Probability II

Stochastic Processes

Jürgen Potthoff
Universität Mannheim

. . . But what really enriches and enlivens things is that we deal with lots of σ–algebras, not just the one σ–algebra which is the concern of measure theory.

D. WILLIAMS
Probability with Martingales, (1991)

Preface

The present lecture notes are based on the pencil notes that I used over the last two decades or so at Mannheim university for the lecture *Stochastische Prozesse* (which recently has been renamed to *Wahrscheinlichkeitstheorie II — Stochastische Prozesse*). As the reader can see, the name of the lecture is in German, while the notes are in English. Why so? As a consequence of the *Bologna process*, which drastically reshaped the German university landscape, our department set up new Bachelor and Master programs, and it was decided that each semester we offer at least two Master courses in English. Therefore, eventually also the lecture on stochastic processes will be held in English, and so it is quite natural or maybe even necessary to have the lecture notes for this course in English, too.

The very first version of my private notes was based on H. Bauer's masterful book [3], and even though over the years the notes changed a lot, Bauer's influence is certainly still noticeable. Other important sources were the books [12, 15, 16, 28, 31].

It is a pleasant duty to thank all the students who participated in these lectures during the last two decades for their comments, and for pointing out my errors. I am especially grateful to the students of my courses in the spring terms 2013, 2016. Great thanks I owe to Johannes Berger, Thomas Deck, Oliver Falkenburg, Karen Greive, Nils Harberts, Daniel Heck, Annika Lang, Kerstin Lux, Markus Huggenberger, and Dimitri Schwab for their comments. I feel especially indebted to Florian Werner for his extraordinary effort in proofreading the first draft of this manuscript and for his many helpful suggestions.

Jürgen Potthoff
Mannheim, June 2016

Contents

Preface **i**

1 Introduction **1**

2 Monotone Class Theorem **4**

3 Construction of Stochastic Processes **7**
3.1 Definition of Stochastic Processes 7
3.2 Path Space 10
3.3 Canonical Construction 15
3.4 Kernels and Semigroups 18
3.5 Convolution Semigroups of Measures 25

4 Continuity of Paths **31**
4.1 Hölder Continuous Functions 32
4.2 The Kolmogorov–Chentsov–Theorem 32
4.3 Application to Brownian Motion, Wiener Space 38

5 Conditional Expectation **42**
5.1 Definition of Conditional Expectation 42
5.2 Properties of the Conditional Expectation 45
5.3 Conditional Densities 50

6 Filtrations and Stopping Times **52**
6.1 Filtrations 52
6.2 Stopping Times 55

7 Martingales **60**
7.1 Martingales 60
7.2 Optional Stopping Theorem 65
7.3 Doob's Inequalities 67
7.4 Doob's Downcrossing Inequality 71
7.5 Submartingale Convergence Theorem 74

8 Markov Processes **76**
8.1 The Simple Markov Property 76
8.2 Markov Property and Markovian Semigroups 80
8.3 The Universal Markov Property 82
8.4 Strong Markov Property and Feller Semigroups 86
8.5 Reflection Principle for Brownian Motion 90

9 Fine Structure of Brownian Motion **95**
9.1 Gaussian Processes 95
9.2 Brownian Motion as a Gaussian Process 98
9.3 Law of the Iterated Logarithm 100

10 Introduction to Stochastic Integration **108**
10.1 Non-Rectifiability of Brownian Paths 109
10.2 Stochastic Integral of Elementary Processes 111
10.3 Mean Square Construction 115
10.4 Itô's Formula — A Sketch 118

A Kolmogorov Extension Theorem **123**

B Construction of a Poisson Process **127**

C Borel–Cantelli Lemma **131**

D Hilbert Spaces and Riesz' Theorem **133**

E Fubini's Theorem for Conditional Expectations **135**

F Doob's Factorization Lemma **138**

G Some Banach Spaces of Functions **139**

H On the Laplace Transform **142**

Chapter 1

Introduction

Stochastic processes are models for systems developing in time which are subject to random influences. Obviously, phenomena of this type are quite ubiquitious: They may range from so trivial things like the popularity curves of politicians over stock market prizes to the level of a water reservoir or the temperature dynamics over a certain period at some location — for all these examples curves which could look like the one in figure 1.1 might appear on the pages of the local newspaper (most of the time with a lower resolution of the (horizontal) time–axis).

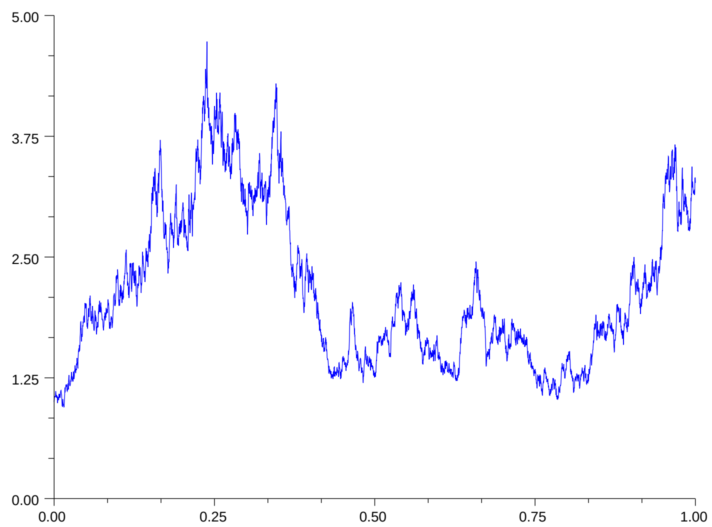

Figure 1.1: A path of a geometric Brownian motion

But the importance of the theory of stochastic processes goes well beyond its relevance to applications. The interaction with other fields of mathematics, such as differential geometry, functional analysis, (partial) differential equations, number theory, to name just a few, has been and still is intense, inspiring and fruitful. One important instance of this interaction is the central role that "the" Laplace operator plays in many domains of mathematics, and the fact that typically associated with a Laplace operator is a "Brownian motion", which gives a kind of "microscopic picture" of the Laplacian.

Indeed the observations of the Scottish botanist Robert Brown in 1827 stimulated the curiosity and fantasy of many researchers, and it took well into the second half of the 19th century until the first realistic ideas about this incessant, irregular motions of small particles in fluids were available. The first mathematical theories were due

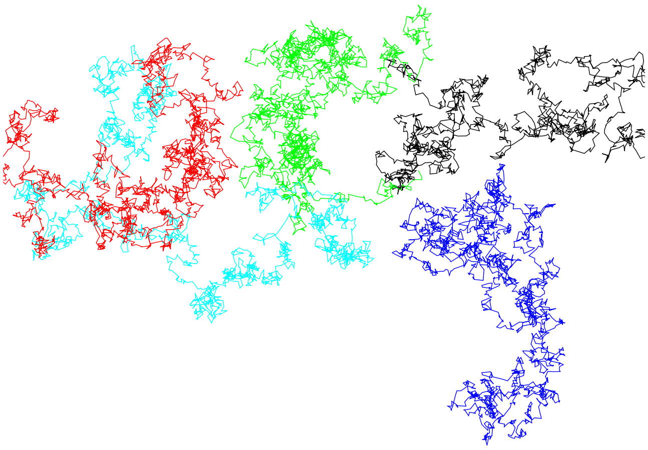

Figure 1.2: Five paths of a two-dimensional Brownian motion

to L. Bachelier (1900) and A. Einstein (1905), a construction and analysis with full mathematical rigor was published by N. Wiener in 1923. After this began, what M. Loève calls the "heroic period" of probability theory, and in particular of the theory of stochastic processes. It is connected with the names of A.N. Kolmogorov, P. Lévy, W. Doeblin, A.J. Khintchine, W. Feller, J. Doob, E.B. Dynkin, G.A. Hunt, D.B. Ray, K. Itô, H.P. McKean. Since the middle of the last century, there is a tremendous amount of activity in the field of stochastic processes, and the names of the mathematicians involved and their results are too numerous to be mentioned here.

The scope of applications includes practically all branches of (quantitative) science: Physics, chemistry, biology, statistics, economy, mathematical finance, engineering, medicine, social sciences, astronomy...

Nowadays one distinguishes three important classes of stochastic processes (not at all distinct — for example, Brownian motion belongs to all of them): *Lévy processes* — processes with independent, stationary increments, *Markov processes* — processes without memory, and *martingales* — processes modeling fair games. In these lectures we shall meet all three of them (of course, with more precise definitions than above).

The theory of stochastic processes is not an easy subject. Roughly speaking, the interplay between the underlying measure theory, which is based on countability, and the idea of time modeled by an uncountable set, entails a large number of serious technical difficulties. (But also in the case of a discrete time parameter, one runs into very challenging problems. For example, there are discrete-time Markov chains with very complicated behavior.) On the other hand, once these difficulties are mastered, one is rewarded with a rich, interesting theory, with which one can attack many important problems in applications.

Chapter 2

Monotone Class Theorem

The monotone class theorem (and its variants) are extremely useful for the reduction of calculations with complicated sets to computations with simple sets. For example, if one wants to show that two probability measures defined on a measurable space $(\Omega, \mathcal{A})$ are equal, then the monotone class theorem shows that it suffices to only prove equality on a $\cap$–stable generator of $\mathcal{A}$ (cf. theorem 2.5 below).

Throughout we suppose that Ω is a non-empty set.

2.1 Definition A non-empty family $\mathcal{S}$ of subsets of Ω is called a

(a) *π–system*, if $\mathcal{S}$ is $\cap$–stable;

(b) *d-system*, if the following hold true

(i) $\Omega \in \mathcal{S}$,
(ii) $A, B \in \mathcal{S}, A \subset B \Rightarrow B \setminus A \in \mathcal{S}$,
(iii) if $(A_n, n \in \mathbb{N})$ is a sequence in $\mathcal{S}$ which increases to A, then $A \in \mathcal{S}$.

2.2 Lemma *$\mathcal{S}$ is a d-system over Ω, if and only if the following hold true:*

(i) $\Omega \in \mathcal{S}$,

(ii) $A \in \mathcal{S} \Rightarrow \complement A \in \mathcal{S}$,

(iii) *for every pairwise disjoint sequence* $(A_n, n \in \mathbb{N})$ *in* $\mathcal{S}$, $\uplus_n A_n \in \mathcal{S}$.

Proof [1] Exercise! □

2.3 Theorem *$\mathcal{S}$ is a σ–algebra over Ω, if and only if $\mathcal{S}$ is a π–system and a d-system.*

Proof Exercise! □

[1] The end of a proof is indicated by the symbol □.

2.4 Theorem (Monotone Class Theorem) *Suppose that $\mathcal{C}$ is a π–system, and that $\mathcal{S}$ is a d-system such that $\mathcal{C} \subset \mathcal{S}$. Then $\sigma(\mathcal{C}) \subset \mathcal{S}$ holds true.*

Proof
Step 1: As an *exercise* shows, the intersection of an arbitrary family of d-systems is again a d-system. Let $d(\mathcal{C})$ denote the d-system which is the intersection of all d-systems containing $\mathcal{C}$: it is hence the smallest d-system containing $\mathcal{C}$, called the *d-system generated by* $\mathcal{C}$. Therefore it is sufficient to prove that $\sigma(\mathcal{C}) \subset d(\mathcal{C})$. Actually, we shall even show that equality holds true. In view of theorem 2.3 this is true, if we can show that $d(\mathcal{C})$ is a π–system: Because then $d(\mathcal{C})$ is a σ–algebra containing $\mathcal{C}$, and hence $d(\mathcal{C}) \supset \sigma(\mathcal{C})$ — on the other hand $\sigma(\mathcal{C})$ is a d-system containing $\mathcal{C}$, and $d(\mathcal{C})$ is by definition the smallest such: $d(\mathcal{C}) = \sigma(\mathcal{C})$. That is, we want to prove that $d(\mathcal{C})$ inherits from $\mathcal{C}$ the stability under intersections. This we prove in the next two steps.

Step 2: We show that $d(\mathcal{C})$ is stable by intersections with sets in $\mathcal{C}$. To this end, define

$$\mathcal{D}_1 = \{B \in d(\mathcal{C}),\ \forall\, C \in \mathcal{C} : B \cap C \in d(\mathcal{C})\}.$$

An *exercise* shows that $\mathcal{D}_1$ is a d-system. On the other hand we have that $\mathcal{D}_1 \supset \mathcal{C}$, because $\mathcal{C}$ is $\cap$–stable. Consequently $d(\mathcal{C}) \subset \mathcal{D}_1$. But by construction of $\mathcal{D}_1$ we also have that $\mathcal{D}_1 \subset d(\mathcal{C})$, so that we get $\mathcal{D}_1 = d(\mathcal{C})$.

Step 3: We prove that $d(\mathcal{C})$ is $\cap$–stable. Set

$$\mathcal{D}_2 = \{B \in d(\mathcal{C}),\ \forall\, C \in d(\mathcal{C}) : B \cap C \in d(\mathcal{C})\}.$$

In another *exercise* the reader checks that $\mathcal{D}_2$ is a d-system. By construction, $\mathcal{D}_2 \subset d(\mathcal{C})$, and in step 2 we proved that $\mathcal{C} \subset \mathcal{D}_2$. Because $d(\mathcal{C})$ is the smallest d-system which contains $\mathcal{C}$ we obtain $d(\mathcal{C}) \subset \mathcal{D}_2$. Thus we proved that $\mathcal{D}_2 = d(\mathcal{C})$, which in turn shows that $d(\mathcal{C})$ is $\cap$–stable. $\square$

One of the typical applications of the monotone class theorem 2.4 is the following result.

2.5 Theorem *Let $(\Omega, \mathcal{A})$ be a measurable space, $\mathcal{C}$ a $\cap$–stable generator of $\mathcal{A}$, and let P_1, P_2 be two probability measures on $(\Omega, \mathcal{A})$ which coincide on $\mathcal{C}$. Then $P_1 = P_2$ holds true.*

Proof Define

$$\mathcal{S} = \{A \in \mathcal{A},\ P_1(A) = P_2(A)\}.$$

A by now routine *exercise* shows that $\mathcal{S}$ is a d-system, and by hypothesis $\mathcal{S}$ contains the π-system $\mathcal{C}$. By theorem 2.4 we obtain $\mathcal{S} \supset \sigma(\mathcal{C}) = \mathcal{A}$. Hence $P_1(A) = P_2(A)$ for all $A \in \mathcal{A}$. $\square$

2.6 Exercise Suppose that $(\Omega, \mathcal{A}, P)$ is a probability space, and that $\{\mathcal{F}_i,\ i \in I\}$ is an independent family of sub–σ–algebras of $\mathcal{A}$, where I is some index set. Show that for every $i_0 \in I$, the σ–algebras $\mathcal{F}_{i_0}$ and $\sigma(\mathcal{F}_i,\ i \in I,\ i \neq i_0)$ are independent. (Hint: Choose as an $\cap$–stable generator $\mathcal{C}$ of $\sigma(\mathcal{F}_i,\ i \in I,\ i \neq i_0)$ the family of sets of the form $A_{i_1} \cap \cdots \cap A_{i_n}$, $n \in \mathbb{N}$, $i_1, \ldots, i_n \in I \setminus \{i_0\}$, $A_{i_j} \in \mathcal{F}_{i_j}$, $j = 1, \ldots, n$. Define $\mathcal{D}$ as the family of all subsets A of $\mathcal{A}$, so that A and every $A_{i_0} \in \mathcal{F}_{i_0}$ are independent. Show that $\mathcal{D}$ is a d-system containing $\mathcal{C}$, therefore by theorem 2.4 containing $\sigma(\mathcal{C}) = \sigma(\mathcal{F}_i,\ i \in I,\ i \neq i_0)$.)

There are a number of theorems of the monotone class type for measurable functions (instead of sets as above). A relatively simple one, which will be sufficient for the purposes of these lectures is the following:

2.7 Theorem (Monotone Class Theorem for Functions) *Suppose that $\mathcal{H}$ is a vector space of real valued (bounded real valued, respectively) functions defined on a set E, such that*

(i) *the constant function* 1 *belongs to $\mathcal{H}$,*

(ii) *if $(f_n,\ n \in \mathbb{N})$ is an increasing sequence of non-negative functions in $\mathcal{H}$ which converges pointwise to a real valued (bounded real valued, respectively) function f, then $f \in \mathcal{H}$.*

If $\mathcal{H}$ contains all indicators 1_A of sets A in a π–system $\mathcal{C}$ of subsets of E, then $\mathcal{H}$ contains all real valued (respectively bounded real valued) functions which are $\sigma(\mathcal{C})/\mathcal{B}(\mathbb{R})$–measurable.

Proof Set

$$\mathcal{S} = \{A \subset E,\ 1_A \in \mathcal{H}\}.$$

Then by assumption $\mathcal{S}$ contains the π–system $\mathcal{C}$. It is a straightforward *exercise* to check that $\mathcal{S}$ is a d-system. Thus by theorem 2.4 $\mathcal{S} \supset \sigma(\mathcal{C})$, i.e., $\mathcal{H}$ contains all indicators 1_A with $A \in \sigma(\mathcal{C})$. Since $\mathcal{H}$ is a vector space, this implies that $\mathcal{H}$ contains all elementary $\sigma(\mathcal{C})$–functions (i.e., linear combinations of indicators of sets in $\sigma(\mathcal{C})$ with non-negative coefficients). Since every non-negative $\sigma(\mathcal{C})$–measurable function can be written as an increasing pointwise limit of $\sigma(\mathcal{C})$–elementary functions, and because $\mathcal{H}$ is stable under increasing limits of non-negative fucntions, it follows that all non-negative $\sigma(\mathcal{C})$–measurable functions belong to $\mathcal{H}$. Since any $\sigma(\mathcal{C})$–measurable function can be decomposed into its positive and negative parts, which are $\sigma(\mathcal{C})$–measurable, too, every $\sigma(\mathcal{C})$–measurable function belongs to $\mathcal{H}$. Clearly, the same argument applies to the case of bounded, measurable real valued functions. □

The reader who is interested in the more sophisticated versions of the monotone class theorem for functions is referred to [8, 14-I ff].

Chapter 3

Construction of Stochastic Processes

3.1 Definition of Stochastic Processes

Roughly speaking, a stochastic process is a family of random variables indexed by a totally ordered set taking values in the same measurable space. More precisely we make the following

3.1 Definition Suppose that $(\Omega, \mathcal{A}, P)$ is a probability space, that $(E, \mathcal{E})$ is a measurable space, and that T is a totally ordered set. An *E–valued stochastic process* is a family $X = (X_t,\ t \in T)$ of random variables X_t, $t \in T$, with values in E. $(E, \mathcal{E})$ is called the *state space* of X, T is called the *(time) parameter domain* of X.

For large parts of these lectures, it will be sufficient to make no additional assumptions about the state space other than that $(E, \mathcal{E})$ is a measurable space. However, sometimes it will be necessary to assume some underlying topological structure, that is, we need to assume that $(E, \mathcal{O})$ is a topological space with certain properties, and that $\mathcal{E}$ is its *Borel–σ–algebra*: $\mathcal{E} = \sigma(\mathcal{O})$. Hence in this case $\mathcal{E}$ is the smallest σ–algebra over E containing all open sets. A very nice class of topologial spaces, which is defined below (cf. 3.15) and which enters the hypothesis of *Kolmogorov's extension theorem* (theorem 3.16) are the *polish spaces*. Moreover, sometimes we shall need to have a linear structure on E, for example, in order to be able to discuss the increments of stochastic processes (i.e., *differences* of E–valued random variables). In this case, we shall simply choose $E = \mathbb{R}^d$, $d \in \mathbb{N}$, with Borel–σ–algebra $\mathcal{B}(\mathbb{R}^d)$.

The time parameter domain T will typically be chosen as $T = \mathbb{N}$, $T = \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $T = [0, 1]$, or $T = \mathbb{R}_+$.

3.2 Remark We shall often deal with finite subsets $\{t_1, t_2, \ldots, t_n\} \subset T$, $n \in \mathbb{N}$, of the time parameter domain T. Since T is totally ordered, all these finite subsets are totally ordered, and it will be convenient to make throughout the convention that this ordering is indicated by lexicographical order of the indices. That is, for the subset above we have $t_1 < t_2 < \ldots < t_n$.

One can look at a stochastic process from various points of view:

(i) directly as in definition 3.1 as a mapping from the set T into the space $\mathcal{L}^0((\Omega, \mathcal{A}, P); (E, \mathcal{E}))$ of E–valued random variables;

(ii) as a mapping

$$\begin{aligned} X : \Omega &\to E^T \\ \omega &\mapsto X.(\omega) \end{aligned}$$

from Ω into the *path space* E^T, i.e., the set of all mappings from T into E, by setting

$$X.(\omega)(t) = X_t(\omega), \qquad \omega \in \Omega,\ t \in T,$$

and for each $\omega \in \Omega$, the mapping $t \mapsto X_t(\omega)$ is called a *path* of X;

(iii) or as a mapping

$$\begin{aligned} X : T \times \Omega &\to E \\ (t, \omega) &\mapsto X_t(\omega). \end{aligned}$$

The measurability of these mapping will be discussed later on, and it turns out to be a non-trivial question. Each of these points of view has a certain advantage, and it pays off to keep them in mind simultaneously. As for the notation, for $X_t(\omega)$, $t \in T$, $\omega \in \Omega$, we shall also interchangeably use $X(t, \omega)$, or $X(t)(\omega)$, and often — as it is custom in probability — suppress the argument $\omega \in \Omega$.

3.3 Example (Finite Random Walk) Choose $E = \mathbb{Z}$, $T = \{0, 1, \dots, n\}$, $x \in \mathbb{Z}$, $p \in [0, 1]$. Heuristically the process $X = (X_0, \dots, X_n)$ is constructed as follows: The random walker starts at x, $X_0 = x$, and throws a coin which with probability p shows heads and with probability $1 - p$ shows tails. If heads appear the random walker makes a step to the right in $\mathbb{Z}$, otherwise she makes a step to the left. Then she does the same experiment again for the next step and so on, until n steps have been made. Thus

$$X_t = X_0 + Y_1 + \dots + Y_t, \qquad t = 1, \dots, n, \tag{3.1}$$

where $(Y_1, \dots, Y_n)$ are *iid* (independent, identically distributed) Bernoulli random variables with values ± 1, $P(Y_t = 1) = p$, $t = 1, \dots, n$.

In order to formalize this example, we obviously only need n iid Bernoulli random variables. For example, we could choose

$$\begin{aligned} \Omega &= \{-1, 1\}^n \ni \omega = (\omega_1, \dots, \omega_n), \quad \omega_k = \pm 1, \\ \mathcal{A} &= \mathcal{P}(\Omega), \\ P(\{\omega\}) &= p^{\nu(\omega)} q^{n-\nu(\omega)}, \quad q = 1 - p, \\ Y_t(\omega) &= \omega_t, \qquad t = 1, \dots, n, \end{aligned}$$

where $\nu(\omega)$ denotes the number of 1's in ω:

$$\nu(\omega) = \frac{1}{2}\big(n + \sum_{t=1}^{n} \omega_t\big).$$

In an *exercise* the reader is invited to check that this model together with equation (3.1) for the definition of the actual process X is indeed doing the job.

Note that we have for the increments of the process X

$$X_t - X_{t-1} = Y_t, \qquad t = 1, \dots, n,$$

and

$$P_{Y_t} = P_{Y_{t'}} = p\varepsilon_1 + (1-p)\varepsilon_{-1}.$$

That is, the random walk has independent increments whose law is *stationary*: it does not change in time. Later we shall take these properties as the definition of the important class of *Lévy processes* (with a little addition).

We carry out another construction of the finite random walk. This time we choose as the underlying sample space $\tilde{\Omega}$ the space of all paths of the finite random walk:

$$\begin{aligned}
\tilde{\Omega} &= \mathbb{Z}^{n+1} \ni \omega = (\omega_0, \omega_1, \dots, \omega_n), \quad \omega_0 = x,\ \omega_t \in \mathbb{Z},\ t = 1, \dots, n \\
\tilde{\mathcal{A}} &= \mathcal{P}(\tilde{\Omega}), \\
\tilde{X}_t(\omega) &= \omega_t, \qquad t = 0, 1, \dots, n,
\end{aligned}$$

and set

$$\tilde{P}(\{\omega\}) = \begin{cases} 0, & \exists\, t \in \{1, \dots, n\}\colon |\omega_t - \omega_{t-1}| \neq 1, \\ p^{(n+\omega_n - x)/2}(1-p)^{(n-\omega_n+x)/2}, & \text{otherwise.} \end{cases}$$

Again we invite the reader to check in an *exercise* that also this model formalizes the finite random walk, that is, that this stochastic process has the same law as the previous one: [1]

$$P_{X_0, X_1, \dots, X_n} = \tilde{P}_{\tilde{X}_0, \tilde{X}_1, \dots, \tilde{X}_n}. \qquad \diamond$$

We just saw that a heuristic model can have different mathematical realizations. The terminology to deal with this is the following:

3.4 Definition Let $X = (X_t,\ t \in T)$ be a stochastic process with values in $(E, \mathcal{E})$ defined on $(\Omega, \mathcal{A}, P)$, and let $X' = (X'_t,\ t \in T)$ be a stochastic process with values in $(E, \mathcal{E})$ defined on $(\Omega', \mathcal{A}', P')$.

(a) X and X' are called *equivalent* or *versions of each other*, if all their *finite dimensional distributions* coincide, that is, if for all $n \in \mathbb{N}$, $\{t_1, \dots, t_n\} \subset T$, and all $B \in \mathcal{E}^{\otimes n}$,

$$P\big((X_{t_1}, \dots, X_{t_n}) \in B\big) = P'\big((X'_{t_1}, \dots, X'_{t_n}) \in B\big) \tag{3.2}$$

holds.

[1] The end of an example is indicated by the symbol $\diamond$.

(b) Assume that $(\Omega, \mathcal{A}, P) = (\Omega', \mathcal{A}', P')$. X and X' are called *modifications of each other*, if

$$P(X_t = X'_t) = 1, \quad \text{for all } t \in T.$$

That is, if for every $t \in T$ there exists a P–null set N_t so that for all $\omega \in \complement N_t$, $X_t(\omega) = X'_t(\omega)$ holds true.

(c) Assume that $(\Omega, \mathcal{A}, P) = (\Omega', \mathcal{A}', P')$. X and X' are called *indistinguishable*, if

$$P(X_t = X'_t, \forall\, t \in T) = 1.$$

I.e., there exists a P–null set N so that for all $t \in T$ and all $\omega \in \complement N$ the equality $X_t(\omega) = X'_t(\omega)$ holds true.

3.5 Remark Of course, equality (3.2) is the same as

$$P_{X_{t_1}\otimes\cdots\otimes X_{t_n}}(B) = P'_{X'_{t_1}\otimes\cdots\otimes X'_{t_n}}(B), \tag{3.3}$$

where, as usual,

$$\begin{aligned} X_{t_1} \otimes \cdots \otimes X_{t_n} : \Omega &\to E^n \\ \omega &\mapsto \big(X_{t_1}(\omega), \ldots, X_{t_n}(\omega)\big). \end{aligned}$$

By theorem 2.5 we only need to check relation (3.2) or (3.3) for B being a Cartesian product of sets in $\mathcal{E}$, since the family of these sets forms (by definition) a $\cap$–stable generator of $\mathcal{E}^{\otimes n}$.

3.2 Path Space

In this section we study the path space of stochastic processes, and the related questions of σ-algebras and measurability in more detail. In particular, this will serve as basis for the canonical construction of stochastic processes which we shall carry out in the next section.

We begin with an extremely useful, well-known lemma from measure theory which will be proved here for the convenience of the reader. Recall that if T is a mapping from a non-empty set M into N, $(N, \mathcal{N})$ a measurable space, then $\sigma(T)$ is the smallest σ–algebra on M, so that T becomes $\sigma(T)/\mathcal{N}$–measurable. It is easy to see that $\sigma(T) = T^{-1}(\mathcal{N})$. More generally, let I be any index set, $\big((N_i, \mathcal{N}_i),\ i \in I\big)$ a family of measurable spaces, and consider a family $(T_i,\ i \in I)$ of mappings such that T_i is a mapping from M into N_i. Then $\sigma(T_i,\ i \in I)$ is by definition the smallest σ–algebra on M so that for every $i \in I$, T_i is $\sigma(T_i,\ i \in I)/\mathcal{N}_i$–measurable:

$$\sigma(T_i,\ i \in I) = \sigma\Big(\bigcup_{i\in I} T_i^{-1}(\mathcal{N}_i)\Big).$$

3.6 Lemma *Assume that M is non-empty set, $(N, \mathcal{N})$ a measurable space, $\mathcal{C}$ a generator of $\mathcal{N}$, i.e., $\mathcal{N} = \sigma(\mathcal{C})$. Let T be a mapping from M into N.*

(a) $\sigma(T) = \sigma\big(T^{-1}(\mathcal{C})\big)$.

(b) *If $\mathcal{M}$ is a σ–algebra over M, then T is $\mathcal{M}/\mathcal{N}$–measurable, if and only if $T^{-1}(\mathcal{C}) \subset \mathcal{M}$.*

Proof
(a): The inclusion $\sigma(T) \supset \sigma\big(T^{-1}(\mathcal{C})\big)$ is almost obvious: $\sigma(T) = T^{-1}(\mathcal{N}) \supset T^{-1}(\mathcal{C})$, and since $\sigma\big(T^{-1}(\mathcal{C})\big)$ is the smallest σ–algebra containing $T^{-1}(\mathcal{C})$, we get $\sigma(T) \supset \sigma\big(T^{-1}(\mathcal{C})\big)$. To prove the converse inclusion set

$$\mathcal{S} = \big\{A \subset N,\ T^{-1}(A) \in \sigma\big(T^{-1}(\mathcal{C})\big)\big\}.$$

It is an easy *exercise* to show that $\mathcal{S}$ is a σ–algebra over N. Moreover, by construction we have $\mathcal{S} \supset \mathcal{C}$. Since $\mathcal{N}$ is the smallest σ–algebra containing $\mathcal{C}$, we get $\mathcal{S} \supset \mathcal{N}$. Therefore $T^{-1}(\mathcal{N}) \subset \sigma\big(T^{-1}(\mathcal{C})\big)$.

(b): T is $\mathcal{M}/\mathcal{N}$–measurable, if and only if $T^{-1}(\mathcal{N}) \subset \mathcal{M}$. From (a) we have that $\sigma(T) = T^{-1}(\mathcal{N}) = \sigma\big(T^{-1}(\mathcal{C})\big)$. Thus if $T^{-1}(\mathcal{C}) \subset \mathcal{M}$, then $\sigma\big(T^{-1}(\mathcal{C}\big) \subset \mathcal{M}$, and hence T is $\mathcal{M}/\mathcal{N}$–measurable. The converse is trivial. □

3.7 Remark Since $\sigma(T) = T^{-1}(\mathcal{N}) = T^{-1}(\sigma(\mathcal{C}))$, statement (a) of the lemma can also be written as $T^{-1}(\sigma(\mathcal{C})) = \sigma(T^{-1}(\mathcal{C}))$.

