ are the probabilities of the option expiring in-the-money under the equivalent exponential martingale probability measure (numéraire = stock) and the equivalent martingale probability measure (numéraire = risk free asset), respectively. The equivalent martingale probability measure is also called the risk-neutral probability measure. Note that both of these are probabilities in a measure theoretic sense, and neither of these is the true probability of expiring in-the-money under the real probability measure.

Derivation. We now show how to get from the general Black-Scholes PDE to a specific valuation for an option. Consider as an example the Black-Scholes price of a call option, for which the PDE above has boundary conditions

$$\begin{aligned} C(0,t) &= 0 \text{ for all } t \\ C(S,t) &\to S \text{ as } S \to \infty \\ C(S,T) &= \max(S-K,0). \end{aligned}$$

The last condition gives the value of the option at the time that the option matures. The solution of the PDE gives the value of the option at any earlier time, $\mathbb{E}\left[\max(S-K,0)\right]$. In order to solve the PDE we transform the equation into a diffusion equation which may be solved using standard methods. To this end we introduce the change-of-variable transformation

$$\tau = T - t, \quad u(x,\tau) = C\left(Ke^{x-\left(r-\frac{1}{2}\sigma^2\right)\tau}, T-\tau\right)e^{r\tau}, \text{ and } x = \log\frac{S}{K} + (r - \frac{\sigma^2}{2})\tau.$$

Note: in fact in case we consider a call option we replace $V\left(S(t),t\right)$ with $C\left(S(t),t\right)$. Instead of u we may also consider

$$v(x,t) = V\left(Ke^{x-\left(r-\frac{1}{2}\sigma^2\right)(T-t)}, t\right)e^{r(T-t)}.$$

In case we consider a European call option we take as final value for v: $v(x,T) = V\left(Ke^x,T\right) = C\left(Ke^x,T\right) = \max\left(Ke^x - K, 0\right) = K\max\left(e^x-1,0\right)$. Then the Black-Scholes PDE becomes a diffusion equation

$$\frac{\partial u}{\partial \tau} = \frac{1}{2}\sigma^2\frac{\partial^2 u}{\partial x^2}.$$

The terminal condition $C(S,T) = \max(S-K,0)$ now becomes an initial condition

$$u(x,0) = u_0(x) \equiv K\max\left(e^x-1,0\right).$$

Using the standard method for solving a diffusion equation we have

$$u(x,\tau) = \frac{1}{\sigma\sqrt{2\pi\tau}}\int_{-\infty}^{\infty} u_0(y)e^{-(x-y)^2/(2\sigma^2\tau)}\,dy.$$

After some calculations we obtain

$$u(x,\tau) = Ke^{x+\sigma^2\tau/2}N\left(d_1\right) - KN\left(d_2\right)$$

where

$$d_1 = \frac{x+\sigma^2\tau}{\sigma\sqrt{\tau}} \quad \text{and} \quad d_2 = \frac{x}{\sigma\sqrt{\tau}}.$$

Substituting for u, x, and τ, we obtain the value of a call option in terms of the Black-Scholes parameters is given by

$$C(S,t) = SN(d_1) - Ke^{-r(T-t)}N(d_2),$$

where d_1 and d_2 are as in (3.216). The price of a put option may be computed from this by the put-call parity and simplifies to

$$P(S,t) = Ke^{-r(T-t)}N(-d_2) - SN(-d_1).$$

Risk neutral measure. Suppose our economy consists of 2 assets, a stock and a risk-free bond, and that we use the Black-Scholes model. In the model the evolution of the stock price can be described by Geometric Brownian Motion:

$$dS(t) = \mu S(t)\,dt + \sigma S(t)\,dW(t)$$

where $W(t)$ is a standard Brownian motion with respect to the physical measure. If we define

$$\widetilde{W}(t) = W(t) + \frac{\mu - r}{\sigma}t,$$

Girsanov's theorem states that there exists a measure Q under which $\widetilde{W}(t)$ is a standard Brownian motion, *i.e.*, a Brownian motion without a drift term and such that $\mathbb{E}_Q\left[\widetilde{W}(t)^2\right] = t$. For a more thorough discussion on the Girsanov's theorem, which is in fact (much) more general, see assertion (4) in Proposition 4.24 in Chapter 4 Section 3. The quantity $\dfrac{\mu - r}{\sigma}$ is known as the market price of risk. Differentiating and rearranging yields:

$$dW(t) = d\widetilde{W}(t) - \frac{\mu - r}{\sigma}\,dt.$$

Put this back in the original equation:

$$dS(t) = rS(t)\,dt + \sigma S(t)\,d\widetilde{W}(t).$$

The probability Q is the unique risk-neutral measure for the model. The (discounted) payoff process of a derivative on the stock $H(t) = \mathbb{E}_Q\left(H(T) \mid \mathcal{F}_t\right)$ is a martingale under Q. Since S and H are Q-martingales we can invoke the martingale representation theorem to find a replicating strategy – a holding of stocks and bonds that pays off $H(t)$ at all times $t \leqslant T$. The measure Q is given by $Q(A) = \mathbb{E}\left[e^{-Z(T)}\mathbf{1}_A\right]$, $A \in \mathcal{F}_T$, where

$$Z(t) = \frac{1}{2}\left(\frac{\mu - r}{\sigma}\right)^2 t + \frac{\mu - r}{\sigma}W(t).$$

In fact a more general result is true. Let $s \mapsto h(s)$ be a predictable process such that $\mathbb{E}\left[\exp\left(\dfrac{1}{2}\displaystyle\int_0^T |h(s)|^2\,ds\right)\right] < \infty$. Put

$$Z_h(t) = \int_0^t h(s)dW(s) + \frac{1}{2}\int_0^t |h(s)|^2\,ds.$$

Define the measure Q_h by $Q_h(A) = \mathbb{E}\left[e^{-Z_h(T)}\mathbf{1}_A\right]$, $A \in \mathcal{F}_T$. Put $W_h(t) = W(t) + \int_0^t h(s)\,ds$. Then the process W_h is a Brownian motion relative to the measure Q_h. The proof of this result uses Lévy's characterization of Brownian motion: see Corollary 4.7. It says that a process W_h is a Q_h-Brownian motion if and only if the following two conditions are satisfied:

(1) The quadratic variation of W_h satisfies $\langle W_h, W_h\rangle (t) = t$.
(2) The process W_h is a local martingale relative to the measure Q_h.

(For a proof of this result see Theorem 4.5.) In our case we have $\langle W_h, W_h\rangle (t) = \langle W, W\rangle (t) = t$, and so (1) is satisfied. In order to establish (2) we use Itô calculus to obtain:

$$e^{-Z_h(t)}W_h(t) = \int_0^t e^{-Z_h(s)}dW(s) - \int_0^t e^{-Z_h(s)}W_h(s)h(s)\,dW(s).$$

Since the process $t \mapsto e^{-Z_h(t)}$ is a martingale we see that the process W_h is a local Q_h-martingale.

We like to spend more time on the Black-Sholes model and the corresponding risk-neutral measure. Again we have trading strategy $(a(t), b(t))$ of a financial asset and a bond. Its portfolio value $V(t) := V(S(t), t)$ is given by $V(t) = a(t)S(t) + b(t)\beta(t)$. Here $S(t)$ is the price of the option at time t and $\beta(t)$ is the price of the bond at time t. It is assumed that the process $t \mapsto S(t)$ follows a geometric Brownian motion: $dS(t) = \mu S(t)\,dt + \sigma S(t)\,dW(t)$, or $S(t) = S(0)e^{\sigma W(t)+\left(\mu-\frac{1}{2}\sigma^2\right)t}$. Let $\widetilde{S}(t)$ be the discounted price of the option, *i.e.*,

$$\widetilde{S}(t) = \frac{\beta(0)}{\beta(t)}S(t). \tag{3.218}$$

Put $\widetilde{W}(t) = W(t) + \int_0^t q(s)\,ds$, where

$$q(s) = \frac{1}{\sigma}\left(\mu - \frac{\beta'(s)}{\beta(s)}\right).$$

Then the process $\widetilde{S}(t)$ satisfies the equation

$$d\widetilde{S}(t) = \sigma\frac{\beta(0)}{\beta(t)}S(t)\,d\left(\frac{1}{\sigma}\int_0^t\left(\mu - \frac{\beta'(s)}{\beta(s)}\right)ds + W(t)\right) = \sigma\widetilde{S}(t)\,d\widetilde{W}(t). \tag{3.219}$$

Put $Z_q(t) = \displaystyle\int_0^t q(s)\,dW(s) + \frac{1}{2}\int_0^t q(s)^2\,ds$. By Girsanov's theorem the process $t \mapsto \widetilde{W}(t)$ is a (standard) Brownian motion under the measure $\mathbb{Q}_q$ given by $\mathbb{Q}_q(A) = \mathbb{E}\left[e^{-Z_q(T)}\mathbf{1}_A\right]$, $A \in \mathcal{F}_T$. The solution $\widetilde{S}(t)$ of the SDE in (3.219) can be written in the form

$$\widetilde{S}(t) = \widetilde{S}(0)e^{\sigma\widetilde{W}(t)-\frac{1}{2}\sigma^2 t}. \tag{3.220}$$

Assume that the portfolio is self-financing we will show that

$$V(t) = \mathbb{E}^{\mathbb{Q}_q}\left[\frac{\beta(t)}{\beta(T)}h(S(T)) \,\middle|\, \mathcal{F}_t\right], \quad t \in [0, T], \tag{3.221}$$

where $V(T)$ is equal to the contingent claim $h(S(T))$ at the time of maturity T. Of course, $\mathbb{E}^{\mathbb{Q}_q}\left[F \,\middle|\, \mathcal{F}_t\right]$ denotes the conditional expectation of F relative $\mathbb{Q}_q$, given the σ-field $\mathcal{F}_t = \sigma(W(s) : s \leqslant t)$ of the variable $F \in L^1(\Omega, \mathcal{F}_T, \mathbb{Q}_q)$ with respect to the probability measure $\mathbb{Q}_q$. Another application of Itô's lemma together with the definition of $\widetilde{S}(t)$, $\widetilde{W}(t)$ and $\widetilde{V}(t) = \dfrac{\beta(0)}{\beta(t)}V(t)$ shows the following result

$$d\widetilde{V}(t) = -\frac{\beta(0)\beta'(t)}{\beta(t)^2}V(t)\,dt + \frac{\beta(0)}{\beta(t)}\,dV(t)$$

(the hedging strategy $(a(t), b(t))$ is self-financing)

$$= -\frac{\beta(0)\beta'(t)}{\beta(t)^2}V(t)\,dt + \frac{\beta(0)}{\beta(t)}\left(a(t)\,dS(t) + b(t)d\beta(t)\right)$$

(employ the equation for the option price $S(t)$)

$$\begin{aligned}
&= -\frac{\beta(0)\beta'(t)}{\beta(t)^2}V(t)\,dt + \frac{\beta(0)}{\beta(t)}S(t)\left(\mu a(t)\,dt + b(t)\frac{\beta'(t)}{\beta(t)}\,dt + \sigma dW(t)\right)\\
&= -\frac{\beta(0)\beta'(t)}{\beta(t)^2}\left(a(t)S(t) + b(t)\beta(t)\right)\,dt\\
&\qquad + \frac{\beta(0)}{\beta(t)}S(t)\left(\mu a(t)\,dt + b(t)\frac{\beta'(t)}{\beta(t)}\,dt + \sigma dW(t)\right)\\
&= \sigma a(t)\frac{\beta(0)}{\beta(t)}S(t)\,d\left\{\frac{1}{\sigma}\int_0^t\left(\mu - \frac{\beta'(s)}{\beta(s)}\right)\,ds + W(t)\right\}\\
&= \sigma a(t)\widetilde{S}(t)\,d\widetilde{W}(t).
\end{aligned} \tag{3.222}$$

In fact the equality in (3.222) could also have been obtained by observing that

$$d\widetilde{V}(t) = a(t)\, d\widetilde{S}(t), \quad \text{and} \quad d\widetilde{S}(t) = \sigma d\widetilde{W}(t). \tag{3.223}$$

From (3.222) we infer

$$\widetilde{V}(t) = \widetilde{V}(0) + \sigma \int_0^t a(s)\, d\widetilde{W}(s), \tag{3.224}$$

and hence, the process $t \mapsto \widetilde{V}(t)$ is a martingale with respect to the measure $\mathbb{Q}_q$. So from (3.224) we get

$$\frac{\beta(0)}{\beta(t)} V(t) = \widetilde{V}(t) = \mathbb{E}^{\mathbb{Q}_q}\left[\widetilde{V}(T) \mid \mathcal{F}_t\right] = \mathbb{E}^{\mathbb{Q}_q}\left[\frac{\beta(0)}{\beta(T)} V(T) \mid \mathcal{F}_t\right], \tag{3.225}$$

and hence

$$V(t) = \mathbb{E}^{\mathbb{Q}_q}\left[\frac{\beta(t)}{\beta(T)} V(T) \mid \mathcal{F}_t\right] = \mathbb{E}^{\mathbb{Q}_q}\left[\frac{\beta(t)}{\beta(T)} h\left(S(T)\right) \mid \mathcal{F}_t\right]. \tag{3.226}$$

In addition, we observe that

$$\begin{aligned}
&S(t)\frac{\beta(T)}{\beta(t)} e^{-\frac{1}{2}\sigma^2(T-t)+\sigma\left(\widetilde{W}(T)-\widetilde{W}(t)\right)} \\
&= S(t)\exp\left(\int_t^T \left(\frac{\beta'(s)}{\beta(s)} - \frac{1}{2}\sigma^2\right) ds + \sigma\left(\widetilde{W}(T) - \widetilde{W}(t)\right)\right) \\
&= S(0)\exp\left(\sigma W(t) + \left(\mu - \frac{1}{2}\sigma^2\right) t\right) \\
&\quad\times \exp\left(\int_t^T \left(\frac{\beta'(s)}{\beta(s)} - \frac{1}{2}\sigma^2\right) ds + \sigma\left(W(T) - W(t) + \int_t^T q(s)\, ds\right)\right) \\
&= S(0)e^{\left(\mu-\frac{1}{2}\sigma^2\right)T+\sigma W(T)} = S(T).
\end{aligned} \tag{3.227}$$

Inserting the equality for $S(T)$ from (3.227) into (3.226) yields

$$V(t) = \mathbb{E}^{\mathbb{Q}_q}\left[\frac{\beta(t)}{\beta(T)} h\left(S(t)\frac{\beta(T)}{\beta(t)} e^{-\frac{1}{2}\sigma^2(T-t)+\sigma\left(\widetilde{W}(T)-\widetilde{W}(t)\right)}\right) \mid \mathcal{F}_t\right]. \tag{3.228}$$

Since the variable $S(t)$ is measurable with respect to $\mathcal{F}_t$, Since process $t \mapsto \widetilde{W}(t)$ is a $\mathbb{Q}_q$-Brownian motion, the variable $\widetilde{W}(T) - \widetilde{W}(t)$ and the σ-field $\mathcal{F}_t$ are $\mathbb{Q}_q$-independent. Moreover, the process $t \mapsto \beta(t)$ is supposed to deterministic. Hence, since the variable $S(t)$ is measurable with respect to $\mathcal{F}_t$, we deduce that

$$\begin{aligned}
V(t) &= V\left(S(t), t\right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\beta(t)}{\beta(T)} h\left(x\frac{\beta(T)}{\beta(t)} e^{-\frac{1}{2}\sigma^2(T-t)+\sigma\sqrt{T-t}y}\right) e^{-\frac{1}{2}y^2}\, dy\, \Big|_{x=S(t)}.
\end{aligned} \tag{3.229}$$

Hence if the pay-off, i.e. the value of the call option at expiry (time of maturity T), is given by $h\left(S(T)\right) = \max\left\{S(T) - K, 0\right\}$, then the value of the portfolio at time $t \leqslant T$ is given by the formula in (3.229). If $\beta(t) = \beta(0)e^{rt}$, then this integral can be rewritten as in (3.215) with $C\left(S, t\right) = C\left(S(t), t\right) = V\left(S(t), t\right) = V(t)$. Similarly, if $h\left(S(T)\right) = \max\left\{K - S(T), 0\right\}$, then $P\left(S, t\right) = P\left(S(t), t\right) = V\left(S(t), t\right) = V(t)$ is the price of a European put option: see the somewhat more explicit expression in (3.217). For a modern treatment of several stock price models see, *e.g.*, Gulisashvili [**60**].

8. An Ornstein-Uhlenbeck process in higher dimensions

Part of this text is taken from [**146**]. Let $C(t,s)$, $t \geqslant s$, t, $s \in \mathbb{R}$, be a family of $d \times d$ matrices with real entries, with the following properties:

(a) $C(t,t) = I$, $t \in \mathbb{R}$, (I stands for the identity matrix).
(b) The following identity holds: $C(t,s)C(s,\tau) = C(t,\tau)$ holds for all real numbers t, s, τ for which $t \geqslant s \geqslant \tau$.
(c) The matrix valued function $(t,s,x) \mapsto C(t,s)x$ is continuous as a function from the set $\left\{(t,s) \in \mathbb{R}^d \times \mathbb{R}^d : t \geqslant s\right\} \times \mathbb{R}^d$ to $\mathbb{R}^d$.

Define the backward propagator Y_C on $C_b\left(\mathbb{R}^d\right)$ by $Y_C(s,t)f(x) = f\left(C(t,s)x\right)$, $x \in \mathbb{R}^d$, $s \leqslant t$, and $f \in C_b\left(\mathbb{R}^d\right)$. Then Y_C is a backward propagator on the space $C_b\left(\mathbb{R}^d\right)$, which is $\sigma\left(C_b\left(\mathbb{R}^d\right), M\left(\mathbb{R}^d\right)\right)$-continuous. Here the symbol $M\left(\mathbb{R}^d\right)$ stands for the vector space of all signed measures on $\mathbb{R}^d$. The operator family $\{Y_C(s,t) : s \leqslant t\}$ satisfies $Y_C\left(s_1,s_2\right) Y_C\left(s_2,s_3\right) = Y_C\left(s_1,s_3\right)$, $s_1 \leqslant s_2 \leqslant s_3$.