3.8 Exercise Suppose that I is an index set, that L, M are non-empty sets, and that $\big((N_i, \mathcal{N}_i),\ i \in I\big)$ is a family of measurable spaces. Assume furthermore that S is a mapping from L into M, and that $(T_i,\ i \in I)$, is a family of mappings so that T_i, $i \in I$, maps M into N_i. Use lemma 3.6.(a) to prove that

$$S^{-1}\big(\sigma(T_i,\ i \in I)\big) = \sigma(T_i \circ S,\ i \in I). \tag{3.4}$$

From now on we fix a measurable space $(E, \mathcal{E})$, and an index set I. E^I denotes the set of all mappings from I into E. As explained in section 3.1, if I is a time parameter domain then E^I is the *path space* of E–valued stochastic processes. Thus we shall call elements $\xi : I \to E$ in E^I also *paths in* E.

For any subset J of I the projection π_I^J from E^I onto E^J is defined as follows

$$\begin{aligned} \pi_I^J : E^I &\to E^J, \\ \xi &\mapsto \pi_I^J \xi = \xi\big|_J, \end{aligned}$$

where $\xi|_J$ means the restriction of ξ to J. If $J = \{i\}$, $i \in J$, then we simply write π_I^i, and in this case we may and will identify $\pi_I^i \xi$ with $\xi(i)$, $\xi \in E^I$. We define a σ–algebra $\mathcal{E}^I$ over E^I as follows

$$\mathcal{E}^I = \sigma\big(\pi_I^i,\ i \in I\big). \tag{3.5}$$

Thus, by definition $\mathcal{E}^I$ is the smallest σ–algebra over E^I making all one dimensional projections π_I^i, $i \in I$, measurable.

Later it will be very useful to have different representations of this σ–algebra. To this end, we now introduce the family $\mathcal{Z}^I$ of *cylinder sets* on E^I: Let $\mathcal{I}_0$ denote the family of all finite subsets of I. Then $\mathcal{Z}^I$ is defined as the family of all subsets Z of E^I for which there exists $J \in \mathcal{I}_0$, $J = \{i_1, \ldots, i_n\}$ (lexicographically ordered), $n \in \mathbb{N}$, and $B_1, \ldots, B_n \in \mathcal{E}$, so that

$$\begin{aligned} Z &= (\pi_I^J)^{-1}\big(B_1 \times B_2 \times \cdots \times B_n\big), \\ &= \big\{\xi \in E^I,\, \xi(i_1) \in B_1, \ldots, \xi(i_n) \in B_n\big\}. \end{aligned} \tag{3.6}$$

Figure 3.1 shows three paths in blue which belong to the cylinder set defined by $J = \{i_1, i_2, i_3, i_4\}$ and $B_J = B_1 \times B_2 \times B_3 \times B_4$, while the red path does not belong to $(\pi_I^J)^{-1}(B_J)$. [2]

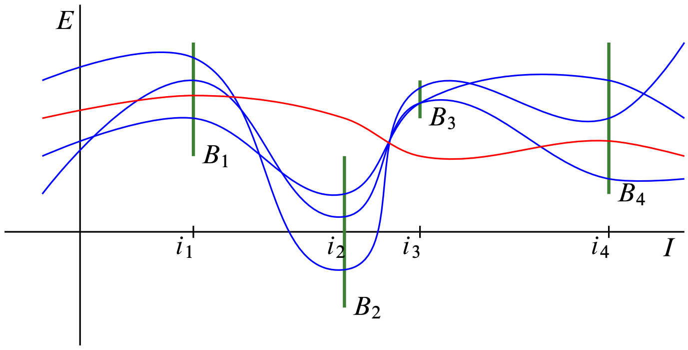

Figure 3.1: Three paths (blue) in the cylinder set $(\pi_I^J)^{-1}(B_J)$, the red path does not belong to $(\pi_I^J)^{-1}(B_J)$.

We claim that $\mathcal{Z}^I$ is a generator of $\mathcal{E}^I$ (which obviously is $\cap$–stable). To see this, we first remark that every $Z \in \mathcal{Z}^I$ belongs to $\mathcal{E}^I$, because Z of the form as in (3.6) can be written as follows

$$Z = (\pi_I^{i_1})^{-1}(B_1) \cap \cdots \cap (\pi_I^{i_n})^{-1}(B_n) \tag{3.7}$$

which obviously belongs to $\sigma(\pi_I^i,\, i \in I)$. On the other hand, by definition $\mathcal{E}^I$ is the smallest σ–algebra containing all sets of the form $(\pi_I^i)^{-1}(B)$ with $i \in I$, and $B \in \mathcal{E}$. But these sets are particular cylinder sets, and therefore $\sigma(\mathcal{Z}^I) \supset \mathcal{E}^I$. Hence our claim is proved, and from now on we shall also call $\mathcal{E}^I$ the *σ–algebra generated by the cylinder sets*.

[2]The reader should beware to take the illustration of cylinder sets of paths in figure 3.1 at face value. The "typical" paths in E^I are by far not as smooth as in the graphics — as a matter of fact they are extremely irregular.

3.9 Exercise Check that if I is finite, $|I| = n$, then $\mathcal{E}^I$ is just the usual product σ–algebra $\mathcal{E}^{\otimes n}$ of n copies of $\mathcal{E}$.

3.10 Exercise Denote by $\mathcal{C}^I$ the family of all subsets C of E^I of the following form: There exist $J \in \mathcal{J}_0$ and $B \in \mathcal{E}^J$ so that $C = (\pi_I^J)^{-1}(B)$. Obviously we have that $\mathcal{Z}^I \subset \mathcal{C}^I$. [3] Show that $\mathcal{C}^I$ is an algebra over E^I, i.e., it contains E^I, and it is $\cap$– as well as $\cup$–stable.

Our next claim is that for every $J \subset I$, the projection π_I^J is $\mathcal{E}^I/\mathcal{E}^J$–measurable. By lemma 3.6 we only have to prove that for every $Z \in \mathcal{Z}^J$ we have $(\pi_I^J)^{-1}(Z) \in \mathcal{E}^I$, because $\mathcal{Z}^J$ is a generator of $\mathcal{E}^J$. Such Z is of the form

$$Z = \big(\pi_J^K\big)^{-1}(B_1 \times \cdots \times B_k)$$

with $K = \{j_1, \ldots, j_k\} \subset J$, $B_1, \ldots, B_k \in \mathcal{E}$. But then

$$\begin{aligned}\big(\pi_I^J\big)^{-1}(Z) &= \big(\pi_J^K \circ \pi_I^J\big)^{-1}(B_1 \times \cdots \times B_k)\\ &= \big(\pi_I^K\big)^{-1}(B_1 \times \cdots \times B_k)\\ &\in \mathcal{Z}^I,\end{aligned}$$

and since $\mathcal{Z}^I \subset \mathcal{E}^I$, our claim is shown.

A particular case of the previous argument is that $J \in \mathcal{J}_0$, i.e., that π_I^J is a finite dimensional projection. Then we have derived that $\sigma(\pi_I^J,\ J \in \mathcal{J}_0) \subset \mathcal{E}^I$. On the other hand, the family of one dimensional projections, which generate $\mathcal{E}^I$, is a subfamily of $\{\pi_I^J,\ J \in \mathcal{J}_0\}$, and therefore we get equality:

$$\mathcal{E}^I = \sigma\big(\pi_I^J,\ J \in \mathcal{J}_0\big).$$

We can even go one interesting step further. Let $\mathcal{J}_1$ denote the family of all at most countable subsets of I, and let $J \in \mathcal{J}_1$. Above we have already proved that π_I^J is $\mathcal{E}^I/\mathcal{E}^J$–measurable. Hence for every $B \in \mathcal{E}^J$ we get $(\pi_I^J)^{-1}(B) \in \mathcal{E}^I$. Therefore we find $\mathcal{E}_\sigma^I \subset \mathcal{E}^I$, where

$$\mathcal{E}_\sigma^I = \big\{A \subset E^I,\ A = \big(\pi_I^J\big)^{-1}(B),\ J \in \mathcal{J}_1,\ B \in \mathcal{E}^J\big\}. \tag{3.8}$$

(Sometimes sets in $\mathcal{E}_\sigma^I$ are called *σ–cylinders*.) Now we prove that $\mathcal{E}_\sigma^I$ is a σ–algebra over E^I: $E^I \in \mathcal{E}_\sigma^I$ is obvious — we just have to choose any $J \in \mathcal{J}_1$, and $B = E^J$. If $A \in \mathcal{E}_\sigma^I$, say of the form as in (3.8), then

$$\complement A = \big(\pi_I^J\big)^{-1}\big(\complement B\big)$$

[3] In some part of the literature, also the family $\mathcal{C}^I$ is called the family of *cylinder sets*, e.g., [3]. We shall make systematic use of $\mathcal{C}^I$ in appendix A only.

which clearly belongs to $\mathcal{E}_\sigma^I$, because $\complement B \in \mathcal{E}^J$. Finally, let $(A_n,\, n \in \mathbb{N})$ be a sequence in $\mathcal{E}_\sigma^I$. Then we have for every $n \in \mathbb{N}$, $J_n \in \mathcal{J}_1$, and $B_n \in \mathcal{E}^{J_n}$ so that $A_n = (\pi_I^{J_n})^{-1}(B_n)$. Set $J = \cup_n J_n$, and note that $J \in \mathcal{J}_1$. Define

$$C_n = \left(\pi_J^{J_n}\right)^{-1}(B_n) \in \mathcal{E}^J, \qquad n \in \mathbb{N}.$$

Then

$$\left(\pi_I^J\right)^{-1}(C_n) = \left(\pi_J^{J_n} \circ \pi_I^J\right)^{-1}(B_n) = A_n.$$

Therefore we obtain

$$\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} \left(\pi_I^J\right)^{-1}(C_n) = \left(\pi_I^J\right)^{-1}\Big(\bigcup_{n\in\mathbb{N}} C_n\Big).$$

Thus $\cup_n A_n$ belongs to $\mathcal{E}_\sigma^I$, and the proof that $\mathcal{E}_\sigma^I$ is a σ–algebra over Ω is finished. On the other hand, $\mathcal{E}_\sigma^I$ contains all cylinder sets, and therefore $\mathcal{E}_\sigma^I \supset \mathcal{E}^I$. Hence $\mathcal{E}_\sigma^I = \mathcal{E}^I$. We wrap up our discussion in the following

3.11 Lemma *The following equalities hold true:*

$$\mathcal{E}^I = \sigma\big(\pi_I^i,\, i \in I\big) = \sigma\big(\pi_I^J,\, J \in \mathcal{J}_0\big) = \sigma\big(\mathcal{Z}^I\big) = \mathcal{E}_\sigma^I. \tag{3.9}$$

Consider the following situation, which is slightly more general than in the previous section: Ω is a non-empty set, $(E, \mathcal{E})$ is measurable space, I is an arbitrary set, and $(X_i,\, i \in I)$ is a family of mappings from Ω into E. Define a mapping X from Ω into E^I, by $X(\omega)(i) = X_i(\omega)$. For $J \in \mathcal{J}_0$ we shall write [4]

$$X_J = \pi_I^J \circ X = (X_{i_1}, \ldots, X_{i_n}) = X_{i_1} \otimes \cdots \otimes X_{i_n}, \quad J = \{i_1, \ldots, i_n\}. \tag{3.10}$$

As before, E^I is equipped with the σ–algebra $\mathcal{E}^I$.

3.12 Lemma *The following σ–algebras over Ω coincide with $\sigma(X) = X^{-1}(\mathcal{E}^I)$: $\sigma(X_i,\, i \in I)$, $\sigma(X_J,\, J \in \mathcal{J}_0)$, $\sigma(X^{-1}(\mathcal{Z}^I))$, $X^{-1}(\mathcal{E}_\sigma^I)$.*

Proof This follows directly from lemmas 3.6, 3.11, and exercise 3.8, if we observe that $\pi_I^J \circ X = X_J$, and in particular, $\pi_I^i \circ X = X_i$. □

It is worthwhile to write out the meaning of $\sigma(X^{-1}(\mathcal{Z}^I))$: If $Z \in \mathcal{Z}^I$, then Z is of the form (3.6). Thus

$$X^{-1}(Z) = \{\omega \in \Omega,\; X_{i_1}(\omega) \in B_1, \ldots, X_{i_n}(\omega) \in B_n\}. \tag{3.11}$$

Thus $\sigma(X)$ can be viewed as the smallest σ–algebra over Ω containing all sets of the form (3.11), which we shall also call *(X–) cylinder sets*.

[4] This is an attempt to bring in an extremely useful device from computer science, namely *operator overloading*. That is, to employ the same symbol with different meaning and different arguments depending on the context. Of course, this is somewhat dangerous and prone to misunderstandings. But I hope that the attentive reader will have no problem to understand the correct meaning within the given context — whenever there is danger of confusion, I will write everything out in detail.

3.3 Canonical Construction

In this section we address the question whether with stochastic processes, as defined in section 3.1 (cf. definition 3.1), we do not deal with the empty set, i.e., whether stochastic processes exist at all. In order to make this question more concrete and more precise, let us imagine that we have a heuristic stochastic process $\hat{X} = (\hat{X}_t,\, t \in T)$ with values in some measurable space $(E, \mathcal{E})$ that we observe. An *observation* of this process is given by determining where $\hat{X}$ is at finitely (different) many moments of time, namely it is given by a choice $n \in \mathbb{N}$, $t_1, t_2, \ldots, t_n \in T$, $B_{t_1}, B_{t_2}, \ldots, B_{t_n} \in \mathcal{E}$ so that

$$\hat{X}_{t_1} \in B_{t_1},\ \hat{X}_{t_2} \in B_{t_2},\ \ldots,\ \hat{X}_{t_n} \in B_{t_n}.$$

A mathematical model for these observations is readily constructed: We use the notation from the previous section with I, $\mathcal{I}_0$, $\mathcal{I}_1$ replaced by T, $\mathcal{T}_0$, $\mathcal{T}_1$ respectively, and set

$$\begin{aligned} J &= \{t_1, \ldots, t_n\}, \\ \Omega_J &= E^J \ni \omega = (\omega_{t_1}, \ldots, \omega_{t_n}), \\ \mathcal{A}_J &= \mathcal{E}^J, \\ X_{t_k}(\omega) &= \omega_{t_k}, \qquad k = 1, \ldots, n. \end{aligned}$$

Then the observation above becomes the event (and the $\hat{\cdot}$ is left out in order to distinguish the heuristic process (with $\hat{\cdot}$) and its mathematical model (without $\hat{\cdot}$))

$$\begin{aligned} &\big\{\omega \in \Omega_J,\ X_{t_k}(\omega) \in B_{t_k},\ k = 1, \ldots, n\big\} \\ &\qquad\qquad = \big\{\omega \in \Omega_J,\ \omega_{t_k} \in B_{t_k},\ k = 1, \ldots, n\big\}. \end{aligned}$$

Suppose that we have — still for fixed $n \in \mathbb{N}$, $J = \{t_1, \ldots, t_n\} \subset T$, but for varying $B_{t_1}, \ldots, B_{t_n} \in \mathcal{E}$ — enough observations to make statistics of these observations, leading us to the determination of a probability measure $P_J \equiv P_{\{t_1,\ldots,t_n\}}$ on $(\Omega_J, \mathcal{A}_J)$. Then we have a complete model for n observations at the times $t_1, \ldots, t_n$, and for $B = B_{t_1} \times \cdots \times B_{t_n}$,

$$\begin{aligned} P_J(B) &\equiv P_{\{t_1,\ldots,t_n\}}\big(B_{t_1} \times \cdots \times B_{t_n}\big) \\ &= P_{\{t_1,\ldots,t_n\}}\big(X_{t_1} \in B_{t_1}, \ldots X_{t_n} \in B_{t_n}\big) \end{aligned}$$

is the probability of the above observation. [5]

As we make more and more observations at various moments of time, we collect a family of probability spaces $\{(\Omega_J, \mathcal{A}_J, P_J),\ J \in \mathcal{T}_0\}$. Now we can make our question above more precise: Given a family of probability spaces $\{(\Omega_J, \mathcal{A}_J, P_J),\ J \in \mathcal{T}_0\}$ as above, is there *one* probability space $(\Omega, \mathcal{A}, P)$, and a stochastic process $X = (X_t,\, t \in T)$ defined thereon, such that we have

$$P\big(X_{t_k} \in B_{t_k},\ k = 1, \ldots, n\big) = P_J\big(B_{t_1} \times \cdots \times B_{t_n}\big) \tag{3.12}$$

[5] Since Cartesian products of n sets in $\mathcal{E}$ form a $\cap$-stable generator of $\mathcal{E}^J$, theorem 2.5 implies that indeed these observations uniquely determine the probability measure P_J.

for all $n \in \mathbb{N}$, $J = \{t_1, \dots, t_n\} \in \mathcal{T}_0$, and $B_{t_1}, \dots, B_{t_n} \in \mathcal{E}$?

The natural choice for the underlying sample space Ω and its σ–algebra $\mathcal{A}$ of events is the path space $(E^T, \mathcal{E}^T)$, which we have discussed in section 3.2.

On our basic measurable space $(E^T, \mathcal{E}^T)$ we define the following stochastic process (without yet having a probability thereon!):

3.13 Definition The family $X = (X_t, t \in T)$ of mappings

$$\begin{aligned} X_t : E^T &\to E \\ \omega &\mapsto X_t(\omega) = \omega_t \end{aligned}$$

is called the *canonical coordinate process* on $(E^T, \mathcal{E}^T)$.

It follows directly from lemma 3.11 for every $t \in T$, $X_t : E^T \to E$ is $\mathcal{E}^T/\mathcal{E}$–measurable. This justifies that we called the family $(X_t, t \in T)$ a "process" in definition 3.13. In terms of the canonical coordinate process X on $(E^T, \mathcal{E}^T)$ the desired relation (3.12) reads as follows

$$P\Big(\big\{\omega \in E^T, \omega_{t_k} \in B_{t_k}, k = 1, \dots, n\big\}\Big) = P_J\big(B_{t_1} \times \cdots \times B_{t_n}\big) \tag{3.13}$$

Consider again the family $\{P_J, J \in \mathcal{T}_0\}$ of probability measures on the measurable spaces $(\Omega_J, \mathcal{A}_J)$, coming from our observations of the heuristic model. This system of probability measures cannot be an arbitrary one: Let $J = \{t_1, \dots, t_n\}$, and assume that among the sets $B_{t_1}, \dots, B_{t_n}$ one is equal to E, say $B_{t_k} = E$. Then with (3.13) we get

$$\begin{aligned} &P_J(B_{t_1} \times \cdots \times B_{t_k} \times \cdots \times B_{t_n}) \\ &\quad = P\Big(\big\{\omega \in E^T, \omega_{t_1} \in B_{t_1}, \dots, \omega_{t_k} \in E, \dots, \omega_{t_n} \in B_{t_n}\big\}\Big) \\ &\quad = P\Big(\big\{\omega \in E^T, \omega_{t_1} \in B_{t_1}, \dots, \omega_{t_{k-1}} \in B_{t_{k-1}}, \omega_{t_{k+1}} \in B_{t_{k+1}}, \dots, \omega_{t_n} \in B_{t_n}\big\}\Big) \\ &\quad = P_{J\setminus\{t_k\}}(B_{t_1} \times \cdots \times B_{t_{k-1}} \times B_{t_{k+1}} \times \cdots \times B_{t_n}). \end{aligned}$$

Thus we obtain a consistency relation between the probability measures P_J and P_I on $(\Omega_J, \mathcal{A}_J)$, and on $(\Omega_I, \mathcal{A}_I)$ respectively, where $I = J \setminus \{t_k\}$. Next we have to systematize this kind of consistency relations.

Suppose that $I, J \in \mathcal{T}_0$ with

$$I = \{t_{i_1}, \dots, t_{i_m}\} \subset J = \{t_1, \dots, t_n\}.$$

Let

$$B_I = \bigtimes_{k=1}^{m} B_{t_{i_k}},$$

with $B_{t_{i_k}} \in \mathcal{E}, k = 1, \dots, m$. Then

$$\big(\pi_J^I\big)^{-1}(B_I) = \bigtimes_{k=1}^{n} C_{t_k},$$

where

$$C_{t_k} = \begin{cases} B_{t_k}, & \text{if } t_k \in I, \\ E, & \text{otherwise.} \end{cases}$$

Now consider the image $\pi^I_J P_J$ of the measure P_J under the measurable mapping π^I_J evaluated on B_I. With a similar computation as above, we obtain

$$\begin{aligned} \pi^I_J P_J(B_I) &= P_J\Big(\big(\pi^I_J\big)^{-1}(B_I)\Big) \\ &= P_J(C_{t_1} \times \cdots \times C_{t_n}) \\ &= P_{J \setminus I^c}(B_{t_{i_1}} \times \cdots \times B_{t_{i_m}}) \\ &= P_I(B_I). \end{aligned}$$

Thus P_I and $\pi^I_J P_J$ have to coincide on all Cartesian products of m sets in $\mathcal{E}$. Since by definition of $\mathcal{E}^I$ these sets form a generator which is $\cap$–stable, theorem 2.5 implies their equality on $\mathcal{E}^I$. Hence the condition we look for is given in the following definition.

3.14 Definition Assume that $\{(E^J, \mathcal{E}^J, P_J),\ J \in \mathcal{T}_0\}$ is a family of probability spaces. It is called *projective* if the *consistency condition* $\pi^I_J P_J = P_I$ holds true for all $I, J \in \mathcal{T}_0$ such that $I \subset J$.

We have already found that the projectivity of the family $\{(E^J, \mathcal{E}^J, P_J),\ J \in \mathcal{T}_0\}$ is a necessary condition for the existence of a probability measure P on $(E^T, \mathcal{E}^T)$ so that for every $J \in \mathcal{T}_0$ we have $\pi^J_T P = P_J$. For a special class of measurable spaces $(E, \mathcal{E})$, the central result below, theorem 3.16, states that this condition is also sufficient. First we make the

3.15 Definition A topological space $(E, \mathcal{O})$ is called *polish*, if it is metrizable (the topology $\mathcal{O}$ can be obtained from a metric), separable (it has a countable dense subset), and complete for a metric defining the topology (every Cauchy sequence for this metric converges in E with respect to this metric).

We continue to use the notation developed above.

3.16 Theorem (Kolmogorov) *Assume that $(E, \mathcal{O})$ is a polish space with Borel–σ–algebra $\mathcal{E} = \sigma(\mathcal{O})$. Suppose furthermore that $\{(E^J, \mathcal{E}^J, P_J),\ J \in \mathcal{T}_0\}$ is a projective family of probability spaces. Then there exists a unique probability measure P on $(E^T, \mathcal{E}^T)$ so that for every $J \in \mathcal{T}_0$, $\pi^J_T P = P_J$. The probability space $(E^T, \mathcal{E}^T, P)$ is called the* projective limit *of $\{(E^J, \mathcal{E}^J, P_J),\ J \in \mathcal{T}_0\}$.*

The proof of theorem 3.16 is deferred to appendix A.

3.17 Remark The following observation is quite useful: It suffices to check the consistency condition $\pi^I_J P_J = P_I$ for the case that $J \setminus I$ contains only one element.

Suppose that this is verified, and consider the general case $I \subset J$, $I, J \in \mathcal{T}_0$. Then there are elements $H_1, \dots, H_n \in \mathcal{T}_0$, such that

$$J \supset H_n \supset H_{n-1} \supset \cdots \supset H_1 \supset I$$

and each inclusion differs by one element. But then

$$\pi^I_J = \pi^I_{H_1} \circ \pi^{H_1}_{H_2} \circ \cdots \circ \pi^{H_{n-1}}_{H_n} \circ \pi^{H_n}_J .$$

Therefore

$$\begin{aligned}
\pi^I_J P_J &= \pi^I_{H_1} \circ \pi^{H_1}_{H_2} \circ \cdots \circ \pi^{H_{n-1}}_{H_n} \circ \pi^{H_n}_J P_J \\
&= \pi^I_{H_1} \circ \pi^{H_1}_{H_2} \circ \cdots \circ \pi^{H_{n-1}}_{H_n} P_{H_n} \\
&= \cdots = P_I .
\end{aligned}$$

3.18 Corollary *Let $(E, \mathcal{E})$ be as in the hypothesis of theorem 3.16, and assume that $\{(E^J, \mathcal{E}^J, P_J),\ J \in \mathcal{T}_0\}$ is a projective family of probability spaces. Then there is a stochastic process $X = (X_t,\ t \in T)$ defined on the projective limit $(E^T, \mathcal{E}^T, P)$, such that its finite dimensional distributions are given by the probability measures $(P_J,\ J \in \mathcal{T}_0)$.*

Proof As the process X we just choose the canonical coordinate process of definition 3.13. If then $J = \{t_1, \dots, t_n\} \in \mathcal{T}_0$ and $B_{t_1}, \dots, B_{t_n} \in \mathcal{E}$ are given, equation (3.13) reads

$$P\big(X_{t_k} \in B_{t_k},\ k = 1, \dots, n\big) = P_J\big(B_{t_1} \times \cdots \times B_{t_n}\big),$$

and the proof is done. □

3.19 Definition Suppose that $Y = (Y_t,\ t \in T)$ is a stochastic process with values in $(E, \mathcal{E})$ defined on $(\Omega', \mathcal{A}', P')$, and let its family of finite distributions on $\{(E^J, \mathcal{E}^J),\ J \in \mathcal{T}_0\}$ be denoted by $(P_J,\ J \in \mathcal{T}_0)$. Then X constructed from the family $(P_J,\ J \in \mathcal{T}_0)$ on $(E^T, \mathcal{E}^T, P)$ is called the *canonical version of Y*.

3.4 Kernels and Semigroups

Let us go back to our motivating gedankenexperiment of observations of a heuristic stochastic process. Clearly, the most basic type of observation is to consider a process which starts at a given point x, and to observe where it is located after a certain fixed amount of time has passed. That is, to model the probability that it has reached a certain set A. This kind of observation is formalized with the notion of a transition kernel which we define next.

3.20 Definition Assume that $(E_i, \mathcal{E}_i)$, $i = 1, 2$, are two measurable spaces. A mapping

$$K : E_1 \times \mathcal{E}_2 \to [0, +\infty]$$

is called a *kernel (from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$)* if

(i) for every $A \in \mathcal{E}_2$ the mapping

$$\begin{aligned} K(\,\cdot\,, A) : E_1 &\to [0, +\infty] \\ x &\mapsto K(x, A) \end{aligned}$$

is $\mathcal{E}_1/\mathcal{B}(\overline{\mathbb{R}}_+)$–measurable;

(ii) for every $x \in E_1$ the mapping

$$\begin{aligned} K(x, \,\cdot\,) : \mathcal{E}_2 &\to [0, +\infty] \\ A &\mapsto K(x, A) \end{aligned}$$

is a measure on $(E_2, \mathcal{E}_2)$.

The kernel K is called *(sub-) Markovian*, if for all $x \in E_1$, $K(x, E_2) = 1$ (≤ 1) holds true. A Markovian kernel is also called a *transition kernel*.

3.21 Example Assume that $(E_1, \mathcal{E}_1) = (E_2, \mathcal{E}_2)$. Then

$$I(x, A) = \varepsilon_x(A), \qquad x \in E_1,\ A \in \mathcal{E}_1, \tag{3.14}$$

where ε_x is the Dirac measure in x, defines a Markovian kernel on $(E_1, \mathcal{E}_1)$ called the *unit kernel*. ◇

3.22 Example Let $E_1 = E_2 = \mathbb{R}$, $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{B}(\mathbb{R})$, then

$$K(x, A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-(y-x)^2/2}\, dy, \qquad x \in \mathbb{R},\ A \in \mathcal{B}(\mathbb{R}),$$

defines a Markovian kernel. ◇

3.23 Example Let $n \in \mathbb{N}$, and let $E_1 = E_2 = \{1, 2, \ldots, n\}$, $\mathcal{E}_1 = \mathcal{E}_2 = \mathcal{P}(E_1)$. Suppose that P is an $n \times n$ matrix with positive matrix elements p_{ij}, $i, j = 1, \ldots, n$. Then

$$K(i, A) = \sum_{j \in A} p_{ij}, \qquad i \in \{1, \ldots, n\},\ A \subset \{1, \ldots, n\},$$

defines a kernel. The kernel is Markovian, if and only if the matrix P is a stochastic matrix, i.e., all its row sums are equal to 1. ◇

From now on we shall always consider Markovian or transition kernels. Let K be a Markovian kernel from $(E_1, \mathcal{E}_2)$ to $(E_2, \mathcal{E}_2)$. Then K defines a mapping from the space of bounded measurable functions $\mathcal{M}_b(E_2, \mathcal{E}_2)$ on E_2 to $\mathcal{M}_b(E_1, \mathcal{E}_1)$ as follows: [6]

$$(Kf)(x) = \int_{E_2} K(x, dy)\, f(y), \qquad x \in E_1,\ f \in \mathcal{M}_b(E_2, \mathcal{E}_2). \tag{3.15}$$

(Observe that K is Markovian, if and only if $K1 = 1$.)

3.24 Exercise Use the monotone class theorem for functions, theorem 2.7, to prove that for all $f \in \mathcal{M}_b(E_2, \mathcal{E}_2)$, $x \mapsto (Kf)(x)$ is $\mathcal{E}_1/\mathcal{B}(\mathbb{R})$–measurable.

Via

$$(\mu K)(A) = \int_{E_1} d\mu(x)\, K(x, A), \qquad A \in \mathcal{E}_2,\ \mu \in \mathcal{P}(E_1, \mathcal{E}_1) \tag{3.16}$$

K defines a mapping from the space $\mathcal{P}(E_1, \mathcal{E}_1)$ of probability measures on $(E_1, \mathcal{E}_1)$ to $\mathcal{P}(E_2, \mathcal{E}_2)$.

We also can combine both operations: Suppose that $f \in \mathcal{M}_b(E_2, \mathcal{E}_2)$, and that $\mu \in \mathcal{P}(E_1, \mathcal{E}_1)$. Then we can set

$$(\mu K)(f) = \int_{E_2} (\mu K)(dx_2)\, f(x_2),$$

and μK is the probability measure defined by (3.16). Just as well we can define

$$\mu(Kf) = \int_{E_1} \mu(dx_1)\, (Kf)(x_1),$$

where Kf is given in (3.15). An *exercise* with theorem 2.7 shows that both expressions are equal, so that we can also write them as

$$\mu K f = \int_{E_1 \times E_2} \mu(dx_1)\, K(x_1, dx_2)\, f(x_2). \tag{3.17}$$

In the sequel we shall only consider the case where $(E_1, \mathcal{E}_1) = (E_2, \mathcal{E}_2) = (E, \mathcal{E})$. Suppose then that K_1 and K_2 are two Markovian kernels on $(E, \mathcal{E})$. Then we can compose them by

$$K_2 \circ K_1(x, A) = \int_E K_2(x, dy)\, K_1(y, A), \qquad x \in E,\ A \in \mathcal{E}. \tag{3.18}$$

[6] The reader might feel uneasy that I write the integrating measure to the left of the function to be integrated. There is nothing special about this or the usual opposite convention, and I will write such integrals in any of these ways, whichever is more convenient in the given situation. However, the way they are written here will turn out to yield quite intuitive formulae.

It is not hard to see (*exercise*) that $K_2 \circ K_1$ is again a Markovian kernel. Furthermore, the following hold true for $\mu \in \mathcal{P}(E, \mathcal{E})$, $f \in \mathcal{M}_b(E, \mathcal{E})$ (*exercise*, K_3 is another kernel on $(E, \mathcal{E})$):

$$\begin{aligned} I \circ K_1 &= K_1 \circ I = K_1 \\ (K_3 \circ K_2) \circ K_1 &= K_3 \circ (K_2 \circ K_1) \\ \mu(K_2 \circ K_1) &= (\mu K_2) K_1 \\ (K_2 \circ K_1) f &= K_2(K_1 f). \end{aligned}$$

We have already seen (e.g., example 3.22) that a kernel $K(x, A)$ might be given by an appropriate function $k : E \times E \to \mathbb{R}_+$, that is,

$$K(x, A) = \int_A k(x, y)\, d\mu(y), \tag{3.19}$$

where μ is some measure on $(E, \mathcal{E})$. (In example 3.22 we had $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and μ was just Lebesgue measure.) In this case the function k is called the *(transition) density (with respect to μ)* of the kernel K.

Now we bring in the time parameter, and we suppose throughout that the time parameter domain T is either $\mathbb{N}_0$ or $\mathbb{R}_+$. (Observe that in both cases T is a *commutative semigroup*, that is, a set with the commutative inner composition "$+$" and the neutral element "0".)

3.25 Definition Suppose that $(U_t,\, t \in T)$ is a family of Markovian kernels on $(E, \mathcal{E})$, such that for all $s, t \in T$, the *semigroup relation*

$$U_{s+t} = U_s \circ U_t \tag{3.20}$$

holds true. Then $(U_t,\, t \in T)$ is called a *Markovian semigroup of kernels*. If $U_0 = I$ (I is the unit kernel (3.14)), then the semigroup is called *normal*.

3.26 Remarks

1. From now on we shall always assume that semigroups of kernels $(U_t,\, t \in T)$ are normal: $U_0 = I$.

2. If $(U_t,\, t \in T)$ is a semigroup of kernels, it is automatically *commutative*: $U_t \circ U_s = U_s \circ U_t$, $s, t \in T$.

3. In the context of probability theory, equation (3.20) is called the *Chapman–Kolmogorov–equation*, and its explicit form is

$$U_{s+t}(x, A) = \int_E U_t(x, dy)\, U_s(y, A), \qquad x \in E,\ A \in \mathcal{E}. \tag{3.21}$$

Its obvious interpretation is: The probability to move from x to the set A in time $s + t$ is the same as the probability of first moving from x in time t to the infinitesimal volume element dy at y, and then from y to A in the remaining time s, where one has to "sum over all possible intermediate states" y.

4. It is easy to check that if the kernels U_t, $t \in T$, are determined by a family $(u(t,\cdot,\cdot),\ t \in T)$, of densities with respect to a measure μ on $(E,\mathcal{E})$, that the semigroup relation (3.20) follows if
$$\int_E u(s,x,z)\,u(t,z,y)\,d\mu(z) = u(s+t,x,y) \tag{3.22}$$
holds for all $x, y \in E$, $s, t \in T$.

Suppose that $U = (U_t,\ t \in T)$ is a normal semigroup of kernels on $(E,\mathcal{E})$. Recall that $\mathcal{T}_0$ denotes the family of all finite subsets of T, ordered in the natural way, and we shall write $\mathcal{T}_{0+}$ for those $J = \{t_1,\dots,t_n\} \in \mathcal{T}_0$ so that $t_1 > 0$. In the sequel we put $t_0 = 0$. We generalize the operations in equations (3.15), (3.16), and (3.17) in the following way. Let $J = \{t_1,\dots,t_n\} \in \mathcal{T}_{0+}$, $\mu \in \mathcal{P}(E,\mathcal{E})$, and $f \in \mathcal{M}_b(E^{n+1},\mathcal{E}^{\otimes(n+1)})$. Set [7]
$$\begin{aligned}\mu U_J f = \int_{E^{n+1}} &\mu(dx_0)\,U_{t_1}(x_0,dx_1)\,U_{t_2-t_1}(x_1,dx_2)\\ &\cdots U_{t_n-t_{n-1}}(x_{n-1},dx_n)\,f(x_0,x_1,\dots,x_n),\end{aligned} \tag{3.23}$$
where *a priori* the multiple integral on the right hand side is defined by starting with the integration with respect to x_n, then the integration with respect to x_{n-1} is done, and so on. (Here one needs one of the statements of the Fubini–Tonelli–theorem, namely that a measurable function of several variables is also measurable function of some of the variables, the others being fixed.) As in an exercise above, one can then use theorem 2.7 to show that any order of integration yields the same result, so that we can omit putting brackets.

If $J = \emptyset$, i.e., $n = 0$, we interpret the right hand side of (3.23) as
$$\int_E \mu(dx_0)\,f(x_0)$$
The special case $\mu = \varepsilon_x$, $x \in E$, reads as follows
$$\begin{aligned}\varepsilon_x U_J f = \int_{E^n} &U_{t_1}(x,dx_1)\,U_{t_2-t_1}(x_1,dx_2)\\ &\cdots U_{t_n-t_{n-1}}(x_{n-1},dx_n)\,f(x,x_1,\dots,x_n).\end{aligned} \tag{3.24}$$

Choose $J = \{t_1,\dots,t_n\} \in \mathcal{T}_{0+}$, $n \in \mathbb{N}$, $B_{t_0}, B_{t_1}, \dots, B_{t_n} \in \mathcal{E}$ with $B_{t_k} = E$ for some $k \in \{1,\dots,n\}$. Then
$$\begin{aligned}&\mu U_J 1_{B_{t_0}\times\cdots\times B_{t_n}}\\ &\quad= \int_{B_{t_0}\times B_{t_1}\times\cdots\times B_{t_n}} \mu(dx_0)\,U_{t_1}(x_0,dx_1)\cdots U_{t_n-t_{n-1}}(x_{n-1},dx_n).\end{aligned}$$

[7] Here is another instance of "operator overloading": This notation – even though formally not completely correct — will have a nice payoff later on.

As mentioned above, we can do the k-th integration first, and the corresponding term in the above integral is given by

$$\begin{aligned}\int_E U_{t_k-t_{k-1}}(x_{k-1}, dx_k)\, U_{t_{k+1}-t_k}(x_k, dx_{k+1}) \\ &= U_{t_k-t_{k-1}} \circ U_{t_{k+1}-t_k}(x_{k-1}, dx_{k+1}) \\ &= U_{t_{k+1}-t_{k-1}}(x_{k-1}, dx_{k+1}),\end{aligned}$$

where we used the semigroup property of U. Inserting this above, we obtain

$$\mu U_J 1_{B_{t_0}\times\cdots\times B_{t_n}} = \mu U_{J\setminus\{t_k\}} 1_{B_{t_0}\times\cdots\times \cancel{B_{t_k}}\times\cdots\times B_{t_n}}.$$

Thus, if for

$$B = B_{t_0} \times B_{t_1} \times \cdots \times B_{t_n}, \quad B_{t_0}, \ldots, B_{t_n} \in \mathcal{E}, \tag{3.25a}$$

we define

$$P_J^\mu(B) = \mu U_J 1_B, \tag{3.25b}$$

then — in view of remark 3.17 — the family $(P_J^\mu,\ J \in \mathcal{T}_0)$ is a projective family. With theorem 3.16, and corollary 3.18 we obtain

3.27 Theorem *Suppose that $(E, \mathcal{E})$ is a polish space, that μ is a probability measure on $(E, \mathcal{E})$, and that $U = (U_t,\ t \in T)$ is a semigroup of normal Markovian kernels on $(E, \mathcal{E})$. Then equations* (3.25) *define a projective family of probability measures on $((E^J, \mathcal{E}^J),\ J \in \mathcal{T}_0)$, and therefore a unique probability measure P^μ on $(E^T, \mathcal{E}^T)$. The measures P_J^μ, $J \in \mathcal{T}_0$, are the finite dimensional distributions of the canonical stochastic process $X = (X_t,\ t \in T)$ defined in corollary 3.18.*

Thus, with the above notation, we have

$$\begin{aligned}P^\mu\big(X_{t_0} \in B_{t_0}, X_{t_1} \in B_{t_1}, \ldots, X_{t_n} \in B_{t_n}\big) &= \mu U_J 1_B \\ &= \int_{B_{t_0}\times B_{t_1}\times\cdots\times B_{t_n}} \mu(dx_0)\, U_{t_1}(x_0, dx_1)\, U_{t_2-t_1}(x_1, dx_2) \\ &\qquad \times \cdots \times U_{t_n-t_{n-1}}(x_{n-1}, dx_n).\end{aligned} \tag{3.26}$$

In particular, the canonical process X has its initial law given by μ, and if we set $J = \{t_1\}$, $\mu = \varepsilon_x$, $x \in E$, $B_{t_0} = E$, we find

$$P_x\big(X_{t_1} \in B_{t_1}\big) = U_{t_1}(x, B_{t_1}), \tag{3.27}$$

i.e., U_t, $t > 0$, is its transition kernel. Here and in what follows we write P_x for P^{ε_x}. Finally, with a routine application of the monotone class theorem 2.7 (*exercise!*), we obtain for every $f \in \mathcal{M}_b(E^{n+1}, \mathcal{E}^{\otimes(n+1)})$

$$\begin{aligned}E^\mu\big(f(X_{t_0}, X_{t_1}, \ldots, X_{t_n})\big) &= \mu U_J f \\ &= \int \mu(dx_0)\, U_{t_1}(x_0, dx_1)\, U_{t_2-t_1}(x_1, dx_2) \\ &\qquad \times \cdots \times U_{t_n-t_{n-1}}(x_{n-1}, dx_n)\, f(x_0, x_1, \ldots, x_n),\end{aligned} \tag{3.28}$$

where E^{μ} denotes expectation with respect to P^{μ}. In the case $\mu = \varepsilon_x$, $x \in E$, we also write E_x for E^{ε_x}.

3.28 Example (Brown–Semigroup) Let $T = \mathbb{R}_+$, $d \in \mathbb{N}$, $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, and for $t > 0$, $x, y \in \mathbb{R}^d$ set

$$u(t, x, y) = \frac{1}{(2\pi t)^{d/2}} \exp\Bigl(-\frac{|x-y|^2}{2t}\Bigr), \tag{3.29}$$

where $|\cdot|$ is the Euclidean norm in $\mathbb{R}^d$. As an *exercise* shows, one has

$$\int_{\mathbb{R}^d} u(s, x, z)\, u(t, z, y)\, dz = u(s + t, x, y), \qquad s, t > 0,\ x, y \in \mathbb{R}^d. \tag{3.30}$$

(It is of advantage to use Fourier transforms for this exercise.) Therefore $u(t, x, y)$, $t > 0$, x, $y \in \mathbb{R}^d$ is the density of a Markovian semigroup $U = (U_t,\ t \geq 0)$ of kernels, called the *Brown-semigroup*. The associated canonical process is called a *pre-Brownian motion* — the "pre" is used to indicate that the definition of stochastic processes called *Browian motion* usually includes a condition of path continuity, but the canonical process does not have this property. (However, we are going to discuss continuity of paths in the next chapter.)

In another *exercise* the reader is asked to prove that for $f \in C_b(\mathbb{R}^d)$, the function v, defined on $(0 + \infty) \times \mathbb{R}^d$ by

$$v(t, x) = U_t f(x), \qquad t > 0,\ x \in \mathbb{R}^d,$$

is a solution of the following Cauchy problem of the *heat equation*:

$$\partial_t v(t, x) = \frac{1}{2}\, \Delta v(t, x), \tag{3.31a}$$

$$\lim_{t \to 0,\, t > 0} v(t, x) = f(x), \tag{3.31b}$$

where Δ is the *Laplace operator* in $\mathbb{R}^d$:

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}.$$

(To prove (3.31a), again it pays off to use the Fourier transform.) Since by construction of the canonical process X we have that $U_t(x, dy)$ is the law of X_t with start in x, by the transformation theorem for Lebesgue integrals we get that

$$v(t, x) = E_x\bigl(f(X_t)\bigr). \tag{3.32}$$

Equations of the type (3.32) are often referred to as a *Feynman–Kac–formula*. ◇

3.29 Example (Ornstein–Uhlenbeck–Semigroup) For the time parameter domain T choose $\mathbb{R}_+$, for the state space $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and for $t > 0$ let the kernel U_t be defined by the following density

$$u(t, x, y) = \frac{1}{\sqrt{\pi(1 - e^{-2t})}} \exp\Big(-\frac{\big(y - e^{-t}x\big)^2}{1 - e^{-2t}}\Big). \tag{3.33}$$

An exercise (somewhat tedious — this time Fourier transforms do not help) shows the semigroup property (3.30) with $d = 1$ holds for this density, too. The associated semigroup and canonical process are the *Ornstein–Uhlenbeck–semigroup* and the *pre–Ornstein–Uhlenbeck–process*. (The "pre" is for the same reason as above.)

For $f \in C_b(\mathbb{R})$,

$$v(t, x) = U_t f(x) = E_x\Big(f(X_t)\Big), \qquad t > 0,\ x \in \mathbb{R}$$

solves

$$\partial_t v(t, x) = \frac{1}{2}\,\partial_x^2 v(t, x) - x\partial_x v(t, x), \tag{3.34a}$$

$$\lim_{t\to 0,\, t>0} v(t, x) = f(x). \tag{3.34b}$$

as the reader shows in an *exercise*. ◇

3.30 Example (Poisson–Semigroup) Let $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $T = \mathbb{R}_+$, $\lambda > 0$, and consider kernels

$$U_t(x, B) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!}\, \varepsilon_{x+n}(B), \qquad t \in \mathbb{R}_+,\ x \in \mathbb{R}. \tag{3.35}$$

(Note that by definition $U_0 = \varepsilon_0$.) Then $U = (U_t,\ t \in \mathbb{R}_+)$ is a normal Markovian semigroup of kernels (*exercise*), which is called the *Poisson–semigroup (with parameter* λ), the associated canonical process is called the *pre–Poisson–process*. (Again, the "pre" is for the same reason as before.)

It is easy to check that $v(t, x) = U_t f(x)$, $t \geq 0$, $x \in \mathbb{R}$, $f \in C_b(\mathbb{R})$, solves

$$\partial_t v(t, x) = \lambda\big(v(t, x + 1) - v(t, x)\big), \tag{3.36a}$$

$$\lim_{t\to 0,\, t>0} v(t, x) = f(x). \tag{3.36b}$$

Details are left to an *exercise*. ◇

3.5 Convolution Semigroups of Measures

The examples 3.28 and 3.30 have a special important feature which we are going to analyze in this section. To this end, we need that the state space E has additionally a linear structure (actually a notion of translation would be enough), and hence for simplicity we shall assume for the remainder of this chapter that $E = \mathbb{R}^d$, $d \in \mathbb{N}$, $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$. But notationally it will be convenient to occasionally continue to write E and $\mathcal{E}$.

For $a \in \mathbb{R}^d$ let τ_a denote translation by a: $x \mapsto \tau_a x = x + a$, $x \in \mathbb{R}^d$. Also, if $B \subset \mathbb{R}^d$, $\tau_a B = \{\tau_a x,\ x \in B\}$. It is clear that for every $a \in \mathbb{R}^d$, τ_a is a measurable bijection of $\mathbb{R}^d$ onto itself.

3.31 Definition A kernel K on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is called *invariant under translations* if $K(\tau_a x, \tau_a B) = K(x, B)$ holds for all $a, x \in \mathbb{R}^d$, $B \in \mathcal{B}(\mathbb{R}^d)$.

Suppose that μ is a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, and that f is a bounded, measurable function. Then the transformation theorem for Lebesgue integrals immediately yields

$$\int f(y)\, d(\tau_a\mu)(y) = \int f(a+y)\, d\mu(y). \tag{3.37}$$

3.32 Lemma *Every Markovian kernel K on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ which is invariant under translations is in one-to-one correspondence with a probability measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.*

Proof If K is a translation invariant kernel, then $\mu = K(0, \cdot)$ uniquely defines a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Conversely, if μ is a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, then

$$K(x, B) = (\tau_x\mu)(B), \qquad x \in \mathbb{R}^d,\ B \in \mathcal{B}(\mathbb{R}^d), \tag{3.38}$$

defines a kernel thereon, where — as usual — $\tau_x\mu = \mu \circ \tau_x^{-1}$ is the image of μ under τ_x (*exercise* (hint: write $\tau_x\mu(B) = \int 1_B(x+y)\, d\mu(y)$, and use Fubini's theorem)). □

3.33 Exercise Suppose that K_i, $i = 1, 2$, are two translation invariant Markovian kernels on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ associated with the probability measures μ_i, $i = 1, 2$. Show that $K_1 \circ K_2$ corresponds to the measure $\mu_1 * \mu_2$.

3.34 Example The kernel U_t, $t > 0$, of the Brown–semigroup in example 3.28 is translation invariant, and its associated probability measure μ_t is given by

$$\mu_t(B) = (2\pi t)^{-d/2} \int_B e^{-|y|^2/2t}\, dy. \tag{3.39}$$

The reader takes care of the details in an *exercise*. ◇

3.35 Example The kernel U_t, $t > 0$, of the Poisson–semigroup, formula (3.35), is translation invariant, and its associated probability measure μ_t is given by

$$\mu_t = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^n}{n!}\, \varepsilon_n. \tag{3.40}$$

Again the details are left to the reader. ◇

3.36 Definition A family $(\mu_t,\ t \in T)$ of probability measures is called a *convolution semigroup* of probability measures, if for all $s, t \in T$, the equality $\mu_s * \mu_t = \mu_{s+t}$ holds true.

3.37 Remark One can show that for every convolution semigroup $(\mu_t,\ t \in \mathbb{R}_+)$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ one has $\mu_0 = \varepsilon_0$.

Lemma 3.32, together with exercise 3.33, implies

3.38 Corollary *Every semigroup of Markovian kernels* $(U_t,\ t \in T)$ *on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ *which is invariant under translations is via equation* (3.38) *in one-to-one correspondence with a convolution semigroup* $(\mu_t,\ t \in T)$ *of probability measures.*

Next we shall compute the expression of $\mu U_J f$ as in (3.23) for the case that the semigroup $(U_t,\ t \in T)$ is translation invariant and corresponds to $(\mu_t,\ t \in T)$: $U_t(x, dy) = \tau_x \mu_t(dy)$, $x,\ y \in \mathbb{R}^d$, $t \in T$. With (3.37) for the integration with respect to x_1 we find

$$\begin{aligned}\mu U_J f &= \int \mu(dx_0)\, \tau_{x_0}\mu_{t_1}(dx_1)\, \tau_{x_1}\mu_{t_2-t_1}(dx_2) \\ &\qquad\qquad \times \cdots \times \tau_{x_{n-1}}\mu_{t_n-t_{n-1}}(dx_n)\, f(x_0, x_1, \ldots, x_n) \\ &= \int \mu(dx_0)\, \mu_{t_1}(dx_1)\, \tau_{x_0+x_1}\mu_{t_2-t_1}(dx_2) \\ &\qquad\qquad \times \cdots \times \tau_{x_{n-1}}\mu_{t_n-t_{n-1}}(dx_n)\, f(x_0, x_0+x_1, x_2, \ldots, x_n).\end{aligned}$$

Now we can do this successively for remaining integrations, and obtain

$$\begin{aligned}\mu U_J f = \int \mu(dx_0)\, \mu_{t_1}(dx_1)\, \mu_{t_2-t_1}(dx_2) \\ \times \cdots \times \mu_{t_n-t_{n-1}}(dx_n)\, f \circ \varphi(x_0, x_1, \ldots, x_n),\end{aligned} \tag{3.41a}$$

where

$$\varphi(x_0, x_1, \ldots, x_n) = (x_0, x_0 + x_1, \ldots, x_0 + x_1 + \cdots + x_n). \tag{3.41b}$$

Equation (3.41a) can be compactly written as follows

$$\mu U_J f = \int d(\mu \otimes \mu_{t_1-t_0} \otimes \mu_{t_2-t_1} \otimes \cdots \otimes \mu_{t_n-t_{n-1}})\, f \circ \varphi. \tag{3.42}$$

Obviously, the linear transformation φ from $\mathbb{R}^{d(n+1)}$ into itself corresponds to the (block-) matrix

$$\phi = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix}. \tag{3.43}$$

If, beginning with the last row, we subtract from a row the one above it we get the identity matrix. Hence ϕ is invertible, i.e., φ is a bijection, and the inverse of ϕ is given by

$$\phi^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}. \tag{3.44}$$

In other words,

$$\varphi^{-1}(x_0, x_1, \ldots, x_n) = (x_0, x_1 - x_0, x_2 - x_1, \ldots, x_n - x_{n-1}). \tag{3.45}$$

Now let $g \in \mathcal{M}_b(\mathbb{R}^{d(n+1)}, \mathcal{B}(\mathbb{R}^{d(n+1)}))$, $J \in \mathcal{T}_{0+}$ as above. Set $f = g \circ \varphi^{-1}$. Then by (3.42) (see also formula (3.28))

$$\begin{aligned} E^{\mu}\big(g(X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}})\big) \\ &= E^{\mu}\big(g \circ \varphi^{-1}(X_{t_0}, X_{t_1}, \ldots, X_{t_n}\big) \\ &= \mu U_J f \\ &= \int d(\mu \otimes \mu_{t_1} \otimes \mu_{t_2 - t_1} \otimes \cdots \otimes \mu_{t_n - t_{n-1}})\, f \circ \varphi \\ &= \int d(\mu \otimes \mu_{t_1} \otimes \mu_{t_2 - t_1} \otimes \cdots \otimes \mu_{t_n - t_{n-1}})\, g. \end{aligned}$$

Thus we have proved the following formula

$$\begin{aligned} P^{\mu}_{X_{t_0} \otimes (X_{t_1} - X_{t_0}) \otimes \cdots \otimes (X_{t_n} - X_{t_{n-1}})} &= P^{\mu}_{\varphi^{-1} \circ X_J} \\ &= \mu \otimes \mu_{t_1 - t_0} \otimes \cdots \otimes \mu_{t_n - t_{n-1}}. \end{aligned} \tag{3.46}$$

Before we cast this result in form of a theorem, we make the following

3.39 Definition Assume that $X = (X_t,\, t \in T)$ is a stochastic process with values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The family of *increments* of X is the family of random variables $(X_t - X_s,\ s, t \in T, s < t)$. The family of increments is called *independent* if for every choice of $J \in \mathcal{T}_{0+}$, $J = \{t_1, t_2, \ldots, t_n\}$, $n \in \mathbb{N}$, the family $(X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}})$ is independent. The family of increments is called *stationary*, if for every choice of $h, s, t \in T$, $s < t$, the increment $X_{t+h} - X_{s+h}$ has the same law as the increment $X_t - X_s$.

Above we have proved the first half of the following

3.40 Theorem *Suppose that $U = (U_t,\, t \in T)$ is a semigroup of Markovian kernels on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ which are invariant under translations. Then the associated canonical stochastic process $X = (X_t,\, t \in T)$ has independent and stationary increments.*

If $(\mu_t,\ t \in T)$ is the convolution semigroup of probability measures corresponding to U, then the increment $X_t - X_s$, $s, t \in T$, $s < t$, has the law μ_{t-s}.

Conversely, assume that $Y = (Y_t,\ t \in T)$ is a stochastic process with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ defined on some probability space $(\Xi, \mathcal{C}, Q)$ such that Y has stationary, independent increments. Then there is a convolution semigroup of probability measures and an associated translation invariant semigroup U of Markovian kernels on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that the canonical stochastic process X constructed from U and the initial law of Y has the same finite dimensional distributions as Y. In particular, X is a version of Y, called the canonical version *of Y.*

Proof We only have to prove the second statement. To this end, for $t > 0$, $B \in \mathcal{B}(\mathbb{R}^d)$, set

$$\mu(B) = Q_{Y_0}(B) = Q(Y_0 \in B),$$

and

$$\mu_t(B) = Q_{Y_t - Y_0}(B) = Q(Y_t - Y_0 \in B).$$

First we show that $(\mu_t,\ t \in T)$ forms a semigroup. For $s, t \in T$ we compute in the following way:

$$\begin{aligned}
\mu_{s+t} &= Q_{Y_{s+t} - Y_0} \\
&= Q_{Y_{s+t} - Y_t + Y_t - Y_0} \\
&= Q_{Y_{s+t} - Y_t} * Q_{Y_t - Y_0} \\
&= Q_{Y_s - Y_0} * Q_{Y_t - Y_0} \\
&= \mu_s * \mu_t,
\end{aligned}$$

where we used the independence of the increments in the third, and their stationarity in the fourth step. Then we put $U_0 = I$, and

$$U_t(x, B) = (\tau_x \mu_t)(B), \qquad x \in \mathbb{R}^d,\ t > 0,\ B \in \mathcal{B}(\mathbb{R}^d). \tag{3.47}$$

The semigroup property of $(\mu_t,\ t \in T)$ entails the semigroup property of the normal family of Markovian kernels $U = (U_t,\ t \in T)$. With the initial law μ and the semigroup U we can now enter the canonical construction of a probability measure P^μ, and the stochastic process X on the path space $((\mathbb{R}^d)^T, \mathcal{B}(\mathbb{R}^d)^T)$ as in theorem 3.27. Clearly, for the stochastic process X we get for any choice of $J \in \mathcal{T}_{0+}$ the formula (3.46) for the joint distribution of its increments. On the other hand, by definition of μ, μ_t, $t > 0$, and the independence of the increments of Y we find

$$\begin{aligned}
Q_{Y_{t_0} \otimes (Y_{t_1} - Y_{t_0}) \otimes \cdots \otimes (Y_{t_n} - Y_{t_{n-1}})} &= Q_{\varphi^{-1} \circ Y_J} \\
&= \mu \otimes \mu_{t_1 - t_0} \otimes \cdots \otimes \mu_{t_n - t_{n-1}}.
\end{aligned} \tag{3.48}$$

Note that $Q_{\varphi^{-1} \circ Y_J} = \varphi^{-1} Q_{Y_J}$, and $P^\mu_{\varphi^{-1} \circ X_J} = \varphi^{-1} P^\mu_{X_J}$. Thus we obtain

$$\varphi^{-1} P^\mu_{X_J} = \varphi^{-1} Q_{Y_J}.$$

Since φ^{-1} and its inverse φ are measurable bijections from $\mathbb{R}^d$ onto itself, we find for every $J \in \mathcal{T}_{0+}$ the equality $P^{\mu}_{X_J} = Q_{Y_J}$, that is, equality of all finite dimensional distributions, as claimed. $\square$

We end this chapter with a few additional definitions.

3.41 Definition A stochastic process $X = (X_t,\, t \in \mathbb{R}_+)$ with values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is called *continuous in probability*, if for all $t \in \mathbb{R}_+$, and every sequence $(t_n,\, n \in \mathbb{N})$ which converges to t, the sequence $(X_{t_n},\, n \in \mathbb{N})$ converges stochastically (that is, in probability) to X_t:

$$\forall\, \varepsilon > 0 : \quad \lim_{n\to\infty} P(|X_t - X_{t_n}| \geq \varepsilon) = 0.$$

In an *exercise* the reader will verify that the pre–Brownian motion and the pre–Poisson process are stochastically continuous.

3.42 Definition A stochastic process with values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is called a *Lévy process*, if it has independent, stationary increments, it is continuous in probability, and it has right continuous paths.

3.43 Definition A stochastic process with values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is called a *d–dimensional Brownian motion*, if it is a version of the pre–Brownian motion with continuous paths. A Brownian motion is called *standard* if it (a.s.) starts in the origin.

For ease of later reference, let us record here explicitly what this definition means:

3.44 Corollary *A stochastic process $B = (B_t,\, t \in \mathbb{R}_+)$ with a.s. continuous paths is a one dimensional Brownian motion, if and only if it has independent, stationary increments, such that $B_t - B_s$, $s, t \in \mathbb{R}_+$, $s < t$, has the law $N(0, t-s)$. It is normal if $B_0 = 0$ a.s. A d–dimensional Brownian motion, $d \in \mathbb{N}$, is a stochastic process with values in $\mathbb{R}^d$ such that its Cartesian components are independent one dimensional Brownian motions.*

In the next chapter we shall show that the pre–Brownian motion has a modification with continuous paths, i.e., that there exists Brownian motions as defined above.

3.45 Definition A stochastic process with values in $(\mathbb{R}, \mathcal{B}(\mathbb{R})$ is called a *Poisson process*, if it is a version of the pre–Poisson process with (a.s) right continous paths.

In a similar vein as for the Brownian motion we record here

3.46 Corollary *A stochastic process $X = (X_t,\, t \in \mathbb{R}_+)$ with a.s. right continuous paths is a Poisson process with rate $\lambda > 0$, if and only if it has independent, stationary increments, such that $X_t - X_s$, $s, t \in \mathbb{R}_+$, $s < t$, has the Poisson law with parameter $\lambda(t - s)$ as its distribution.*

An instructive construction of a Poisson process with the help of an independent sequence of identically distributed exponential holding times is carried out in appendix B.

Chapter 4

Continuity of Paths

In the last chapter we discussed the canonical construction of a stochastic process $X = (X_t, t \in T)$ on the path space $(E^T, \mathcal{E}^T)$, based on a consistent family of finite dimensional distributions ("observations"). For the remainder of this chapter we consider T to be $\mathbb{R}_+$ or an interval in $\mathbb{R}_+$. As the path space *is* the space of all paths of the resulting stochastic process, all possible mappings $t \mapsto E$, $t \in T$, appear as sample paths of X. Thus there is no hope to be able to control any analytic properties of the paths. In fact, the following very simple example shows that a stochastic process which has exclusively continuous paths might have a modification with no continuous path, and in particular, both processes being modifications of each other have the same finite dimensional distributions.

4.1 Example Choose $([0,1], \mathcal{B}([0,1]), \lambda)$ as the underlying probability space (λ denotes the Lebesgue measure), $T = [0,1]$, and define the following stochastic processes X, $\tilde{X}$, by

$$\begin{aligned} X_t(\omega) &= 1_{\{t\}}(\omega) \\ \tilde{X}_t(\omega) &= 0, \end{aligned}$$

where $t \in [0,1]$, $\omega \in [0,1]$. Obviously, X has no continuous path, while $\tilde{X}$ has only continuous paths. Moreover, they are modifications of each other: For $t \in [0,1]$ set $N = \{t\}$ so that N is a null set for λ. If $\omega \in \complement N$, i.e., $\omega \neq t$, then $\tilde{X}_t(\omega) = 0 = X_t(\omega)$. $\diamond$

Thus, if a stochastic process has been constructed from its finite dimensional distributions, the only statement one can expect is that it has a modification with paths which are continuous, right continuous etc. In this chapter we shall give a version of a celebrated theorem of Kolmogorov and Chentsov on the existence of Hölder continuous modifications. We begin with a preparation on Hölder continuous functions.