Let $W(t)$ be standard m-dimensional Brownian motion on $(\Omega, \mathcal{F}_t, \mathbb{P})$ and let $\sigma(\rho)$ be a deterministic continuous function which takes its values in the space of $d \times m$-matrices. Put $Q(\rho) = \sigma(\rho)\sigma(\rho)^*$. Another interesting example is the following:

$$\begin{aligned}
&Y_{C,Q}\left(s,t\right) f(x) \\
&= \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} f\left(C(t,s)x + \left(\int_s^t C(t,\rho)Q(\rho)C(t,\rho)^* d\rho\right)^{1/2} y\right) dy \\
&= \mathbb{E}\left[f\left(C(t,s)x + \int_s^t C(t,\rho)\sigma(\rho)\, dW(\rho)\right)\right], \qquad (3.230)
\end{aligned}$$

where $Q(\rho) = \sigma(\rho)\sigma(\rho)^*$ is a positive-definite $d \times d$ matrix. Then the propagators $Y_{C,Q}$ and $Y_{C,S}$ are backward propagators on $C_b\left(\mathbb{R}^d\right)$. We will prove this. The equality of the expressions in (3.230) is a consequence of the following arguments. Let the variable $\xi \in \mathbb{R}^d$ have the standard normal distribution. Fix $t \geqslant \tau$. Both variables

$$\begin{aligned}
&X^{\tau,x}(t) := C\left(t,\tau\right) x + \int_\tau^t C\left(t,\rho\right) \sigma(\rho) dW(\rho), \quad t \geqslant \tau, \text{ and} \\
&C(t,\tau)x + \left(\int_\tau^t C(t,\rho)Q(\rho)C(t,\rho)^* d\rho\right)^{1/2} \xi, \quad t \geqslant \tau, \qquad (3.231)
\end{aligned}$$

are $\mathbb{R}^d$-valued Gaussian vectors. A calculation shows that they have the same expectation and the same covariance matrix with entries given by (3.242) below with $s = t$.

Next suppose that the forward propagator C on $\mathbb{R}^d$ consists of contractive operators, i.e. $C(t,s)C(t,s)^* \leqslant I$ (this inequality is to be taken in matrix sense).

Choose a family $S(t,s)$ of square $d \times d$-matrices such that $C(t,s)C(t,s)^* + S(t,s)\,S(t,s)^* = I$, and put

$$Y_{C,S}(s,t)\,f(x) = \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} f\left(C(t,s)x + S(t,s)y\right) dy. \tag{3.232}$$

In fact the example in (3.232) is a special case of the example in (3.230) provided $Q(\rho)$ is given by the following limit:

$$Q(\rho) = \lim_{h \downarrow 0} \frac{I - C(\rho - h)\,C(\rho - h)^*}{h} \tag{3.233}$$

If $Q(\rho)$ is as in (3.233), then

$$S(t,s)\,S(t,s)^* = I - C(t,s)\,C(t,s)^* = \int_s^t C(t,\rho)\,Q(\rho) C(t,\rho)^*\,d\rho.$$

The following auxiliary lemma will be useful. Condition (3.234) is satisfied if the three pairs (C_1, S_1), (C_2, S_2), and (C_3, S_3) satisfy: $C_1C_1^* + S_1S_1^* = C_2C_2^* + S_2S_2^* = C_3C_3^* + S_3S_3^* = I$. It also holds if $C_2 = C(t_2, t_1)$, and

$$S_jS_j^* = \int_{t_{j-1}}^{t_j} C(t_j,\rho)\,\sigma(\rho)\sigma(\rho)^* C(t_j,\rho)^*\,d\rho, \quad j = 1,\ 2, \quad \text{and}$$

$$S_3S_3^* = \int_{t_0}^{t_2} C(t_2,\rho)\,\sigma(\rho)\sigma(\rho)^* C(t_2,\rho)^*\,d\rho.$$

3.89. Lemma. *Let C_1, S_1, C_2, S_2, and C_3, S_3 be $d\times d$-matrices with the following properties:*

$$C_3 = C_2 C_1, \quad \text{and} \quad C_2 S_1 S_1^* C_2^* + S_2 S_2^* = S_3 S_3^*. \tag{3.234}$$

Let $f \in C_b\left(\mathbb{R}^d\right)$, and put

$$Y_{1,2}f(x) = \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} f\left(C_1 x + S_1 y\right) dy; \tag{3.235}$$

$$Y_{2,3}f(x) = \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} f\left(C_2 x + S_2 y\right) dy; \tag{3.236}$$

$$Y_{1,3}f(x) = \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} f\left(C_3 x + S_3 y\right) dy. \tag{3.237}$$

Then $Y_{1,2}Y_{2,3} = Y_{1,3}$.

Proof. Let the matrices C_j and S_j, $1 \leqslant j \leqslant 3$, be as in (3.234). Let $f \in C_b\left(\mathbb{R}^d\right)$. First we assume that the matrices S_1 and C_2 are invertible, and we put $A_3 = S_1^{-1} C_2^{-1} S_3$, and $A_2 = S_1^{-1} C_2^{-1} S_2$. Then, using the equalities in (3.234) we see $A_3 A_3^* = I + A_2 A_2^*$. We choose a $d \times d$-matrix A such that $A^* A = I + A_2^* A_2$, and we put $D = \left(A^{-1}\right)^* A_2^* A_3$. Then we have $A_3^* A_3 = I + D^* D$. Let $f \in C_b\left(\mathbb{R}^d\right)$. Let the vectors $(y_1, y_2) \in \mathbb{R}^d \times \mathbb{R}^d$ and $(y, z) \in \mathbb{R}^d \times \mathbb{R}^d$ be such that

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} A_3 & -A_2 A^{-1} \\ 0 & A^{-1} \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}. \tag{3.238}$$

Since

$$A_2 A_2^* \left(I + A_2 A_2^*\right)^{-1} = A_2 \left(I + A_2^* A_2\right)^{-1} A_2^*,$$

we obtain

$$\det\left(I + A_2 A_2^*\right) = \det\left(I + A_2^* A_2\right).$$

Hence, the absolute value of the determinant of the matrix in the right-hand side of (3.238) can be rewritten as:

$$\begin{aligned} \left|\det \begin{pmatrix} A_3 & -A_2 A^{-1} \\ 0 & A^{-1} \end{pmatrix}\right|^2 &= \left|\det A_3 \left(\det A\right)^{-1}\right|^2 \\ = \frac{\det\left(A_3 A_3^*\right)}{\det\left(A^* A\right)} &= \frac{\det\left(I + A_2 A_2^*\right)}{\det\left(I + A_2^* A_2\right)} = 1. \end{aligned} \tag{3.239}$$

From (3.238) and (3.239) it follows that the corresponding volume elements satisfy: $dy_1\, dy_2 = dy\, dz$. We also have

$$\left|y_1\right|^2 + \left|y_2\right|^2 = |y|^2 + |z - Dy|^2. \tag{3.240}$$

Employing the substitution (3.238) together with the equalities $dy_1\, dy_2 = dy\, dz$ and (3.240) and applying Fubini's theorem we obtain:

$$\begin{aligned} Y_{1,2}Y_{2,3}f(x) &= \frac{1}{(2\pi)^d} \iint e^{-\frac{1}{2}\left(|y_1|^2 + |y_2|^2\right)} f\left(C_2 C_1 x + C_2 S_1 y_1 + S_2 y_2\right) dy_1 dy_2 \\ &= \frac{1}{(2\pi)^d} \iint e^{-\frac{1}{2}\left(|y|^2 + |z - Dy|^2\right)} f\left(C_3 x + S_3 y\right) dy\, dz \end{aligned}$$

$$= \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} f\left(C_3 x + S_3 y\right) dy = Y_{1,3} f(x) \tag{3.241}$$

for all $f \in C_b\left(\mathbb{R}^d\right)$. If the matrices S_1 and C_2 are not invertible, then we replace the C_1 with $C_{1,\varepsilon} = e^{-\varepsilon} C_1$ and $S_{1,\varepsilon}$ satisfying $C_{1,\varepsilon} C_{1,\epsilon}^* + S_{1,\varepsilon} S_{1,\varepsilon}^* = I$, and $\lim_{\varepsilon \downarrow 0} S_{1,\varepsilon} = S_1$. We take $S_{2,\varepsilon} = e^{-\varepsilon} S_2$ instead of S_2. In addition, we choose the matrices $C_{2,\varepsilon}$, $\varepsilon > 0$, in such a way that $C_{2,\varepsilon} C_{2,\epsilon}^* + S_{2,\varepsilon} S_{2,\varepsilon}^* = I$, and $\lim_{\varepsilon \downarrow 0} C_{2,\varepsilon} = C_2$.

This completes the proof of Lemma 3.89. □

We formulate a proposition in which an Ornstein-Uhlenbeck process plays a central role. Here $\rho \mapsto \sigma(\rho)$ is a deterministic square matrix function, and $Q(\rho) = \sigma(\rho)\sigma(\rho)^*$.

3.90. PROPOSITION. *Put $X^{\tau,x}(t) = C(t,\tau)\, x + \int_\tau^t C(t,\rho)\, \sigma(\rho) dW(\rho)$. Then the process $X^{\tau,x}(t)$ is Gaussian. Its expectation is given by $\mathbb{E}\left[X^{\tau,x}(t)\right] = C(t,\tau)\, x$, and its covariance matrix has entries (s, $t \geqslant \tau$*

$$\mathbb{P}\text{-cov}\left(X_j^{\tau,x}(s), X_k^{\tau,x}(t)\right) = \left(\int_\tau^{\min(s,t)} C(s,\rho)\, Q(\rho) C(t,\rho)^*\, d\rho\right)_{j,k} \tag{3.242}$$

Let $\left\{(\Omega, \mathcal{F}, \mathbb{P}_{\tau,x}), (X(t), t \geqslant 0), \left(\mathbb{R}^d, \mathcal{B}^d\right)\right\}$ be the corresponding time-inhomogeneous Markov process. By definition, the $\mathbb{P}$-distribution of the process $t \mapsto X^{\tau,x}(t)$, $t \geqslant \tau$, is the $\mathbb{P}_{\tau,x}$-distribution of the process $t \mapsto X(t)$, $t \geqslant \tau$. Then this process is generated by the family operators $L(t)$, $t \geqslant 0$, where

$$L(t) f(x) = \frac{1}{2} \sum_{j,k=1}^d Q_{j,k}(t) D_j D_k f(x) + \langle \nabla f(x), A(t) x \rangle. \tag{3.243}$$

Here the matrix-valued function $A(t)$ is given by $A(t) = \lim_{h \downarrow 0} \dfrac{C(t+h,t) - I}{h}$. The semigroup $e^{sL(t)}$, $s \geqslant 0$, is given by

$$\begin{aligned}
& e^{sL(t)} f(x) \\
&= \mathbb{E}\left[f\left(e^{sA(t)} x + \int_0^s e^{(s-\rho)A(t)} \sigma(t) dW(\rho)\right)\right] \\
&= \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} f\left(e^{sA(t)} x + \left(\int_0^s e^{\rho A(t)} Q(t) e^{\rho A(t)^*} d\rho\right)^{1/2} y\right) dy \\
&= \int p(s,x,y;t)\, f(y) dy
\end{aligned} \tag{3.244}$$

where, with $Q_{A(t)}(s) = \displaystyle\int_0^s e^{\rho A(t)} Q(t) e^{\rho A(t)^*} d\rho$, the integral kernel $p(s,x,y;t)$ is given by

$$\begin{aligned}
& p(s,x,y;t) \\
&= \frac{1}{(2\pi)^{d/2} \left(\det Q_{A(t)}(s)\right)^{d/2}} e^{\left(-\frac{1}{2}\left\langle \left(Q_{A(t)}(s)\right)^{-1} \left(y - e^{sA(t)} x\right), y - e^{sA(t)} x\right\rangle\right)}.
\end{aligned}$$

If all eigenvalues of the matrix $A(t)$ have strictly negative real part, then the measure

$$B \mapsto \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} \mathbf{1}_B \Bigg(\Bigg(\int_0^\infty e^{\rho A(t)} Q(t) e^{\rho A(t)^*} d\rho \, y \Bigg) \, dy$$

defines an invariant measure for the semigroup $e^{sL(t)}$, $s \geqslant 0$.

A Markov process of the form $\left\{ (\Omega, \mathcal{F}, \mathbb{P}_{\tau,x}), (X(t), t \geqslant 0), (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}) \right\}$ is called a (generalized) Ornstein-Uhlenbeck process. It is time-homogeneous by putting $C(t,s) = e^{-(t-s)A}$, where A is a square $d \times d$-matrix. We will elaborate on the time-homogeneous case. In this case we write, for x, $b \in \mathbb{R}^d$,

$$S(t)f(x) := \mathbb{E}\left[f\left(e^{-tA}x + \left(I - e^{-tA}\right) b + \int_0^t e^{-(t-s)A} \sigma \, dB(s) \right) \right], \tag{3.245}$$

where $f : \mathbb{R}^d \to \mathbb{C}$ is a bounded Borel measurable function. If f belongs to $C_0\left(\mathbb{R}^d\right)$, then $S(t)f$ does so as well. For brevity we write

$$X^x(t) = e^{-tA}x + \left(I - e^{-tA}\right) b + \int_0^t e^{-(t-s)A} \sigma \, dW(s).$$

It also follows that for such functions $\lim_{t\downarrow 0} S(t)f(x) = f(x)$ for all $x \in \mathbb{R}^d$. Since we also have the semigroup property $S(t_1+t_2)f = S(t_1)S(t_2)f$ for all t_1, $t_2 \geqslant 0$, it follows that the semigroup $t \mapsto S(t)$ is in fact a Feller semigroup. Theorem 3.37 implies that there exists a time-homogeneous Markov process

$$\left\{(\Omega, \mathcal{F}, \mathbb{P}_x)_{x\in\mathbb{R}^d}, (X(t), t \geqslant 0), (\vartheta_t, t \geqslant 0), \left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)\right\}$$

such that for a bounded Borel function f we have

$$\mathbb{E}_x\left[f(X(t))\right] = \mathbb{E}\left[f(X^x(t))\right] = S(t)f(x), \ x \in \mathbb{R}^d. \tag{3.246}$$

Nest we prove the semigroup property. First we observe that, for $x \in \mathbb{R}^d$ and t_1, $t_2 \geqslant 0$,

$$X^x(t_1+t_2) = e^{-t_2A}X^x(t_1) + \left(I - e^{-t_2A}\right)b + \int_0^{t_2} e^{-(t_2-s)A}\sigma\, dW(s+t_1). \tag{3.247}$$

Let $\left(\Omega^W, \mathcal{F}^W, \mathbb{P}\right)$ be the probability space on which the process $t \mapsto W(t)$ is a Brownian motion. Let $\left(\mathcal{F}_t^W\right)_{t\geqslant 0}$ be the internal history of the Brownian motion $\{W(t) : t \geqslant 0\}$, so that $\mathcal{F}_t^W = \sigma(W(s) : s \leqslant t)$. Then by the equality in (3.247) we have

$$\begin{aligned}&\mathbb{E}\left[f(X^x(t_1+t_2)) \bigm| \mathcal{F}_{t_1}^W\right] &(3.248)\\ &= \mathbb{E}\left[f\left(e^{-t_2A}X^x(t_1) + \left(I - e^{-t_2A}\right)b + \int_0^{t_2} e^{-(t_2-s)A}\sigma\, dW(s+t_1)\right) \Bigm| \mathcal{F}_{t_1}^W\right]\end{aligned}$$

We employ the fact that the state variable $X^x(t_1)$ is $\mathcal{F}_{t_1}^W$-measurable, and that

$$\int_{t_1}^{t_1+t_2} e^{-(t_1+t_2-s)A}\sigma\, dW(s) = \int_0^{t_2} e^{-(t_2-s)A}\sigma\, d\left\{W(s+t_1) - W(t_1)\right\}$$

is $\mathbb{P}$-independent of $\mathcal{F}_{t_1}^W$ and possesses the same $\mathbb{P}$-distribution as the variable $\int_0^{t_2} e^{-(t_2-s)A}\sigma\, dW(s)$ to conclude from (3.248) the following equality:

$$\begin{aligned}&\mathbb{E}\left[f(X^x(t_1+t_2)) \bigm| \mathcal{F}_{t_1}^W\right]\\ &= \mathbb{E}\left[f\left(e^{-t_2A}z + \left(I - e^{-t_2A}\right)b + \int_0^{t_2} e^{-(t_2-s)A}\sigma\, dW(s)\right)\right]\Big|_{z=X^x(t_1)}\\ &= \mathbb{E}\left[f(X^z(t_2))\right]\Big|_{z=X^x(t_1)}. &(3.249)\end{aligned}$$

From (3.249) it follows that the process $t \mapsto X^x(t)$ is a Markov process and that, by the definition of the operators $S(t)$, $t \geqslant 0$,

$$\begin{aligned}S(t_1+t_2)f(x) &= \mathbb{E}\left[f(X^x(t_1+t_2))\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[f(X^z(t_2))\right]\Big|_{z=X^x(t_1)}\right] = \mathbb{E}\left[S(t_2)f(X^x(t_1))\right]\\ &= S(t_1)S(t_2)f(x). &(3.250)\end{aligned}$$

We calculate the differential $dX^x(t)$ and the covariation process $\left\langle X_{j_1}^x, X_{j_2}^x\right\rangle(t)$:

$$dX^x(t) = -A(X^x(t) - b)\,dt + \sigma\, dW(s), \quad \text{and} \tag{3.251}$$

$$\left\langle X_{j_1}^x, X_{j_2}^x\right\rangle(t) = \int_0^t \left(e^{-sA}\sigma\sigma^* e^{-sA^*}\right)_{j_1,j_2} ds = \operatorname{cov}\left(X_{j_1}^x(t), X_{j_2}^x(t)\right). \tag{3.252}$$

In other words the process $t \mapsto X^x(t)$ satisfies the equation

$$X^x(t) = x + \int_0^t A\left(b - X^x(s)\right) ds + \int_0^t \sigma\, dW(s). \tag{3.253}$$

Since its covariation is deterministic we have that the covariation coincides with its covariance: see (3.252). Let $f : \mathbb{R}^d \to \mathbb{C}$ be a bounded continuous function with bounded and continuous first and second order derivatives. Next we apply Itô's lemma, and employ (3.251) and (3.252) to obtain

$$\begin{aligned} f\left(X^x(t)\right) - f\left(X^x(0)\right) &= \int_0^t \nabla f\left(X^x(s)\right) \cdot \left\{-A\left(X^x(s) - b\right)\right\} ds \\ &\quad + \frac{1}{2} \sum_{j_1,j_2=1}^d \int_0^t D_{j_1} D_{j_2} f\left(X^x(s)\right) \left(e^{-sA}\sigma\sigma^* e^{-sA^*}\right)_{j_1,j_2} ds \\ &\quad + \int_0^t \nabla f\left(X^x(s)\right) \cdot \sigma\, dW(s). \end{aligned} \tag{3.254}$$

Upon taking expectations in the right-hand and left-hand sides of (3.254), using the fact that the stochastic integral in (3.254) ia a martingale, and letting $t \downarrow 0$ shows:

$$\begin{aligned} L_A f(x) &:= \lim_{t\downarrow 0} \frac{S(t)f(x) - f(x)}{t} = \lim_{t\downarrow 0} \frac{\mathbb{E}\left[f\left(X^x(t)\right) - f\left(X^x(0)\right)\right]}{t} \\ &= -\left(A(x-b)\right) \cdot \nabla f(x) + \frac{1}{2} \sum_{j_1,j_2=1}^d \left(\sigma\sigma^*\right)_{j_1,j_2} D_{j_1} D_{j_2} f(x). \end{aligned} \tag{3.255}$$

In the following proposition we collect the main properties of the time-homogeneous Ornstein-Uhlenbeck process $t \mapsto X^x(t)$. It is adapted from Proposition 3.90. In adition, $\sigma = \sigma(\rho)$ is independent of ρ.