4.1 Hölder Continuous Functions

Consider a real valued function f defined on an interval $[a,b]$. To say that f is continuous on $[a,b]$ means that $|f(t)-f(s)|$, $s,t \in [a,b]$, vanishes as $|t-s|$ tends to zero. Often it is useful to have a more quantitative statement, like that $|f(t)-f(s)|$ vanishes as $C\,|\ln(|t-s|)|^{-1}$, C some constant, or as $C\;\exp(-|t-s|^{-1})$, or — to give another example — as $C\,|t-s|^{\gamma}$, $\gamma > 0$, when $|t-s| \to 0$. A little thought shows that of these the first is the weakest, the second the strongest statement, and the third is in between. It is the last statement, that we are going to systematize now, because it reflects the continuity behaviour of paths of many processes.

4.2 Definition Suppose that f is a real valued function on an interval $[a,b]$. f is called *Hölder continuous of order* γ*,* $0 < \gamma \le 1$*,* if there exist δ, $\eta > 0$, so that

$$\sup_{s,t\in[a,b],\,0<|t-s|<\delta} \frac{|f(t)-f(s)|}{|t-s|^{\gamma}} \le \eta$$

holds true. If f is defined on $\mathbb{R}$ or $\mathbb{R}_+$, and the above property is true for every choice of an interval $[a,b]$ in $\mathbb{R}$, $\mathbb{R}_+$ respectively, then f is called *locally Hölder continuous of order* γ.

4.3 Remark In the case that $\gamma = 1$, one says that f is *Lipshitz continuous*.

4.2 The Kolmogorov–Chentsov–Theorem

First we consider the case $T = [0,1]$. Set

$$D_n = \Big\{\frac{k}{2^n},\; k = 0, 1, \dots, 2^n\Big\},\; n \in \mathbb{N}, \quad D = \bigcup_{n=1}^{\infty} D_n. \tag{4.1}$$

Observe that for all $n \in \mathbb{N}$ we have that $D_n \subset D_{n+1}$, and that D is dense in $[0,1]$. D is called the set of *dyadic rationals* in $[0,1]$.

4.4 Proposition *Assume that* $X = (X_t,\, t \in [0,1])$ *is a real valued stochastic process on* $(\Omega, \mathcal{A}, P)$*, and that there exist* β*,* γ*,* $C > 0$ *so that*

$$P\big(|X_t - X_s| > |t-s|^{\gamma}\big) \le C|t-s|^{1+\beta}, \qquad \text{for all } s,t \in [0,1]. \tag{4.2}$$

Then X *has a modification* $\tilde{X}$ *with paths which are Hölder continuous of order* γ*. In more detail: There exist a constant* $\eta > 0$*, and a non-negative random variable* δ *which is strictly positive on the complement* $\complement N$ *of a* P*–null set* $N \in \mathcal{A}$*, so that for all* $\omega \in \complement N$

$$\sup_{s,t\in[0,1],\,0<|t-s|<\delta(\omega)} \frac{|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)|}{|t-s|^{\gamma}} \le \eta \tag{4.3}$$

holds. Moreover, $\tilde{X}$ can be chosen in such a way that

$$P\big(\tilde{X}_t = X_t,\, t \in D\big) = 1,$$

and $\tilde{X} = 0$ on N.

We carry out the proof in a sequence of lemmas. Throughout we suppose that the process X is as in the hypothesis of proposition 4.4.

4.5 Lemma *X is continuous in probability.*

Proof Given $\varepsilon > 0$, consider $s, t \in [0,1]$ so that $|t-s| \le \varepsilon^{1/\gamma}$. Then

$$\begin{aligned} P(|X_t - X_s| \ge \varepsilon) &\le P(|X_t - X_s| \ge |t-s|^{\gamma}) \\ &\le C|t-s|^{1+\beta}, \end{aligned}$$

which converges to zero as $|t-s| \to 0$. □

The next lemma reduces the proof of the Hölder continuity to a dense subset Q of $[0,1]$.

4.6 Lemma *Suppose that there exists a P–null set N, a set Q which is dense in $[0,1]$, a random variable δ, and a constant $\eta > 0$ so that for every $\omega \in \complement N$, $\delta(\omega) > 0$ and*

$$|X_t(\omega) - X_s(\omega)| \le \eta |t-s|^{\gamma} \tag{4.4}$$

holds for all $s, t \in Q$, $0 < |t-s| < \delta(\omega)$. Then X has a modification $\tilde{X}$ which is Hölder continuous of order γ. Moreover, $P(\tilde{X}_t = X_t,\, t \in Q) = 1$.

Proof First we construct the modification $\tilde{X}$: For $\omega \in N$ we put $\tilde{X}_t(\omega) = 0$ for all $t \in [0,1]$. If $\omega \in \complement N$, $t \in Q$, we set $\tilde{X}_t(\omega) = X_t(\omega)$. If $\omega \in \complement N$, $t \notin Q$, we choose a sequence $(t_n,\, n \in \mathbb{N})$ in Q which converges to t. By assumption, $s \mapsto X_s(\omega)$ is uniformly continuous on Q, and therefore $(X_{t_n}(\omega),\, n \in \mathbb{N})$ is a Cauchy sequence. We define $\tilde{X}_t(\omega) = \lim_n X_{t_n}(\omega)$. Thus, by construction $\tilde{X}$ has continuous paths. Next we show that inequality (4.4) holds for $\tilde{X}$ as well. Let $\omega \in \complement N$, $s, t \in [0,1]$ with $0 < |t-s| < \delta(\omega)$, and let $(s_n,\, n \in \mathbb{N})$, $(t_n,\, n \in \mathbb{N})$ be sequences in Q converging to s, t respectively. Choose $\varepsilon > 0$, then for all n large enough we get from the construction of $\tilde{X}$

$$|\tilde{X}_t(\omega) - X_{t_n}(\omega)| \le \frac{\varepsilon}{2}, \quad |\tilde{X}_s(\omega) - X_{s_n}(\omega)| \le \frac{\varepsilon}{2}.$$

Therefore we find

$$|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)| \le \eta |t_n - s_n|^{\gamma} + \varepsilon.$$

By letting $n \to +\infty$ we get

$$|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)| \le \eta |t-s|^{\gamma} + \varepsilon,$$

and now we make $\varepsilon \downarrow 0$ to obtain

$$|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)| \leq \eta |t - s|^\gamma.$$

Thus $\tilde{X}$ is Hölder continuous of order γ. Finally we show that $\tilde{X}$ is a modification of X: For $t \in [0, 1]$ let $(t_n, n \in \mathbb{N})$ be a sequence in Q converging to t. By construction we have that $(X_{t_n}, n \in \mathbb{N})$ converges a.s. to $\tilde{X}_t$, and therefore it also converges in probability to $\tilde{X}_t$. On the other hand, because X is continuous in probability (cf. lemma 4.5), $(X_{t_n}, n \in \mathbb{N})$ converges to X_t in probability. Since limits in probability are a.s. unique, we proved that $P(\tilde{X}_t = X_t) = 1$. □

In the next three lemmas we prove that the hypothesis of the last lemma indeed follows from the hypothesis of proposition 4.4.

4.7 Lemma *There exists a P–null set N so that for all $\omega \in \complement N$ exists an $n(\omega) \in \mathbb{N}$ such that for all $n \geq n(\omega)$*

$$\max_{1 \leq k \leq 2^n} \left| X_{k/2^n}(\omega) - X_{(k-1)/2^n}(\omega) \right| \leq 2^{-\gamma n} \tag{4.5}$$

holds true.

Proof Define a sequence $(A_n, n \in \mathbb{N})$ of events by

$$A_n = \left\{ \omega \in \Omega, \max_{1 \leq k \leq 2^n} \left| X_{k/2^n}(\omega) - X_{(k-1)/2^n}(\omega) \right| > 2^{-\gamma n} \right\}.$$

From the hypothesis, inequality (4.2), we get for $n \in \mathbb{N}$, $k = 1, \ldots, 2^n$,

$$P\left(\left| X_{k/2^n} - X_{(k-1)/2^n} \right| > 2^{-\gamma n} \right) \leq C 2^{-n(1+\beta)}.$$

Hence

$$\begin{aligned} P(A_n) &= P\Big(\bigcup_{k=1}^{2^n} \{ |X_{k/2^n} - X_{(k-1)/2^n} > 2^{-\gamma n} \} \Big) \\ &\leq \sum_{k=1}^{2^n} P\left(\left| X_{k/2^n} - X_{(k-1)/2^n} \right| > 2^{-\gamma n} \right) \\ &\leq C 2^n 2^{-n(1+\beta)} \\ &= C 2^{-n\beta}. \end{aligned}$$

Since $\beta > 0$, we find $\sum_n P(A_n) < +\infty$. The Borel–Cantelli–Lemma, lemma C.1, entails that $P(\limsup_n A_n) = 0$. Set $N = \limsup_n A_n$. Then N is a P–null set, and for every $\omega \in \complement N$ there exists an index $n(\omega) \in \mathbb{N}$ such that for all $n \geq n(\omega)$, $\omega \in \complement A_n$. That is, for all $n \geq n(\omega)$ we find that inequality (4.5) holds. □

From now on we choose the set of dyadic rationals $D = \cup_n D_n$, see (4.1), as the dense set Q.

4.8 Lemma *Let N, $n(\omega)$, $\omega \in \complement N$, be as in lemma 4.7. For all $\omega \in \complement N$, $n \geq n(\omega)$, $m \in \mathbb{N}$, $m > n$, and all $s, t \in D_m$ with $0 \leq |t-s| < 2^{-n}$ one has*

$$|X_t(\omega) - X_s(\omega)| \leq 2 \sum_{j=n+1}^{m} 2^{-\gamma j}. \tag{4.6}$$

Proof We prove estimate (4.6) by induction on m. First consider $m = n+1$: Clearly, for $s = t$ we have nothing to prove. For $s \neq t$ we have to have $s = l/2^{n+1}$, $t = k/2^{n+1}$, for some $l, k \in \{0, 1, \ldots, 2^{n+1}\}$, with

$$0 < \left| \frac{l}{2^{n+1}} - \frac{k}{2^{n+1}} \right| < \frac{1}{2^n}.$$

This implies $0 < |l-k| < 2$, i.e., $|l-k| = 1$. Therefore, without loss of generality, $s = (k-1)/2^{n+1}$, $t = k/2^{n+1}$, $k \in \{1, \ldots, 2^{n+1}\}$. We use (4.5) which yields

$$|X_t(\omega) - X_s(\omega)| \leq 2^{-\gamma(n+1)} \leq 2\, 2^{-\gamma(n+1)}.$$

This proves the claim for $m = n+1$. Now assume that inequality (4.6) holds true for $m-1 \in \mathbb{N}$, $m-1 > n$, and suppose that $s, t \in D_m$, with $0 < |t-s| < 2^{-n}$. Without loss of generality we may assume in addition that $s < t$. Choose $s' \in D_{m-1} \subset D_m$ as the smallest number larger than s, and $t' \in D_{m-1} \subset D_m$ as the largest number less than t. Then we obtain

$$s \leq s' \leq t' \leq t, \quad s' - s \leq 2^{-m}, \quad t - t' \leq 2^{-m},$$

and in particular we also have $t' - s' < 2^{-n}$. From (4.5) we get

$$|X_{s'}(\omega) - X_s(\omega)| \leq 2^{-\gamma m}, \quad |X_t(\omega) - X_{t'}(\omega)| \leq 2^{-\gamma m},$$

and the induction hypothesis gives the estimate

$$|X_{t'}(\omega) - X_{s'}(\omega)| \leq 2 \sum_{j=n+1}^{m-1} 2^{-\gamma j}.$$

Therefore

$$\begin{aligned}
|X_t(\omega) - X_s(\omega)| &\leq |X_t(\omega) - X_{t'}(\omega)| + |X_{t'}(\omega) - X_{s'}(\omega)| + |X_{s'}(\omega) - X_s(\omega)| \\
&\leq 2\, 2^{-\gamma m} + 2 \sum_{j=n+1}^{m-1} 2^{-\gamma j} \\
&\leq 2 \sum_{j=n+1}^{m} 2^{-\gamma j}
\end{aligned}$$

and the proof is finished. □

4.9 Lemma *Let N, $n(\omega)$, $\omega \in \complement N$, be as in lemma 4.7. There exists a constant $\eta > 0$, and a random variable $\delta \geq 0$, so that $\delta(\omega) > 0$ for every $\omega \in \complement N$ and for all $s, t \in D$ with $0 < |t - s| < \delta(\omega)$,*

$$|X_t(\omega) - X_s(\omega)| \leq \eta |t - s|^\gamma \tag{4.7}$$

holds true.

Proof On N set $\delta = 1$, and for $\omega \in \complement N$ define $\delta(\omega) = 2^{-n(\omega)}$, where $n(\omega)$ is as in the statement of lemma 4.7. Suppose that $s, t \in D$ are such that $0 < t - s < \delta(\omega)$. Let n be the largest natural number so that $0 < t - s < 2^{-n}$. Then we get

$$2^{-(n+1)} \leq t - s < 2^{-n}.$$

Moreover, since $s, t \in D$, there exists $m \in \mathbb{N}$, $m > n$, so that $s, t \in D_m$. With lemma 4.8 we can estimate in the following way:

$$\begin{aligned} |X_t(\omega) - X_s(\omega)| &\leq 2 \sum_{j=n+1}^{m} 2^{-\gamma j} \\ &\leq 2 \sum_{j=n+1}^{\infty} 2^{-\gamma j} \\ &= \eta \, 2^{-\gamma(n+1)} \\ &\leq \eta \, |t - s|^\gamma, \end{aligned}$$

where we have set

$$\eta = \frac{2}{1 - 2^{-\gamma}}. \tag{4.8}$$

Thus the lemma is proved. □

Since the statements of lemmas 4.5, 4.7–4.9 all follow from the hypothesis of proposition 4.4, we have proved proposition 4.4.

Now suppose that the time parameter domain of the process X is $\mathbb{R}_+$, and suppose that inequality (4.2) holds true for all $s, t \in \mathbb{R}_+$. Decompose $\mathbb{R}_+$ into the intervals $[0, 1]$, $(n, n + 1]$, $n \in \mathbb{N}$. Then on each interval $[n, n + 1]$, $n \in \mathbb{N}_0$, we obtain the statement of proposition 4.4, and a glance at equation (4.8) shows that for all of them we can choose the same η. Moreover, for the modifications $\tilde{X}$ of X we find for every $n \in \mathbb{N}$, $P(\tilde{X}_n = X_n, n \in \mathbb{N}_0) = 1$, because above we have that $0, 1 \in D$ (cf. lemma 4.6). Thus we obtain a sequence $(N_n, n \in \mathbb{N}_0)$ of P–null sets, and a sequence $(\delta_n, n \in \mathbb{N}_0)$ of strictly positive random variables δ_n so that for every $\omega \in \complement N$, $N = \cup_n N_n$,

$$\sup_{s,t \in [n,n+1],\, 0<|t-s|<\delta_n(\omega)} \frac{|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)|}{|t - s|^\gamma} \leq \eta,$$

holds true for every $n \in \mathbb{N}_0$. Note that N is a P–null set, too. Consider now $\omega \in \complement N$, $n \in \mathbb{N}$, and $s \in [n-1, n]$, $t \in [n, n+1]$, $|t-s| < \min(\delta_{n-1}(\omega), \delta_n(\omega))$, then

$$\begin{aligned} |\tilde{X}_t(\omega) - \tilde{X}_s(\omega)| &\le |\tilde{X}_t(\omega) - \tilde{X}_n(\omega)| + |\tilde{X}_n(\omega) - \tilde{X}_s(\omega)| \\ &\le \eta \left(|t-n|^\gamma + |n-s|^\gamma\right) \\ &\le 2\eta\, |t-s|^\gamma. \end{aligned}$$

On the other hand, for each interval $[n, n+1]$, $n \in \mathbb{N}_0$, and all s, $t \in [n, n+1]$, $|t-s| < \delta_n(\omega)$, $\omega \in \complement N$, we have

$$|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)| \le \eta |t-s|^\gamma \le 2\eta |t-s|^\gamma.$$

Therefore, replacing below 2η by η, we altogether have proved the following

4.10 Theorem (Kolmogorov–Chentsov) *Assume that $X = (X_t,\, t \in \mathbb{R}_+)$ is a real valued stochastic process on $(\Omega, \mathcal{A}, P)$, and that there exist β, γ, $C > 0$ so that*

$$P\big(|X_t - X_s| > |t-s|^\gamma\big) \le C|t-s|^{1+\beta}, \qquad \text{for all } s, t \in \mathbb{R}_+. \tag{4.9}$$

Then X has a modification $\tilde{X}$ with paths which are locally Hölder continuous of order γ. That is, there exist a constant $\eta > 0$, a P–null set $N \in \mathcal{A}$, and for every compact subset K of $\mathbb{R}_+$ a non-negative random variable δ_K, which is strictly positive on $\complement N$, so that for every $\omega \in \complement N$

$$\sup_{s,t \in K,\, 0<|t-s|<\delta_K(\omega)} \frac{|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)|}{|t-s|^\gamma} \le \eta, \tag{4.10}$$

and $\tilde{X} = 0$ on N hold true.

We end this section with a corollary to theorem 4.10, which very often is used in applications.

4.11 Corollary *Suppose that there exist constants α, ρ, $C > 0$ so that for all s, $t \in \mathbb{R}_+$,*

$$E\big(|X_t - X_s|^\alpha\big) \le C|t-s|^{1+\rho} \tag{4.11}$$

holds. Then X has a modification with paths which are locally Hölder continuous of order γ for any $\gamma \in (0, \rho/\alpha)$.

Proof Choose $\gamma \in (0, \rho/\alpha)$, and put $\beta = \rho - \alpha\gamma$ so that $\beta > 0$. Chebyshev's inequality entails

$$\begin{aligned} P\big(|X_t - X_s| > |t-s|^\gamma\big) &\le |t-s|^{-\alpha\gamma} E\big(|X_t - X_s|^\alpha\big) \\ &\le C|t-s|^{1+\rho-\alpha\gamma} \\ &= C|t-s|^{1+\beta}. \end{aligned}$$

Now we can apply theorem 4.10 to finish the proof. $\square$

4.3 Application to Brownian Motion, Wiener Space

Consider a real valued stochastic process $X = (X_t,\, t \in \mathbb{R}_+)$ with independent, stationary increments, so that $X_t - X_s \sim N(0, \rho(t-s))$, $s, t \in \mathbb{R}_+$, $s < t$. Then for all $n \in \mathbb{N}$,

$$E\big((X_t - X_s)^{2n}\big) = (2n-1)!!\, \rho(t-s)^n.$$

In particular, for a pre–Brownian motion X, we have $\rho(u) = u$. Thus we get for all $s, t \in \mathbb{R}_+$, $n \in \mathbb{N}$,

$$E\big((X_t - X_s)^{2n}\big) = (2n-1)!!\, |t-s|^{1+(n-1)}.$$

For the order γ of the local Hölder continuity of the previous section we therefore obtain the condition $\gamma < 1/2(1 - 1/n)$. Since we can choose n as large as we want, we find that X has a modification with paths which are locally Hölder continuous of order γ for every $\gamma \in (0, 1/2)$. In particular, we have shown that a Brownian motion as in definition 3.43 exists.

The case of an $\mathbb{R}^d$–valued Brownian motion, $d \in \mathbb{N}$, is done in the same way — one just has to construct a modification for each of its d independent components.

In later chapters we shall make use of Brownian motions which have exclusively (instead of P–a.s.) continuous sample paths. As long as we have one Brownian motion B in $\mathbb{R}^d$ with a *fixed* starting point, say $x \in \mathbb{R}^d$, one just has to remove a P–null set N from Ω, take the trace σ–algebra of $\mathcal{A}$ on $\Omega \setminus N$, and restrict P and B on this new sample space to achieve this. However, we shall consider a whole family of Brownian motions B with probability measures P_x, $x \in \mathbb{R}^d$, so that $P_x(B_0 = x) = 1$. In this situation we cannot simply remove all P_x–null sets, because we do not know whether the uncountable union of these null sets for each P_x form a null set for all these probability measures. Therefore we need a different approach. One might be tempted to use the following idea: Consider the Brownian motion $B = (B_t,\, t \in T)$, $T = \mathbb{R}_+$, as mapping from Ω into the path space $(\mathbb{R}^d)^T$ with its σ–algebra $\mathcal{B}(\mathbb{R}^d)^T$ generated by the cylinder sets (see section 3.2). We know that the mapping B is measurable. The space $C(T, \mathbb{R}^d)$ of $\mathbb{R}^d$–valued continuous functions on $T = \mathbb{R}_+$ is a subset of the path space, and one might think of taking the pre-image of $C(T, \mathbb{R}^d)$ under B, and replace Ω by this subset of Ω. However, $C(T, \mathbb{R}^d)$ is *not* a set in the σ–algebra $\mathcal{B}(\mathbb{R}^d)^T$, and therefore this does not make sense. To see that $C(T, \mathbb{R}^d) \notin \mathcal{B}(\mathbb{R}^d)^T$, recall that we have shown in lemma 3.11 that every $\Lambda \in \mathcal{B}(\mathbb{R}^d)^T$ is such that there exists a countable subset J of $T = \mathbb{R}_+$, and an element $B \in \mathcal{B}(\mathbb{R}^d)^J$ so that $\Lambda = (\pi^J_T)^{-1}(B)$. If ξ is a continuous function from T into $\mathbb{R}^d$ in Λ, choose any discontinuous function $\eta \in (\mathbb{R}^d)^T$ which coincides with ξ on J. Then also $\eta \in \Lambda$, so that Λ cannot consist exclusively of continuous functions.

It is convenient to solve the above described problem in an abstract setting. Suppose that $X = (X_t,\, t \in T)$ is an E–valued stochastic process, $(E, \mathcal{E})$ some measurable space, defined on a probability space $(\Omega, \mathcal{A}, P)$. Assume furthermore that there exists some subset Θ of the path space E^T, so that P–a.s. the paths of X belong to Θ. We shall *not* assume that $\Theta \in \mathcal{E}^T$. As usual, P_X denotes the image measure of P

on $(E^T, \mathcal{E}^T)$ under X, considered as a (measurable, see lemma 3.12) mapping from Ω into $\mathcal{E}^T$.

4.12 Lemma *For every $\Lambda \in \mathcal{E}^T$ so that $\Lambda \supset \Theta$, $P_X(\Lambda) = 1$, and for every $\Lambda \in \mathcal{E}^T$ so that $\Lambda \cap \Theta = \emptyset$, $P_X(\Lambda) = 0$.*

Proof That the paths of X belong P–a.s. to Θ means that there is a P–null set N so that for all $\omega \in \complement N$, $(X_t(\omega),\ t \in T) \in \Theta$. If $\Lambda \supset \Theta$, then we get that $X^{-1}(\Lambda) \supset \Omega \setminus N$. Therefore $P_X(\Lambda) = P(X^{-1}(\Lambda)) \geq P(\Omega \setminus N) = 1$. That is, $P_X(\Lambda) = 1$. $\Lambda \cap \Theta = \emptyset$ implies that $\complement\Lambda \supset \Theta$, and therefore by the first statement $P_X(\complement\Lambda) = 1$. Hence $P_X(\Lambda) = 0$. $\square$

We equip $\Theta \subset \mathcal{E}^T$ with the trace σ–algebra $\mathcal{T} = \Theta \cap \mathcal{E}^T$ of $\mathcal{E}^T$ on Θ. On the measurable space $(\Theta, \mathcal{T})$ we define

$$Q_X(A) = P_X(\Lambda), \quad \text{for } A \in \mathcal{T},\ A = \Theta \cap \Lambda,\ \Lambda \in \mathcal{E}^T. \tag{4.12}$$

4.13 Lemma *Q_X is a well-defined probability measure on $(\Theta, \mathcal{T})$.*

Proof First we show that Q_X is well-defined. To this end assume that $A = \Theta \cap \Lambda_1 = \Theta \cap \Lambda_2$, $\Lambda_1, \Lambda_2 \in \mathcal{E}^T$. Then

$$\emptyset = (\Theta \cap \Lambda_1) \triangle (\Theta \cap \Lambda_1) = \Theta \cap (\Lambda_1 \triangle \Lambda_2).$$

Thus by lemma 4.12 we find that $P_X(\Lambda_1 \triangle \Lambda_2) = 0$. Hence $P_X(\Lambda_1) = P_X(\Lambda_2)$, [1] and therefore Q_X is well-defined.

Now we prove that Q_X is a probability measure on $(\Theta, \mathcal{T})$. Let $\Lambda \in \mathcal{E}^T$ be such that $\Lambda \cap \Theta = \emptyset$. Then $Q_X(\emptyset) = P_X(\Lambda) = 0$ by lemma 4.12. To show that $Q_X(\Theta) = 1$, write $\Theta = \Theta \cap E^T$, so that $Q_X(\Theta) = P_X(E^T) = 1$. Finally, suppose that $(A_n,\ n \in \mathbb{N})$ is a pairwise disjoint sequence in $\mathcal{T}$. Then there exists a sequence $(\Lambda_n,\ n \in \mathbb{N})$ (not necessarily pairwise disjoint), such that for every $n \in \mathbb{N}$ we have $A_n = \Theta \cap \Lambda_n$. We want to show that

$$Q_X\Big(\biguplus_{n \in \mathbb{N}} A_n\Big) = \sum_{n=1}^{\infty} Q_X(A_n). \tag{4.13}$$

First we claim that

$$P_X\Big(\biguplus_{n \in \mathbb{N}} \Lambda_n\Big) = \sum_{n=1}^{\infty} P_X(\Lambda_n). \tag{4.14}$$

For $m, n \in \mathbb{N}$, $m \neq n$, we find

$$\emptyset = A_m \cap A_m = (\Lambda_m \cap \Lambda_n) \cap \Theta,$$

[1] $P_X(\Lambda_1) = P_X(\Lambda_1 \setminus \Lambda_1 \cap \Lambda_2) + P_X(\Lambda_1 \cap \Lambda_2) = P_X(\Lambda_2 \setminus \Lambda_1 \cap \Lambda_2) + P_X(\Lambda_1 \cap \Lambda_2) = P_X(\Lambda_2)$

so that lemma 4.12 implies $P_X(\Lambda_m \cap \Lambda_n) = 0$. We have

$$P_X\Big(\Lambda_1 \cup \Big(\bigcup_{n\geq 2} \Lambda_n\Big)\Big) = P_X(\Lambda_1) + P_X\Big(\bigcup_{n\geq 2} \Lambda_n\Big) - P_X\Big(\bigcup_{n\geq 2}(\Lambda_1 \cap \Lambda_n)\Big),$$

and

$$P_X\Big(\bigcup_{n\geq 2}(\Lambda_1 \cap \Lambda_n)\Big) \leq \sum_{n\geq 2} P_X(\Lambda_1 \cap \Lambda_n) = 0.$$

Therefore

$$P_X\Big(\Lambda_1 \cup \Big(\bigcup_{n\geq 2} \Lambda_n\Big)\Big) = P_X(\Lambda_1) + P_X\Big(\bigcup_{n\geq 2} \Lambda_n\Big),$$

and by iteration we find that for every $m \in \mathbb{N}$,

$$P_X\Big(\bigcup_{n\in\mathbb{N}} \Lambda_n\Big) = \sum_{n=1}^{m} P_X(\Lambda_n) + P_X\Big(\bigcup_{n\geq m} \Lambda_n\Big).$$

Since the left hand side is bounded by 1, the series on the right hand side converges. But

$$P_X\Big(\bigcup_{n\geq m} \Lambda_n\Big) \leq \sum_{n=m}^{\infty} P_X(\Lambda_n),$$

and the last term converges to zero as $m \to \infty$. Thus we get proved that indeed equation (4.14) holds true. Now

$$\begin{aligned} Q_X\Big(\biguplus_{n\in\mathbb{N}} A_n\Big) &= P_X\Big(\bigcup_{n\in\mathbb{N}} \Lambda_n\Big) \\ &= \sum_{n=1}^{\infty} P_X(\Lambda_n) \\ &= \sum_{n=1}^{\infty} Q_X(A_n) \end{aligned}$$

concludes the proof. □

On $(\Theta, \mathcal{T})$ define $Y = (Y_t,\, t \in T)$ by $Y_t(\eta) = \eta(t)$, $\eta \in \Theta$, $t \in T$. With the notation of section 3.2 we get for $B \in \mathcal{E}$ that $Y_t^{-1}(B) = \Theta \cap (\pi^t_T)^{-1}(B) \in \mathcal{T}$. Thus Y is an E–valued stochastic process.

4.14 Lemma *Under Q_X the stochastic process Y has the same finite dimensional distributions as X, i.e., X and Y are versions of each other.*

Proof For $t_1, \ldots, t_n \in \mathbb{R}_+$, $B_1, \ldots, B_n \in \mathcal{E}$, set

$$\Lambda = \{\eta \in E^T,\ \eta(t_1) \in B_1, \ldots, \eta(t_n) \in B_n\} \in \mathcal{E}^T.$$

Then

$$\begin{aligned} Q_X\big(Y_{t_1} \in B_1, \dots, Y_{t_n} \in B_n\big) &= Q_X\big(\Theta \cap \Lambda\big) \\ &= P_X(\Lambda) \\ &= P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n), \end{aligned}$$

and the proof is finished. □

Altogether we have proved the following

4.15 Theorem *Suppose that $(E, \mathcal{E})$ is a measurable space, $(\Omega, \mathcal{A}, P)$ a probability space, and that $X = (X_t,\, t \in T)$ is an E–valued stochastic process, with paths which P–a.s. belong to some subset Θ of E^T. Then there exists a version of X on some probability space exclusively with paths in Θ. In particular, there exists a probability measure Q_X on $(\Theta, \mathcal{T})$, where $\mathcal{T}$ is the σ–algebra over Θ generated by the cylinder sets of Θ, so that the canonical coordinate process on $(\Theta, \mathcal{T})$ is a version of X.*

We return to our discussion of Brownian motion. Suppose that $(\Omega, \mathcal{A})$ is a measurable space with a family $(P_x,\, x \in \mathbb{R}^d)$, $d \in \mathbb{N}$, of probability measures defined thereon. Suppose furthermore that $B = (B_t,\, t \in \mathbb{R}_+)$ is a family of R^d–valued random variables, so that for each $x \in \mathbb{R}^d$, B is a Brownian motion on $(\Omega, \mathcal{A}, P_x)$ which P_x–a.s. starts in x: $P_x(B_0 = x) = 1$. For example, this situation could have been canonically constructed as in section 3.4 with the Brown-semigroup, and initial laws $\mu = \varepsilon_x$, $x \in \mathbb{R}^d$. Our application of the Kolmogorov–Chentsov theorem at the beginning of this section gives us the following statement: For every $x \in \mathbb{R}^d$, P_x–a.s. the paths of B belong to $C(\mathbb{R}_+, \mathbb{R}^d)$. $\mathcal{C}$ denotes the σ–algebra over $C(\mathbb{R}_+, \mathbb{R}^d)$ generated by cylinder sets of $C(\mathbb{R}_+, \mathbb{R}^d)$, and W denotes the canonical coordinate process on $C(\mathbb{R}_+, \mathbb{R}^d)$. $\big(C(\mathbb{R}_+, \mathbb{R}^d), \mathcal{C}\big)$ is called the *Wiener space*, W the *Wiener process*. Theorem 4.15 states that there exists a family $(Q_x,\, x \in \mathbb{R}^d)$ of probability measures on Wiener space, called *Wiener measures*, such that for every $x \in \mathbb{R}^d$, W is under Q_x a Brownian motion starting Q_x–a.s. at x. Trivially, for every $x \in \mathbb{R}^d$ these Brownian motions have only continuous sample paths.

Chapter 5

Conditional Expectation

The conditional expecation of a random variable is a central notion in probability theory which allows to formulate, analyze and compute interesting properties of a random variable or a stochastic process. The basic idea is to define the expectation value of a random variable given the additional knowlegde of certain information. It turns out that this "additional information" is best encoded in form of a sub–σ–algebra $\mathcal{A}_0$ of the underlying σ–algebra $\mathcal{A}$ of the probability space. For example, $\mathcal{A}_0$ could be the σ–algebra generated by a collection of random variables.

5.1 Definition of Conditional Expectation

Throughout we continue to assume that $(\Omega, \mathcal{A}, P)$ is the underlying probability space. Suppose that we are given a sub–σ–algebra $\mathcal{A}_0$ of $\mathcal{A}$.

5.1 Definition If X is a positive or integrable random variable, any random variable Y, which is positive, respectively integrable, and such that

(i) Y is $\mathcal{A}_0$–measurable,

(ii) for every $A \in \mathcal{A}_0$

$$\int_A Y\,dP = \int_A X\,dP \tag{5.1}$$

hold true, is called a *conditional expectation of X given $\mathcal{A}_0$*.

An *exercise* shows that, if such a Y exists (which we shall prove below), then it is a.s. unique. Therefore we shall henceforth speak of *the* conditional expectation (meaning actually the P–class of Y), and denote it by $E(X \mid \mathcal{A}_0)$.

5.2 Exercise Suppose that $\Omega = [0, 1]$, $\mathcal{A} = \mathcal{B}([0, 1])$, $P = \lambda$, and let X be any $\mathcal{B}([0, 1]/\mathcal{B}(\mathbb{R})$–measurable function which is positive or λ-integrable. Fix $t \in (0, 1)$, and let $\mathcal{A}_0$ be the sub–σ–algebra of $\mathcal{B}([0, 1])$ generated by $\mathcal{B}([0, t])$. Show the following statements:

(a) $A \in \mathcal{A}_0$, if and only if there exists $\tilde{A} \in \mathcal{B}([0,t])$ so that $A = \tilde{A} \cup (t,1]$ or $A = \tilde{A}$.

(b) X is $\mathcal{A}_0$–measurable, if and only if X is constant on $(t,1]$.

(c) $E(X \mid \mathcal{A}_0) = X 1_{[0,t]} + c 1_{(t,1]}$ with

$$c = \frac{1}{1-t}\int_t^1 X\, d\lambda.$$

5.3 Theorem *Suppose that X is a positive or integrable random variable, and that $\mathcal{A}_0$ is a sub–σ–algebra of $\mathcal{A}$. Then there exists an a.s. unique conditional expectation $E(X \mid \mathcal{A}_0)$ of X.*

Proof For the proof we first assume that X belongs to $\mathcal{L}^2(\Omega, \mathcal{A}, P)$, and we base our proof on Riesz' representation theorem for Hilbert spaces, cf. appendix D.

Consider the measure space $(\Omega, \mathcal{A}_0, P_0)$, where P_0 is the restriction of P to $\mathcal{A}_0$. For $Z \in \mathcal{L}^2(\Omega, \mathcal{A}_0, P_0)$ we denote by $[Z]$ the P_0–equivalence class of Z in the Hilbert space $L^2(\Omega, \mathcal{A}_0, P_0)$. Consider the mapping

$$\begin{aligned} T_X : L^2(\Omega, \mathcal{A}_0, P_0) &\to \mathbb{R} \\ [Z] &\mapsto T_X([Z]) = \int_\Omega XZ\, dP. \end{aligned} \tag{5.2}$$

The reader will check that this mapping is well-defined (i.e., does not depend on the choice of the representative of $[Z]$). Moreover, Schwarz' inequality immediately gives the estimate

$$\left|T_X([Z])\right| \le \|X\|_2 \|Z\|_2.$$

Therefore, T_X defines a continuous linear functional on $L^2(\Omega, \mathcal{A}_0, P_0)$. Riesz' representation theorem, theorem D.5, implies that there exists $[Y] \in L^2(\Omega, \mathcal{A}_0, P_0)$ so that for every Z in $\mathcal{L}^2(\Omega, \mathcal{A}_0, P_0)$ we have

$$T_X([Z]) = ([Y],[Z])_{L^2(\Omega,\mathcal{A}_0,P_0)} = \int_\Omega YZ\, dP_0, \tag{5.3}$$

where Y is any representative of $[Y]$. Clearly, Y is $\mathcal{A}_0$–measurable, and by choosing $Z = 1_A$ with $A \in \mathcal{A}_0$ in (5.2), (5.3), we get that

$$\int_A Y\, dP = \int_A X\, dP,$$

since P and P_0 coincide on $\mathcal{A}_0$. Thus Y is a conditional expectation of X given $\mathcal{A}_0$, and it is unique up to P_0–null functions.

Next consider the case that $X \ge 0$. Define $X_n = X \wedge n$, [1] $n \in \mathbb{N}$, so that for every $n \in \mathbb{N}$, X_n is bounded and hence in $\mathcal{L}^2(\Omega, \mathcal{A}, P)$. Then by the preceding case,

[1] $a \wedge b \equiv \min\{a,b\}$

for all $n \in \mathbb{N}$ there exists a conditional expectation Y_n of X_n given $\mathcal{A}_0$. Moreover, it is easy to check that $(X_n,\ n \in \mathbb{N})$ is pointwise increasing to X, while $(Y_n,\ n \in \mathbb{N})$ is P_0–a.s. increasing. Define

$$Y = \limsup_n Y_n.$$

Then Y is $\mathcal{A}_0$ measurable, and moreover $(Y_n,\ n \in \mathbb{N})$ increases P_0–a.s. to Y. Thus for every $A \in \mathcal{A}_0$ we get by the monotone convergence theorem that

$$\begin{aligned}\int_A X\,dP &= \lim_n \int_A X_n\,dP \\ &= \lim_n \int_A Y_n\,dP \\ &= \int_A Y\,dP.\end{aligned}$$

Hence Y is a conditional expectation of X relative to $\mathcal{A}_0$. Observe that the last equality also implies that P–a.s the random variable Y is non-negative, and that if the expectation of X is finite, then so is the expecation of Y.

Finally, suppose that X is integrable, and decompose $X = X^+ - X^-$ into its positive and negative parts. Then we have from above two $\mathcal{A}_0$–measurable random variables $Y^\pm$ being a conditional expectation of $X^\pm$, respectively. Then $Y^+ - Y^-$ is a conditional expectation of X. $\square$

The next lemma will prove very useful.

5.4 Lemma *Suppose that X is a positive or integrable random variable, and that $\mathcal{A}_0$ is a sub–σ–algebra of $\mathcal{A}$. Assume furthermore that $\mathcal{E}_0$ is a generator of $\mathcal{A}_0$ which is $\cap$–stable and contains Ω. A positive or integrable, $\mathcal{A}_0$–measurable random variable is a version of the conditional expectation $E(X \mid \mathcal{A}_0)$, if and only if equation* (5.1) *is true for all $A \in \mathcal{E}_0$.*

Proof If Y is a version of $E(X \mid \mathcal{A}_0)$, then equation (5.1) is true for all $A \in \mathcal{E}_0$ because $\mathcal{E}_0 \subset \mathcal{A}_0$. For the proof of the converse implication let $\mathcal{D}$ be the family of sets A in $\mathcal{A}$ so that (5.1) holds true. By hypothesis $\mathcal{D} \supset \mathcal{E}_0$. We show that $\mathcal{D}$ is a d-system: $\Omega \in \mathcal{D}$ is true by assumption. Suppose that $A, B \in \mathcal{D}$ with $A \subset B$. Then $B = A \uplus (B \setminus A)$, and hence (for convenience we no longer distinguish between P and P_0)

$$\begin{aligned}\int_{B\setminus A} Y\,dP + \int_A Y\,dP &= \int_B Y\,dP \\ &= \int_B X\,dP \\ &= \int_{B\setminus A} X\,dP + \int_A X\,dP.\end{aligned}$$

It follows that (5.1) holds with A replaced by $B \setminus A$, i.e., that $B \setminus A \in \mathcal{D}$. Finally, let $(A_n,\, n \in \mathbb{N})$ be an increasing sequence in $\mathcal{D}$, and let A be its limit. Then we have for all $n \in \mathbb{N}$,

$$\int 1_{A_n} X \, dP = \int 1_{A_n} Y \, dP.$$

X and Y are P–integrable, and $1_{A_n} \leq 1$, so that the dominated convergence theorem entails that

$$\int 1_A X \, dP = \int 1_A Y \, dP,$$

so that $A \in \mathcal{D}$. Thus $\mathcal{D}$ is a d-system, and by the monotone class theorem, theorem 2.4, we get that $\mathcal{D} \supset \sigma(\mathcal{E}_0) = \mathcal{A}_0$. □

Notation For $A \in \mathcal{A}$, and any positive or integrable random variable R, write

$$E(R; A) = E\big(R\, 1_A\big). \tag{5.4}$$

With this notation we can conveniently write equation (5.1) as

$$E(Y; A) = E(X; A), \qquad A \in \mathcal{A}_0. \tag{5.5}$$

5.5 Definition Assume that $\mathcal{A}_0$ is a sub–σ–algebra of $\mathcal{A}$, and that $A \in \mathcal{A}$. Then $E(1_A \,|\, \mathcal{A}_0)$ is called the *conditional probability of A given $\mathcal{A}_0$*, and it is denoted by $P(A \,|\, \mathcal{A}_0)$.

5.6 Exercise Suppose that $A,\, B \in \mathcal{A}$, and that $P(B) \in (0, 1)$. Let $\mathcal{A}_0$ be the σ–algebra generated by B. Show that the random variable $P(A \,|\, \mathcal{A}_0)$ has on B the value $P(A \,|\, B)$, while on $\complement B$ the value $P(A \,|\, \complement B)$, where $P(A \,|\, B)$ and $P(A \,|\, \complement B)$ are defined in the usual elementary way.

5.2 Properties of the Conditional Expectation

Since the conditional expectation $E(X \,|\, \mathcal{A}_0)$ is a random variable, statements about the conditonal expectation only make sense P–a.s. For convenience we omit the "P–a.s." everywhere below.

Throughout this section we fix a sub–σ–algebra $\mathcal{A}_0$ of $\mathcal{A}$. Moreover, whenever a random variable appears in a conditional expectation it is assumed to be positive or integrable.

The statements of the first theorem below follow directly from the definition of the conditional expectation and its a.s. uniqueness. They can be left as an *exercise* to the reader (the arguments are almost the same as those used in the introductory course [26]).

5.7 Theorem

(a) $E(\cdot \,|\, \mathcal{A}_0)$ *is linear.*

(b) $X \leq Y \Rightarrow E(X \mid \mathcal{A}_0) \leq E(Y \mid \mathcal{A}_0)$.

(c) $E(1 \mid \mathcal{A}_0) = 1$.

(d) $X = Y \Rightarrow E(X \mid \mathcal{A}_0) = E(Y \mid \mathcal{A}_0)$.

(e) $E\big(E(X \mid \mathcal{A}_0)\big) = E(X)$.

(f) *X is $\mathcal{A}_0$–measurable* $\Rightarrow E(X \mid \mathcal{A}_0) = X$.

5.8 Theorem *Suppose that X is a positive or integrable random variable. A positive, respectively integrable, $\mathcal{A}_0$–measurable random variable Y is a version of $E(X \mid \mathcal{A}_0)$, if and only if for all positive, respectively bounded, $\mathcal{A}_0$–measurable random variables Z the equation*

$$E(ZY) = E(ZX) \tag{5.6}$$

holds true.

Proof If (5.6) holds true for all positive, bounded $\mathcal{A}_0$–measurable random variables Z, then it holds in particular for all choices $Z = 1_A$ with $A \in \mathcal{A}_0$. Hence we obtain relation (5.1) for all $A \in \mathcal{A}_0$, and consequently Y is a version of $E(X \mid \mathcal{A}_0)$.

Now suppose that Y is a version of $E(X \mid \mathcal{A}_0)$. Assume first that X and Y are (a.s.) positive. From equation (5.1) and the linearity of the conditional expecatation (cf. theorem 5.7.(b)) we conclude that (5.6) holds true for all $\mathcal{A}_0$–elementary functions, i.e., all Z which are positive linear combinations of indicator functions. For a general positive $\mathcal{A}_0$–measurable Z, there exists a sequence $(Z_n,\, n \in \mathbb{N})$ of $\mathcal{A}_0$–elementary functions, which increases pointwise to Z. An application of the monotone convergence theorem shows that (5.6) holds true for Z.

Next consider the case where X, Y are integrable. If Z is bounded and positive, then — as before — it is pointwise the (increasing) limit of $\mathcal{A}_0$–elementary random variables Z_n for which (5.6) holds true with the same argument as above. Since $|Z_n X| \leq |ZX|$, $|ZX|$ is an integrable majorant which is uniform in n, and similarly $|Z_n Y| \leq |ZY|$, $ZY \in \mathcal{L}^1(P)$. Hence an application of the dominated convergence shows that (5.6) holds true for bounded, positive Z. For general bounded Z, decompose $Z = Z^+ - Z^-$, and apply the preceding argument to both parts. Then also in this case (5.6) follows. □

5.9 Theorem

(a) *Suppose that X and Y are positive or that X and XY are integrable. Suppose furthermore that Y is $\mathcal{A}_0$–measurable. Then $E(XY \mid \mathcal{A}_0) = Y E(X \mid \mathcal{A}_0)$.*

(b) $E(\,\cdot \mid \mathcal{A}_0)$ *is* idempotent, *i.e.,* $E(E(X \mid \mathcal{A}_0) \mid \mathcal{A}_0) = E(X \mid \mathcal{A}_0)$.

(c) $E(\,\cdot \mid \mathcal{A}_0)$ *is* symmetric*: if X and Y are positive or X, Y are square integrable, then*

$$E\big(E(X \mid \mathcal{A}_0)\, Y\big) = E\big(X\, E(Y \mid \mathcal{A}_0)\big) \tag{5.7}$$

holds.

(d) *Suppose that* $\mathcal{A}_1 \subset \mathcal{A}_0$. *Then*

$$E(X \mid \mathcal{A}_1) = E\big(E(X \mid \mathcal{A}_0) \mid \mathcal{A}_1\big) = E\big(E(X \mid \mathcal{A}_1) \mid \mathcal{A}_0\big) \tag{5.8}$$

holds true.

(e) *If* X *is independent of* $\mathcal{A}_0$, *then* $E(X \mid \mathcal{A}_0) = E(X)$.

Proof
(a): Without loss of generality we consider $X, Y, Z \geq 0$, Y and Z $\mathcal{A}_0$–measurable. By theorem 5.8

$$\begin{aligned} E\big(E(XY \mid \mathcal{A}_0)\, Z\big) &= E\big((XY)\, Z\big) \\ &= E\big(X\,(YZ)\big) \\ &= E\big(E(X \mid \mathcal{A}_0)\, YZ\big) \end{aligned}$$

where we used theorem 5.8 once more.

(b): Since $E(X \mid \mathcal{A}_0)$ is $\mathcal{A}_0$–measurable, we obtain the statement by replacing X with $E(X \mid \mathcal{A}_0)$ in (f) of theorem 5.7.

(c): For simplicity we consider again the case where $X, Y \geq 0$. We use (a) and (e) in theorem 5.7 to compute as follows

$$\begin{aligned} E\big(E(X \mid \mathcal{A}_0)\, Y\big) &= E\big(E\big(E(X \mid \mathcal{A}_0)\, Y \bigm| \mathcal{A}_0\big)\big) \\ &= E\big(E(X \mid \mathcal{A}_0)\, E(Y \mid \mathcal{A}_0)\big) \\ &= E\big(E\big(X\, E(Y \bigm| \mathcal{A}_0) \mid \mathcal{A}_0\big)\big) \\ &= E\big(X\, E(Y \bigm| \mathcal{A}_0)\big) \end{aligned}$$

(d): Let Z be bounded and $\mathcal{A}_1$–measurable. Note that Z is then also $\mathcal{A}_0$–measurable. With theorem 5.7 (e) and (f) we can compute as follows

$$\begin{aligned} E(ZX) &= E\big(ZE(X \mid \mathcal{A}_0)\big) \\ &= E\big(ZE\big(E(X \mid \mathcal{A}_0) \mid \mathcal{A}_1\big)\big). \end{aligned}$$

On the other hand, from theorem 5.8 we get

$$E(ZX) = E\big(ZE(X \mid \mathcal{A}_1)\big),$$

and the fact that $E(X \mid \mathcal{A}_1)$ is uniquely determined by this equation and its $\mathcal{A}_1$–measurability. Thus we find

$$E(X \mid \mathcal{A}_1) = E\big(E(X \mid \mathcal{A}_0) \mid \mathcal{A}_1\big).$$

Since $E(X \mid \mathcal{A}_1)$ is $\mathcal{A}_1$– and therefore $\mathcal{A}_0$–measurable, we also obtain

$$E(X \mid \mathcal{A}_1) = E\big(E(X \mid \mathcal{A}_1) \mid \mathcal{A}_0\big),$$

by theorem 5.7.(f).

(e): Let Z be bounded, $\mathcal{A}_0$–measurable. Then X and Z are independent, and hence by linearity

$$E(XZ) = E(X)E(Z) = E\big(E(X)\,Z\big),$$

and the statement follows from theorem 5.8. □

5.10 Remark Statements (b) and (c) actually mean that $E(\cdot\,|\,\mathcal{A}_0)$ is the orthogonal projection onto the (closed) subspace $L^2(\Omega, \mathcal{A}_0, P)$ of $L^2(\Omega, \mathcal{A}, P)$. (However, this is not quite clean: In order to consider $L^2(\Omega, \mathcal{A}_0, P)$ as a subspace of $L^2(\Omega, \mathcal{A}, P)$ one has to suppose that $\mathcal{A}_0$ is augmented by all P–null sets in $\mathcal{A}$. But most of the time this can be assumed without any harm.) Therefore, $E(X|\mathcal{A}_0)$ is the minimizer of the L^2–distance of X to $L^2(\Omega, \mathcal{A}_0, P)$ — a result which is the basis for the method of regression, and which is known from the introductory course [26]. (And here we have identified X with its class in $L^2(\Omega, \mathcal{A}, P)$ for simplicity.)

5.11 Theorem

(a) **(Monotone Convergence Theorem for Conditional Expectations)** *Suppose that $(X_n,\ n \in \mathbb{N})$ is a sequence of non-negative random variables which a.s. increases to a random variable X. Then $E(X_n\,|\,\mathcal{A}_0)$ a.s. increases to $E(X\,|\,\mathcal{A}_0)$.*

(b) **(Dominated Convergence Theorem for Conditional Expectations)** *Suppose that $(X_n,\ n \in \mathbb{N})$ is a sequence which converges a.s. to a random variable X, and that there exists an integrable random variable Y so that $|X_n| \le Y$ a.s. for all $n \in \mathbb{N}$. Then $E(X_n\,|\,\mathcal{A}_0)$ converges a.s. to $E(X\,|\,\mathcal{A}_0)$.*

Proof

(a): By theorem 5.7.(b) also the sequence $\big(E(X_n\,|\,\mathcal{A}_0),\ n \in \mathbb{N}\big)$ is a.s. increasing. Define

$$Y = \sup_n E(X_n\,|\,\mathcal{A}_0),$$

so that a.s. $E(X_n\,|\,\mathcal{A}_0)$ increases to Y. As a pointwise supremum of $\mathcal{A}_0$–measurable non-negative random variables, Y is $\mathcal{A}_0$–measurable, and $Y \ge 0$. Let $A \in \mathcal{A}_0$, then

$$\begin{aligned}
\int_A Y\,dP &= \int_A \sup_n E(X_n\,|\,\mathcal{A}_0)\,dP \\
&= \sup_n \int_A E(X_n\,|\,\mathcal{A}_0)\,dP \\
&= \sup_n \int_A X_n\,dP \\
&= \int_A \sup_n X_n\,dP \\
&= \int_A X\,dP,
\end{aligned}$$

where we have twice used of the monotone convergence theorem. Thus we have proved that $Y = E(X \mid \mathcal{A}_0)$.

(b): For $n \in \mathbb{N}$ set

$$\hat{X}_n = \sup_{k \geq n} X_k, \quad \check{X}_n = \inf_{k \geq n} X_k.$$

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

$$-Y \leq \check{X}_n \leq X_n \leq \hat{X}_n \leq Y.$$

Moreover, the sequences $(\check{X}_n,\ n \in \mathbb{N})$, $(\hat{X}_n,\ n \in \mathbb{N})$ are monotone increasing, respectively decreasing. By hypothesis we get (a.s.)

$$\begin{aligned}
\inf_n \hat{X}_n &= \inf_n \sup_{k \geq n} X_k = \limsup_n X_n = \lim_n X_n = X, \\
\sup_n \check{X}_n &= \sup_n \inf_{k \geq n} X_k = \liminf_n X_n = \lim_n X_n = X.
\end{aligned}$$

The sequences $(Y - \hat{X}_n,\ n \in \mathbb{N})$, $(Y + \check{X}_n,\ n \in \mathbb{N})$ are non-negative and monotone increasing, and they (a.s) increase to $Y - X$, $Y + X$ respectively. Thus part (a) gives (a.s.)

$$\begin{aligned}
\lim_n E(Y - \hat{X}_n \mid \mathcal{A}_0) &= E(Y - X \mid \mathcal{A}_0), \\
\lim_n E(Y + \check{X}_n \mid \mathcal{A}_0) &= E(Y + X \mid \mathcal{A}_0).
\end{aligned}$$

Thus we find (a.s.)

$$\lim_n E(\hat{X}_n \mid \mathcal{A}_0) = E(X \mid \mathcal{A}_0), \quad \lim_n E(\check{X}_n \mid \mathcal{A}_0) = E(X \mid \mathcal{A}_0).$$

On the other hand, $\check{X}_n \leq X_n \leq \hat{X}_n$, and theorem 5.7.(b) implies

$$E(\check{X}_n \mid \mathcal{A}_0) \leq E(X_n \mid \mathcal{A}_0) \leq E(\hat{X}_n \mid \mathcal{A}_0), \qquad n \in \mathbb{N}.$$

But then $\lim_n E(X_n \mid \mathcal{A}_0) = E(X \mid \mathcal{A}_0)$ a.s., as claimed. □

There also is a theorem of Fubini–Tonelli type for conditional expectations, see theorem E.1. Since its formulation and its proof need some special care, and since it does not play a central role in these lectures, it is deferred to appendix E.

A function φ defined on an interval $I \subset \mathbb{R}$ is called *convex*, if every line segment between two points of the graph of φ lies above the graph, or — what is the same — if for any $x, y \in I$, $t \in [0, 1]$, the inequality $\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$ holds. The following theorem is essentially based on the fact that at every point x in the interior of I, the right and left tangents at $(x, \varphi(x))$ (which exist) are below the graph of φ. We state the theorem without proof, instead refer to [3, Satz 15.3] or [6, Theorem 9.1.4].

5.12 Theorem (Jensen's Inequality for Conditional Expectations) *Assume that φ is a convex function on $\mathbb{R}$, and that X and $\varphi \circ X$ are positive or integrable. Then*

$$\varphi\big(E(X \mid \mathcal{A}_0)\big) \le E(\varphi \circ X \mid \mathcal{A}_0) \tag{5.9}$$

holds true.

5.13 Remark By choosing the trivial sub–σ–algebra $\mathcal{A}_0 = \{\emptyset, \Omega\}$, we obtain Jensen's inequality in the form $\varphi(E(X)) \le E(\varphi(X))$.

Notation If $(Y_i,\ i \in I)$ is a family of random variables, Y_i, $i \in I$, taking values in a measurable space $(E_i, \mathcal{E}_i)$, and if $\mathcal{A}_0 = \sigma(Y_i,\ i \in I)$, then we write $E(X \mid Y_i,\ i \in I)$ for $E(X \mid \mathcal{A}_0)$.

5.3 Conditional Densities

5.14 Lemma *Suppose that $(E, \mathcal{E})$ is a measurable space, and that Y is an E–valued random variable. Suppose furthermore that X is positive or integrable. Then there exists a measurable function g on E so that for all $B \in \mathcal{E}$*

$$\int_{\{Y \in B\}} X \, dP = \int_B g \, dP_Y. \tag{5.10}$$

Moreover, g is P_Y–a.s. unique. Conversely, if there exists a measurable function g on E so that for all $B \in \mathcal{E}$ equation (5.10) *holds, then $g \circ Y$ is a version of $E(X \mid Y)$.*

Proof By the factorization lemma F.1, there exists a measurable function g on E so that $E(X \mid Y) = g \circ Y$. Therefore by the definition of the conditional expectation we get for $B \in \mathcal{E}$,

$$\begin{aligned}
\int_{\{Y \in B\}} X \, dP &= \int_{\{Y \in B\}} E(X \mid Y) \, dP \\
&= \int_{\{Y \in B\}} g \circ Y \, dP \\
&= \int_B g \, dP_Y,
\end{aligned}$$

where the last equality follows from the transformation theorem for Lebesgue integrals. P_Y–a.s. uniqueness is easy, and left to the reader as an *exercise*. Conversely, assume that (5.10) holds true for all $B \in \mathcal{E}$. Then by the last calculation we find that $g \circ Y$ is a version of the conditional expectation $E(X \mid Y)$. □

Notation Suppose that X, Y, and g are as above, then we define

$$E(X \mid Y = y) = g(y), \qquad y \in E.$$

With this notation, equation (5.10) reads as follows

$$\int_{\{Y \in B\}} X \, dP = \int_B E(X \mid Y = y) \, dP_Y(y). \tag{5.11}$$

And the usual routine arguments yield for a bounded measurable function f the formula

$$E\big(f(Y)X\big) = \int f(y) \, E(X \mid Y = y) \, dP_Y(y). \tag{5.12}$$

5.15 Exercise Suppose that X, Y are two real valued random variables which have a joint density $\varphi(x, y)$, $x, y \in \mathbb{R}$. Suppose furthermore that X is integrable, and that the marginal density φ_Y of Y is strictly positive on $\mathbb{R}$. Show that P_Y–a.s.,

$$g(y) = \int_{\mathbb{R}} x\varphi(x|y) \, dx$$

where $\varphi(x|y)$ is the conditional density (cf. also, e.g., [26, Kapitel 5])

$$\varphi(x|y) = \frac{\varphi(x, y)}{\varphi_Y(y)}.$$

Chapter 6

Filtrations and Stopping Times

An important concept which enters the definition of a martingale, and as well that of a Markov process is that of a *filtration*. It is the probabilistic formulation of the idea of information or knowledge which is accumulated as time goes by.

6.1 Filtrations

Throughout this chapter we fix some probability space $(\Omega, \mathcal{A}, P)$ ($(\Omega, \mathcal{A})$ large enough so that all following concepts, processes etc. can be defined thereon). As before, T will be a time parameter domain, most typically (but not necessarily) $T = \mathbb{N}_0$ or $T = \mathbb{R}_+$.

6.1 Definition A *filtration* $\mathcal{F} = (\mathcal{F}_t, t \in T)$ is an increasing family of σ–algebras over Ω: $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s, t \in T$, $s \leq t$. The quadruple $(\Omega, \mathcal{A}, \mathcal{F}, P)$ is called a *filtered probability space*.

6.2 Remarks If nothing else is said, we shall always assume that the σ–algebras $\mathcal{F}_t$, $t \in T$, are sub–σ–algebras of $\mathcal{A}$. The σ–algebra over Ω generated by $\mathcal{F}$ is denoted by $\mathcal{F}_\infty$. That is, $\mathcal{F}_\infty$ is the smallest σ–algebra over Ω containing all $\mathcal{F}_t$, $t \in T$.

6.3 Definition Assume that $X = (X_t, t \in T)$ is an $(E, \mathcal{E})$–valued stochastic process, and that $\mathcal{F} = (\mathcal{F}_t, t \in T)$ is a filtration over Ω. X is called *adapted to* $\mathcal{F}$, if for every $t \in T$, X_t is $\mathcal{F}_t$–measurable.

For the next definition we assume that $T = \mathbb{R}_+$, and we consider $X = (X_t, t \in \mathbb{R}_+$ as a mapping from $T \times \Omega$ to E. Temporarily we denote this mapping as $\hat{X}$:

$$\begin{aligned} \hat{X} : T \times \Omega &\to E, \\ (t, \omega) &\mapsto X_t(\omega). \end{aligned}$$

6.4 Definition Assume that $X = (X_t,\ t \in \mathbb{R}_+)$ is an $(E, \mathcal{E})$–valued stochastic process. X is called *measurable*, if the mapping $\hat{X}$ is $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{A}/\mathcal{E}$–measurable. If $\mathcal{F} = (\mathcal{F}_t,\ t \in \mathbb{R}_+)$ is a filtration over Ω, then X is called *progressively measurable relative to* $\mathcal{F}$, if for every $t \in \mathbb{R}_+$, the restriction of the mapping $\hat{X}$ from $T \times \Omega$ into E to $[0,t] \times \Omega$ is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t/\mathcal{E}$–measurable.

6.5 Exercise Check that if a process is progressive with respect to a filtration $\mathcal{F}$ of σ–algebras of $\mathcal{A}$, then it is measurable.

6.6 Remark If the underlying filtration is understood from the context we shall just say that X is adapted, respectively progressive.

Thus if $T = \mathbb{R}_+$, then X being adapted to $\mathcal{F}$ means that for any given $t \in \mathbb{R}_+$, $B \in \mathcal{E}$,

$$\{\omega \in \Omega,\ X_t(\omega) \in B\} \in \mathcal{F}_t$$

while X is progressive if

$$\{(s,\omega) \in [0,t] \times \Omega,\ X_s(\omega) \in B\} \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t.$$

6.7 Exercise

(a) Use Fubini's theorem to prove that if $X = (X_t,\ t \in \mathbb{R}_+)$ is progressive relative to $\mathcal{F} = (\mathcal{F}_t,\ t \in \mathbb{R}_+)$ then it is also adapted.