3.91. Proposition. *Put $X^x(t) = e^{-tA}x + \left(I - e^{-tA}\right) b + \int_0^t e^{-(t-\rho)A}\sigma\, dW(\rho)$. Then the process $X^x(t)$ is Gaussian. Its expectation is given by $\mathbb{E}\left[X^x(t)\right] = e^{-tA}x + \left(I - e^{-tA}\right) b$, and its covariance matrix has entries*

$$\mathbb{P}\text{-cov}\left(X^x_{j_1}(s), X^x_{j_2}(t)\right) = \left(\int_0^{\min(s,t)} e^{-(s-\rho)A}\sigma\sigma^* e^{-(t-\rho)A^*}\, d\rho\right)_{j_1,j_2} \tag{3.256}$$

Let $\left\{\left(\Omega, \mathcal{F}, \mathbb{P}_{\tau,x}\right), \left(X(t), t \geqslant 0\right), \left(\mathbb{R}^d, \mathcal{B}^d\right)\right\}$ be the corresponding time-inhomogeneous Markov process. By definition, the $\mathbb{P}$-distribution of the process $t \mapsto X^x(t)$, $t \geqslant \tau$, is the $\mathbb{P}_x$-distribution of the process $t \mapsto X(t)$, $t \geqslant 0$. Then this process is generated by the operator L_A, $t \geqslant 0$, where

$$L_A f(x) = \frac{1}{2} \sum_{j_1,j_2=1}^d \left(\sigma\sigma^*\right)_{j_1,j_2} D_{j_1} D_{j_2} f(x) - \left\langle \nabla f(x), A(x-b)\right\rangle. \tag{3.257}$$

The semigroup e^{sL_A}, $s \geqslant 0$, is given by

$$\begin{aligned} &e^{sL_A} f(x) \\ &= \mathbb{E}\left[f\left(e^{-sA}(x-b) + b + \int_0^s e^{-(s-\rho)A}\sigma\, dW(\rho)\right)\right] \end{aligned}$$

$$= \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} f\left(e^{-sA}(x-b) + b + \left(\int_0^s e^{-\rho A}\sigma\sigma^* e^{-\rho A^*}\,d\rho\right)^{1/2} y\right) dy$$
$$= \int p_A(s,x,y)\, f(y)\, dy \tag{3.258}$$

where, with $Q_A(s) = \displaystyle\int_0^s e^{-\rho A}\sigma\sigma^* e^{-\rho A^*}\,d\rho$*, the integral kernel* $p_A(s,x,y)$ *is given by*

$$p_A(s,x,y) = \frac{1}{(2\pi)^{d/2}\left(\det Q_A(s)\right)^{d/2}} e^{\left(-\frac{1}{2}\left\langle (Q_A(s))^{-1}\left(y-e^{-sA}(x-b)-b\right), y-e^{-sA}(x-b)-b\right\rangle\right)}.$$

If all eigenvalues of the matrix A have strictly positive real part, then the measure

$$B \mapsto \frac{1}{(2\pi)^{d/2}} \int e^{-\frac{1}{2}|y|^2} \mathbf{1}_B\left(b + \int_0^\infty e^{-\rho A}\sigma\sigma^* e^{-\rho A^*} y\, d\rho\right) dy \tag{3.259}$$

defines an invariant measure for the semigroup e^{sL_A}*,* $s \geqslant 0$*.*

Proof. The results in Proposition 3.91 follow more or less directly from those in Proposition 3.90. The result in (3.259) follows by letting $s \to \infty$ in the second equality of (3.258) or in the definition of the probability density $p_A(s,x,y)$. □

For more information about invariant, or stationary, measures see, *e.g.*, [**146**] (Chapter 10) and the references therein like Meyn and Tweedie [**97**].

In order to apply our results on the Ornstein-Uhlenbeck process to bond pricing and determining interest rates in financial mathematics the identities and results in the following proposition are very useful. It will be applied in the context of the Vasicek model.

3.92. Proposition. *Let the notation and hypotheses be as in Proposition 3.91. Put*

$$A(t,T) = \int_0^{T-t} e^{-\rho A}\, d\rho = \int_t^T e^{-\rho s}\, ds = A^{-1}\left(I - e^{-(T-t)A}\right), \quad 0 \leqslant t \leqslant T,$$

where the last equality is only valid if A is invertible. Let y be a vector in $\mathbb{R}^d$. The following assertions hold true.

(1) *The following identity is true for* $0 \leqslant t < T$*:*

$$\int_t^T X^x(s)\, ds = A(t,T)\left(X^x(t) - b\right) + (T-t)b + \int_t^T A(\rho,T)\,\sigma\, dW(\rho). \tag{3.260}$$

(2) *The random vector* $\int_t^T X^x(s)\, ds$ *is Gaussian (or, what is the same, multivariate normally distributed) with conditional expectation given by*

$$\mathbb{E}\left[\int_t^T X^x(s)\, ds \,\middle|\, \mathcal{F}_t\right] = \mathbb{E}\left[\int_t^T X^x(s)\, ds \,\middle|\, X^x(t)\right] = A(t,T)\left(X^x(t)-b\right) + (T-t)b, \tag{3.261}$$

and covariance matrix given by $(1 \leqslant j_1,\, j_2 \leqslant d)$

$$cov\left(\int_t^T X_{j_1}^x(s)\,ds \mid X^x(t), \int_t^T X_{j_2}^x(s)\,ds \mid X^x(t)\right)$$
$$= \left(\int_t^T A(\rho,T)\,\sigma\sigma^* A(\rho,T)^*\,d\rho\right)_{j_1,j_2}. \tag{3.262}$$

(3) *The random variable* $\left\langle y, \int_t^T X^x(s)\,ds\right\rangle$ *is normally distributed with conditional expectation given by*

$$\mathbb{E}\left[\left\langle y, \int_t^T X^x(s)\,ds\right\rangle \mid \mathcal{F}_t\right] = \mathbb{E}\left[\left\langle y, \int_t^T X^x(s)\,ds\right\rangle \mid X^x(t)\right]$$
$$= \langle y, A(t,T)\,(X^x(t)-b)\rangle + (T-t)\,\langle y, b\rangle\,, \tag{3.263}$$

and variance given by

$$var\left(\left\langle y, \int_t^T X^x(s)\,ds\right\rangle \mid X^x(t)\right) = \int_t^T \left|\sigma^* A(\rho,T)^*\,y\right|^2\,d\rho. \tag{3.264}$$

(4) *The conditional expectation of* $\exp\left(-\left\langle y, \int_t^T X^x(s)\,ds\right\rangle\right)$ *given* $\mathcal{F}_t$ *is log-normal, and*

$$\mathbb{E}\left[\exp\left(-\left\langle y, \int_t^T X^x(s)\,ds\right\rangle\right) \mid \mathcal{F}_t\right] = \mathbb{E}\left[\exp\left(-\left\langle y, \int_t^T X^x(s)\,ds\right\rangle\right) \mid X^x(t)\right]$$
$$= \exp\left(-\langle y, A(t,T)\,(X^x(t)-b)\rangle - (T-t)\,\langle y,b\rangle + \frac{1}{2}\int_t^T \left|\sigma^* A(\rho,T)^*\,y\right|^2\,d\rho\right). \tag{3.265}$$

Proof. (1) From (3.247) we see, for $s \geqslant t$,

$$X^x(s) = e^{-(s-t)A}\left(X^x(t) - b\right) + b + \int_t^s e^{-(s-\rho)A}\sigma\, dW(\rho). \tag{3.266}$$

Then we integrate the expressions in (3.266) against s for $t \leqslant s \leqslant T$, and we interchange the integrals with respect to ds and $dW(\rho)$ to obtain the equality in (3.260). This proves assertion (1).

(2) Although the process $t \mapsto \int_t^T A(\rho, T)\, dW(\rho)$ is not a martingale, it has enough properties of a martingale that its expectation is 0, and that its quadratic covariation matrix is given by the expression in (3.262). The reason for all this relies on the equality:

$$\int_t^T A(\rho, A)\, dW(\rho) = \int_0^T A(\rho, T)\, dW(\rho) - \int_0^t A(\rho, T)\, dW(\rho), \tag{3.267}$$

combined with the fact that the process $t \mapsto \int_0^t A(\rho, T)\, dW(\rho)$ is a martingale. So we can apply the Itô isometry and its consequences to complete the proof of assertion (2). An alternative way of understanding this reads as follows. Processes of the form $s \mapsto X^x(s)$, $s \leqslant 0$, and $t \mapsto \int_t^T X^x(s)\, ds$, $0 \leqslant t \leqslant T$, consist of Gaussian vectors with known means and variances. For $s \geqslant t$ we use the representation in (3.266) for $X^x(s)$, and for $\int_t^T X^x(s)\, ds$ we employ (3.260).

(3) The proof of this assertion follows the same line as the proof of the assertion in (2).

(4) If the stochastic variable Z is normally distributed with expectation μ and variance $v^2 = \mathbb{E}\left[(Z-\mu)^2\right]$, then $\mathbb{E}\left[e^Z\right] = e^{\mu + \frac{1}{2}v^2}$. This result is applied to the variable $Z = -\left\langle y, \int_t^T X^x(s)\, ds\right\rangle$ to obtain the equality in (3.265).

This completes the proof of Proposition 3.92. $\square$

3.93. Lemma. *Let the notation and hypotheses be as in the proposition 3.91 and 3.92. Suppose that the matrix A is invertible. The following equality holds for $0 \leqslant t \leqslant T$:*

$$\begin{aligned}
&\int_t^T A(\rho, T)\,\sigma\sigma^* A(\rho, T)^*\, d\rho \\
&= (T-t)A^{-1}\sigma\sigma^*\left(A^*\right)^{-1} - A(t,T)\,A^{-1}\sigma\sigma^*\left(A^*\right)^{-1} - A^{-1}\sigma\sigma^*\left(A^*\right)^{-1} A(t,T)^* \\
&\quad + \int_0^{T-t} e^{-\rho A} A^{-1}\sigma\sigma^*\left(A^*\right)^{-1} e^{-\rho A^*}\, d\rho.
\end{aligned} \tag{3.268}$$

If the invertible matrix A is such that $A\sigma\sigma^* = \sigma\sigma^* A^*$, then the following equality is valid for $0 \leqslant t \leqslant T$:

$$\begin{aligned}
&\int_t^T A(\rho, T)\,\sigma\sigma^* A(\rho, T)^*\, d\rho \\
&= (T-t)A^{-1}\sigma\sigma^*\left(A^*\right)^{-1} - A(t,T)\,A^{-1}\sigma\sigma^*\left(A^*\right)^{-1} - \frac{1}{2}\left(A(t,T)\right)^2\sigma\sigma^*\left(A^*\right)^{-1}
\end{aligned}$$

$$= \left(T - t - A(t,T) - \frac{1}{2} A \left(A(t,T) \right)^2 \right) A^{-1} \sigma\sigma^* \left(A^* \right)^{-1}. \tag{3.269}$$

Observe that an equality of the form $A\sigma\sigma^* = \sigma\sigma^* A^*$ holds whenever $A = A^*$ and the matrix $\sigma\sigma^*$ is a "function" of A. In particular this is true when $d = 1$ and $A = a$ is a real number.

Proof. Since A is invertible we have $A(\rho, T) = \left(I - e^{-(T-\rho)A} \right) A^{-1}$, and so

$$\begin{aligned}
&\int_t^T A(\rho,T)\, \sigma\sigma^* A(\rho,T)^*\, d\rho \\
&= \int_t^T \left(I - e^{-(T-\rho)A} \right) A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} \left(I - e^{-(T-\rho)A^*} \right) d\rho \\
&= \int_0^{T-t} \left(I - e^{-\rho A} \right) A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} \left(I - e^{-\rho A^*} \right) d\rho \\
&= (T-t) A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} - A(t,T)\, A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} - A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} A(t,T)^* \\
&\quad + \int_0^{T-t} e^{-\rho A} A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} e^{-\rho A^*}\, d\rho.
\end{aligned} \tag{3.270}$$

The final equality in (3.270) proves (3.268). Next we also assume that $A\sigma\sigma^* = \sigma\sigma^* A^*$. Then $\sigma\sigma^* e^{-\rho A^*} = e^{-\rho A} \sigma\sigma^*$, and hence

$$\begin{aligned}
&\int_0^{T-t} e^{-\rho A} A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} e^{-\rho A^*}\, d\rho \\
&= \int_0^{T-t} e^{-2\rho A}\, d\rho\, A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} e^{-\rho A^*} \\
&= \frac{1}{2} \left(I - e^{2(T-t)A} \right) A^{-2} \sigma\sigma^* \left(A^* \right)^{-1} e^{-\rho A^*}.
\end{aligned} \tag{3.271}$$

A simple calculation shows

$$\frac{1}{2} \left(I - e^{-2(T-t)A} \right) = A(t,T) A - \frac{1}{2} \left(A(t,T) \right)^2 A^2, \tag{3.272}$$

and so the equalities in (3.271) show

$$\begin{aligned}
&\int_0^{T-t} e^{-\rho A} A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} e^{-\rho A^*}\, d\rho = \int_0^{T-t} e^{-2\rho A}\, d\rho\, A^{-1} \sigma\sigma^* \left(A^* \right)^{-1} \\
&= \left(A(t,T) - \frac{1}{2} \left(A(t,T) \right)^2 A \right) A^{-1} \sigma\sigma^* \left(A^* \right)^{-1}.
\end{aligned} \tag{3.273}$$

A combination of (3.270) and (3.273) together with the equality $\sigma\sigma^* A(t,T)^* = A(t,T)\, \sigma\sigma^*$ then yields the equality in (3.269), completing the proof of Lemma 3.93. □

Before we discuss the Vasicek model we insert Girsanov's theorem formulated in a way as we will use it in Theorem 3.101. In fact we will formulate it in a multivariate context.

3.94. THEOREM. *Let $\{X(t) : 0 \leqslant t \leqslant t\}$ be an Itô process satisfying*

$$dX(t) = v(t)\,dt + u(t)\,dW(t). \quad 0 \leqslant t \leqslant T.$$

Suppose there exists a process $\{\theta(t) : 0 \leqslant t \leqslant T\}$, with the property that

$$\mathbb{P}\left[\int_0^T |\vartheta(t)|^2 \, dt < \infty\right] = 1,$$

such that the process $v(t) - u(t)\theta(t)$ has this property as well. Assume furthermore that the process $t \mapsto \mathcal{E}(t)$, $0 \leqslant t \leqslant T$, defined by

$$\mathcal{E}(t) = \exp\left(-\int_0^t \theta(s)\,dW(s) - \frac{1}{2}\int_0^t |\theta(s)|^2 \, ds\right) \tag{3.274}$$

is a $\mathbb{P}$-martingale, which is guaranteed provided $\mathbb{E}[\mathcal{E}(t)] = 1$ for $0 \leqslant t \leqslant T$. Define the measure $\mathbb{P}^*$ such that $\dfrac{d\mathbb{P}^*}{d\mathbb{P}} = \mathcal{E}(T)$. Then

$$t \mapsto W^*(t) := W(t) + \int_0^t \theta(s)\,ds, \quad t \in [0, T],$$

is a Brownian motion w.r.t. $\mathbb{P}^*$ and the process $\{X(t) : 0 \leqslant t \leqslant T\}$ has a representation w.r.t. $W^*(t)$ given by

$$dX(t) = (v(t) - u(t)\theta(t))\,dt + u(t)\,dW^*(t).$$

We shortly show that $\{\mathcal{E}(t) : 0 \leqslant t \leqslant T\}$ is a diffusion process. Set $Y(t) = \int_0^t \theta(s)\,dW(s)$, $0 \leqslant t \leqslant T$, and consider the function $f(t,x) \in C^2([0,T],\mathbb{R})$ defined by

$$f(t,x) = \exp\left(-x - \frac{1}{2}\int_0^t |\theta(s)|^2\,ds\right).$$

Then we clearly have that $\mathcal{E}(t) = f(t, Y(t))$. By Itô's formula we have

$$\begin{aligned} d\mathcal{E}(t) &= -\frac{1}{2}|\theta(t)|^2\,\mathcal{E}(t)\,dt - \mathcal{E}(t)\theta(t)\,dW(t) + \frac{1}{2}\mathcal{E}(t)\,d\langle Y,Y\rangle(t) \\ &= -\frac{1}{2}|\theta(t)|^2\,\mathcal{E}(t)\,dt - \mathcal{E}(t)\theta(t)\,dW(t) + \frac{1}{2}\mathcal{E}(t)\,|\theta(t)|^2\,dt \\ &= -\theta(t)\mathcal{E}(t)\,dW(t). \end{aligned} \tag{3.275}$$

Hence, it follows that

$$\mathcal{E}(t) = \mathcal{E}(0) - \int_0^t \theta(s)\mathcal{E}(s)\,dW(s),$$

which in general is a local martingale for which $\mathbb{E}[\mathcal{E}(t)] \leqslant 1$. It is a sub-martingale, but not necessarily a martingale. If, for $0 \leqslant t \leqslant T$, the expectation $\mathbb{E}[\mathcal{E}(t)] = 1$, then $t \mapsto \mathcal{E}(t)$, $0 \leqslant t \leqslant T$, is a martingale. If Novikov's condition, *i.e.*, if $\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T |\theta(t)|^2\,dt\right)\right] < \infty$ is satisfied, then the process $\{\mathcal{E}(t) : 0 \leqslant t \leqslant T\}$ is a martingale. For details on this condition, see Corollary 4.27 in Chapter 4. For more results on (local) exponential martingales see subsection 1.3 of Chapter 4 as well. In section 3 of the same chapter the reader may find some more information on Girsanov's theorem. In particular, see assertion (4) of Proposition 4.24 and Theorem 4.25.

8.1. The Vasicek model. In this subsection we want to employ the results in Proposition 3.92 with $d = 1$ to find the bond prices in the Vasicek model. Until now we were always working in the physical probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In order to calculate the fair price of a financial instrument one often uses the method of risk-neutral pricing. Through this technique the price of a financial asset is the expectation of its discounted pay-off at the so-called risk-neutral measure Q. The risk-neutral measure is equivalent to the physical measure $\mathbb{P}$. Suppose for example that $\{S(t)\}_{s\geqslant 0}$ is the price of a certain asset at time t. The price of our asset at time t discounted to time 0 is then given by $\widetilde{S}(t) := e^{-\int_0^t r(u)\,du} S(t)$. As a main property of the risk-neutral measure, the family of discounted prices $\left\{\widetilde{S}(t)\right\}_{t\geqslant 0}$ is a Q-martingale. This means that for every s, $0 \leqslant s \leqslant t$, we have

$$\mathbb{E}\left[\widetilde{S}(t) \mid \mathcal{F}_s\right] = \mathbb{E}\left[e^{-\int_0^t r(u)\,du} S(t) \mid \mathcal{F}_s\right] = e^{-\int_0^s r(u)\,du} S(s) = \widetilde{S}(s), \tag{3.276}$$

where expectations $\mathbb{E}$ are with respect to Q. Because of this property, a risk-neutral measure is also called an equivalent martingale measure. Roughly speaking, the existence of such a measure is equivalent with the no-arbitrage assumption. We will use this martingale property to price a zero-coupon bond. That is a financial debt instrument that pays the holder a fixed amount named the *face value at maturity* T. For simplicity we take 1 as face value. The price of a zero-coupon bond is then given by the following theorem.

3.95. Theorem. *Consider a zero-coupon bond which pays an amount of 1 at maturity* T*. The price at time* $t \leqslant T$ *is then*

$$P(t,T) = \mathbb{E}\left[e^{-\int_t^T r(s)\,ds} \mid \mathcal{F}_t\right]. \tag{3.277}$$

Proof. We use the above explained property that the discounted price $e^{-\int_0^t r(s)\,ds}P(t,T)$ is a martingale, and the trivial fact that $P(T,T) = 1$,

$$\begin{aligned} e^{-\int_0^t r(s)\,ds}P(t,T) &= \mathbb{E}\left[e^{-\int_0^T r(s)\,ds}P(T,T) \mid \mathcal{F}_t\right] = \mathbb{E}\left[e^{-\int_0^T r(s)\,ds} \mid \mathcal{F}_t\right] \\ &= e^{-\int_0^t r(s)\,ds}\mathbb{E}\left[e^{-\int_t^T r(s)\,ds} \mid \mathcal{F}_t\right]. \end{aligned} \tag{3.278}$$

The bond's price can thus be written as $P(t,T) = \mathbb{E}\left[e^{-\int_t^T r(s)\,ds} \mid \mathcal{F}_t\right]$. This completes the proof of Theorem 3.95. □

Formula (3.277) is an expression of the bond's price for an arbitrary chosen interest rate process. We will now apply this to our Vasicek model $\{r(t)\}_{t\geqslant 0}$. We will investigate three methods that all lead to the same result stated in the following theorem. We follow the approach of Mamon in [**94**]. For an alternative approach see [**114**] as well.

3.96. Theorem. *Consider a zero-coupon bond which pays an amount of 1 at maturity* T*. Suppose that under the risk-neutral measure the short rate follows an Ornstein-Uhlenbeck process:* $dr(t) = a\,(b - r(t))\,dt + \sigma\,dW(t)$*. The fair price of the bond at time* $t \leqslant T$ *is then given by*

$$P(t,T) = e^{-A(t,T)r(t)+D(t,T)}, \tag{3.279}$$

where

$$\begin{aligned} A(t,T) &= \frac{1-e^{-a(T-t)}}{a}, \quad \textit{and} \\ D(t,T) &= \frac{(A(t,T)-T+t)\,(a^2b-\sigma^2/2)}{a^2} - \frac{\sigma^2 A(t,T)^2}{4a}. \end{aligned} \tag{3.280}$$

Equation (3.279) is an affine term structure model. In fact, the bond yield $y_t(T)$ is defined as the constant interest rate at which the price of the bond grows to it's face value, *i.e.*, $P(t,T)e^{y_t(T)(T-t)} = 1$. We thus find that

$$y_t(T) = \frac{-\log P(t,T)}{T-t} = \frac{A(t,T)r(t) - D(t,T)}{T-t},$$

which is indeed affine in $r(t)$. The yield curve or term structure at time t is the graph $(T, y_t(T))$.

8.1.1. *Bond price implied by the distribution of the short rate.* The first method to calculate the bond price is quite straightforward. It calculates the conditional expectation in formula (3.277) by determining the distribution of $\mathbb{E}\left[\int_t^T r(s)\,ds \mid \mathcal{F}_t\right]$.

First proof of Theorem 3.96. Because formula (3.277) shows that the bond's price at time t is conditional on $\mathcal{F}_t$, we may assume that $r(t)$ is a parameter. Using formula (3.266) (for $d = 1$, and $A = 1$) with starting time t we find, for $s > t$,

$$r(s) = r(t)e^{-a(s-t)} + b\left(1 - e^{-a(s-t)}\right) + \int_t^s e^{-a(s-\rho)}\sigma\,dW(\rho).$$

We want to determine the distribution of $e^{-\int_t^T r(s)\,ds}$ conditioned by $\mathcal{F}_t$. Note that because of the Markov property of the Otnstein-Uhlenbeck process (or more generally for diffusion processes: see the equality in (3.249)), this distribution will only depend on $r(t)$. Let's start by determining the distribution of $\int_t^T r(s)\,ds$ given $\mathcal{F}_t$. This distribution is normal, and essentially speaking this it follows from Proposition 3.92 and Lemma 3.93. First of all from assertion (3) in Proposition 3.92 we get by (3.263)

$$\mathbb{E}\left[\int_t^T r(s)\,ds \mid \mathcal{F}_t\right] = \mathbb{E}\left[\int_t^T r(s)\,ds \mid r(t)\right] = A(t,T)\left(r(t) - b\right) + (T-t)b. \tag{3.281}$$

Secondly from (3.264) and (3.269) in Lemma 3.93 we get

$$\begin{aligned}\operatorname{var}\left(\int_t^T X^x(s)\,ds \mid X^x(t)\right) &= \int_t^T \sigma^2\left(A(\rho,T)\right)^2\,d\rho \\ &= \frac{\sigma^2}{a^2}\left(T - t - A(t,T) - \frac{a}{2}\left(A(t,T)\right)^2\right).\end{aligned} \tag{3.282}$$

The equality in (3.279) of Theorem 3.96 then follows from (3.282) and (3.265) in (4) of Proposition 3.92. $\square$

8.1.2. *Bond price by solving the PDE.* A second method that is proposed to calculate the bond's price in the Vasicek model, is by solving partial differential equations. More precisely, we will derive a PDE for the bond's price by using martingales.

Taking into account the Markov property of the process $\{r(t)\}_{t\geqslant 0}$ (see equality in (3.249)) one can introduce the following variable:

$$\begin{aligned}P(t,T) &= \mathbb{E}\left[e^{-\int_t^T r(s)\,ds} \mid \mathcal{F}_t\right] = \mathbb{E}\left[e^{-\int_t^T r(s)\,ds} \mid r(t)\right] \\ &= \mathbb{E}\left[e^{-\int_t^T r(s)(z)\,ds}\right]\Big|_{z=r(t)} =: P\left(t,T,r(t)\right).\end{aligned} \tag{3.283}$$

Here $r(s)$, $s > t$, is the function of $r(t)$ given by

$$r(s) = r(t)e^{-a(s-t)} + b\left(1 - e^{a(s-t)}\right) + \int_t^s e^{-a(s-\rho)}\,dW(\rho). \tag{3.284}$$

We now provide a second proof of Theorem 3.96.

Second proof of Theorem 3.96. We will apply Itô's formula to the function $f(t,x) = e^{-\int_0^t r(s)\,ds} P(t,T,x)$. Then we obtain

$$
\begin{aligned}
&\mathbb{E}\left[e^{-\int_0^T r(s)\,ds} - P\left(0,T,r(0)\right) \,\middle|\, \mathcal{F}_t\right] = e^{-\int_0^t r(s)\,ds} P\left(t,T,r(t)\right) - P\left(0,T,r(0)\right) \\
&= \int_0^t \left[-r(u) e^{-\int_0^u r(s)\,ds} P\left(u,T,r(u)\right) + e^{-\int_0^u r(s)\,ds} \frac{\partial P\left(u,T,r(u)\right)}{\partial u}\right] du \\
&\quad + \int_0^t \left[e^{-\int_0^u r(s)\,ds} \frac{\partial P\left(u,T,r(u)\right)}{\partial r(u)}\right] \left(a\left(b - r(u)\right) du + \sigma dW(u)\right) \\
&\quad + \frac{\sigma^2}{2} \int_0^t \left[e^{-\int_0^u r(s)\,ds} \frac{\partial^2 P\left(u,T,r(u)\right)}{\partial r(u)^2}\right] du. \qquad (3.285)
\end{aligned}
$$

Put

$$
\begin{aligned}
f(t) &= -r(t) e^{-\int_0^t r(s)\,ds} P\left(t,T,r(t)\right) + e^{-\int_0^t r(s)\,ds} \frac{\partial P\left(t,T,r(t)\right)}{\partial t} \\
&\quad + \left[e^{-\int_0^t r(s)\,ds} \frac{\partial P\left(t,T,r(t)\right)}{\partial r(u)}\right] \left(a\left(b - r(t)\right)\right) \\
&\quad + \frac{\sigma^2}{2} e^{-\int_0^t r(s)\,ds} \frac{\partial^2 P\left(t,T,r(t)\right)}{\partial r(t)^2}. \qquad (3.286)
\end{aligned}
$$

From the equality in (3.285) it follows that the process $t \mapsto \int_0^t f(u)\,du$ is a martingale. By Lemma 3.97 below it follows that $f(t) = 0$ $\mathbb{P}$-almost surely. From (3.286) it then follows that the function $P(t,T,x)$ satisfies the following differential equation:

$$-xP\left(t,T,x\right) + \frac{\partial P\left(t,T,x\right)}{\partial t} + \frac{\partial P\left(t,T,x\right)}{\partial x}\left(a\left(b-x\right)\right) + \frac{\sigma^2}{2}\frac{\partial^2 P\left(t,T,x\right)}{\partial x^2} = 0. \tag{3.287}$$

From (3.283) and (3.284) it follows that

$$\frac{\partial P(t,T,x)}{\partial x} = \frac{-1}{a}\left(1-e^{-a(T-t)}\right)P(t,T,x) = -A(t,T)P(t,T,x). \tag{3.288}$$

From (3.288) we easily infer that

$$P\left(t,T,x\right) = C\left(t,T\right)e^{-A(t,T)x}. \tag{3.289}$$

Inserting this expression for $P\left(t,T,x\right)$ into (3.287) yields the first order equation

$$-x + \frac{1}{C(t,T)}\frac{\partial C(t,T)}{\partial t} - \frac{\partial A(t,T)}{\partial t}x - A(t,T)\left\{a(b-x)\right\} + \frac{\sigma^2}{2}A(t,T)^2 = 0. \tag{3.290}$$

Because $-1 - \dfrac{\partial A(t,T)}{\partial t} + aA(t,T) = 0$, the equality in (3.290) implies:

$$\frac{1}{C(t,T)}\frac{\partial C(t,T)}{\partial t} - abA(t,T) + \frac{\sigma^2}{2}A(t,T)^2 = 0. \tag{3.291}$$

Since $C(T,T) = P\left(T,T,0\right) = 1$ from (3.291) we infer $C(t,T) = e^{D(t,T)}$ and hence

$$P(t,T) = P\left(t,T,r(t)\right) = e^{-A(t,T)r(t)+D(t,T)},$$

which completes the proof of Theorem 3.96 by employing the PDE as formulated in (3.287). □

The equation in (3.287) is called the *PDE for the bond price in the Vasicek model.*

3.97. Lemma. *Let $\left(\Omega, \left(\mathcal{F}_t\right)_{t\geqslant 0}, \mathbb{P}\right)$ be filtered probability space, and let the right-continuous adapted process $\left\{f(t)\right\}_{t\geqslant 0}$ be such that for some sequence of stopping times $\left(\tau_n\right)_{n\in\mathbb{N}}$, which increases to ∞, the integrals $\int_0^t |f(s)|\,\mathbf{1}_{[1,\tau_n]}\,ds$ are finite $\mathbb{P}$-almost surely. If the process $t \mapsto \int_0^t f(s)\,ds$ is a local martingale, then $f(t) = 0$ $\mathbb{P}$-almost surely for almost all t.*

Proof. Fix $0 < T < \infty$. By localizing at stopping times $\left(\tau_n'\right)_{n\in\mathbb{N}}$, $\tau_n' \leqslant \tau_n$, $n \in \mathbb{N}$, $\tau_n \uparrow \infty$ $(n \to \infty)$ we may assume that

$$\mathbb{E}\left[\int_0^T |f(s)|\,ds\right] < \infty. \tag{3.292}$$

Otherwise we replace $f(t)$ with $f\left(t\right)\mathbf{1}_{[0,\tau_n']}(t)$, and prove that $f\left(t\right)\mathbf{1}_{[0,\tau_n']}(t) = 0$ for all $n \in \mathbb{N}$. But then $f(t) = 0$, by letting $n \to \infty$. So we assume that (3.292) is satisfied. Then for $0 \leqslant s < t \leqslant T$ we have

$$\int_0^s f(\rho)\,d\rho + \mathbb{E}\left[\int_s^t f(\rho)\,d\rho \;\middle|\; \mathcal{F}_s\right] = \mathbb{E}\left[\int_0^t f(\rho)\,d\rho \;\middle|\; \mathcal{F}_s\right] = \int_0^s f(\rho)\,d\rho. \quad (3.293)$$

From (3.293) we infer that $\mathbb{E}\left[\int_s^t f(\rho)\,d\rho \;\middle|\; \mathcal{F}_s\right] = 0$, $\mathbb{P}$-almost surely, for all $0 \leqslant s < t < T$. differentiating with respect to t then results in $\mathbb{E}\left[f(t) \;\middle|\; \mathcal{F}_s\right] = 0$ $\mathbb{P}$-almost surely for all $0 \leqslant s < t < T$. But then, by the right-continuity of the process $\{f(t)\}_{t\geqslant 0}$ it follows that

$$f(s) = \lim_{t\downarrow s} \mathbb{E}\left[f(t) \;\middle|\; \mathcal{F}_s\right] = 0, \quad \mathbb{P}\text{-almost surely.}$$

This completes the proof of Lemma 3.97. □

8.1.3. *Bond prices using forward rates.* The third and last method to calculate a bond's price in the Vasicek model, is based upon the concept of forward rates. Indeed, in the Heath-Jarrow-Morton pricing paradigm the closed-form of the bond's price follows directly from the short rate dynamics under the so-called forward measure. Suppose we are at time t. We want to know the rate of interest in the period of time between T_1 en T_2 with $t < T_1 < T_2$. This is called the forward rate for the period between T_1 and T_2 and we denote it by $f(t, T_1, T_2)$. When the rates between time t and T_1 and between time t and T_2 are known - write R_1 and R_2 - we must have:

$$e^{R_1(T_1-t)} e^{f(t,T_1,T_2)(T_2-T_1)} = e^{R_2(T_2-t)}.$$

Hence, we find for the forward rate

$$f(t, T_1, T_2) = \frac{R_2(T_2-t) - R_1(T_1-t)}{T_2 - T_1}.$$

Applying this in our framework of bond prices, R_1 and R_2 equal the bond yields:

$$R_1 = \frac{-\log P(t, T_1)}{T_1 - t}, \quad R_2 = \frac{-\log P(t, T_2)}{T_2 - t},$$

such that the forward rate is given by

$$f(t, T_1, T_2) = \frac{-\log P(t, T_2) - \log P(t, T_1)}{T_2 - T_1}.$$

When T_1 and T_2 come infinitesimally close to each other, we obtain a so-called instantaneous forward rate. The instantaneous forward rate at time $T > t$ is

$$f(t, T) = -\lim_{t'\to T} \frac{\log P(t, T) - \log P(t, t')}{T - t'} = -\frac{\partial \log P(t, t')}{\partial t'}\Big|_{t'=T}.$$

Solving this partial differential equation for $P(t, T)$ on $[t, T]$ we find immediately that

$$P(t, T) = e^{-\int_t^T f(t,s)\,ds}. \quad (3.294)$$

Later on we will see that the link between the instantaneous forward rate and the short rate is the so-called forward measure. In the sequel, we will need two properties of conditional expectations under change of measure. These results can be found in [**32**]. In the following theorems $\mathbb{P}$ is a probability measure on a

σ-algebra $\mathcal{F}$, the probability measure $Q \ll \mathbb{P}$ is such that $\frac{dQ}{d\mathbb{P}} = Z$. Furthermore $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$. The symbol $\mathbb{E}$ denotes expectation w.r.t. $\mathbb{P}$, while $\mathbb{E}^Q$ stands for expectation w.r.t. Q.