(b) Define the notion of progressive measurability for the case that $T = \mathbb{N}_0$, and show that "X adapted" is equivalent to "X progressive".

Suppose that $X = (X_t,\ t \in \mathbb{R}_+)$ is a stochastic process with values in $\mathbb{R}^d$, $d \in \mathbb{N}$. Then a convenient criterium for X being measurable, respectively $\mathcal{F}$–progressive, is given in the following

6.8 Lemma *Assume that $X = (X_t,\ t \in \mathbb{R}_+)$ is a stochastic process with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $d \in \mathbb{N}$, and with left- or right-continuous paths. Then X is measurable. If in addition X is adapted to a filtration $\mathcal{F} = (\mathcal{F}_t,\ t \in \mathbb{R}_+)$, then X is progressive with respect to $\mathcal{F}$.*

Proof We only prove the second statement, the proof of measurability is similar but easier. So assume that X is $\mathcal{F}$–adapted. Consider the case that the paths of X are continuous from the right, the other case is proved in an analogous way. We have to show that for every $t > 0$, the restriction of X to $[0,t]\times\Omega$ is $(\mathcal{B}([0,t])\otimes\mathcal{F}_t)/\mathcal{B}(\mathbb{R}^d)$–measurable. Fix $t > 0$, and for the rest of the proof identify X with its restriction to $[0,t]\times\Omega$. Let $a, b \le t$ with $a < b$, and define a stochastic process $Y = (Y_s,\ s \in [0,t])$ by $Y_s = 1_{(a,b]}(s)\, X_b$. We show that Y, as a mapping from $[0,t] \times \Omega$ into $\mathbb{R}^d$, is $(\mathcal{B}([0,t]) \otimes \mathcal{F}_t)/\mathcal{B}(\mathbb{R}^d)$–measurable. Indeed, for $B \in \mathcal{B}(\mathbb{R}^d)$ we get

$$Y^{-1}(B) = \begin{cases} (a,b] \times X_b^{-1}(B) \cup \big([0,t] \setminus (a,b]\big) \times \Omega, & \text{if } 0 \in B, \\ (a,b] \times X_b^{-1}(B), & \text{otherwise.} \end{cases}$$

Obviously, these sets belong to $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$, because X_b is $\mathcal{F}_b$–measurable, and, since $b \le t$, therefore also $\mathcal{F}_t$–measurable. Now define for $n \in N$, $X_0^n = X_0$, and for $s \in (0,t]$,

$$X_s^n = \sum_{k=0}^{2^n-1} X_{(k+1)t/2^n} 1_{(kt/2^n,(k+1)t/2^n]}(s).$$

Then X^n is a sum of processes of the same form as Y above, each of which being $(\mathcal{B}([0,t]) \otimes \mathcal{F}_t)/\mathcal{B}(\mathbb{R}^d)$–measurable. Therefore X^n is $(\mathcal{B}([0,t]) \otimes \mathcal{F}_t)/\mathcal{B}(\mathbb{R}^d)$–measurable. But due to the continuity to right of X, X^n converges pointwise on $[0,t] \times \Omega$ to X as $n \to \infty$. Thus, X is $(\mathcal{B}([0,t]) \otimes \mathcal{F}_t)/\mathcal{B}(\mathbb{R}^d)$–measurable. □

Very often a filtration $\mathcal{F} = (\mathcal{F}_t,\ t \in T)$ is generated by a stochastic process:

6.9 Definition Suppose that $X = (X_t,\ t \in T)$ is an $(E, \mathcal{E})$–valued stochastic process. Then the *filtration generated by* X is the filtration $\mathcal{F}^X = (\mathcal{F}_t^X,\ t \in T)$ consisting of the increasing family of smallest σ–algebras $\mathcal{F}_t^X$ so that X becomes adapted to $\mathcal{F}^X$. It is also called the *natural filtration of* X.

Fix $t \in T$, and suppose that $s \le t$. Then X_s is $\mathcal{F}_s^X/\mathcal{E}$–measurable, and since $\mathcal{F}_s^X \subset \mathcal{F}_t^X$, X_s is also $\mathcal{F}_t^X/\mathcal{E}$–measurable. Thus, we have $\mathcal{F}_t^X = \sigma(X_s,\ s \le t)$, $t \in T$, and this is the smallest σ–algebra over Ω so that for every $s \le t$, X_s is $\mathcal{F}_t^X/\mathcal{E}$–measurable. Clearly, $\mathcal{F}^X$ is a filtration of sub–σ–algebras of $\mathcal{A}$. Moreover, the statement that X is adapted to $\mathcal{F}$ is equivalent to $\mathcal{F}^X \subset \mathcal{F}$, where the last inclusion means by definition that $\mathcal{F}_t^X \subset \mathcal{F}_t$ for every $t \in T$. From lemma 3.12 we immediately get

$$\mathcal{F}_t^X = \sigma\big((X_{t_0}, \dots, X_{t_n}),\ n \in \mathbb{N}_0,\ t_0 < t_1 < \dots < t_n \le t\big). \tag{6.1}$$

Now suppose that $(\varphi_n,\ n \in \mathbb{N})$ is a sequence of bijective, bi-measurable mappings from E^{n+1} onto itself. Then for every $J \in \mathcal{J}_0$ of the form $J = \{t_0, t_1, \dots, t_n\}$ the mapping X_J from Ω into E^J is $\mathcal{A}/\mathcal{E}^J$–measurable, if and only if $\varphi_n \circ X_J$ or equivalently $\varphi_n^{-1} \circ X_J$ is $\mathcal{A}/\mathcal{E}^J$–measurable: Suppose that $B \in \mathcal{E}^J$, and that $\varphi_n \circ X_J$ is $\mathcal{A}/\mathcal{E}^J$–measurable. Set $C = \varphi_n(B) \in \mathcal{E}^J$, then

$$X_J^{-1}(B) = (\varphi_n \circ X_J)^{-1}(C) \in \mathcal{A},$$

and similarly for $\varphi_n^{-1} \circ X_J$. The converse is easy.

For the remainder of this section let assume that $E = \mathbb{R}^d$, $\mathcal{E} = \mathcal{B}(\mathbb{R}^d)$, $d \in \mathbb{N}$. For φ_n, φ_n^{-1} choose the mappings (3.41b) and (3.45) from chapter 3. Then, the last argument combined with equation (6.1) gives us the following

6.10 Lemma *Suppose that* $X = (X_t,\ t \in T)$ *is an* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$*–valued stochastic process,* $d \in \mathbb{N}$*. Then for every* $t \in T$

$$\mathcal{F}_t^X = \sigma\big((X_{t_0}, X_{t_1} - X_{t_0} \dots, X_{t_n} - X_{t_{n-1}}),\ n \in \mathbb{N}_0,\ t_0 < t_1 < \dots < t_n \le t\big) \tag{6.2}$$

holds.

Usually, and as we have done in chapter 3, we shall set $t_0 = 0$. Then we have shown that in the case where we consider an $\mathbb{R}^d$–valued stochastic process X, the filtration generated by X is the same as the filtration generated by the increments of X. We shall simplify the notation a bit by henceforth writing

$$\sigma\big(X_{t_0}, X_{t_1} - X_{t_0} \dots, X_{t_n} - X_{t_{n-1}},\ n \in \mathbb{N}_0,\ t_0 < t_1 < \cdots < t_n \le t\big)$$

for the right hand side of equation (6.2).

6.2 Stopping Times

In this section we introduce another central concept of the theory of stochastic processes, the idea of a random time which respects the flow of information as time passes by. We base our discussion on a general time parameter domain T, and on a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$. $\overline{T}$ denotes the set $T \cup \{+\infty\}$.

6.11 Definition A *stopping time* τ is a random variable with values in $\overline{T}$, such that for every $t \in T$ the event $\{\tau \le t\}$ belongs to $\mathcal{F}_t$.

6.12 Remark Some authors (e.g., [3]) also consider the condition $\{\tau < t\} \in \mathcal{F}_t$, but here we shall not do so.

6.13 Exercise

(a) Show that every constant time $t_0 \in T$ is a stopping time, and that if $t_0 \in T$, and τ is a stopping time, then $\tau + t_0$ is a stopping time.

(b) Show that if τ, σ are stopping times, then $\tau \vee \sigma$ and $\tau \wedge \sigma$ are stopping times.

(c) Show that for every $t \in \mathbb{R}_+$, and $B \in \mathcal{B}([0,t])$ one has $\{\tau \in B\} \in \mathcal{F}_t$. (Hint: Use lemma 3.6, and choose a suitable generator of $\mathcal{B}([0,t])$.)

In the next exercise the reader proves a technical device which is used in many proofs involving stopping times:

6.14 Exercise Assume that $T = \mathbb{R}_+$, and let τ be a stopping time relative to $\mathcal{F}$. Then there exists a sequence $(\tau_n,\ n \in \mathbb{N})$ of stopping times which decreases to τ, and is such that for every $n \in \mathbb{N}$, the range of τ_n is finite.

6.15 Example Suppose that $X = (X_t,\ t \in \mathbb{R}_+)$ is a stochastic process with values in a metric space (E, d), equipped with its Borel–σ–algebra, which has exclusively continuous paths. Let $A \subset E$ be closed, and define

$$\tau_A(\omega) = \inf\{t \ge 0,\ X_t(\omega) \in A\}, \qquad \omega \in \Omega, \tag{6.3}$$

where we make the convention that $\inf \emptyset = +\infty$. τ_A is called the *entry time of* A. We show that τ_A is a stopping time relative to the natural filtration $\mathcal{F}^X$ of X.

Fix $t \in \mathbb{R}_+$. We have to show that $\{\tau_A \le t\} \in \mathcal{F}_t^X$. Consider first the case $t = 0$, that is $\tau_A = 0$. So let $\omega \in \Omega$ be such that $\tau_A(\omega) = 0$. By definition of τ_A as the infimum of the set on the right hand side of (6.3), we find that there exists a sequence $(t_n,\, n \in \mathbb{N})$ decreasing to zero, such that for every $n \in \mathbb{N}$, $X_{t_n}(\omega) \in A$. By hypothesis, $t \mapsto X_t(\omega)$ is continuous, and therefore we get that $X_{t_n}(\omega) \to X_0(\omega)$ as $n \to \infty$. Since A is closed we obtain that $X_0(\omega) \in A$. Thus we have proved $\{\tau_A = 0\} \subset X_0^{-1}(A)$. On the other we clearly have that $X_0^{-1}(A) \subset \{\tau_A = 0\}$, so that $\{\tau_A = 0\} = \{\tau_A \le 0\} = X_0^{-1}(A) \in \mathcal{F}_0^X$.

Now assume that $t > 0$. First we remark that since A is closed, we can write

$$\tau_A(\omega) = \inf\{t \ge 0,\, d(X_t(\omega), A) = 0\}, \qquad \omega \in \Omega,$$

where $d(x, B) = \inf\{d(x, b),\, b \in B\}$ for $x \in E$, $B \subset E$. Note that $x \mapsto d(x, B)$, $B \subset E$, is continuous on E. We claim that

$$\{\tau_A \le t\} = \big\{\omega \in \Omega,\, \exists\, s \in [0, t] : d(X_s(\omega), A) = 0\big\}. \tag{6.4}$$

That the right hand side is a subset of the set on the left hand side is obvious. Assume that ω does not belong to the set on the right. To prove our claim we show that then $\tau_A(\omega) > t$. Indeed, we get that for all $s \in [0, t]$, $d(X_s(\omega), A) > 0$. Since $s \mapsto d(X_s(\omega), A)$ is continuous, and real valued continous functions have a minimum on every compact, we get that there exists $\varepsilon > 0$ so that $d(X_s(\omega), A) \ge \varepsilon$ for all $s \in [0, t]$. Again by the continuity of $s \mapsto d(X_s(\omega), A)$ we find that there exists $\delta > 0$ so that $d(X_s(\omega), A) \ge \varepsilon/2 > 0$ for all $s \in [0, t + \delta]$. Therefore we get that $\tau_A(\omega) \ge t + \delta > t$, and our claim is proved. Next we show that

$$\{\tau_A \le t\} = \bigcap_{n \in \mathbb{N}} \bigcup_{s \in [0,t] \cap \mathbb{Q}} \Big\{d(X_s, A) \le \frac{1}{n}\Big\} \tag{6.5}$$

Suppose that ω belongs to the right hand side of (6.5). Then this implies that there exists a sequence $(s_n,\, n \in \mathbb{N})$ in $[0, t] \cap \mathbb{Q}$ so that $d(X_{s_n}(\omega), A)$ converges to zero as $n \to \infty$. As a sequence in the compact interval $[0, t]$, $(s_n,\, n \in \mathbb{N})$ has a convergent subsequence, which we denote again by $(s_n,\, n \in \mathbb{N})$, i.e., that there exists $s \in [0, t]$ so that $s_n \to s$ with $n \to \infty$. By continuity of $s \mapsto d(X_s(\omega), A)$ we therefore obtain $d(X_s(\omega), A)) = 0$. Hence $\tau_A(\omega) \le s \le t$. For the converse inclusion suppose that $\tau_A(\omega) \le t$. From (6.4) we find that there exists $s \in [0, t]$ so that $d(X_s(\omega), A) = 0$. The continuity of $s \mapsto d(X_s(\omega), A)$ entails that for every $n \in \mathbb{N}$, there exists $\delta > 0$ so that for every $s' \in (s - \delta, s + \delta) \cap [0, t]$, we have $d(X_{s'}(\omega), A) \le 1/n$. But the set $(s - \delta, s + \delta) \cap [0, t]$ contains a rational point, and therefore ω belongs to set on the right hand side of (6.5).

But now we are done: $x \mapsto d(x, A)$ is continuous and therefore measurable, so that the random variable $d(X_s, A)$ is $\mathcal{F}_s^X$–measurable. Consequently it is also $\mathcal{F}_t^X$–measurable for all $s \le t$. Thus on the right hand side of equation (6.5) we have a set belonging to $\mathcal{F}_t^X$. $\diamond$

6.16 Exercise Suppose that $X = (X_n,\, n \in \mathbb{N}_0)$ is a discrete time stochastic process with values in E, $(E, \mathcal{E})$ a measurable space. Prove by an elementary argument (i.e., without using example 6.15) that for every $A \in \mathcal{E}$ the entry time $\tau_A = \inf\{n \in \mathbb{N}_0,\ X_n \in A\}$ is a stopping time relative to $\mathcal{F}^X$.

Consider a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$ with any time parameter domain T. Suppose that τ is an $\mathcal{F}$–stopping time, and recall that $\mathcal{F}_\infty$ is the σ–algebra over Ω generated by the family $(\mathcal{F}_t,\, t \in T)$. Define

$$\mathcal{F}_\tau = \big\{A \in \mathcal{F}_\infty,\ \forall\, t \in T : A \cap \{\tau \le t\} \in \mathcal{F}_t\big\}. \tag{6.6}$$

6.17 Lemma *$\mathcal{F}_\tau$ is a σ–algebra over Ω.*

Proof $\Omega \in \mathcal{F}_\infty$, and for $t \in T$, $\Omega \cap \{\tau \le t\} = \{\tau \le t\} \in \mathcal{F}_t$, so that $\Omega \in \mathcal{F}_\tau$. Next suppose that $A \in \mathcal{F}_\tau$, then $\complement A \in \mathcal{F}_\infty$, and for $t \in T$,

$$\complement A \cap \{\tau \le t\} = \{\tau \le t\} \setminus (A \cap \{\tau \le t\}) \in \mathcal{F}_t,$$

and we find that $\complement A \in \mathcal{F}_\tau$. Finally let $(A_n,\, n \in \mathbb{N})$ be a sequence in $\mathcal{F}_\tau$. Then also $\cup_n A_n \in \mathcal{F}_\infty$, and for $t \in T$,

$$\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) \cap \{\tau \le t\} = \bigcup_{n \in \mathbb{N}} (A_n \cap \{\tau \le t\}).$$

The union on the right hand side is a union of sets in $\mathcal{F}_t$, and therefore it belongs to $\mathcal{F}_t$. Thus $\cup_n A_n \in \mathcal{F}_\tau$. □

6.18 Exercise Prove that τ is measurable with respect to $\mathcal{F}_\tau$.

6.19 Remark $\mathcal{F}_\tau$ is interpreted as the information available up to the random time τ.

The following alternative representation of $\mathcal{F}_\tau$ is sometimes useful:

$$\mathcal{F}_\tau = \big\{A \subset \Omega,\ \forall\, t \in \overline{T} : A \cap \{\tau \le t\} \in \mathcal{F}_t\big\}. \tag{6.7}$$

We leave the proof as an *exercise* to the reader.

6.20 Lemma *Suppose that σ, τ are two $\mathcal{F}$–stopping times, such that $\sigma \le \tau$. Then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.*

Proof If $A \in \mathcal{F}_\sigma$, and $t \in T$, then $\tau \le t$ implies $\sigma \le t$, and therefore

$$A \cap \{\tau \le t\} = (A \cap \{\sigma \le t\}) \cap \{\tau \le t\}.$$

Now $A \cap \{\sigma \le t\} \in \mathcal{F}_t$, and also $\{\tau \le t\} \in \mathcal{F}_t$, so that indeed $A \cap \{\tau \le t\} \in \mathcal{F}_t$, and therefore $A \in \mathcal{F}_\tau$. □

Assume that we are given in addition a stochastic process $X = (X_t,\, t \in T)$. We would like to define the value X_τ of X at the random time τ or the *stopped process* $t \mapsto X_{t\wedge\tau}$. In the first case we face the problem that τ might be equal $+\infty$, and X_∞ is not defined. Thus from now on we suppose that Δ is some ideal point, separate from the state space E, and we put $X_\infty(\omega) = \Delta$, $\omega \in \Omega$. A σ–algebra $\overline{\mathcal{E}}$ on $E \cup \{\Delta\}$ which is a minimal extension of $\mathcal{E}$ can be defined as follows: A set A in $\overline{\mathcal{E}}$ is either a set in $\mathcal{E}$ or a set of the form $A' \cup \{\Delta\}$ with $A' \in \mathcal{E}$. The case of discrete time parameter domain is easy:

6.21 Exercise Suppose that $X = (X_n,\, n \in \overline{\mathbb{N}}_0)$ is a stochastic process on a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$ with values in E, $(E, \mathcal{E})$ a measurable space. Assume furthermore that τ is an $\mathcal{F}$–stopping time.

(a) Show that $(X_\tau)(\omega) = (X_{\tau(\omega)})(\omega)$, $\omega \in \Omega$, defines a random variable.

(b) Assume in addition that X is adapted to $\mathcal{F}$. Prove that X_τ is $\mathcal{F}_\tau$–measurable, and that $(X_{n\wedge\tau},\, n \in \overline{\mathbb{N}}_0)$ is adapted to the filtration $(\mathcal{F}_{n\wedge\tau},\, n \in \overline{\mathbb{N}}_0)$.

The case of a time parameter domain which is uncountable is more involved, and we shall treat $T = \mathbb{R}_+$ below. To keep the notation simple, in the sequel we shall not distinguish between $\mathcal{E}$ and $\overline{\mathcal{E}}$, nor between E and $E \cup \{\Delta\}$, and similarly for $\overline{\mathbb{R}}_+$, $\mathbb{R}_+$, $\overline{\mathbb{N}}_0$, $\mathbb{N}_0$. Most of the time there will be no danger of confusion.

6.22 Lemma *Assume that $X = (X_t,\, t \in \mathbb{R}_+)$ is a stochastic process, $\mathcal{F} = (\mathcal{F}_t,\, t \in \mathbb{R}_+)$ is a filtration, and τ is an $\mathcal{F}$–stopping time.*

(a) *If X is measurable (see definition 6.4), then X_τ, defined by*

$$X_\tau(\omega) = (X_{\tau(\omega)})(\omega), \qquad \omega \in \Omega,$$

is a random variable.

(b) *If X is progressive relative to $\mathcal{F}$ (see definition 6.4), then X_τ restricted to $\{\tau < +\infty\}$ is $\mathcal{F}_\tau$–measurable, and $(X_{t\wedge\tau},\, t \in \mathbb{R}_+)$ is adapted to $(\mathcal{F}_{t\wedge\tau},\, t \in \mathbb{R}_+)$.*

Proof We only prove statement (b), the proof of (a) is similar to (and much simpler than) the proof of the first statement in (b), and can be left as an *exercise* to the reader. For the first statement of (b), we have to prove that for every $A \in \mathcal{E}$ we have that $\{X_\tau \in A, \tau < +\infty\} \in \mathcal{F}_\tau$. To show this, we have to prove that this event belongs to $\mathcal{F}_\infty$, and that for all $s \in \mathbb{R}_+$ the event $\{X_\tau \in A\} \cap \{\tau \le s\}$ belongs to $\mathcal{F}_s$. First we show the second of these statements, and suppose that $s \in \mathbb{R}_+$ and $A \in \mathcal{E}$. Define the following mapping

$$\begin{aligned} {}^s(\tau \otimes \mathrm{id}) : \{\tau \le s\} &\to [0, s] \times \Omega \\ \omega &\mapsto (\tau(\omega), \omega), \end{aligned}$$

where we consider Ω as equipped with the σ–algebra $\mathcal{F}_s$, and $\{\tau \le s\}$ equipped with $\mathcal{F}_s^\tau = \mathcal{F}_s \cap \{\tau \le s\}$. Note that $\mathcal{F}_s^\tau$ is a sub–σ–algebra of $\mathcal{F}_s$, because $\{\tau \le s\} \in \mathcal{F}_s$.

First we show that this mapping is $\mathcal{F}_s^\tau/\mathcal{B}([0,s]) \otimes \mathcal{F}_s$–measurable. By lemma 3.6 it is sufficient to show that each element of a generator of $\mathcal{B}([0,s]) \otimes \mathcal{F}_s$ has a pre-image under ${}^s(\tau \otimes \mathrm{id})$ belonging to $\mathcal{F}_s^\tau$. As a generator we choose $\mathcal{B}([0,s]) \times \mathcal{F}_s$, and consider $B \in \mathcal{B}([0,s])$, $C \in \mathcal{F}_s$. Then

$$ {}^s(\tau \otimes \mathrm{id})^{-1}(B \times C) = \{\tau \in B\} \cap C. $$

With exercise 6.13.(c) we get $\{\tau \in B\} \in \mathcal{F}_s \cap \{\tau \le s\} = \mathcal{F}_s^\tau$, and therefore $\tau \otimes \mathrm{id}$ is $\mathcal{F}_s^\tau/\mathcal{B}([0,s]) \otimes \mathcal{F}_s$–measurable. Now denote the restriction of the mapping $X : \mathbb{R}_+ \times \Omega \to E$ to $[0,s] \times \Omega$ by ${}^s\hat{X}$. Since X is $\mathcal{F}$–progressive, ${}^s\hat{X}$ is $\mathcal{B}([0,s]) \otimes \mathcal{F}_s/\mathcal{E}$–measurable. Thus for every $A \in \mathcal{E}$ we find that

$$ \left({}^s\hat{X} \circ {}^s(\tau \otimes \mathrm{id})\right)^{-1}(A) \in \mathcal{F}_s^\tau \subset \mathcal{F}_s. $$

But

$$ \begin{aligned} \left({}^s\hat{X} \circ {}^s(\tau \otimes \mathrm{id})\right)^{-1}(A) &= {}^s(\tau \otimes \mathrm{id})^{-1}\left({}^s\hat{X}^{-1}(A)\right) \\ &= \left\{\omega \in \{\tau \le s\}, (\tau(\omega), \omega) \in {}^s\hat{X}^{-1}(A)\right\} \\ &= \left\{\omega \in \{\tau \le s\},\ X_{\tau(\omega)}(\omega) \in A\right\} \\ &= \{X_\tau \in A\} \cap \{\tau \le s\}, \end{aligned} $$

and we have just seen that this set belongs to $\mathcal{F}_s$.

Next we show that $\{X_\tau \in A,\ \tau < +\infty\} \in \mathcal{F}_\infty$:

$$ \{X_\tau \in A,\ \tau < +\infty\} = \bigcup_{n \in \mathbb{N}} \left(\{X_\tau \in A\} \cap \{\tau \le n\}\right), $$

and we have just proved that the sets under the union on the right hand side belong to $\mathcal{F}_n \subset \mathcal{F}_\infty$. Hence we find that $\{X_\tau \in A,\ \tau < +\infty\} \in \mathcal{F}_\infty$. Thus X_τ is $\mathcal{F}_\tau$–measurable on $\{\tau < +\infty\}$, and the first statement in (b) is proved.

If we replace in the first statement of (b) the stopping time τ by the (finite) stopping time $t \wedge \tau$, $t \in \mathbb{R}_+$, then we get that for every $t \in \mathbb{R}_+$, $X_{t\wedge\tau}$ is $\mathcal{F}_{t\wedge\tau}$–measurable, proving the second statement of (b). $\square$

Chapter 7

Martingales

One of the three central classes of stochastic processes is the class of martingales, which we shall discuss in this chapter. Martingales are probabilistic models of "fair games", but their importance goes far beyond this: They are present throughout practically all topics of the modern theory of stochastic processes. Throughout this chapter we only consider real-valued stochastic processes.

7.1 Martingales

Suppose that we model a game via a probability space $(\Omega, \mathcal{A}, P)$. Then we may consider a given filtration $\mathcal{F} = (\mathcal{F}_t, t \in T)$ as a model of observations of the game (not necessarily made by one of the gamblers — there might be a person or a device which is able to obtain more information than the gamblers). A *fair* game could be defined as one in which it is in the average impossible to change one's luck through observations. This is expressed in the following

7.1 Definition Given a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$, a stochastic process $X = (X_t, t \in T)$ is called a *martingale relative to* $\mathcal{F}$, if X is $\mathcal{F}$–adapted, for every $t \in T$, X_t is integrable, and if for all $s, t \in T$, $s \leq t$,

$$E(X_t \mid \mathcal{F}_s) = X_s \tag{7.1}$$

holds true. If (7.1) holds for all $s, t \in T$, $s \leq t$, with "=" replaced by "$\leq$" ("$\geq$", respectively), then X is called a *supermartingale* (*submartingale*, respectively).

If no filtration is explicitly mentioned for a (sub-, super-) martingale X, the natural filtration $\mathcal{F}^X$ is meant.

7.2 Remark The origin of the name "martingale" is uncertain (cf., e.g., the remarks in [3, p. 144]). It seems that this name appears for the first time in the mathematical literature in [29]. There the author defined a martingale to be a "system of games", which may be interpreted as a strategy for a game. The origin of the name "supermartingale", however, is well documented — here is a quotation from [10, p. 808]:

> *... This obviously inappropriate nomenclature was chosen under the malign influence of the noise level of radio's SUPERman program, a favourite supper-time program of Doob's son during the writing of [9].*

The statement of the following lemma is obvious

7.3 Lemma *The expectation value of a martingale is constant, of a submartingale increasing, and of a supermartingale decreasing.*

7.4 Exercise

(a) Suppose that $T = \mathbb{N}_0$. Show that if for every $n \in \mathbb{N}_0$

$$E(X_{n+1} \mid \mathcal{F}_n^X) \;\square\; X_n$$

where $\square$ stands for "$=$", "$\geq$", "$\leq$" respectively, then X is a martingale, submartingale, supermartingale, respectively.

(b) Suppose that X is an integrable random variable, that $T = \{0, 1, \dots, N\}$, and that $\mathcal{F} = (\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_N)$ is any filtration. Show that $X_n = E(X \mid \mathcal{F}_n)$, $n \in T$, defines a martingale.

7.5 Lemma *Suppose that $T = \mathbb{N}_0$ or $T = \mathbb{R}_+$, and assume that X is an integrable stochastic process, which has independent increments. If the expectation value of the increments is non-negative, then X is a submartingale relative to $\mathcal{F}^X$, if they are centered, X is an $\mathcal{F}$–martingale.*

Proof Since the increments of X are independent, exercise 2.6 implies that for s, $t \in T$, with $s \leq t$, the increment $X_t - X_s$ is independent of the σ–algebra

$$\sigma\big(X_{s_0}, X_{s_1} - X_{s_0}, \dots, X_{s_n} - X_{s_{n-1}},\, n \in \mathbb{N}, s_0 < s_1 < \cdots < s_n \leq s\big),$$

with $s_0 = 0$. On the other hand, we have shown in lemma 6.10 that this is just the σ–algebra $\mathcal{F}_s^X$. Thus we have that $X_t - X_s$ is independent of $\mathcal{F}_s^X$. If $E(X_t - X_s) \geq 0$ for all $s, t \in T$, with $s \leq t$, then

$$\begin{aligned} E(X_t \mid \mathcal{F}_s^X) &= X_s + E(X_t - X_s \mid \mathcal{F}_s^X) \\ &= X_s + E(X_t - X_s) \\ &\geq X_s, \end{aligned}$$

so that X is an $\mathcal{F}$–submartingale. If $X_t - X_s$ is centered, this calculation shows that X is an $\mathcal{F}$–martingale. $\square$

7.6 Examples

(a) The symmetric random walk is a martingale.

(b) Every Poisson process $X = (X_t,\, t \in \mathbb{R}_+)$ is a submartingale.

(c) Every one dimensional Brownian motion $B = (B_t,\, t \in \mathbb{R}_+)$ is a martingale relative to its natural filtration $\mathcal{F}^B$. Consider now $X = (X_t,\, t \in \mathbb{R}_+)$ with $X_t = B_t^2 - t$. Then for $s \le t$

$$\begin{aligned} E\big(B_t^2 - t \,\big|\, \mathcal{F}_s^B\big) &= E\big((B_t - B_s)^2 - t + 2B_t B_s - B_s^2 \,\big|\, \mathcal{F}_s^B\big) \\ &= E\big((B_t - B_s)^2\big) + 2B_s E(B_t \mid \mathcal{F}_s^B) - B_s^2 - t \\ &= B_s^2 - s. \end{aligned}$$

Therefore $t \mapsto B_t^2 - t$ is a martingale.

(d) Consider still a one dimensional Brownian motion B as in (c), and for $\lambda \in \mathbb{R}$ set

$$X_t = e^{\lambda B_t - \lambda^2 t/2}, \qquad t \in \mathbb{R}_+. \tag{7.2}$$

The process $X = (X_t,\, t \in \mathbb{R}_+)$ above is called *geometric Brownian motion*. A glance at the law of B_t shows that X_t is integrable, and clearly it is $\mathcal{F}^B$–adapted. We check that it is a martingale ($s \le t$):

$$\begin{aligned} E(X_t \mid \mathcal{F}_s^B) &= e^{-\lambda^2 t/2} e^{\lambda B_s} E\big(e^{\lambda(B_t - B_s)} \,\big|\, \mathcal{F}_s^B\big) \\ &= e^{-\lambda^2 t/2} e^{\lambda B_s} E\big(e^{\lambda(B_t - B_s)}\big) \\ &= e^{-\lambda^2 t/2} e^{\lambda B_s} e^{\lambda^2 (t-s)/2} \\ &= X_s, \end{aligned}$$

where we employed again the independence of $B_t - B_s$ and $\mathcal{F}_s^B$, and in the third step we used the well-known formula for the moment generating function of a normally distributed random variable.