3.98. Theorem. *In the notation of above, it holds that*

$$\frac{dQ\mid_{\mathcal{G}}}{d\mathbb{P}\mid_{\mathcal{G}}} = \mathbb{E}\left[Z \mid \mathcal{G}\right].$$

Proof. Take an arbitrary $B \in \mathcal{G}$. We need to show that

$$Q(B) = \mathbb{E}\left[\mathbb{E}\left[Z \mid \mathcal{G}\right] \mathbf{1}_B\right].$$

Indeed:

$$\mathbb{E}\left[\mathbb{E}\left[Z \mid \mathcal{G}\right] \mathbf{1}_B\right] = \mathbb{E}\left[\mathbb{E}\left[Z\mathbf{1}_B \mid \mathcal{G}\right]\right] = \mathbb{E}\left[Z\mathbf{1}_B\right] = Q(B).$$

This completes the proof of Theorem 3.98. □

3.99. Theorem. *For any $\mathcal{F}$-measurable random variable X :*

$$\mathbb{E}\left[Z \mid \mathcal{G}\right]\mathbb{E}^Q\left[X \mid \mathcal{G}\right] = \mathbb{E}\left[ZX \mid \mathcal{G}\right].$$

Proof. Let $Y = \mathbb{E}\left[Z \mid \mathcal{G}\right]$. Take $B \in \mathcal{G}$ arbitrary, then:

$$\begin{aligned}\mathbb{E}^Q\left[\mathbf{1}_B\mathbb{E}\left[ZX \mid \mathcal{G}\right]\right] &= \mathbb{E}\left[Y\mathbf{1}_B\mathbb{E}\left[ZX \mid \mathcal{G}\right]\right] = \mathbb{E}\left[\mathbb{E}\left[Y\mathbf{1}_B ZX \mid \mathcal{G}\right]\right] \\ &= \mathbb{E}\left[Y\mathbf{1}_B ZX\right] = \mathbb{E}^Q\left[Y\mathbf{1}_B X\right] = \mathbb{E}^Q\left[\mathbb{E}^Q\left[\mathbf{1}_B YX \mid \mathcal{G}\right]\right] \\ &= \mathbb{E}^Q\left[\mathbf{1}_B\mathbb{E}^Q\left[YX \mid \mathcal{G}\right]\right].\end{aligned}$$

In the first step we used that $\mathbf{1}_B\mathbb{E}\left[ZX \mid \mathcal{G}\right]$ is $\mathcal{G}$-mesurable. Hence we could apply Theorem 3.98 which tells us that $dQ\mid_{\mathcal{G}} = Y d\mathbb{P}\mid_{\mathcal{G}}$. Because the previous reasoning holds for all $B \in \mathcal{G}$ we must have:

$$\mathbb{E}\left[ZX \mid \mathcal{G}\right] = \mathbb{E}^Q\left[YX \mid \mathcal{G}\right] = Y\mathbb{E}^Q\left[X \mid \mathcal{G}\right]$$

what proves the claim in Theorem 3.99. □

As well as the economic term forward rates, we introduce the concept of a numéraire. A numéraire is a tradeable economic security in terms of which the relative prices of other assets can be expressed. This allows us not only to compare different financial instruments at a certain moment, it makes it also possible to compare the prices of assets at different times. A typical example of a numéraire is money. The random variable $M(t) = e^{\int_0^t r(s)\,ds}$ represents the value at time t of an asset which was invested in the money market at time 0 with value 1. Recall that in accordance with the definition of a risk-neutral measure Q, the price of an asset relative to the money market is a martingale. In our new notation the expressions in (3.276) become:

$$\mathbb{E}\left[\frac{S(t)}{M(t)} \mid \mathcal{F}_s\right] = \mathbb{E}\left[e^{-\int_0^t r(u)\,du} S(t) \mid \mathcal{F}_s\right] = e^{-\int_0^s r(u)\,du} S(s) = \frac{S(s)}{M(s)},$$

with $0 \leqslant s \leqslant t$. We say that Q is an equivalent martingale measure for the numéraire $\{M(t)\}_{t \geqslant 0}$. Let $N(t)$ be the price at time t of another traded asset.

Suppose that Q^* is an equivalent martingale measure for $\{N(t)\}_{t\geqslant 0}$, i.e. for all $0 \leqslant s \leqslant t$:

$$\mathbb{E}^*\left[\frac{S(t)}{N(t)} \mid \mathcal{F}_s\right] = \frac{S(s)}{N(s)}.$$

We can also define this measure on the basis of the Radon-Nikodym derivative of Q^* w.r.t. Q.

3.100. THEOREM. *Suppose that Q is an equivalent martingale measure for the numéraire $\{M(t)\}_{t\geqslant 0}$. Let Q^* be an absolutely continuous measure w.r.t. Q defined by the Radon-Nikodym derivative:*

$$\Gamma_t := \frac{dQ^*}{dQ}\Big|_{\mathcal{F}_t} = \frac{M(0)}{M(t)}\frac{N(t)}{N(0)}, \tag{3.295}$$

where $N(t) > 0$ is the price at time t of a particular asset. Then Q^* is an equivalent martingale measure for $\{N(t) : t \geqslant 0\}$.

PROOF. Denote expectations w.r.t. Q by $\mathbb{E}$ and w.r.t. Q^* by $\mathbb{E}^*$. Let $S(t)$ be the price of an asset at time $t \geqslant 0$ and assume $S(t) \in L^2(\Omega, \mathcal{F}_t, Q) \cap L^2(\Omega, \mathcal{F}_t, Q^*)$. For $t \geqslant s \geqslant 0$ we find using Theorem 3.99

$$\begin{aligned}\mathbb{E}^*\left[\frac{S(t)}{N(t)} \mid \mathcal{F}_s\right] &= \mathbb{E}\left[\frac{M(0)N(t)}{M(t)N(0)}\frac{S(t)}{N(t)} \mid \mathcal{F}_s\right] / \mathbb{E}\left[\frac{M(0)N(t)}{M(t)N(0)} \mid \mathcal{F}_s\right] \\ &= \frac{M(0)}{N(0)}\mathbb{E}\left[\frac{S(t)}{M(t)} \mid \mathcal{F}_s\right]\frac{N(0)}{M(0)}\frac{M(s)}{N(s)} = \frac{S(s)}{N(s)}.\end{aligned}$$

The proof of Theorem 3.100 is complete now. $\square$

Note that the measures Q and Q^* are equivalent because of the strictly positiveness of the Radon-Nikodym derivative. We already mention the following theorem which transforms the dynamics of a process under Q to a process under Q^*.

3.101. THEOREM. *Let Q be an equivalent martingale measure for $\{M(t)\}_{t\geqslant 0}$ and let Q^* be defined by equation (3.295). Assume $\{X(t) : 0 \leqslant t \leqslant t\}$ a diffusion process with dynamics under Q*

$$dX(t) = b\,(t, X(t))\;dt + \sigma\,(t, X(t))\;dW(t).$$

Let also $M(t)$ and $N(t)$ have dynamics under Q given by

$$dM(t) = m_M\,dt + \sigma_M\,dW(t), \quad dN(t) = m_N\,dt + \sigma_N\,dW(t).$$

Then the dynamics of $\{X(t) : 0 \leqslant t \leqslant t\}$ under Q^* is given by

$$dX(t) = b\,(t,\omega)\;dt - \sigma\,(t,\omega)\left(\frac{\sigma_M}{M(t)} - \frac{\sigma_N}{N(t)}\right)dt + \sigma\,(t,\omega)\;dW^*(t),$$

where

$$W^*(t) = W(t) + \int_0^t \left(\frac{\sigma_M}{M(s)} - \frac{\sigma_N}{N(s)}\right)ds.$$

PROOF. It is clear that we want to apply Girsanov's Theorem 3.94. But then we need to know how $\theta\,(t,\omega)$ in expression (3.274) looks like. From expression (3.275) we know that

$$d\Gamma_t = -\theta\,(t,\cdot)\,\Gamma_t\,dW(t). \tag{3.296}$$

On the other hand:

$$\begin{aligned} d\Gamma_t &= \frac{M(0)}{N(0)}\,d\left(\frac{N(t)}{M(t)}\right) = \frac{M(0)}{N(0)}\frac{dN(t)M(t) - N(t)dM(t)}{M(t)^2} \\ &= \frac{M(0)}{N(0)M(t)^2}\left(\left(m_N\,dt + \sigma_N\,dW(t)\right)M(t) - N(t)\left(m_M\,dt + \sigma_M\,dW(t)\right)\right). \end{aligned} \tag{3.297}$$

Because $\{\Gamma_t : 0 \leqslant t \leqslant T\}$ is a martingale we must have that the coefficient of dt is 0, hence

$$\begin{aligned} d\Gamma_t &= \frac{M(0)}{N(0)M(t)^2}\left(\sigma_N\,dW(t)M(t) - N(t)\sigma_M\,dW(t)\right) \\ &= \left(\frac{\sigma_N}{N(t)} - \frac{\sigma_M}{M(t)}\right)\Gamma_t\,dW(t). \end{aligned}$$

Comparing this with (3.296) we have that

$$\theta(t,\omega) = \frac{\sigma_M}{M(t)} - \frac{\sigma_N}{N(t)}.$$

Finally applying Girsanov's Theorem 3.94 to this we have

$$dX(t) = b\,(t,\omega)\;dt - \sigma\,(t,\omega)\left(\frac{\sigma_M}{M(t)} - \frac{\sigma_N}{N(t)}\right)dt + \sigma\,(t,\omega)\;dW^*(t),$$

with

$$W^*(t) = W(t) + \int_0^t \left(\frac{\sigma_M}{M(s)} - \frac{\sigma_N}{N(s)} \right) ds.$$

Altogether this completes the proof of Theorem 3.101. □

In order to make the link between the short rate $r(t)$ and the instantaneous forward rate $f(t,T)$, we introduce a new measure Q^T. Suppose again that Q is the risk neutral measure w.r.t. the money market and $\mathbb{E}$ the expectation w.r.t. Q.

3.102. DEFINITION. Take $T \geqslant 0$. The forward measure Q^T is defined on $\mathcal{F}_T$ by setting

$$\Gamma_T := \frac{dQ^T}{dQ} = \frac{M(0)}{M(T)} P(T,T) P(0,T) = e^{-\int_0^T r(s)\,ds} P(0,T),$$

where

$$M(t) = e^{\int_0^t r(s)\,ds}, \quad \text{and} \quad P(t,T) = \mathbb{E}\left[e^{-\int_t^T r(s)\,ds} \mid \mathcal{F}_t \right].$$

By the previous theorem we conclude that Q^T is an equivalent martingale measure which has a bond with maturity T as numéraire.

For $t < T$ we can easily calculate Γ_t as follows:

$$\begin{aligned}\Gamma_t &:= \mathbb{E}\left[\Gamma_T \mid \mathcal{F}_t\right] = \frac{M(0)}{P(0,T)} \mathbb{E}\left[\frac{P(T,T)}{M(T)} \mid \mathcal{F}_t \right] \\ &= \frac{1}{P(0,T)} \frac{P(t,T)}{M(t)} = e^{-\int_0^t r(s)\,ds} \frac{P(t,T)}{P(0,T)},\end{aligned}$$

where we used in the second to last equality that Q has $\{M(t)\}_{t\geqslant 0}$ as numéraire.

Now we have all theoretical background information to formulate the third proof of Theorem 3.96.

THIRD PROOF OF THEOREM 3.96. Denote as before the expectation w.r.t. Q by $\mathbb{E}$ and the expectation w.r.t. Q^T by $\mathbb{E}^T$. We have by Theorem 3.99 that for any $\mathcal{F}_T$-measurable random variable X and $t \leqslant T$

$$\begin{aligned}\mathbb{E}^T\left[X j \mathcal{F}_t\right] &= \Gamma_t^{-1} \mathbb{E}\left[X \Gamma_T \mid \mathcal{F}_t\right] = \mathbb{E}\left[X \frac{\Gamma_T}{\Gamma_t} \mid \mathcal{F}_t \right] \\ &= \mathbb{E}\left[X \frac{M(t) P(T,T)}{M(T) P(t,T)} \mid \mathcal{F}_t \right] = \mathbb{E}\left[X \frac{e^{-\int_t^T r(s)\,ds}}{P(t,T)} \right].\end{aligned}$$

We want to express the forward rate in terms of the short rate. We got a formula for the bonds price in function of both of them. Differentiating expression (16) towards T gives

$$\begin{aligned}\frac{\partial P(t,T)}{\partial T} &= \mathbb{E}\left[-r(T) e^{-\int_t^T r(s)\,ds} \mid \mathcal{F}_t \right] = \mathbb{E}^T\left[-r(T) P(t,T) \mid \mathcal{F}_t \right] \\ &= -\mathbb{E}^T\left[r(T) \mid \mathcal{F}_t \right] P(t,T). \qquad (3.298)\end{aligned}$$

In the second step we used the above reasoning with $X = -r(T)P(t,T)$. Differentiating now formula (3.294) with respect to T gives

$$\frac{\partial P(t,T)}{\partial T} = -P(t,T)\, f(t,T). \tag{3.299}$$

Comparing (3.298) en (3.299) we get the link between short rate and forward rate

$$f(t,T) = \mathbb{E}^T\left[r(T) \mid \mathcal{F}_t\right]. \tag{3.300}$$

Considering the right hand side of (3.300) we will need to describe the dynamics of $r(t)$ under Q^T . Applying Itô's formula on $f(t,x) = e^{\int_0^t r(s)\,ds}$ we immediately find that $dM(t) = r(t)M(t)\,dt$. In the notation of Theorem 3.101 we thus have $\sigma_M = 0$. If we then apply this theorem with $X(t) = r(t)$, $Q^* = Q^T$, $\sigma(t, X(t)) = \sigma$, $b(t, X(t)) = a(b - r(t))$ and $\sigma_N = -\sigma A(t,T)\,P(t,T)$ we obtain:

$$\begin{aligned} dr(t) &= \left(ab - \sigma^2 A(t,T) - ar(t)\right)\, dt + \sigma\, dW^T(t) \\ &= a\left(b - \frac{\sigma^2}{a^2}\left(1 - e^{-a(T-t)}\right) - rt\right)\, dt + \sigma\, dW^T(t) \end{aligned} \tag{3.301}$$

where $W^T(t)$ is the Q^T-Brownian motion defined by

$$W^T(t) = W(t) + \sigma \int_0^t A(s,T)\, ds.$$

Expression (3.301) resembles an ordinary Vasicek process, except that the term $b - \frac{\sigma^2}{a^2}\left(1 - e^{-a(T-t)}\right)$ does depend upon t and is thus not a constant. However, we will use a similar reasoning as in the classical situation to solve the SDE for $r(t)$ on the interval $[t,T]$. First, we apply Itô's formula on $g(t,x) = e^{at}x$:

$$\begin{aligned} d\left(e^{at}r(t)\right) &= abe^{at}\,dt - \frac{1 - e^{-a(T-t)}}{a}\sigma^2 e^{at}\,dt + \sigma e^{at}\,dW^T(t) \\ &= abe^{at}\,dt - \frac{\sigma^2}{a}\left(e^{at} - e^{-a(T-2t)}\right)\,dt + \sigma e^{at}\,dW^T(t). \end{aligned}$$

Integrating from t to T gives

$$\begin{aligned} &e^{aT}r(T) - e^{at}r(t) \\ &= ab\int_t^T e^{as}\,ds - \frac{\sigma^2}{a}\int_t^T \left(e^{as} - e^{-a(T-2s)}\right)\,ds + \sigma\int_t^T e^{as}\,dW^T(s) \\ &= b\left(e^{aT} - e^{at}\right) - \frac{\sigma^2}{a}\left[\frac{1}{a}\left(e^{aT} - e^{at}\right) - \frac{1}{2a}\left(e^{aT} - e^{-a(T-2t)}\right)\right] + \sigma\int_t^T e^{as}\,dW^T(s) \\ &= b\left(e^{aT} - e^{at}\right) - \frac{\sigma^2}{2a^2}\left(e^{aT} - 2e^{at} + e^{-a(T-2t)}\right) + \sigma\int_t^T e^{as}\,dW^T(s). \end{aligned}$$

Thus we have that

$$\begin{aligned} r(T) = r(t)e^{-a(T-t)} &+ b\left(1 - e^{-a(T-t)}\right) - \frac{\sigma^2}{2a^2}\left(1 - 2e^{-a(T-t)} + e^{-2a(T-t)}\right) \\ &+ \sigma\int_t^T e^{-a(T-s)}\,dW^T(s). \end{aligned}$$

And hence,

$$\begin{aligned}
f(t,s) &= \mathbb{E}^s\left[r(s) \mid \mathcal{F}_t\right] \\
&= r(t)e^{-a(s-t)} + b\left(1-e^{-a(s-t)}\right) - \frac{\sigma^2}{a^2}\left(1-2e^{-a(s-t)}+e^{-2a(s-t)}\right) \\
&= r(t)e^{-a(s-t)} + \left(b-\frac{\sigma^2}{2a^2}\right)\left(1-e^{-a(s-t)}\right) + \frac{\sigma^2}{2a^2}\left(e^{-a(s-t)}-e^{-2a(s-t)}\right).
\end{aligned}$$

Integrating results in

$$\begin{aligned}
\int_t^T \mathbb{E}^s\left[r(s) \mid \mathcal{F}_t\right] ds &= \frac{r(t)}{a}\left(1-e^{-a(T-t)}\right) + \left(b-\frac{\sigma^2}{2a^2}\right)\left(T-t-\frac{1-e^{-a(T-t)}}{a}\right) \\
&\quad + \frac{\sigma^2}{2a^3}\left(1-e^{-a(T-t)}\right) - \frac{\sigma^2}{4a^3}\left(1-e^{-2a(T-t)}\right) \\
&= r(t)A(t,T) + \left(b-\frac{\sigma^2}{2a^2}\right)\left(T-t-A(t,T)\right) + \frac{\sigma^2}{4a}A(t,T)^2 \\
&= r(t)A(t,T) - D(t,T). \qquad (3.302)
\end{aligned}$$

Reminding formula (3.294) and formula (3.300) we find again that

$$P(t,T) = e^{-A(t,T)rt+D(t,T)}.$$

This completes the third proof of Theorem 3.96. $\square$

9. A version of Fernique's theorem

The following theorem is due to Fernique. We follow the proof of H.H. Kuo [**77**].

3.103. Theorem. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $X : \Omega \to \mathbb{R}^d$ *be a Gaussian vector with mean zero. Put*

$$\alpha = \sup_{u,v>0} \frac{1}{(u+v)^2} \log \frac{\mathbb{P}(|X| \leqslant v)}{\mathbb{P}(|X| > u)}. \tag{3.303}$$

Then $\alpha > 0$ *and*

$$\mathbb{E}\left(\exp\left(\frac{1}{2}\eta |X|^2\right)\right) < \infty \quad \textit{for} \quad \eta < \alpha. \tag{3.304}$$

For the proof we shall need two lemmas. The first one contains the main idea.

3.104. Lemma. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* X *be as in Theorem 3.103. Let* $s > 0$ *be such that* $\mathbb{P}(|X| \leqslant s) > 0$ *and fix* $t > s$. *Then*

$$\frac{\mathbb{P}(|X| > t)}{\mathbb{P}(|X| \leqslant s)} \leqslant \left(\frac{\mathbb{P}\left(|X| > (t-s)/\sqrt{2}\right)}{\mathbb{P}(|X| \leqslant s)}\right)^2. \tag{3.305}$$

Proof. Let $(\Omega \otimes \Omega, \mathcal{F} \otimes \mathcal{F}, \mathbb{P} \otimes \mathbb{P})$ be the tensor product space of $(\Omega, \mathcal{F}, \mathbb{P})$ with itself and define X_i, $i = 1, 2$, by $X_i(\omega_1, \omega_2) = X(\omega_i)$. Then the variables X_1 and X_2 are independent with respect $\mathbb{P} \otimes \mathbb{P}$ and their $\mathbb{P} \otimes \mathbb{P}$-distribution coincides with the $\mathbb{P}$-distribution of X. We shall prove Lemma 3.104. for $s = v$ and $t = u\sqrt{2} + v$. Since the vector (X_1, X_2) is Gaussian with respect to $\mathbb{P} \otimes \mathbb{P}$ and since the components of $X_1 - X_2$ are uncorrelated with the components of $X_1 + X_2$ (with respect to the probability $\mathbb{P} \otimes \mathbb{P}$), it follows that the vectors $X_1 - X_2$ and $X_1 + X_2$ are independent. Notice that $\int X d\mathbb{P} = \int X_1 d\mathbb{P} \otimes \mathbb{P} = \int X_2 d\mathbb{P} \otimes \mathbb{P} = 0$ and that the covariance matrices of X, of $(X_1 - X_2)/\sqrt{2}$ and of $(X_1 + X_2)/\sqrt{2}$ all coincide. It follows that the joint distributions of (X_1, X_2) and of $\left(\dfrac{X_1 - X_2}{\sqrt{2}}, \dfrac{X_1 + X_2}{\sqrt{2}}\right)$ are the same as well. Hence the following (in-)equalities are now self-explanatory:

$$\begin{aligned}
&\mathbb{P}(|X| \leqslant v)\, \mathbb{P}\left(|X| > u\sqrt{2} + v\right) = \mathbb{P} \otimes \mathbb{P}(|X_1| \leqslant v) \times \mathbb{P} \otimes \mathbb{P}\left(|X_2| > u\sqrt{2} + v\right) \\
&\quad = \mathbb{P} \otimes \mathbb{P}\left(|X_1| \leqslant v \quad \text{and} \quad |X_2| > u\sqrt{2} + v\right) \\
&\quad = \mathbb{P} \otimes \mathbb{P}\left(|X_1 - X_2| \leqslant v\sqrt{2} \quad \text{and} \quad |X_1 + X_2| > 2u + v\sqrt{2}\right) \\
&\quad \leqslant \mathbb{P} \otimes \mathbb{P}\left(\big||X_1| - |X_2|\big| \leqslant v\sqrt{2} \quad \text{and} \quad |X_1| + |X_2| > 2u + v\sqrt{2}\right) \\
&\quad \leqslant \mathbb{P} \otimes \mathbb{P}(|X_1| > u \;\; \text{and} \;\; |X_2| > u) = \mathbb{P}(|X| > u)^2.
\end{aligned} \tag{3.306}$$

Inequality (3.305) in Lemma 3.104 follows from (3.306). ☐

3.105. Lemma. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and X be as in Theorem 3.103. Let $v > 0$ be such that $\mathbb{P}(|X| \leqslant v) > 0$ and fix $\ell \in \mathbb{N}$ and fix $u > 0$. Then the following inequality is valid:*

$$\frac{\mathbb{P}\left(|X| > u\left(\sqrt{2}\right)^{\ell} + v\left(\left(\sqrt{2}\right)^{\ell} - 1\right)\left(\sqrt{2}+1\right)\right)}{\mathbb{P}(|X| \leqslant v)} \leqslant \left(\frac{\mathbb{P}(|X| > u)}{\mathbb{P}(|X| \leqslant v)}\right)^{2^{\ell}}. \tag{3.307}$$

Proof. For $\ell = 0$ this assertion is trivial and for $\ell = 1$ it is the same as inequality (3.305) in Lemma 3.104. Next suppose that (3.307) is already established for ℓ. We are going to prove (3.307) with $\ell + 1$ replacing ℓ. Again we invoke inequality (3.305) to obtain

$$\begin{aligned}
&\frac{\mathbb{P}\left(|X| > u\left(\sqrt{2}\right)^{\ell+1} + v\left(\left(\sqrt{2}\right)^{\ell+1} - 1\right)\left(\sqrt{2}+1\right)\right)}{\mathbb{P}(|X| \leqslant v)} \\
&\leqslant \left(\frac{\mathbb{P}\left(|X| > \left(u\left(\sqrt{2}\right)^{\ell+1} + v\left(\left(\sqrt{2}\right)^{\ell+1} - 1\right)\left(\sqrt{2}+1\right) - v\right)/\sqrt{2}\right)}{\mathbb{P}(|X| \leqslant v)}\right)^{2} \\
&= \left(\frac{\mathbb{P}\left(|X| > u\left(\sqrt{2}\right)^{\ell} + v\left(\left(\sqrt{2}\right)^{\ell} - 1\right)\left(\sqrt{2}+1\right)\right)}{\mathbb{P}(|X| \leqslant v)}\right)^{2}
\end{aligned}$$

(induction hypothesis)

$$\leqslant \left(\frac{\mathbb{P}(|X| > u)}{\mathbb{P}(|X| \leqslant v)}\right)^{2^{\ell+1}}. \tag{3.308}$$

The inequality in (3.308) completes the proof of Lemma 3.105. $\square$

Proof of Theorem 3.103. If $X \equiv 0$, then there is nothing to prove. So suppose $X \neq 0$ and choose strictly positive real numbers u and v for which $\dfrac{\mathbb{P}(|X| > u)}{\mathbb{P}(|X| \leqslant v)} < 1$. Put

$$\begin{aligned}
\alpha(u,v) &= \frac{1}{(u+v)^2} \log \frac{\mathbb{P}(|X| \leqslant v)}{\mathbb{P}(|X| > u)} \quad \text{and} \\
\beta(u,v) &= \mathbb{P}(|X| \leqslant v) \left(\frac{\mathbb{P}(|X| > u)}{\mathbb{P}(|X| \leqslant v)}\right)^{\dfrac{v^2\left(1+\sqrt{2}\right)^2}{2(u+v)^2}}.
\end{aligned}$$

Then $\alpha(u,v) > 0$ and $\beta(u,v) = \mathbb{P}(|X| \leqslant v) \exp\left(-\frac{1}{2}\alpha(u,v)v^2\left(1+\sqrt{2}\right)^2\right) < 1$. For $s \geqslant u$ choose $\ell \in \mathbb{N}$ in such a way that

$$\begin{aligned}
&u\left(\sqrt{2}\right)^{\ell+1} + v\left(\left(\sqrt{2}\right)^{\ell+1} - 1\right)\left(\sqrt{2}+1\right) > s \\
&\geqslant u\left(\sqrt{2}\right)^{\ell} + v\left(\left(\sqrt{2}\right)^{\ell} - 1\right)\left(\sqrt{2}+1\right).
\end{aligned}$$

Then

$$2^{\ell} > \frac{\left(s+v\left(1+\sqrt{2}\right)\right)^2}{2\left(u+v\right)^2} > \frac{s^2}{2\left(u+v\right)^2} + \frac{v^2\left(1+\sqrt{2}\right)^2}{2\left(u+v\right)^2}$$

and hence

$$\mathbb{P}\left(|X| > s\right) \leqslant \mathbb{P}\left(|X| > u\left(\sqrt{2}\right)^{\ell} + v\left(\left(\sqrt{2}\right)^{\ell} - 1\right)\left(\sqrt{2}+1\right)\right)$$

(inequality (3.307) in Lemma 3.105)

$$\begin{aligned}
&\leqslant \mathbb{P}\left(|X| \leqslant v\right)\left(\frac{\mathbb{P}\left(|X| > u\right)}{\mathbb{P}\left(|X| \leqslant v\right)}\right)^{2^{\ell}} \\
&\leqslant \beta(u,v)\exp\left(-\frac{1}{2}\alpha(u,v)s^2\right) < \exp\left(-\frac{1}{2}\alpha(u,v)s^2\right). \qquad (3.309)
\end{aligned}$$

If $0 \leqslant \eta < \alpha$, then we choose u, $v > 0$ in such a way that $\alpha > \alpha(u,v) > \eta$. Then, for $s \geqslant u$, $\mathbb{P}\left(|X| > s\right) < \exp\left(-\frac{1}{2}\alpha(u,v)s^2\right)$. Consequently, we get from (3.309):

$$\begin{aligned}
&\mathbb{E}\left(\exp\left(\frac{1}{2}\eta\left|X\right|^2\right) : |X| > u\right) \\
&\quad = \int_0^{\infty} \mathbb{P}\left(\exp\left(\frac{1}{2}\eta\left|X\right|^2\right) > \xi, |X| > u\right) d\xi
\end{aligned}$$

(substitute $\xi = \exp\left(\frac{1}{2}\eta s^2\right)$)

$$\begin{aligned}
&\leqslant \exp\left(\frac{1}{2}\eta u^2\right)\mathbb{P}\left(|X| > u\right) + \eta\int_u^{\infty} \mathbb{P}\left(|X| > s\right)\exp\left(\frac{1}{2}\eta s^2\right) sds \\
&\leqslant \exp\left(\frac{1}{2}\eta u^2\right)\mathbb{P}\left(|X| > u\right) + \eta\int_u^{\infty} \exp\left(-\frac{1}{2}\left(\alpha(u,v)-\eta\right)s^2\right) sds \\
&\leqslant \exp\left(\frac{1}{2}\eta u^2\right) + \frac{\eta}{\alpha(u,v)-\eta}\exp\left(-\frac{1}{2}\left(\alpha(u,v)-\eta\right)u^2\right) \\
&\leqslant \frac{\alpha(u,v)}{\alpha(u,v)-\eta}\exp\left(\frac{1}{2}\eta u^2\right). \qquad (3.310)
\end{aligned}$$

From (3.310) we infer

$$\mathbb{E}\left(\exp\left(\frac{1}{2}\eta\left|X\right|^2\right)\right) \leqslant \frac{2\alpha(u,v)-\eta}{\alpha(u,v)-\eta}\exp\left(\frac{1}{2}\eta u^2\right). \qquad (3.311)$$

Inequality (3.311) yields the desired result in Theorem 3.103. □

10. Miscellaneous

We begin this section with the Doob's optional stopping property for discrete time submartingales. Let $\{X(n) : n \in \mathbb{N}\}$ be a submartingale relative to the filtration $\{\mathcal{F}_n : n \in \mathbb{N}\}$. Here the random variables $X(n)$ are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The following result was used in inequality (3.164), the basic step for the continuous time version of the following proposition.

3.106. Proposition. *Let τ be a stopping time. The process*

$$\{X(\min(n,\tau)) : n \in \mathbb{N}\}$$

is a submartingale with respect to the filtration $\{\mathcal{F}_n : n \in \mathbb{N}\}$ *as well as with respect to the filtration* $\left\{\mathcal{F}_{\min(n,\tau)} : n \in \mathbb{N}\right\}$.

Proof. Let m and n be natural numbers with $m < n$ and let A be a member of $\mathcal{F}_m$. Then we have

$$\begin{aligned}
&\mathbb{E}\left(X\left(\min(n,\tau)\right)1_A\right) - \mathbb{E}\left(X\left(\min(m,\tau)\right)1_A\right) \\
&\quad = \sum_{k=m+1}^{n} \left\{\mathbb{E}\left(X\left(\min(k,\tau)\right)1_A\right) - \mathbb{E}\left(X\left(\min(k-1,\tau)\right)1_A\right)\right\} \\
&\quad = \sum_{k=m+1}^{n} \left\{\mathbb{E}\left(\left(X\left(\min(k,\tau)\right) - X\left(\min(k-1,\tau)\right)\right)1_{A\cap\{\tau\geqslant k\}}\right)\right\} \\
&\quad = \sum_{k=m+1}^{n} \mathbb{E}\left\{\mathbb{E}\left(\left(X\left(\min(k,\tau)\right) - X\left(\min(k-1,\tau)\right)\right)1_{A\cap\{\tau\geqslant k\}}\right) \,\middle|\, \mathcal{F}_{k-1}\right\}
\end{aligned}$$

(the event $A \cap \{\tau \geqslant k\}$ belongs to $\mathcal{F}_{k-1}$ for $k \geqslant m+1$, and the variable $X(k-1)$ is $\mathcal{F}_{k-1}$-measurable)

$$= \sum_{k=m+1}^{n} \mathbb{E}\left(\left(\mathbb{E}\left(X(k) \,\middle|\, \mathcal{F}_{k-1}\right) - X(k-1)\right)1_{A\cap\{\tau\geqslant k\}}\right)$$

(submartingale property of the process $\{X(k) : k \in \mathbb{N}\}$)

$$\geqslant \sum_{k=m+1}^{n} \mathbb{E}\left(0 \times \mathbf{1}_{A\cap\{\tau\geqslant k\}}\right) = 0. \tag{3.312}$$

The inequality in (3.312) proves that the process $\{X(\min(k,\tau)) : k \in \mathbb{N}\}$ is a submartingale for the filtration $\{\mathcal{F}_k : k \in \mathbb{N}\}$. Since the σ-field $\mathcal{F}_{\min(k,\tau}$ is contained in the σ-field $\mathcal{F}_k$, $k \in \mathbb{N}$, it also follows that the process

$$\{X(\min(k,\tau)) : k \in \mathbb{N}\}$$

is also a submartingale with respect to the filtration $\left\{\mathcal{F}_{\min(k,\tau)} : k \in \mathbb{N}\right\}$ because we have

$$\begin{aligned}
&\mathbb{E}\left(X(\min(m+1,\tau)) \,\middle|\, \mathcal{F}_{\min(m,\tau)}\right) = \mathbb{E}\left(\mathbb{E}\left(X(m+1) \,\middle|\, \mathcal{F}_m\right) \,\middle|\, \mathcal{F}_{\min(m,\tau)}\right) \\
&\quad \geqslant \mathbb{E}\left(\mathbb{E}\left(X(\min(m+1,\tau)) \,\middle|\, \mathcal{F}_m\right) \,\middle|\, \mathcal{F}_{\min(m,\tau)}\right)
\end{aligned}$$

(employ (3.312))

$$\geqslant \mathbb{E}\left(X(\min(m,\tau)) \,\middle|\, \mathcal{F}_{\min(m,\tau)}\right) = X(\min(m,\tau)). \tag{3.313}$$

The inequalities (3.312) and (3.313) together prove the results in Theorem 3.106. □

Next we prove Doob's maximal inequality for martingales.