7.7 Exercise Let X be a Poisson process with parameter $\lambda > 0$. Show that $t \mapsto X_t - \lambda t$, $t \in \mathbb{R}_+$, defines a martingale.

7.8 Exercise Suppose that X, Y are (sub-) martingales relative to $\mathcal{F}$.

(a) Show that for any $\alpha, \beta \ge 0$, $\alpha X + \beta Y$ is a (sub-) martingale relative to $\mathcal{F}$.

(b) Show that $X \vee Y$ is an $\mathcal{F}$–submartingale.

(c) Show that X^+ is an $\mathcal{F}$–submartingale. (Hint: Use (b).)

7.9 Lemma *Suppose that $X = (X_t,\, t \in T)$ is an integrable stochastic process, $\mathcal{F} = (\mathcal{F}_t,\, t \in T)$ a filtration, and that φ is a convex function so that $\varphi \circ X = (\varphi \circ X_t,\, t \in T)$ is an integrable process.*

(a) *If X is an $\mathcal{F}$–martingale, then $\varphi \circ X$ is an $\mathcal{F}$–submartingale.*

(b) *If in addition φ is increasing, and X is an $\mathcal{F}$–submartingale, then $\varphi \circ X$ is an $\mathcal{F}$–submartingale.*

Proof This follows directly from Jensen's inequality for conditional expectations, see theorem 5.12. □

For example, if X is a martingale, then for all $p \geq 1$, $\lambda \in \mathbb{R}$, $t \mapsto |X_t|^p$, and $t \mapsto \exp(\lambda X_t)$ are submartingales, provided these random variables are integrable for every $t \in T$.

Next we are heading for one of the central theorems of Doob, namely his decomposition theorem for submartingales. We are only discussing the case where $T = \mathbb{N}_0$, the case of a continuous time parameter domain requires a much greater technical effort (leading to the *Doob–Meyer–decomposition–theorem*) and is outside the scope of these lectures.

To begin with, assume that we are given a filtration $\mathcal{F} = (\mathcal{F}_n, n \in \mathbb{N}_0)$, an $\mathcal{F}$–martingale $(Y_n, n \in \mathbb{N}_0)$, and an integrable, adapted, increasing process $Z = (Z_n, n \in \mathbb{N}_0)$, that is, $Z_0 = 0$, and $Z_{n-1} \leq Z_n$ P–a.s. for all $n \in \mathbb{N}$. Set

$$X_n = Y_n + Z_n, \qquad n \in \mathbb{N}_0.$$

Clearly, $X = (X_n, n \in \mathbb{N}_0)$ is integrable and adapted to $\mathcal{F}$. Moreover, for $n \in \mathbb{N}$,

$$\begin{aligned} E(X_n \,|\, \mathcal{F}_{n-1}) &= E(Y_n \,|\, \mathcal{F}_{n-1}) + E(Z_n \,|\, \mathcal{F}_{n-1}) \\ &= X_{n-1} + E(Z_n - Z_{n-1} \,|\, \mathcal{F}_{n-1}) \\ &\geq X_{n-1}, \end{aligned}$$

so that X is an $\mathcal{F}$–submartingale. Observe that nothing would change the argument if we make the assumption that for every $n \in \mathbb{N}_0$, Z_n is $\mathcal{F}_{n-1}$–measurable, where — for convenience — we set $\mathcal{F}_{-1} = \{\emptyset, \Omega\}$. Note that this is a stronger assumption than that of adaptedness. A process satisfying the latter measurability condition will henceforth be called *predictable*.

For the converse, let $X = (X_n, n \in \mathbb{N}_0)$ be an $\mathcal{F}$–submartingale. We are looking for processes Y and Z as above so that $X = Y + Z$. The basic idea of the following argument is that a martingale has increments which have expectation zero (since the expectation value of a martingale is constant). Thus consider the expectation value of an increment $X_k - X_{k-1}$, $k \in \mathbb{N}$, of X:

$$E(X_k - X_{k-1}) = E\big(E(X_k \,|\, \mathcal{F}_{k-1}) - X_{k-1}\big) \geq 0.$$

Thus if we subtract the random variable in the second expectation from the increment, we get zero for the expecation of the resulting random variable:

$$E\big(X_k - X_{k-1} - \big(E(X_k \,|\, \mathcal{F}_{k-1}) - X_{k-1}\big)\big) = E\big(X_k - E(X_k \,|\, \mathcal{F}_{k-1})\big) = 0.$$

Therefore the idea is to sum up $X_k - E(X_k \mid \mathcal{F}_{k-1})$ to get a martingale, and we decompose the increments of X accordingly. Let $n \in \mathbb{N}$ then

$$\begin{aligned} X_n &= X_0 + \sum_{k=1}^{n} (X_k - X_{k-1}) \\ &= X_0 + \sum_{k=1}^{n} \big(X_k - E(X_k \mid \mathcal{F}_{k-1})\big) + \sum_{k=1}^{n} \big(E(X_k \mid \mathcal{F}_{k-1}) - X_{k-1}\big). \end{aligned}$$

Now define $Y_0 = X_0$, $Z_0 = 0$, and for $n \in \mathbb{N}$,

$$\begin{aligned} Y_n &= X_0 + \sum_{k=1}^{n} \big(X_k - E(X_k \mid \mathcal{F}_{k-1})\big), \\ Z_n &= \sum_{k=1}^{n} \big(E(X_k \mid \mathcal{F}_{k-1}) - X_{k-1}\big). \end{aligned}$$

Obviously, $X_n = Y_n + Z_n$ for all $n \in \mathbb{N}_0$. Moreover, Y is $\mathcal{F}$–adapted, and Z is $\mathcal{F}$–predictable. Also

$$\begin{aligned} E(Y_n \mid \mathcal{F}_{n-1}) &= Y_{n-1} + E\Big(\big(X_n - E(X_n \mid \mathcal{F}_{n-1})\big) \,\Big|\, \mathcal{F}_{n-1}\Big) \\ &= Y_{n-1} \end{aligned}$$

because the conditional expectation is idempotent. Thus Y is an $\mathcal{F}$–martingale. Furthermore

$$Z_n = Z_{n-1} + E(X_n \mid \mathcal{F}_{n-1}) - X_{n-1} \geq Z_{n-1},$$

since X is an $\mathcal{F}$–submartingale. We have proved the first statement of the following

7.10 Theorem (Doob's Decomposition Theorem) *Suppose that X is an $\mathcal{F}$–submartingale. Then there exist an $\mathcal{F}$–martingale Y, and an $\mathcal{F}$–predictable, increasing, integrable process Z with $Z_0 = 0$ so that $X = Y + Z$. Moreover, Y and Z with these properties are (a.s.) unique.*

Proof We only have to prove the uniqueness of Y and Z. Suppose that Y' and Z' are two processes as in the statement, different from Y, Z respectively. Then for $n \in \mathbb{N}_0$ we find

$$X_n = Y_n + Z_n = Y'_n + Z'_n,$$

that is

$$Y_n - Y'_n = Z'_n - Z_n.$$

The left hand side is a martingale evaluated at time $n \in \mathbb{N}_0$, and the right hand side is $\mathcal{F}_{n-1}$–measurable. Therefore we obtain for $n \in \mathbb{N}$,

$$\begin{aligned} Z'_n - Z_n &= Y_n - Y'_n \\ &= E(Y_n - Y'_n \mid \mathcal{F}_{n-1}) \\ &= Y_{n-1} - Y'_{n-1} \\ &= Z'_{n-1} - Z_{n-1}. \end{aligned}$$

Hence $n \mapsto Y_n - Y'_n$, and $n \mapsto Z'_n - Z_n$ are constant. For $n = 1$ we then get $Y_1 - Y'_1 = Y_0 - Y'_0 = 0$, which leads to the contradiction $Z_n = Z'_n$, $Y_n = Y'_n$ for all $n \in \mathbb{N}_0$. □

7.11 Exercise Let $X = (X_n, n \in \mathbb{N}_0)$ be a symmetric random walk. Show that $n \mapsto X_n^2$ is a submartingale without using lemma 7.9, and compute its Doob-decomposition.

7.2 Optional Stopping Theorem

Throughout this section we consider the case $T = \mathbb{N}_0$, and we fix a filtration $\mathcal{F} = (\mathcal{F}_n, n \in \mathbb{N}_0)$.

7.12 Lemma (Martingale Transform of H) *Suppose that $H = (H_n, n \in \mathbb{N}_0)$ is an $\mathcal{F}$–predictable, positive and bounded process. Suppose furthermore that $X = (X_n, n \in \mathbb{N}_0)$ is an $\mathcal{F}$–martingale ($\mathcal{F}$–submartingale, respectively). Define the* martingale transform of H *by*

$$\begin{aligned} Y_0 &= X_0, \\ Y_n &= Y_{n-1} + H_n(X_n - X_{n-1}), \qquad n \in \mathbb{N}. \end{aligned} \tag{7.3}$$

Then Y is an $\mathcal{F}$–martingale ($\mathcal{F}$–submartingale, respectively).

7.13 Remark Clearly, Y defined by (7.3) is given by

$$Y_n = X_0 + \sum_{k=1}^{n} H_k(X_k - X_{k-1}), \qquad n \in \mathbb{N}.$$

This is the discrete time analogue of a *stochastic integral* of H with respect to X.

Proof (of lemma 7.12) It is clear that Y is $\mathcal{F}$–adapted and integrable. For $n \in \mathbb{N}$ we have

$$\begin{aligned} &E(Y_n \mid \mathcal{F}_{n-1}) \\ &\qquad = Y_{n-1} + H_n E(X_n - X_{n-1} \mid \mathcal{F}_{n-1}) \begin{cases} = Y_{n-1} & \text{if } X \text{ is a martingale,} \\ \geq Y_{n-1} & \text{if } X \text{ is a submartingale,} \end{cases} \end{aligned}$$

and the proof is finished. □

7.14 Theorem (Optional Stopping Theorem) *Let $X = (X_n, n \in \mathbb{N}_0)$ be an $\mathcal{F}$–martingale ($\mathcal{F}$–submartingale, respectively), and suppose that τ is a stopping time relative to $\mathcal{F}$. Then the stopped process $X^\tau = (X_{n\wedge\tau}, n \in \mathbb{N}_0)$ is an $\mathcal{F}$–martingale ($\mathcal{F}$–submartingale, respectively).*

Proof Consider $H = (1_{\{n \le \tau\}}, n \in \mathbb{N})$. Clearly, H is positive and bounded. Moreover, $H_n = 1 - 1_{\{\tau \le n-1\}}$, and the event $\{\tau \le n-1\}$ belongs to $\mathcal{F}_{n-1}$, so that H is $\mathcal{F}$–predictable. We show that X^τ is the martingale transform of H: For $n \in \mathbb{N}$,

$$\begin{aligned} X_0 + \sum_{k=1}^{n} H_k(X_k - X_{k-1}) &= X_0 + \sum_{k=1}^{n} 1_{\{k \le \tau\}}(X_k - X_{k-1}) \\ &= X_0 + \sum_{k=1}^{n\wedge\tau}(X_k - X_{k-1}) \\ &= X_0 + X_{n\wedge\tau} - X_0 \\ &= X_n^\tau. \end{aligned}$$

Thus the statement of the theorem follows directly from lemma 7.12. □

7.15 Lemma *Suppose that τ is an $\mathcal{F}$–stopping time, so that $\tau < N$ with $N \in \mathbb{N}_0 \cup \{+\infty\}$. For $A \in \mathcal{F}_\tau$ set*

$$\tau_A = \begin{cases} \tau & \text{on } A, \\ N & \text{on } \complement A. \end{cases}$$

Then τ_A is an $\mathcal{F}$–stopping time.

Proof We have to show that for every $n \in \mathbb{N}_0$ the set $\{\tau_A \le n\}$ belongs to $\mathcal{F}_n$. If $n \ge N$, we have $\{\tau_A \le n\} = \Omega \in \mathcal{F}_n$. On the other hand, if $n < N$, then $\{\tau_A \le n\} \subset \{\tau_A < N\} = A = \{\tau_A = \tau\}$. Thus in this case, $\{\tau_A \le n\} = \{\tau \le n\} \cap A$ which belongs to $\mathcal{F}_n$, because $A \in \mathcal{F}_\tau$. □

7.16 Theorem (Optional Sampling Theorem) *Assume that σ, τ are bounded $\mathcal{F}$–stopping times with $\sigma \le \tau$. Suppose furthermore that $X = (X_n, n \in \mathbb{N}_0)$ is an $\mathcal{F}$–martingale (respectively an $\mathcal{F}$–submartingale). Then $E(X_\tau \,|\, \mathcal{F}_\sigma) = X_\sigma$ ($E(X_\tau \,|\, \mathcal{F}_\sigma) \ge X_\sigma$, respectively) holds true.*

Proof We only do the proof for the case of a submartingale, the case of a martingale follows along the same lines. First we show the inequality $E(X_\sigma) \le E(X_\tau)$. To this end set

$$H_n = 1_{\{n \le \tau\}} - 1_{\{n \le \sigma\}} = 1_{\{\sigma < n \le \tau\}}, \qquad n \in \mathbb{N}.$$

As in the proof of theorem 7.14, H is $\mathcal{F}$–predictable, non-negative and bounded. Put $Y_0 = X_0$, and for $n \in \mathbb{N}$,

$$\begin{aligned} Y_n &= X_0 + \sum_{k=1}^{n} H_n\,(X_k - X_{k-1}) \\ &= X_0 + \sum_{k=\sigma+1}^{\tau}(X_k - X_{k-1}) \\ &= X_0 + X_\tau - X_\sigma. \end{aligned}$$

By lemma 7.12 Y is an $\mathcal{F}$–submartingale, and therefore we get $E(Y_n - Y_0) \geq 0$ for all $n \in \mathbb{N}$. Hence we proved $E(X_\tau) \geq E(X_\sigma)$.

Let $N \in \mathbb{N}$ be large enough so that $\sigma \leq \tau \leq N$. For $A \in \mathcal{F}_\sigma \subset \mathcal{F}_\tau$ (see lemma 6.20) define

$$\sigma_A = \sigma\, 1_A + N\, 1_{\complement A},$$
$$\tau_A = \tau\, 1_A + N\, 1_{\complement A}.$$

By lemma 7.15 σ_A and τ_A are again bounded $\mathcal{F}$–stopping times with $\sigma_A \leq \tau_A$. Thus we we may apply the above proved inquality with these stopping times: $E(X_{\sigma_A}) \leq E(X_{\tau_A})$. But this entails

$$E\big(X_\sigma 1_A\big) \leq E\big(X_\tau 1_A\big)$$

for all $A \in \mathcal{F}_\sigma$, that is

$$\int_A X_\sigma \, dP \leq \int_A X_\tau \, dP = \int_A E(X_\tau \mid \mathcal{F}_\sigma)\, dP.$$

Since this is true for all $A \in \mathcal{F}_\sigma$, we get that $E(X_\tau \mid \mathcal{F}_\sigma) \geq X_\sigma$, P–a.s., finishing the proof. □

Combining the last theorem with part (a) of exercise 7.4, we immediately get the following

7.17 Corollary *Suppose that $X = (X_k,\ k \in \mathbb{N})$ is an $\mathcal{F}$–martingale ($\mathcal{F}$–submartingale, respectively), and that for $n \in \mathbb{N}$, $\tau_1, \ldots, \tau_n$ are bounded $\mathcal{F}$–stopping times with $0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_n$. Consider the process $Y = (Y_k,\ k = 1, \ldots, n)$ and the filtration $\mathcal{F}^\tau = (\mathcal{F}^\tau_k,\ k = 1, \ldots, n)$ defined by*

$$Y_k = X_{\tau_k}, \quad \mathcal{F}^\tau_k = \mathcal{F}_{\tau_k}, \qquad k = 1, \ldots, n.$$

Then Y is an $\mathcal{F}^\tau$–martingale ($\mathcal{F}^\tau$–submartingale, respectively).

7.3 Doob's Inequalities

7.18 Lemma *Consider a finite time parameter domain $T = \{0, 1, \ldots, N\}$, $N \in \mathbb{N}$, and let $X = (X_n,\ n \in T)$ be an integrable submartingale relative to a filtration $\mathcal{F} = (\mathcal{F}_n,\ n \in T)$. Then the following inequalities hold true for every $\lambda > 0$:*

$$\lambda P\big(\sup_{n\in T} X_n \geq \lambda\big) \leq E\big(X_N\, 1_{\{\sup_{n\in T} X_n \geq \lambda\}}\big) \leq E\big(|X_N|\, 1_{\{\sup_{n\in T} X_n \geq \lambda\}}\big). \tag{7.4}$$

Proof Define

$$\tau = \begin{cases} \inf\{n \in T,\ X_n \geq \lambda\}, & \text{if } \{n \in T,\ X_n \geq \lambda\} \neq \emptyset, \\ N, & \text{otherwise.} \end{cases}$$

Then it follows from exercise 6.16 that τ is an $\mathcal{F}$–stopping time. (In the formulation of that exercise, $\tau = \tau_{[\lambda,+\infty)} \wedge N$.) The optional sampling theorem 7.16 implies

$$\begin{aligned} E(X_N) &\geq E(X_\tau) \\ &= E(X_\tau\, 1_{\{\sup_{n\in T} X_n \geq \lambda\}}) + E(X_\tau\, 1_{\{\sup_{n\in T} X_n < \lambda\}}) \\ &\geq \lambda P\big(\sup_{n\in T} X_n \geq \lambda\big) + E(X_\tau\, 1_{\{\sup_{n\in T} X_n < \lambda\}}) \\ &= \lambda P\big(\sup_{n\in T} X_n \geq \lambda\big) + E(X_N\, 1_{\{\sup_{n\in T} X_n < \lambda\}}). \end{aligned}$$

Therefore

$$E\big(X_N\,(1 - 1_{\{\sup_{n\in T} X_n < \lambda\}})\big) \geq \lambda P\big(\sup_{n\in T} X_n \geq \lambda\big),$$

which proves the first inequality. The second one is trivial. □

7.19 Lemma *Consider a finite time parameter domain $T = \{0, 1, \ldots, N\}$, $N \in \mathbb{N}$, and let $X = (X_n,\ n \in T)$ be an $\mathcal{F} = (\mathcal{F}_n,\ n \in T)$–martingale or a positive $\mathcal{F}$–submartingale. Then for all $\lambda > 0$, $p \geq 1$,*

$$\lambda^p P\big(\sup_{n\in T} |X_n| \geq \lambda\big) \leq E(|X_N|^p), \tag{7.5}$$

and for all $p > 1$

$$E(|X_N|^p) \leq E\big(\sup_{n\in T} |X_n|^p\big) \leq \Big(\frac{p}{p-1}\Big)^p E(|X_N|^p) \tag{7.6}$$

hold true.

Proof Observe that for every $p \geq 1$ the mapping $x \mapsto |x|^p$, $x \in \mathbb{R}$, is convex, and it is increasing on $\mathbb{R}_+$. Therefore with lemma 7.9 we get that under the hypothesis of the lemma, $n \mapsto |X_n|^p$ is an $\mathcal{F}$–submartingale. Then lemma 7.18 gives

$$\begin{aligned} P\big(\sup_{n\in T} |X_n| \geq \lambda\big) &= P\big((\sup_{n\in T} |X_n|)^p \geq \lambda^p\big) \\ &= P\big(\sup_{n\in T} |X_n|^p \geq \lambda^p\big) \\ &\leq \frac{1}{\lambda^p}\, E\big(|X_N|^p\, 1_{\{\sup_{n\in T} |X_n|^p \geq \lambda^p\}}\big) \\ &\leq \frac{1}{\lambda^p}\, E(|X_N|^p), \end{aligned}$$

proving inequality (7.5). Now let $p > 1$. The left hand inequality of (7.6) is trivial, for the right hand inequality set

$$X^* = \sup_{n\leq N} |X_n|.$$

For $K > 0$ consider

$$\begin{aligned}
E\big((X^* \wedge K)^p\big) &= E\Big(\int_0^{X^*\wedge K} p\lambda^{p-1}\,d\lambda\Big)\\
&= pE\Big(\int_0^K 1_{\{X^*\geq\lambda\}}\,\lambda^{p-1}\,d\lambda\Big)\\
&= p\int_0^K \lambda^{p-1}\,E\big(1_{\{X^*\geq\lambda\}}\big)\,d\lambda\\
&= p\int_0^K \lambda^{p-1}\,P(X^*\geq\lambda)\,d\lambda,
\end{aligned}$$

where we made use of Fubini's theorem. Since $n \mapsto |X_n|$ is an $\mathcal{F}$–submartingale, lemma 7.18 entails

$$P(X^*\geq\lambda) \leq \frac{1}{\lambda}\,E\big(|X_N|\,1_{\{X^*\geq\lambda\}}\big),$$

so that

$$\begin{aligned}
E\big((X^* \wedge K)^p\big) &\leq p\int_0^K \lambda^{p-2}\,E\big(|X_N|\,1_{\{X^*\geq\lambda\}}\big)\,d\lambda\\
&= pE\Big(|X_N|\int_0^K \lambda^{p-2}\,1_{\{X^*\geq\lambda\}}\,d\lambda\Big)\\
&= pE\Big(|X_N|\int_0^{X^*\wedge K} \lambda^{p-2}\,d\lambda\Big)\\
&= \frac{p}{p-1}E\big(|X_N|\,(X^*\wedge K)^{p-1}\big)\\
&\leq \frac{p}{p-1}E\big(|X_N|^p\big)^{1/p}E\big((X^*\wedge K)^p\big)^{(p-1)/p},
\end{aligned}$$

with another application of Fubini's theorem, and in the last step we employed Hölder's inequality with exponents p and $p/(p-1)$. Thus

$$\Big(E\big((X^*\wedge K)^p\big)\Big)^{1-(p-1)/p} = \Big(E\big((X^*\wedge K)^p\big)\Big)^{1/p} \leq \frac{p}{p-1}\,\big(E(|X_N|^p\big)^{1/p}.$$

Finally, we let K tend to infinity, and with the dominated convergence theorem arrive at inequality (7.6). □

Now we turn to the case that the time parameter domain T is a finite interval. Let D be a dense subset of T, and let D_n be a sequence of finite subsets of D which increases to D: $D_n \uparrow D$, as $n \uparrow \infty$. Let $X = (X_t,\, t \in T)$ be a martingale or a positive submartingale with respect to a filtration $\mathcal{F} = (\mathcal{F}_t,\, t \in T)$. Suppose that $\lambda > 0$, and choose $\varepsilon > 0$. Then we have

$$\{\sup_{t\in D}|X_t| \geq \lambda+\varepsilon\} \subset \bigcup_{n\in\mathbb{N}}\{\sup_{t\in D_n}|X_t|\geq\lambda\}. \tag{7.7}$$

Indeed, if $\omega \in \Omega$ does not belong to the set on the right hand side of (7.7) then we have for all $n \in \mathbb{N}$ that $\sup_{t\in D_n} |X_t(\omega)| < \lambda$. On the other hand, $D = \cup_n D_n$, and therefore for all $t \in D$ we get $|X_t(\omega)| < \lambda$. It follows that $\sup_{t\in D} |X_t(\omega)| \le \lambda$, and hence ω does not belong to the set on the left hand side of (7.7). Thus the inclusion (7.7) is proved. Then

$$\begin{aligned} P\bigl(\sup_{t\in D} |X_t| \ge \lambda + \varepsilon\bigr) &\le P\Bigl(\bigcup_{n\in\mathbb{N}} \{\sup_{t\in D_n} |X_t| \ge \lambda\}\Bigr) \\ &= \lim_{n\in\mathbb{N}} P\bigl(\sup_{t\in D_n} |X_t| \ge \lambda\bigr), \end{aligned}$$

because the events $\{\sup_{t\in D_n} |X_t| \ge \lambda\}$ are increasing with n. For $n \in \mathbb{N}$, let $t_n = \max D_n$. With lemma 7.19 we can now estimate as follows:

$$\begin{aligned} P\bigl(\sup_{t\in D} |X_t| \ge \lambda + \varepsilon\bigr) &\le \limsup_{n\in\mathbb{N}} \frac{1}{\lambda^p} E\bigl(|X_{t_n}|^p\bigr) \\ &= \sup_{n\in\mathbb{N}} \frac{1}{\lambda^p} E\bigl(|X_{t_n}|^p\bigr), \end{aligned}$$

because t_n is increasing with n, and since $t \mapsto |X_t|^p$ is a submartingale, also the expectations on the right hand side increase with n. Thus we get for every $\varepsilon > 0$ the inequality

$$P\bigl(\sup_{t\in D} |X_t| \ge \lambda + \varepsilon\bigr) \le \frac{1}{\lambda^p} \sup_{t\in T} E\bigl(|X_t|^p\bigr).$$

The events in the probability on the left are increasing with $\varepsilon \downarrow 0$, so that when we let ε decrease to zero through a sequence, we finally obtain the inequality

$$P\bigl(\sup_{t\in D} |X_t| \ge \lambda\bigr) \le \frac{1}{\lambda^p} \sup_{t\in T} E\bigl(|X_t|^p\bigr). \tag{7.8}$$

7.20 Theorem (Doob's Inequalities) *Suppose that T is a finite interval, and that $X = (X_t,\, t \in T)$ is a martingale or a positive submartingale relative to a filtration $\mathcal{F} = (\mathcal{F}_t,\, t \in T)$. Suppose furthermore, that the paths of X are continuous from the left or from the right. Then for all $\lambda > 0,\ p \ge 1$,*

$$\lambda^p P\bigl(\sup_{t\in T} |X_t| \ge \lambda\bigr) \le \sup_{t\in T} E\bigl(|X_t|^p\bigr), \tag{7.9}$$

and for $p > 1$,

$$E\bigl(\sup_{t\in T} |X_t|^p\bigr) \le \Bigl(\frac{p}{p-1}\Bigr)^p \sup_{t\in T} E\bigl(|X_t|^p\bigr) \tag{7.10}$$

hold true

Proof Since X has left or right continuous paths, it follows that for every $\omega \in \Omega$,

$$\sup_{t\in T} |X_t(\omega)| = \sup_{t\in D} |X_t(\omega)|.$$

Therefore inequality (7.9) follows directly from (7.8). To prove (7.10), one has just to replace X^* in the second part of the proof of lemma 7.19 by $X^* = \sup_{t\in T} |X_t|$ and apply inequality (7.9), while the rest of the calculation remains unchanged. □

7.21 Remark Note that because $t \mapsto |X_t|^p$ is a submartingale, if T is closed to the right with endpoint t_0, then we can replace $\sup_{t\in T} E(|X_t|^p)$ in inequalities (7.9) and (7.10) by $E(|X_{t_0}|^p)$.

7.4 Doob's Downcrossing Inequality

In this and the next section we exclusively consider the time parameter domain $T = \mathbb{N}_0$. Let $X = (X_t,\, t \in \mathbb{N}_0)$ be a submartingale. For any time $t_0 \in \mathbb{N}$, and $a, b \in \mathbb{R}$ with $a < b$ we shall consider the downcrossings the paths of the process X perform across the interval $[a, b]$.

Set $n_0 = \lfloor (t_0 + 1)/2 \rfloor$, [1] and define inductively the following times:

$$s_1 = \inf\{t = 0, \dots, t_0,\ X_t \ge b\}, \qquad s_2 = \inf\{t = s_1, \dots, t_0,\ X_t \le a\},$$
$$s_{2j-1} = \inf\{t = s_{2j-2}, \dots, t_0, X_t \ge b\}, \quad s_{2j} = \inf\{t = s_{2j-1}, \dots, t_0,\ X_t \le a\},$$

where $j = 2, \dots, n_0$, and we put $\inf \emptyset = t_0$. Then we define the number $D^{t_0}_{[a,b]}(X)$ of *downcrossings of* X *across* $[a, b]$ *up to time* t_0 by

$$D^{t_0}_{[a,b]}(X) = \max\{n \in \mathbb{N}_0,\ s_{2n} < t_0 \quad \text{or } s_{2n-1} < t_0,\ s_{2n} = t_0,\ \text{and } X_{t_0} \le a\}. \tag{7.11}$$

Of course we have that $D^{t_0}_{[a,b]}(X) \le n_0$.

First we check that $s_1, \dots, s_{2n_0}$ are stopping times relative to the natural filtration $\mathcal{F} = (\mathcal{F}_t,\, t \in \mathbb{N}_0)$ generated by X, and that $D^{t_0}_{[a,b]}(X)$ is a random variable. For the random times s_k, $k = 1, \dots, 2n_0$, it is enough to prove that for every $t \in \mathbb{N}_0$, $\{s_k = t\} \in \mathcal{F}_t$. Now

$$\{s_1 = t\} = \{X_t \ge b\} \setminus \Big(\bigcup_{s<t} \{X_s \ge b\}\Big),$$
$$\{s_2 = t\} = \{X_t \le a\} \setminus \Big(\bigcup_{s<t} \{X_s \le a, s \ge s_1\}\Big),$$

[1] The *floor* $\lfloor x \rfloor$ of a real number x is the greatest integer less or equal to x.

and both sets obviously belong to $\mathcal{F}_t$. But then also for $j \geq 2$ the events

$$\{s_{2j-1} = t\} = \{X_t \geq b, t \geq s_{2j-2}\} \setminus \Big(\bigcup_{s<t}\{X_s \geq b, s \geq s_{2j-2}\}\Big),$$
$$\{s_{2j} = t\} = \{X_t \leq a, t \geq s_{2j-1}\} \setminus \Big(\bigcup_{s<t}\{X_s \leq a, s \geq s_{2j-1}\}\Big),$$

belong to $\mathcal{F}_t$. Observe that for $k \in \mathbb{N}$, the event $\{s_k < t_0\}$ is the same as $\{0 \leq s_1 < s_2 < \cdots < s_k < t_0\}$, and therefore we have for $j \in \mathbb{N}$,

$$\big\{D^{t_0}_{[a,b]} \geq j\big\} = \{s_{2j} < t_0\} \uplus \{s_{2j-1} < t_0, s_{2j} = t_0, X_{t_0} \leq a\}. \tag{7.12}$$

In particular, this formula shows that $D^{t_0}_{[a,b]}$ is a random variable.

7.22 Theorem (Doob's Downcrossing Inequality) *Supose that $(X_t,\ t \in \mathbb{N}_0)$ is a submartingale. For every choice of $t_0 \in \mathbb{N}$, and a, $b \in \mathbb{R}$ with $a < b$, the inequality*

$$E\big(D^{t_0}_{[a,b]}\big) \leq \frac{1}{b-a}\, E\big((X_{t_0} - b)^+\big) \tag{7.13}$$

holds true.