3.107. Proposition. *Let $\{M(n) : n \in \mathbb{N}\}$ be a martingale. Put*

$$M(n)^* = \max_{k\leqslant n} |M(n)|.$$

The following inequalities are valid:

$$\mathbb{P}\left[M(n)^* \geqslant \lambda\right] \leqslant \frac{1}{\lambda}\mathbb{E}\left[|M(n)| : M(n)^* \geqslant \lambda\right]; \tag{3.314}$$

$$\mathbb{P}\left[M(n)^* \geqslant \lambda\right] \leqslant \frac{1}{\lambda^2}\mathbb{E}\left[|M(n)|^2 : M(n)^* \geqslant \lambda\right]. \tag{3.315}$$

Let $\{M(t) : t \geqslant 0\}$ be a continuous time martingale that is right continuous and possesses left limits. Put $M(t)^* = \sup_{0 \leqslant s \leqslant t} |M(s)|$. Again inequalities like (3.314) and (3.315) are true:

$$\mathbb{P}\left\{M(t)^* \geqslant \lambda\right\} \leqslant \frac{1}{\lambda}\mathbb{E}\left\{|M(t)| : M(t)^* \geqslant \lambda\right\}; \tag{3.316}$$

$$\mathbb{P}\left\{M(t)^* \geqslant \lambda\right\} \leqslant \frac{1}{\lambda^2}\mathbb{E}\left\{|M(t)|^2 : M(t)^* \geqslant \lambda\right\}. \tag{3.317}$$

PROOF. We begin by establishing inequality (3.314). Define the events A_k, $1 \leqslant k \leqslant n$, by $A_0 = \{|M(0)| \geqslant \lambda\}$,

$$A_k = \{|M(j)| < \lambda, 0 \leqslant j \leqslant k-1, |M(k)| \geqslant \lambda\}, \quad 1 \leqslant k \leqslant n.$$

Then $\bigcup_{k=0}^n A_k = \{M(n)^* \geqslant \lambda\}$, $A_k \cap A_\ell = \emptyset$, for $k \neq \ell$, $1 \leqslant k$, $\ell \leqslant n$, and A_k is $\mathcal{F}_k$-measurable for $1 \leqslant k \leqslant n$. Moreover on the event A_k the inequality $|M(k)| \geqslant \lambda$ is valid. From the martingale property it then follows that:

$$\begin{aligned}
\mathbb{P}\left(M(n)^* \geqslant \lambda\right) &= \sum_{k=0}^n \mathbb{P}\left(A_k\right) \leqslant \frac{1}{\lambda}\sum_{k=0}^n \mathbb{E}\left(1_{A_k}|M(k)|\right) \\
&= \frac{1}{\lambda}\sum_{k=0}^n \mathbb{E}\left(1_{A_k}\left|\mathbb{E}\left(M(n) \mid \mathcal{F}_k\right)\right|\right) \\
&= \frac{1}{\lambda}\sum_{k=0}^n \mathbb{E}\left(\left|\mathbb{E}\left(1_{A_k}M(n) \mid \mathcal{F}_k\right)\right|\right) \\
&\leqslant \frac{1}{\lambda}\sum_{k=0}^n \mathbb{E}\left(\mathbb{E}\left(1_{A_k}|M(n)| \mid \mathcal{F}_k\right)\right) \\
&= \frac{1}{\lambda}\sum_{k=0}^n \mathbb{E}\left(1_{A_k}|M(n)|\right) = \frac{1}{\lambda}\mathbb{E}\left(|M(n)| : M(n)^* \geqslant \lambda\right).
\end{aligned} \tag{3.318}$$

Notice that inequality (3.318) is the same as (3.314). The proof of (3.315) goes along the same lines. The fact is used that the process $\left\{|M(n)|^2 : n \in \mathbb{N}\right\}$ constitutes a submartingale. The details read as follows. The events A_k, $1 \leqslant k \leqslant n$, are defined as in the proof of (3.314). The argument in (3.318) is adapted as below:

$$\begin{aligned}
\mathbb{P}\left(M(n)^* \geqslant \lambda\right) &= \sum_{k=0}^n \mathbb{P}\left(A_k\right) \leqslant \frac{1}{\lambda^2}\sum_{k=0}^n \mathbb{E}\left(1_{A_k}|M(k)|^2\right) \\
&\leqslant \frac{1}{\lambda^2}\sum_{k=0}^n \mathbb{E}\left(1_{A_k}\mathbb{E}\left(|M(n)|^2 \mid \mathcal{F}_k\right)\right) \leqslant \frac{1}{\lambda^2}\sum_{k=0}^n \mathbb{E}\left(1_{A_k}|M(n)|^2\right) \\
&= \frac{1}{\lambda^2}\sum_{k=0}^n \mathbb{E}\left(1_{A_k}|M(n)|^2\right) = \frac{1}{\lambda^2}\mathbb{E}\left(|M(n)|^2 : M(n)^* \geqslant \lambda\right).
\end{aligned} \tag{3.319}$$

Again we notice that (3.319) is the same as (3.315). The inequalities in (3.316) and (3.317) are based on a time discretization of the martingale $\{M(t) : t \geqslant 0\}$. Therefore we write $N(j) := M\left(j2^{-n}t\right)$ and we notice that $\{N(j) : j \in \mathbb{N}\}$ is a martingale for the filtration $\{\mathcal{F}_{j2^{-n}t} : j \in \mathbb{N}\}$. From (3.314) we obtain the inequality:

$$\mathbb{P}\left(\max_{0\leqslant j\leqslant 2^n} |N(j)| \geqslant \lambda\right) \leqslant \frac{1}{\lambda}\mathbb{E}\left(|M(t)| : M(t)^* \geqslant \lambda\right). \tag{3.320}$$

Inequality (3.317) is obtained from (3.320) upon letting tend n to ∞. The proof of (3.317) follows in the same manner from (3.315). □

We continue with a proof of the (DL)-property of martingales. More precisely we shall prove the following proposition.

3.108. Proposition. *Let $\{M(s) : s \geqslant 0\}$ be a right continuous martingale on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Fix $t \geqslant 0$. Then the collection of random variables*

$$\{M(\tau) : 0 \leqslant \tau \leqslant t,\ \tau \text{ stopping time}\}$$

is uniformly integrable.

Proof. Fix a stopping time $0 \leqslant \tau \leqslant t$ and write $\tau_n = \min\left(2^{-n}\lceil 2^n t\rceil, t\right)$. Then $0 \leqslant \tau_n \leqslant t$ and every τ_n is a stopping time. Moreover $\tau_n \downarrow \tau$ if n tends to ∞. Since the pair $(M(\tau_n), M(t))$ is a martingale for the pair of σ-fields $(\mathcal{F}_{\tau_n}, \mathcal{F}_t)$ (Use Proposition 3.106 for martingales), the pair $\left(\left|M\left(\tau_n\right)\right|, |M(t)|\right)$ is a submartingale with respect to the same pair of σ-fields. As a consequence we obtain:

$$\mathbb{E}\left(|M(\tau)| : |M(\tau)| \geqslant \lambda\right) = \mathbb{E}\left(\liminf_{n\to\infty} \left|M(\tau_n) 1_{\{|M(\tau_n)|\geqslant\lambda\}}\right|\right)$$

(Fatou's lemma)

$$\leqslant \liminf_{n\to\infty} \mathbb{E}\left(|M(\tau_n)|\, 1_{\{|M(\tau_n)|\geqslant\lambda\}}\right)$$

(submartingale property)

$$\begin{aligned}
&\leqslant \liminf_{n\to\infty} \mathbb{E}\left(\mathbb{E}\left(|M(t)| \bigm| \mathcal{F}_{\tau_n}\right) 1_{\{|M(\tau_n)|\geqslant\lambda\}}\right)\\
&= \liminf_{n\to\infty} \mathbb{E}\left(|M(t)|\, 1_{\{|M(\tau_n)|\geqslant\lambda\}}\right)\\
&= \liminf_{n\to\infty} \mathbb{E}\left(|M(t)| : \left|M\left(\tau_n\right)\right| \geqslant \lambda\right)\\
&\leqslant \mathbb{E}\left(|M(t)| : |M(t)| \geqslant \lambda\right).
\end{aligned} \tag{3.321}$$

This proves Proposition 3.108. □

Remark. In the proof of Proposition 3.108. we did use a discrete approximation of a stopping time. However we could have avoided this and consider directly the pair $(M(\tau), M(t))$. From Proposition 3.107 we see that this pair is a martingale with respect to the pair of σ-fields $(\mathcal{F}_\tau, \mathcal{F}_t)$. This will then imply inequality (3.321) with τ replacing τ_n. On the other hand the discrete approximation of stopping times as performed in the proof of Proposition 3.108 is kind of a standard procedure for passing from discrete time valued stopping times to continuous time valued stopping times. This is a good reason to insert this kind of argument.

The main result of Section 3 of this chapter says that linear operators in $C_0(E)$ which maximally solve the martingale problem are generators of Feller semigroups and conversely. In the sequel we want to verify the claim in the example of Section 3. Its statement is correct, but its proof is erroneous. Example 3.49 in Section 3 reads as follows.

3.109. Example. Let L_0 be an unbounded generator of a Feller semigroup in $C_0(E)$ and let μ_k and ν_k, $1 \leqslant k \leqslant n$, be finite (signed) Borel measures on E. Define the operator $L_{\vec{\mu},\vec{\nu}}$ as follows:

$$\begin{aligned} D\left(L_{\vec{\mu},\vec{\nu}}\right) &= \bigcap_{k=1}^{n} \left\{ f \in D(L_0) : \int L_0 f d\mu_k = \int f d\nu_k \right\}, \\ L_{\vec{\mu},\vec{\nu}} f &= L_0 f, \quad f \in D\left(L_{\vec{\mu},\vec{\nu}}\right). \end{aligned}$$

Then the martingale problem is uniquely solvable for $L_{\vec{\mu},\vec{\nu}}$. In fact let

$$\left\{ \left(\Omega, \mathcal{F}, \mathbb{P}_x\right), \left(X(t) : t \geqslant 0\right), \left(\vartheta_t : t \geqslant 0\right), \left(E, \mathcal{E}\right) \right\}$$

be the strong Markov process associated to the Feller semigroup generated by L_0. Then $\mathbb{P} = \mathbb{P}_x$ solves the martingale problem

(a) For every $f \in D(L_{\vec{\mu},\vec{\nu}})$ the process
$$f(X(t)) - f(X(0)) - \int_0^t L_{\vec{\mu},\vec{\nu}} f(X(s)) ds, \quad t \geqslant 0,$$
is a $\mathbb{P}$-martingale;
(b) $\mathbb{P}(X(0) = x) = 1$,

uniquely. In particular we may take $E = [0,1]$, $L_0 f = \frac{1}{2} f''$,

$$D(L_0) = \left\{ f \in C^2[0,1] : f'(0) = f'(1) = 0 \right\},$$

$\mu_k(I) = \int_{\alpha_k}^{\beta_k} 1_I(s) ds$, $\nu_k = 0$, $0 \leqslant \alpha_k < \beta_k \leqslant 1$, $1 \leqslant k \leqslant n$. Then L_0 generates the Feller semigroup of *reflected Brownian motion*: see Liggett [**86**], Example 5.8, p. 45. For the operator $L_{\vec{\mu},\vec{\nu}}$ the martingale problem is uniquely (but not maximally uniquely) solvable. However it does not generate a Feller semigroup.

From the result in Theorem 3.45 this can be seen as follows. Define the functionals $\Lambda_j : D(L_0) \to \mathbb{C}$, $1 \leqslant j \leqslant n$, as follows:

$$\Lambda_j(f) = \int L_0 f d\mu_j - \int f d\nu_j, \quad 1 \leqslant j \leqslant n.$$

We may and do suppose that the functionals Λ_j, $1 \leqslant j \leqslant n$, are linearly independent and that their linear span does not contain linear combinations of Dirac measures. The latter implies that, for every $x_0 \in E$ and for every function $u \in D(L_0)$, the convex subsets

$$D\left(L_1\right) \cap \left\{\left\{g \in C_0(E) : \operatorname{Re} g \leqslant \operatorname{Re} g(x_0)\right\} + u\right\} \quad \text{and}$$

$$D(L_1) \cap \{\{h \in C_0(E) : \text{Re } h \geqslant \text{Re } h(x_0)\} + u\}$$

are non-void. The latter follows from a Hahn-Banach argument. Hopefully, it will also imply that the quantities in (3.322) and (3.323) coincide. Since $D(L_0^2)$ forms a core for L_0 we may choose functions u_k, $1 \leqslant k \leqslant n$, such that $\Lambda_j(u_k) = \delta_{jk}$ and such that every u_k, $1 \leqslant k \leqslant n$ is in the vector of the two spaces

$$\left\{u \in D(L_0^2) : R(1)u \in D(L_1)\right\} \quad \text{and} \quad \left\{u \in D(L_0^2) : R(2)u \in D(L_1)\right\}.$$

As operator L_1 we take $L_1 = L_{\vec{\mu},\vec{\nu}}$ and for T we take $Tf = \sum_{j=1}^n \Lambda_j(f)u_k$, $f \in D(L_0)$.

The remainder of this section is devoted to the proof of the following result. Whenever appropriate we write $R(\lambda)$ for the operator $(\lambda I - L_0)^{-1}$.

3.110. Theorem. *Let L_0 be the generator of a Feller semigroup in $C_0(E)$ and let L_1 and T be linear operators with the following properties: the operator $I-T$ has domain $D(L_0)$ and range $D(L_1)$, L_1 verifies the maximum principle, the vector sum of the spaces $R(I-T)$ and $R(L_1(I-T))$ is dense in $C_0(E)$, and the operator $L_1(I-T)-(I-T)L_0$ can be considered as a continuous linear operator in the domain of L_0. More precisely, it is assumed that*

$$\limsup_{\lambda\to\infty} \left\| \left(L_1(I-T)-(I-T)L_0\right) R(\lambda)\right\| < 1.$$

Then there exists at most one linear extension L of the operator L_1 for which LT is bounded and that generates a Feller semigroup. In particular, if the martingale problem is solvable for L_1, then it is uniquely solvable for L_1.

Before we actually prove this result we like to make some comments. In order to have existence and uniqueness for the extension L on $R(T)$ it suffices that for every $v \in R(T)$ and for every $x_0 \in E$ the following two expressions are equal:

$$\lim_{\epsilon\downarrow 0} \inf_{f\in D(L_1)} \left\{ \text{Re } L_1 f(x_0) : \inf_{y\in E} \text{Re } \left(f(y)-v(y)\right) > \text{Re } \left(f(x_0)-v(x_0)\right)-\epsilon \right\}; \tag{3.322}$$

$$\lim_{\epsilon\downarrow 0} \sup_{f\in D(L_1)} \left\{ \text{Re } L_1 f(x_0) : \sup_{y\in E} \text{Re } \left(f(y)-v(y)\right) < \text{Re } \left(f(x_0)-v(x_0)\right)+\epsilon \right\}. \tag{3.323}$$

This common value is then by definition Re $L_2 v(x_0)$. The value of $L_2 v(x_0)$ is then given by $[L_2 v](x_0) = \text{Re } [L_2 v](x_0) - i\text{Re } [L_2(iv)](x_0)$ for $v \in R(T)$. Let Λ_j, $1 \leqslant j \leqslant n$, be as in the example of section 1. For every $x_0 \in E$ and for every $1 \leqslant k \leqslant n$ there exist functions g_k and $h_k \in D(L_0)$ with the following properties: $\Lambda_\ell(g_k) = \Lambda_\ell(h_k) = \delta_{k,\ell}$, Re $g_k(x) \leqslant$ Re $g_k(x_0)$ and Re $h_k(x_0) \leqslant h_k(x)$ for all $x \in E$ and Re $L_1(h_k - g_k)(x_0) = 0$. It then readily follows that the two expressions in (3.322) and (3.323) are equal for functions v in the linear span of $u_1, \dots, u_n$. Notice that the function Re h_k attains its minimum at x_0 and that the function Re g_k attains its maximum at x_0. In order to define $[L_1 u_k](x_0)$ we choose functions g_k and h_k with $\Lambda_\ell(g_k) = \text{Re } \Lambda_\ell(h_k) = -\delta_{k,\ell}$ in such a way that the function Re g_k attains its maximum at x_0 and that the function Re h_k attains its minimum at the same point x_0. Moreover we may and do suppose that Re $L_1(g_k - h_k)(x_0) = 0$. The value $[L_2 u_k](x_0)$ is then given by $[L_2 u_k](x_0) = [L_1(g_k + u_k)](x_0) = [L_1(h_k + u_k)](x_0)$.

Proof of Theorem 3.110. Let $\widetilde{L}$ be any linear operator which extends L_1 and that has the property that its domain $D\left(\widetilde{L}\right)$ contains $R(T) = TD(L_0)$. We also suppose that $\widetilde{L}$ verifies the maximum principle. Let L_1 be the restriction of $\widetilde{L}$ to $R(I-T)$ and let L_2 be the operator $\widetilde{L}$ confined to $R(T)$. We shall prove that the operator L_1 has a unique extension that generates a Feller semigroup. We start with the construction of a family of kind of intertwining operators

$\{V(\lambda) : \lambda > 0 \text{ and large}\}$. This is done as follows. The symbol $R(\lambda)$ is always used to denote the operator $R(\lambda) = (\lambda I - L_0)^{-1}$. Define the operator V by

$$V = L_1(I-T) - (I-T)L_0 \tag{3.324}$$

and define the operator $V(\lambda)$, $\lambda > \|L_2 T\|_\beta$, via the equality

$$V(\lambda) = \lambda\,(\lambda I - L_2 T)^{-1}\,V. \tag{3.325}$$

Then we have:

$$(\lambda I - L_1)\,(I-T) + (\lambda I - L_2)\,T\frac{V(\lambda)}{\lambda} = (I-T)\,(\lambda I - L_0 - V(\lambda))\,. \tag{3.326}$$

An equivalent form of (3.326) is the equality

$$\begin{aligned}
&(\lambda I - L_1)\,(I-T) + (\lambda I - L_2)\,T\frac{V(\lambda)}{\lambda} = (I-T)\,((\lambda I - L_0) - V(\lambda)) \qquad (3.327)\\
&= (I-T)\,(\lambda I - L_0) - (I-T)V(\lambda) = (I-T)\,(I - V(\lambda)R(\lambda))\,(\lambda I - L_0)\,.
\end{aligned}$$

Next we shall prove that the martingale problem is solvable for L_1. We do this by showing that the operator L_1 extends to a generator L of a Feller semigroup. For large positive lambda we define the operators $G(\lambda)$ in $C_0(E)$ as follows. For f of the form $f = (I-T)g$, with $g = (I - V(\lambda)R(\lambda))(\lambda I - L_0)h$, we write

$$G(\lambda)f = G(\lambda)(I-T)g = \left(I - T + T\frac{V(\lambda)}{\lambda}\right)h. \tag{3.328}$$

and if the function f is of the form $f = (\lambda I - L_1)\,(I-T)g$ we write

$$G(\lambda)f = G(\lambda)(\lambda I - L_1)(I-T)g = (I-T)g. \tag{3.329}$$

If $(\lambda I - L_1)\,(I-T)g_1 = (I-T)g_2$, then, since $I-T$ is mapping attaining values in the domain of L_1, we see that $(I-T)g_1 - (I-T)g_2$ belongs to $D(L_1)$ and hence the following identities are mutually equivalent (we write $g_2 = (I - V(\lambda)R(\lambda))\,(\lambda I - L_0)\,h_2$):

$$\begin{aligned}
(I-T)g_1 &= \left(I - T + T\frac{V(\lambda)}{\lambda}\right)h_2;\\
(I-T)g_1 - (I-T)h_2 &= T\frac{V(\lambda)}{\lambda}h_2;\\
(\lambda I - L_1)\,((I-T)g_1 - (I-T)h_2) &= (\lambda I - L_1)\,T\frac{V(\lambda)}{\lambda}h_2;\\
(I-T)g_2 &= (\lambda I - L_1)\left(I - T + T\frac{V(\lambda)}{\lambda}\right)h_2;\\
(I-T)\,(I - V(\lambda)R(\lambda))\,(\lambda I - L_0)\,h_2 &= (\lambda I - L_1)\left(I - T + T\frac{V(\lambda)}{\lambda}\right)h_2;\\
(\lambda I - L_1)\,(I-T)h_2 + (\lambda I - L_2)\,T\frac{V(\lambda)}{\lambda}h_2 &= (\lambda I - L_1)\left(I - T + T\frac{V(\lambda)}{\lambda}\right)h_2.
\end{aligned} \tag{3.330}$$

Since $g_2 = (I - V(\lambda)R(\lambda))(\lambda I - L_0)h_2$ it follows that

$$\begin{aligned}
&\left(\lambda I - \widetilde{L}\right)\left((I-T)\left(g_1 - h_2\right) - T\frac{V(\lambda)}{\lambda}h_2\right) \\
&(\lambda I - L_1)\left((I-T)g_1 - (I-T)h_2\right) - (\lambda I - L_2)\,T\frac{V(\lambda)}{\lambda}h_2 \\
&= (I-T)g_2 - (\lambda I - L_1)\,(I-T)h_2 - (\lambda I - L_2)\,T\frac{V(\lambda)}{\lambda}h_2 \\
&= (I-T)\,(I - V(\lambda)R(\lambda))\,(\lambda I - L_0)\,h_2 - (\lambda I - L_1)\,(I-T)h_2 \\
&\qquad - (\lambda I - L_2)\,T\frac{V(\lambda)}{\lambda}h_2 \\
&= (\lambda I - L_1)\,(I-T)h_2 + (\lambda I - L_2)\,T\frac{V(\lambda)}{\lambda}h_2 - (\lambda I - L_1)\,(I-T)h_2 \\
&\qquad - (\lambda I - L_2)\,T\frac{V(\lambda)}{\lambda}h_2 = 0. \qquad (3.331)
\end{aligned}$$

Since the operator $\widetilde{L}$ verifies the maximum principle, it is dissipative, and so the zero space of $\lambda I - \widetilde{L}$ is trivial. We conclude from (3.331) the identity $TV(\lambda)R(\lambda)h_2 = (I-T)g_1 - (I-T)h_2$ and so the function $TV(\lambda)R(\lambda)h_2$ belongs to $D(L_1)$. Hence it follows that (3.330) is satisfied and consequently that the operator $G(\lambda)$ is well-defined. Next we pick h_1 and h_2 in the domain of L_0 and we write

$$f = \lambda(I-T)\,(\lambda I - L_0 - V(\lambda))\,h_2 + (\lambda I - L_1)\,(I-T)\,(h_1 - \lambda h_2)\,. \qquad (3.332)$$

A calculation will yield the following identities:

$$\begin{aligned}
G(\lambda)f &= (I-T)h_1 + TV(\lambda)h_2; \\
\lambda G(\lambda)f - f &= L_1(I-T)h_1 + L_2TV(\lambda)h_2 = \widetilde{L}\,(G(\lambda)f)\,. \qquad (3.333)
\end{aligned}$$

Consequently we get $\left(\lambda I - \widetilde{L}\right)G(\lambda)f = f$, for f of the form (3.332). Since we know $\|V(\lambda)R(\lambda)\|_\beta < 1$ and since, by assumption the subspace $R(I-T) + R(L_1(I-T))$ is dense in $C_0(E)$, it follows that the range $R(\lambda I - \widetilde{L})$ is dense for $\lambda > 0$, λ large. Since the operator $\widetilde{L}$ satisfies the maximum principle and since $\widetilde{L} = L_1(I-T) + L_2T$ it follows that the operator L that assigns to $G(\lambda)f$ the function $\lambda G(\lambda)f - f$, $f \in R(I-T) + R(L_1(I-T))$, is well defined and satisfies the maximum principle. Below we shall show that the family $\{G(\lambda) : \lambda > 0, \quad \lambda \text{ large}\}$ is a resolvent family indeed: see (3.337). The closure of its graph contains the graph

$$\{((I-T)h_1 + Th_2, L_1(I-T)h_1 + L_2Th_2) : h_1, h_2 \in D(L_0)\}\,.$$

Denote the operator with graph $\{(G(\lambda)f, \lambda G(\lambda)f - f) : f \in C_0(E)\}$ again by L. From the previous remarks it follows that the operator L verifies the maximum principle, $(\lambda I - L)G(\lambda)f = f$ for $f \in C_0(E)$ and that it is densely defined. The latter follows because its domain contains all vectors of the form

$$(I-T)f_1 + L_1(I-T)f_2 = (I-T)\,(f_1 + (I-T)L_1(I-T)f_2 + TL_0f_2) + TVf_2.$$

From a general argument it then follows that the operator L is the generator of a Feller semigroup: for more details see [**141**], Theorem 2.2 page 14. Next let h_1 and h_2 belong to $D(L_0)$. Then we have

$$\begin{aligned}
&\lambda \left\| G(\lambda)\big((I-T)h_1 + (\lambda I - L_1)(I-T)h_2\big)\right\|_\infty \\
&\quad = \lambda \left\| G(\lambda)(I-T)h_1 + (I-T)h_2\right\|_\infty
\end{aligned}$$

(the operator $\widetilde{L}$ is dissipative)

$$\begin{aligned}
&\quad \leqslant \left\| \left(\lambda I - \widetilde{L}\right)\left(G(\lambda)(I-T)h_1 + (I-T)h_2\right)\right\|_\infty \\
&\quad = \left\| (I-T)h_1 + (\lambda I - L_1)(I-T)h_2\right\|_\infty . \qquad (3.334)
\end{aligned}$$

Since the vector sum of the spaces $R(I-T)$ and $R(L_1(I-T))$ is dense it follows from (3.334) that the operator $G(\lambda)$ extends as a continuous linear operator to all of $C_0(E)$. Moreover it is dissipative in the sense that

$$\lambda \left\| G(\lambda)\right\| \leqslant 1. \qquad (3.335)$$

Next we prove that the operator $G(\lambda)$ is positive in the sense that $f \geqslant 0$, $f \in C_0(E)$, implies $G(\lambda)f \geqslant 0$. So let $f \in C_0(E)$ be non-negative. There exist sequences of functions (g_n) and (h_n) in the space $D(L_0)$, for which

$$f = \lim_{n\to\infty}\left((I-T)h_n + (\lambda I - L_1)(I-T)g_n\right).$$

Put $f_n = (I-T)h_n + (\lambda I - L_1)(I-T)g_n$. Then $\left(\lambda I - \widetilde{L}\right) G(\lambda) f_n = f_n$ (see (3.332) and (3.333)). Since the operator $\widetilde{L}$ verifies the maximum principle it follows that

$$\lambda \mathrm{Re}\ G(\lambda) f_n \geqslant \inf_{y\in E} \mathrm{Re}\ \left(\lambda I - \widetilde{L}\right) G(\lambda) f_n(y) = \inf_{y\in E} \mathrm{Re}\ f_n(y) \tag{3.336}$$

and hence $\mathrm{Re}\ \lambda G(\lambda) f = \mathrm{Re}\ \lim_{n\to\infty} \lambda G(\lambda) f_n \geqslant 0$. A similar argument will show that the operator $G(\lambda)$ sends real functions to real function and hence $G(\lambda)$ is positivity preserving. Next we prove that the family $\{G(\lambda) : \lambda > 0, \ \text{large}\}$ is a resolvent family. So let λ and μ be large positive real numbers. We want to prove the identity

$$G(\lambda) - G(\mu) - (\mu - \lambda) G(\mu) G(\lambda) = 0. \tag{3.337}$$

First pick the function $f \in D(L_0)$ and apply the operator in (3.337) to the function $(\lambda I - L_1)(I-T)f$ and employ identity $G(\lambda)\left(\lambda I - \widetilde{L}\right) f = f$, for f belonging to $D\left(\widetilde{L}\right)$ to obtain

$$\left(G(\lambda) - G(\mu) - (\mu-\lambda)G(\mu)G(\lambda)\right)(\lambda I - L_1)(I-T)f = 0.$$

The operator in (3.337) also sends functions in the space $R(I-T)$ to 0, because we may apply (3.333) to see that

$$\left(\mu I - \widetilde{L}\right)\left(G(\lambda) - G(\mu) - (\mu-\lambda)G(\mu)G(\lambda)\right)(I-T)f = 0$$

for $f \in D(L_0)$. Finally we show that the resolvent family $\{G(\lambda) : \lambda > 0 \ \text{large}\}$ is strongly continuous in the sense that $\lim_{\lambda\to\infty} \lambda R(\lambda) f = f$ for all $f \in C_0(E)$. Of course it suffices to prove this equality for a subset with a dense span. Next we consider $f \in D(L_0)$ and we estimate

$$\left\|(I-T)f - \lambda G(\lambda)(I-T)f\right\|_\infty$$

as follows:

$$\begin{aligned}
&\left\|(I-T)f - \lambda G(\lambda)(I-T)f\right\|_\infty \\
&\leqslant \frac{1}{\lambda}\left\|(\lambda I - L_1)\left((I-T)f - \lambda G(\lambda)(I-T)f\right)\right\|_\infty \\
&= \frac{1}{\lambda}\left\|(\lambda I - L_1)(I-T)f - \lambda(I-T)f\right\|_\infty = \frac{1}{\lambda}\left\|L_1(I-T)f\right\|_\infty.
\end{aligned} \tag{3.338}$$

Again this expression tends to zero. For brevity we write

$$F(\lambda) = \left(I - V(\lambda)R(\lambda)\right)^{-1} f.$$

For $f \in D(L_0)$ the following equalities are valid:

$$\begin{aligned}
&(\lambda G(\lambda) - I) L_1(I-T)f - L_1(I-T)f \\
&= \lambda^2 G(\lambda)(I-T)f - \lambda(I-T)f - L_1(I-T)f \\
&= \left\{\lambda^2(I-T)R(\lambda) + TV(\lambda)R(\lambda)^2 - \lambda(I-T)(I - V(\lambda)R(\lambda))\right\} F(\lambda) \\
&\qquad\qquad - L_1(I-T)f \\
&= \left\{(I-T)\lambda L_0 R(\lambda) + \lambda^2 TV(\lambda)R(\lambda)^2 + \lambda(I-T)V(\lambda)R(\lambda)\right\} F(\lambda)
\end{aligned}$$

$$\begin{aligned} &- L_1(I-T)f \\ \to &\left\{(I-T)L_0 + TV + (I-T)V\right\} f - L_1(I-T)f = 0. \end{aligned} \tag{3.339}$$

From (3.338) together with (3.339) we conclude that $\lim_{\lambda\to\infty}(\lambda G(\lambda)f - f) = 0$ for all in the span of $R(I-T)$ and $R(L_1(I-T))$. By assumption this span is dense and consequently the resolvent family $\{G(\lambda) : \lambda > 0,\ \lambda \text{ large}\}$ is strongly continuous. In order to conclude the proof of the existence result we choose f_1 and f_2 in the space $D(L_0)$ and we notice the following identities:

$$\begin{aligned} &G(\lambda)\left\{(I-T)\left(I - V(\lambda)R(\lambda)\right)(\lambda I - L_0)f_1 + (\lambda I - L_1)(I-T)f_2\right\} \\ &\qquad = (I-T)(f_1+f_2) + TV(\lambda)R(\lambda)f_1, \end{aligned}$$

and so the space $G(\lambda)C_0(E)$ contains the linear span of the spaces $R(I-T)$ and $R(TV(\lambda)R(\lambda))$. From the resolvent equation it is clear that the space $G(\lambda)C_0(E)$ does not depend on the variable λ. So we see that the space $G(\alpha)C_0(E)$ contains, for a given function $f \in D(L_0)$ the family $\{\lambda TV(\lambda)R(\lambda)f : \lambda \geqslant \alpha\}$. Hence the function $TVf = \lim_{\lambda\to\infty} \lambda TV(\lambda)R(\lambda)f$ belongs to the closure of the space $G(\alpha)C_0(E)$. Since $L_1(I-T)f = TVf + (I-T)\left(L_1(I-T)f + TL_0 f\right)$, for $f \in D(L_0)$, we conclude that the range of $L_1(I-T)$ is contained in the closure of $G(\alpha)C_0(E)$. Since the latter space also contains $R(I-T)$ it follows from the density of the space

$$R(I-T) + R\left(L_1(I-T)\right)$$

in $C_0(E)$ that the domain of the resolvent, i.e. $G(\alpha)C_0(E)$ is dense in $C_0(E)$. From the previous discussion it also follows that the operator which assigns to $G(\lambda)f$ the function $\lambda G(\lambda)f - f$ extends the operator L_1 restricted to $R(I-T)$. It is now also clear that the subspace $\{G(\alpha)f : f \in C_0(E)\}$ is dense and so it is clear that the there exists a Feller semigroup generated by the operator L with graph $\{(G(\alpha)f, \alpha G(\alpha)f - f) : f \in C_0(E)\}$.

For the uniqueness we proceed as follows. Let $\mathbb{P}_x^1$ and $\mathbb{P}_x^2$ be two solutions for the martingale problem. We define the family of operators $\{S(t) : t \geqslant 0\}$ as follows: $S(t)f(x) = \mathbb{E}_x^1 f(X(t) - \mathbb{E}_x^2 f(X(t))$ from the martingale property, it then follows that $S'(t)f = S(t)\widetilde{L}f$ for f belonging to the subspace $R(I-T) + R(L_1(I-T))$. Moreover we have $S(0)f(x) = 0$ for all functions $f \in C_0(E)$. Then we write (for f_1 and $f_2 \in D(L_0)$)

$$\begin{aligned} 0 &= \int_0^\infty \left(-\frac{\partial}{\partial t}\right) e^{-\lambda t} S(t) G(\lambda)\left((I-T)f_1 + TL_1(I-T)f_2\right) dt \\ &= \int_0^\infty e^{-\lambda t} S(t)\left(\lambda I - \widetilde{L}\right) G(\lambda)\left((I-T)f_1 + TL_1(I-T)f_2\right) dt \\ &= \int_0^\infty e^{-\lambda t} S(t)\left((I-T)f_1 + TL_1(I-T)f_2\right) dt. \end{aligned} \tag{3.340}$$

Consequently $S(t)(I-T)f_1 = S(t)TL_1(I-T)f_2 = 0$ for all functions f_1 and f_n in the space $D(L_0)$. We also have, upon using (3.340) the following equality:

$$\int_0^t S(\tau)L_1(I-T)f\,d\tau = S(t)(I-T)f - S(0)f = 0. \tag{3.341}$$

Since by assumption the sum of the vector spaces $R(I-T)$ and $R(L_1)$ is dense in the space $C_0(E)$, we conclude $S(t) \equiv 0$ and hence from a general result on uniqueness of the martingale problem, we finally obtain that $\mathbb{P}_x^1 = \mathbb{P}_x^2$ for all $x \in E$. For more details see Proposition 2.9 (Corollary p. 206 of Ikeda and Watanabe [**61**]). This completes the proof of Theorem 3.110. $\square$

Index

To See Part 2 download
Advanced stochastic processes: Part 2