Proof We use the notation from above, then $s_1, \ldots, s_{2n_0}$ are increasing stopping times, which are bounded by t_0. By theorem 7.16 we then know that $(X_{s_k},\ k = 0, \ldots, 2n_0)$ is also a submartingale, where we have set $s_0 = 0$. The submartingale property reads here as follows

$$\int_{A_k} X_{s_{k+1}}\, dP \geq \int_{A_k} X_{s_k}\, dP, \tag{7.14}$$

for every $k = 0, \ldots, 2n_0 - 1$, and every $A_k \in \mathcal{F}_{s_k}$. We consider especially the events $A_k = \{s_k < t_0\}$, which obviously belong to $\mathcal{F}_{s_k}$. Note that $A_{k+1} \subset A_k$, and that formula (7.12) can then be written as

$$\big\{D^{t_0}_{[a,b]} \geq j\big\} = A_{2j} \uplus \big(A_{2j-1} \cap \{s_{2j} = t_0, X_{t_0} \leq a\}\big). \tag{7.15}$$

On A_{2j-1} we have that $X_{2j-1} \geq b$, while on on $\{D^{t_0}_{[a,b]} \geq j\}$ we have $X_{s_{2j}} \leq a$. Thus for $j \in \{1, \ldots, n_0\}$, using the submartingale property (7.14) we can estimate as follows

$$\begin{aligned}
0 &\leq \int_{A_{2j-1}} (X_{s_{2j-1}} - b)\, dP \\
&\leq \int_{A_{2j-1}} (X_{s_{2j}} - b)\, dP \\
&= \int_{\{D \geq j\}} (X_{s_{2j}} - b)\, dP + \int_{A_{2j-1}\setminus\{D\geq j\}} (X_{s_{2j}} - b)\, dP \\
&\leq (a-b)\, P(D \geq j) + \int_{A_{2j-1}\setminus\{D\geq j\}} (X_{s_{2j}} - b)\, dP
\end{aligned}$$

where we abbreviated $D^{t_0}_{[a,b]}$ by D for simplicity, and we used the fact that $A_{2j-1} \supset \{D \geq j\}$ (see (7.15)). Thus we obtain

$$\begin{aligned} P(D \geq j) &\leq \frac{1}{b-a} \int_{A_{2j-1} \setminus \{D \geq j\}} (X_{s_{2j}} - b)\, dP \\ &\leq \frac{1}{b-a} \int_{A_{2j-1} \setminus A_{2j}} (X_{s_{2j}} - b)\, dP. \end{aligned}$$

On the complement of A_{2j} we have $s_{2j} = t_0$, so that we find the inequality

$$P(D \geq j) \leq \frac{1}{b-a} \int_{A_{2j-1} \setminus A_{2j}} (X_{t_0} - b)^+\, dP. \tag{7.16}$$

Observe now that

$$A_{2j-1} \setminus A_{2j} = \{0 \leq s_1 < \cdots < s_{2j-1} < t_0, s_{2j} = t_0\}, \qquad j = 1, \ldots, n_0,$$

so that these sets are pairwise disjoint. Summing the inequalities (7.16) over $j = 1$, $\ldots, n_0$, we find

$$\sum_{j=1}^{n_0} P(D \geq j) \leq \frac{1}{b-a} \int_A (X_{t_0} - b)^+\, dP \leq \frac{1}{b-a} E\big((X_{t_0} - b)^+\big),$$

where

$$A = \bigcup_{j=1}^{n_0} A_{2j-1} \setminus A_{2j}.$$

Finally we use the fact (*exercise*) that

$$E(D) = \sum_{j=1}^{n_0} P(D \geq j)$$

to conclude the proof. □

7.23 Remark Analogously, one can define the upcrossings of X across $[a, b]$, and derive a similar inequality. Moreover, one can treat supermartingales in the same way.

For the aplication in the next section it is useful to remark that

$$(X_{t_0} - b)^+ \leq X_{t_0}^+ + b^+.$$

Hence

7.24 Corollary *Let $(X_t, t \in \mathbb{N}_0)$ be a submartingale. For every choice of $t_0 \in \mathbb{N}$, and $a, b \in \mathbb{R}$ with $a < b$, the inequality*

$$E(D^{t_0}_{[a,b]}) \leq \frac{1}{b-a} \big(b^+ + E(X_{t_0}^+)\big) \tag{7.17}$$

holds true.

7.5 Submartingale Convergence Theorem

7.25 Theorem (Doob's Submartingale Convergence Theorem) *Given a submartingale $(X_n, n \in \mathbb{N}_0)$, which is such that*

$$\sup_{n\in\mathbb{N}_0} E(|X_n|) < +\infty. \tag{7.18}$$

Then $(X_n, n \in \mathbb{N}_0)$ converges almost surely to an integrable, $\mathcal{F}_\infty^X$–measurable random variable X_∞.

7.26 Remark The condition (7.18) is equivalent to

$$\sup_{n\in\mathbb{N}_0} E(X_n^+) < +\infty, \tag{7.19}$$

because, on one hand, $X_n^+ \le |X_n|$, $n \in \mathbb{N}$. And on the other hand, we have $|X_n| \le 2X_n^+ - X_n$, and therefore the submartingale property of X gives

$$E(|X_n|) \le 2E(X_n^+) - E(X_n) \le 2E(X_n^+) - E(X_0).$$

Proof (of theorem 7.25) Let N denote the event on which $(X_n, n \in \mathbb{N})$ does not converge:

$$N = \{\liminf_n X_n < \limsup_n X_n\}.$$

Suppose that we have shown that $P(N) = 0$. Then with probability 1 converges to some random variable X_∞, which is $\mathcal{F}_\infty^X$–measurable, because so is every X_n, $n \in \mathbb{N}$. Moreover, X_∞ is integrable, because by Fatou's lemma we find

$$\begin{aligned} E(|X_\infty|) &= E(\liminf_n |X_n|) \\ &\le \liminf_n E(|X_n|) \\ &\le \liminf_n \sup_n E(|X_n|) \\ &< +\infty. \end{aligned}$$

Thus it remains to show that $P(N) = 0$. Consider $\omega \in N$, then there exist $a, b \in \mathbb{Q}$, $a < b$, so that

$$\liminf_n X_n(\omega) < a < b < \limsup_n X_n(\omega).$$

That is, infinitely many values of $(X_n(\omega), n \in \mathbb{N})$ are below a, and infinitely many are above b. Hence the number

$$D_{[a,b]}(\omega) = \sup_{k\in\mathbb{N}} D_{[a,b]}^k(\omega)$$

of downcrossings across $[a, b]$ is infinite: $D_{[a,b]}(\omega) = +\infty$. As the increasing limit of the downcrossings $D^k_{[a,b]}$, $k \in \mathbb{N}$, $D_{[a,b]}$ is a well-defined random variable with values in $\mathbb{N}_0 \cup \{+\infty\}$. Therefore

$$N \subset \bigcup_{a,b\in\mathbb{Q},\, a<b} \{D_{[a,b]} = +\infty\}.$$

From the monotone convergence theorem we get

$$E(D_{[a,b]}) = \sup_{k\in\mathbb{N}} E(D^k_{[a,b]}),$$

while corollary 7.24 gives

$$E(D_{[a,b]}) \leq \frac{1}{b-a}\bigl(b^+ + \sup_{k\in\mathbb{N}} E(X_k^+)\bigr).$$

The remark 7.26 together with our assumption (7.18) implies that the last expression is finite. Consequently we obtain that for all a, $b \in \mathbb{Q}$ with $a < b$, $P(D_{[a,b]} = +\infty) = 0$. Clearly this implies that $P(N) = 0$. □

7.27 Remark An analogous result also holds for supermartingales.

This convergence theorem, together with its companion theorems for which we refer to the pertinent literature (e.g., [3, 9, 16, 28, 30, 31]), has — apart from very interesting special, concrete applications (e.g., [3]) — far-reaching consequences in the theory of stochastic processes. For example, it allows to establish important regularity properties of the paths of (super-, sub-) martingales and of Markov processes.

Chapter 8

Markov Processes

Markov processes are models of stochastic processes without memory. Throughout we fix a measurable space $(\Omega, \mathcal{A})$, a time parameter domain T, a filtration $\mathcal{F} = (\mathcal{F}_t,\, t \in T)$ of sub–σ–algebras of $\mathcal{A}$, and a state space $(E, \mathcal{E})$. Throughout we shall assume that $T = \mathbb{R}_+$ or $T = \mathbb{N}_0$ — other time parameter domains need here and there some obvious adjustments. At the beginning of this chapter we shall work with one fixed probability measure P on $(\Omega, \mathcal{A})$.

8.1 The Simple Markov Property

8.1 Definition An $\mathcal{F}$–adapted stochastic process $X = (X_t,\, t \in T)$ with state space $(E, \mathcal{E})$ defined on $(\Omega, \mathcal{A}, P)$ is called a *Markov process (relative to $\mathcal{F}$)*, if for all s, $t \in T$ with $s \leq t$, $B \in \mathcal{E}$,

$$P(X_t \in B \mid \mathcal{F}_s) = P(X_t \in B \mid X_s) \tag{8.1}$$

holds true. Relation (8.1) is called the *(simple) Markov property of X*.

If the filtration $\mathcal{F}$ is not explicitly mentioned, we mean by "X is a Markov process" that X is a Markov process with respect to the natural filtration $\mathcal{F}_X$ of X.

Suppose that for every $s \in T$, $\mathcal{C}_s$ is a $\cap$–stable generator of $\mathcal{F}_s$ which contains Ω. Then lemma 5.4 states that an integrable, $\mathcal{F}_s$–measurable random variable Z is a version of the conditional expectation of an integrable random variable Y given $\mathcal{F}_s$, if and only if

$$E(Y; C) = E(Z; C) \tag{8.2}$$

holds for all choices of $C \in \mathcal{C}_s$. Therefore we obtain

8.2 Lemma *Suppose that X is $\mathcal{F}$–adapted, and that for all s, $t \in T$, $s \leq t$, and all $B \in \mathcal{E}$, $C \in \mathcal{C}_s$, the equation*

$$P\big(C \cap \{X_t \in B\}\big) = E\big(P(X_t \in B \mid X_s); C\big) \tag{8.3}$$

holds true. Then X is a Markov process relative to $\mathcal{F}$.

Notation If $(C, \mathcal{C})$ is a measurable space, denote by $\mathcal{M}_b(C, \mathcal{C})$ the space of bounded, real–valued, $\mathcal{C}/\mathcal{B}(\mathbb{R})$–measurable functions on C.

8.3 Lemma *X is a Markov process relative to $\mathcal{F}$ if and only if for every $n \in \mathbb{N}$, all $s, t_1, \dots, t_n \in T$ with $t_i \geq s$, $i = 1, \dots, n$, and all $f_1, \dots, f_n \in \mathcal{M}_b(E, \mathcal{E})$*

$$E\big(f_1(X_{t_1}) \cdots f_n(X_{t_n}) \,|\, \mathcal{F}_s\big) = E\big(f_1(X_{t_1}) \cdots f_n(X_{t_n}) \,|\, X_s\big) \tag{8.4}$$

holds true.

Proof We prove the lemma by induction. For $n = 1$ the statement follows directly from equation (8.1) by the usual three steps of the construction of the Lebesgue integral in combination with linearity and the monotone convergence theorem for conditional expectation, theorem 5.11.(a). Now suppose that equation (8.4) holds true, and assume that we are given $t_{n+1} \in T$, $t_{n+1} \geq s$, and $f_{n+1} \in \mathcal{M}_b(E, \mathcal{E})$. Without loss of generality we may suppose that $t_1 \leq t_2 \leq \cdots \leq t_{n+1}$. Then

$$\begin{aligned}
E\big(f_1(X_{t_1})&\cdots f_{n+1}(X_{t_{n+1}}) \,|\, \mathcal{F}_s\big) \\
&= E\big(E\big(f_1(X_{t_1}) \cdots f_{n+1}(X_{t_{n+1}}) \,|\, \mathcal{F}_{t_1}\big) \,|\, \mathcal{F}_s\big) \\
&= E\big(f_1(X_{t_1})\, E\big(f_2(X_{t_2}) \cdots f_{n+1}(X_{t_{n+1}}) \,|\, \mathcal{F}_{t_1}\big) \,|\, \mathcal{F}_s\big),
\end{aligned}$$

where we used property (d) of the conditional expextation in theorem 5.8, and the fact that $\mathcal{F}_s \subset \mathcal{F}_{t_1}$, as well as that $f_1(X_{t_1})$ is $\mathcal{F}_{t_1}$–measurable. For the inner conditional expectation we use the induction hypothesis, and obtain

$$\begin{aligned}
E\big(f_1(X_{t_1})&\cdots f_{n+1}(X_{t_{n+1}}) \,|\, \mathcal{F}_s\big) \\
&= E\big(f_1(X_{t_1})\, E\big(f_2(X_{t_2}) \cdots f_{n+1}(X_{t_{n+1}}) \,|\, X_{t_1}\big) \,|\, \mathcal{F}_s\big) \\
&= E\big(g(X_{t_1}) \,|\, \mathcal{F}_s\big),
\end{aligned}$$

for some $g \in \mathcal{M}_b(E, \mathcal{E})$. Here we applied the factorization lemma, lemma F.1, which states that the inner conditional expectation can be written as measurable function of X_{t_1}. Now we apply the statement of the lemma for $n = 1$ to conclude that

$$\begin{aligned}
E\big(f_1(X_{t_1})&\cdots f_{n+1}(X_{t_{n+1}}) \,|\, \mathcal{F}_s\big) \\
&= E\big(g(X_{t_1}) \,|\, X_s\big) \\
&= E\big(f_1(X_{t_1})\, E\big(f_2(X_{t_2}) \cdots f_{n+1}(X_{t_{n+1}}) \,|\, X_{t_1}\big) \,|\, X_s\big) \\
&= E\big(f_1(X_{t_1})\, E\big(f_2(X_{t_2}) \cdots f_{n+1}(X_{t_{n+1}}) \,|\, \mathcal{F}_{t_1}\big) \,|\, X_s\big) \\
&= E\big(E\big(f_1(X_{t_1}) f_2(X_{t_2}) \cdots f_{n+1}(X_{t_{n+1}}) \,|\, \mathcal{F}_{t_1}\big) \,|\, X_s\big),
\end{aligned}$$

and in the step before the last we applied the induction hypothesis backwards. Since X is adapted to $\mathcal{F}$, we find that that X_s is $\mathcal{F}_{t_1}$–measurable, that is, $\sigma(X_s) \subset \mathcal{F}_{t_1}$. Therefore we may use again theorem 5.9.(d) to get

$$E\big(f_1(X_{t_1}) \cdots f_{n+1}(X_{t_{n+1}}) \,|\, \mathcal{F}_s\big) = E\big(f_1(X_{t_1}) \cdots f_{n+1}(X_{t_{n+1}}) \,|\, X_s\big),$$

and the proof is done. □

8.4 Corollary *X is a Markov process relative to $\mathcal{F}$ if and only if for every $n \in \mathbb{N}$, all $s, t_1, \ldots, t_n \in T$ with $t_i \geq s$, $i = 1, \ldots, n$, and all $B_1, \ldots, B_n \in \mathcal{E}$*

$$P\big(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n \mid \mathcal{F}_s\big) = P\big(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n \mid X_s\big) \tag{8.5}$$

holds true.

$\mathcal{G}_s^X$, $s \in T$, is defined to be the σ–algebra of the future of X after time s: $\mathcal{G}_s^X = \sigma(X_u,\, u \geq s)$.

8.5 Theorem *Suppose that X is an $\mathcal{F}$–adapted stochastic process. The following statements are equivalent:*

(a) *X is a Markov process relative to $\mathcal{F}$.*

(b) *For every $s \in T$, and all $Z \in \mathcal{M}_b\big(\Omega, \mathcal{G}_s^X\big)$ the following equation holds:*

$$E(Z \mid \mathcal{F}_s) = E(Z \mid X_s). \tag{8.6}$$

(c) *For every $s \in T$, and all $Y \in \mathcal{M}_b(\Omega, \mathcal{F}_s)$, $Z \in \mathcal{M}_b(\Omega, \mathcal{G}_s^X)$*

$$E(YZ \mid X_s) = E(Y \mid X_s)\, E(Z \mid X_s) \tag{8.7}$$

holds.

(d) *For every $s \in T$, and all $A \in \mathcal{F}_s$, $B \in \mathcal{G}_s^X$*

$$P(A \cap B \mid X_s) = P(A \mid X_s)\, P(B \mid X_s) \tag{8.8}$$

holds true.

Proof
"(a) $\Leftrightarrow$ (b)" The implication "(b) $\Rightarrow$ (a)" is trivial: We just have to choose Z as the indicator of $\{X_t \in B\}$, $s \leq t$, which clearly belongs to $\mathcal{M}_b\big(\Omega, \mathcal{G}_s^X\big)$. For the converse implication we apply the monotone class theorem for functions, theorem 2.7: Let $\mathcal{H}$ denote the set of all bounded, real-valued random variables Z which are $\mathcal{G}_s^X$–measurable, and are such that equation (8.6) holds. By linearity of the conditional expectation (see theorem 5.7), $\mathcal{H}$ is a vector space, and by the monotone convergence theorem for conditional expectations, theorem 5.11.(a), $\mathcal{H}$ contains all limits of sequences of positive, monotone increasing $\mathcal{G}_s^X$–measurable random variables to a bounded, positive random variable. Furthermore, since X is a Markov process, corollary 8.4 shows that $\mathcal{H}$ contains all indicators of the form 1_C for $C \in \mathcal{Z}_{\geq s}^X$, where $\mathcal{Z}_{\geq s}^X$ is the family of all sets of the form $\{X_{t_1} \in B_1, \ldots X_{t_n} \in B_n\}$, with $n \in \mathbb{N}$, $t_1, \ldots, t_n \geq s$, and $B_1, \ldots, B_n \in \mathcal{E}$. But $\mathcal{Z}_{\geq s}^X$ is a $\cap$–stable generator of $\mathcal{G}_s^X$, and therefore

the monotone class theorem 2.7 implies that $\mathcal{H}$ contains all bounded, $\mathcal{G}_s^X$–measurable random variables.

"(b) $\Leftrightarrow$ (c)" Assume first that (b) is true. Since X is $\mathcal{F}$–adapted, we have $\sigma(X_s) \subset \mathcal{F}_s$, and therefore can compute with the properties of the conditional expectation stated in theorem 5.7.(e), (f), and theorem 5.9.(c) as follows:

$$\begin{aligned} E(YZ \mid X_s) &= E\big(Y\, E(Z \mid \mathcal{F}_s) \,\big|\, X_s\big) \\ &= E\big(Y\, E(Z \mid X_s) \,\big|\, X_s\big) \\ &= E(Y \mid X_s)\, E(Z \mid X_s). \end{aligned}$$

Now suppose that (c) holds. Then we have to show that $E(Z \mid X_s)$ is a version of $E(Z \mid \mathcal{F}_s)$. Thus by theorem 5.8 we have to prove that for all bounded random variables Y which are $\mathcal{F}_s$–measurable, we have

$$E(YZ) = E\big(Y\, E(Z \mid X_s)\big).$$

But with (c) we get

$$\begin{aligned} E(YZ) &= E\big(E(YZ \mid X_s)\big) \\ &= E\big(E(Y \mid X_s)\, E(Z \mid X_s)\big) \\ &= E\big(E(Y\, E(Z \mid X_s) \mid X_s)\big) \\ &= E\big(Y\, E(Z \mid X_s)\big), \end{aligned}$$

where we used again theorem 5.7.(e), and (f).

"(c) $\Leftrightarrow$ (d)" is similar to "(a) $\Leftrightarrow$ (b)", and left as an *exercise* to the reader. □

8.6 Remark The equivalent description of the Markov property by equation (8.8) is perhaps the most intuitive one, since it says that conditional on the present of X (given by X_s), past and future are independent.

In the case that $\mathcal{F}$ is the natural filtration $\mathcal{F}^X$ of X, we know from lemma 3.12 that a $\cap$–stable generator $\mathcal{C}_s$ of $\mathcal{F}_s^X$ consists of X–cylinder sets up to time s. Moreover, the family of cylinder sets contains Ω. Thus in this case lemma 8.2 entails that the Markov property (8.1) is equivalent to

$$\begin{aligned} E\big(P(X_t \in B \mid X_s);\, X_{s_0} \in B_0, \dots, X_{s_n} \in B_n\big) \\ = P(X_{s_0} \in B_0, \dots, X_{s_n} \in B_n, X_t \in B), \end{aligned} \tag{8.9}$$

for all choices of $n \in \mathbb{N}$, $s_0, \dots, s_n \in T$, $s_0, \dots, s_n \le s$, $B_0, \dots, B_n \in \mathcal{E}$. More generally,

8.7 Lemma *An E–valued stochastic process is a Markov process, if and only if for all $s, t \in T$, with $s \le t$, and all choices of $n \in \mathbb{N}$, $s_0, \dots, s_n \in T$ with $s_0, \dots, s_n \le s$, and all $g \in \mathcal{M}_b(E^{n+1}, \mathcal{E}^{n+1})$, $f \in \mathcal{M}_b(E, \mathcal{E})$, the relation*

$$E\big(g(X_{s_0}, \dots, X_{s_n})\, E(f(X_t) \mid X_s)\big) = E\big(g(X_{s_0}, \dots, X_{s_n}) f(X_t)\big) \tag{8.10}$$

holds true.

Sketch of the Proof Choosing g as the indicator of a Cartesian product of sets in $\mathcal{E}$, and f as the indicator of a set in $\mathcal{E}$, we get equation (8.9). For the converse implication one carries out an argument on the basis of the monotone class theorem for functions, theorem 2.7, where one considers $\mathcal{H}$ as the space of all bounded, measurable functions g on $(E^{n+1}, \mathcal{E}^{n+1})$ so that (8.10) holds true with $f = 1_B$, $B \in \mathcal{E}$. Then one makes another application of the monotone class theorem 2.7 to generalize from $f = 1_B$ to $f \in \mathcal{M}_b(E, \mathcal{E})$. □

8.2 Markov Property and Markovian Semigroups

Consider a stochastic process $X = (X_t, t \in T)$ with values in a polish space $(E, \mathcal{E})$ which is equivalent to a stochastic process which has been canonically constructed with a normal semigroup $U = (U_t, t \in T)$ of Markovian kernels, and an initial law μ (see chapter 3, section 3.3). In particular, the finite dimensional distributions of X are given by formula (3.26). As a filtration we choose the natural filtration $\mathcal{F}^X$ of X.

8.8 Theorem *X is a Markov process relative to $\mathcal{F}^X$. Moreover, for all $s, t \in T$, $s \leq t$, $B \in \mathcal{E}$,*

$$P(X_t \in B \mid \mathcal{F}_s^X) = U_{t-s}(X_s, B) \tag{8.11}$$

holds true.

Proof We verify equation (8.10) in lemma 8.7. To this end, let g and f be as there. We set $s = s_n$, $J = \{s_0, s_1, \ldots, s_n\}$, $(J, t) = \{s_0, s_1, \ldots, s_n, t\}$, and use the notation of section 3.3. Moreover, we denote

$$\begin{aligned}(g \otimes f)&(x_0, x_1, \ldots, x_n, x_{n+1}) \\ &= g(x_0, x_1, \ldots, x_n) f(x_{n+1}), \qquad x_0, \ldots, x_n, x_{n+1} \in E.\end{aligned}$$

The right hand side of equation (8.10) is

$$\begin{aligned}E\big(g(X_J) f(X_t)\big) &= \mu U_{(J,t)}(g \otimes f) \\ &= \int_{E^{n+2}} \mu(dx_0)\, U_{s_1}(x_0, dx_1) \cdots U_{s_n - s_{n-1}}(x_{n-1}, dx_n) \\ &\qquad \times U_{t-s_n}(x_n, dx_{n+1})\, g(x_0, \ldots, x_n) f(x_{n+1}) \\ &= \int_{E^{n+1}} \mu(dx_0)\, U_{s_1}(x_0, dx_1) \cdots U_{s_n - s_{n-1}}(x_{n-1}, dx_n) \\ &\qquad \times g(x_0, \ldots, x_n) \big(U_{t-s_n} f\big)(x_n) \\ &= E\big(g(X_J)\, (U_{t-s} f)(X_s)\big) \end{aligned} \tag{8.12}$$

where we used formula (3.26) twice. Moreover we used the fact that we can do the integrations in any order, as had been argued in section 3.4 of chapter 3. If we choose g as the indicator of Cartesian products of sets in $\mathcal{E}$, $g(X_J)$ becomes the indicator of an X–cylinder set (with time points less or equal than s), and we can apply lemma 3.12 to obtain

$$E\big(f(X_t)\,\big|\,\mathcal{F}_s^X\big) = U_{t-s}f(X_s). \tag{8.13}$$

For the choice $f = 1_B$, $B \in \mathcal{E}$, we get formula (8.11). Taking the conditional expectation with respect to X_s on both sides of equation (8.13), we find with theorem 5.9.(d) that

$$E\big(f(X_t)\,\big|\,X_s\big) = U_{t-s}f(X_s), \tag{8.14}$$

because $\sigma(X_s) \subset \mathcal{F}_s^X$. Inserting this on the right hand side of (8.12) we get equation (8.10), and therefore the Markov property of X. □

From theorem 3.40 we immediately obtain

8.9 Corollary *Every stochastic process with values in $\mathbb{R}^d$ with independent, stationary increments is a Markov process, and in particular all Lévy processes are Markov processes.*

8.10 Examples Random walks, Brownian motions, and Poisson processes are Markov processes. ◇

We somewhat specialize the situation described at the beginning of this section: Let X the E–valued canonical stochastic process constructed with the normal Markovian semigroup U, and initial law μ. That is, we do *not* consider a process equivalent to X but the canonical coordinate process X on path space, $X_t(\omega) = \omega(t)$, $\omega \in \Omega = E^T$, subject to the probability measure P^μ constructed as in section 3.3 from μ and U. Let us consider the special case where $\mu = \varepsilon_x$, $x \in E$, and write P_x for P^{ε_x}, as we did in section 3.3. Then relation (8.11) reads as follows

$$P_x(X_t \in B \,|\, \mathcal{F}_s^X) = U_{t-s}(X_s, B), \tag{8.15}$$

and we also have (see formula (3.27))

$$U_{t-s}(x, B) = P_x(X_{t-s} \in B). \tag{8.16}$$

Thus we can write formula (8.15) as follows

$$P_x(X_t \in B \,|\, \mathcal{F}_s^X) = P_{X_s}(X_{t-s} \in B). \tag{8.17}$$

The right hand side makes sense, because relation (8.16) shows that for every $B \in \mathcal{E}$ the mapping $x \mapsto P_x(X_{t-s} \in B)$ is measurable, so that its composition with X_s defines an $\mathcal{F}_s^X$–measurable random variable. In order to generalize this we prove now that the mapping $x \mapsto P_x(\Lambda)$ for $\Lambda \in \mathcal{E}^T$ is measurable from $(E, \mathcal{E})$ into $([0,1], \mathcal{B}([0,1]))$, or equivalently, that $(x, \Lambda) \mapsto P_x(\Lambda)$ is a kernel from $(E, \mathcal{E})$ into its path space over T (see definition 3.20).

First consider $\Lambda \in \mathcal{Z}^T$, i.e., $\Lambda = \{X_J \in B\}$ with $J = \{t_0 = 0, t_1, \dots, t_n\} \subset T$, $n \in \mathbb{N}$, $B = B_0 \times \cdots \times B_n$, and $B_0, \dots, B_n \in \mathcal{E}$. In exercise 3.24 it was shown that the action of a Markovian kernel maps a bounded, measurable function into a bounded, measurable function, and in the discussion of equation (3.23) it was argued that in multiple applications of this kind one can — similarly as in Fubini's theorem — do the integrations in any order. Therefore, by doing the integration with respect to x_0 first and with respect to x_1 last, we obtain

$$\begin{aligned} P_x(\Lambda) & \\ &= \int \varepsilon_x(dx_0)\, U_{t_1}(x_0, dx_1) \cdots U_{t_n - t_{n-1}}(x_{n-1}, dx_n)\, 1_B(x_0, \ldots, x_n) \\ &= \int U_{t_1}(x, dx_1)\, U_{t_2 - t_1}(x_1, dx_2) \cdots U_{t_n - t_{n-1}}(x_{n-1}, dx_n)\, 1_B(x, x_1 \ldots, x_n) \\ &= 1_{B_0}(x) \int U_{t_1}(x, dx_1)\, F(x_1), \end{aligned}$$

for some bounded, measurable function F. Another application of exercise 3.24 implies that the last expression is measurable with respect to x. Thus so is $x \mapsto P_x(\Lambda)$ for every $\Lambda \in \mathcal{Z}^T$. $\mathcal{S}$ denotes the family of all sets in $\mathcal{E}^T$ so that $x \mapsto P_x(A)$ is measurable. We have just shown that $\mathcal{S} \supset \mathcal{Z}^T$, and we recall that $\mathcal{Z}^T$ is $\cap$–stable. $\mathcal{S}$ is a d-system: That $\Omega \in \mathcal{S}$ is obvious, and for A, $B \in \mathcal{S}$ with $A \subset B$, we find $P_x(B \setminus A) = P_x(B) - P_x(A)$, so that $x \mapsto P_x(B \setminus A)$ clearly is measurable. Finally let $(A_n,\, n \in \mathbb{N})$ be a sequence in $\mathcal{S}$ which increases to A. Then for every $x \in E$, $P_x(A_n)$ converges (actually increases) to $P_x(A)$, due to the continuity of P_x as a probability measure. Therefore $A \in \mathcal{S}$, and the monotone class theorem, theorem 2.4, implies that $\mathcal{S} \supset \mathcal{E}^T$. (Recall that $\mathcal{E}^T$ is generated by $\mathcal{Z}^T$, cf. lemma 3.11.)

We have proved the following

8.11 Theorem *Suppose that $(E, \mathcal{E})$ is a polish space, and that $(P_x,\, x \in E)$ is a family of probability measures on $(E^T, \mathcal{E}^T)$ constructed from the family $(\varepsilon_x,\, x \in E)$ of initial laws and a Markovian semigroup $U = (U_t,\, t \in T)$. Then $(x, \Lambda) \mapsto P_x(\Lambda)$ defines a Markovian kernel from $(E, \mathcal{E})$ to $(E^T, \mathcal{E}^T)$. Moreover, for the associated Markovian canonical coordinate process $X = (X_t,\, t \in T)$,*

$$P_x(X_{s+t} \in B \mid \mathcal{F}_s^X) = P_{X_s}(X_t \in B) \tag{8.18}$$

holds true for all $x \in E$, $s, t \in T$, $B \in \mathcal{E}$.

8.3 The Universal Markov Property

The special properties of stochastic processes, which are canonically constructed with a Markovian semigroup, described in theorem 8.11, give rise to the following definition.

8.12 Definition Assume that $(E, \mathcal{E})$ is a measurable space, and that T is a time parameter domain. A quintuple $\big(\Omega, \mathcal{A}, P = (P_x,\, x \in E), \mathcal{F} = (\mathcal{F}_t,\, t \in T), X = (X_t,\, t \in T)\big)$ is called a *universal Markov process*, if the following statements hold true:

(i) For every $x \in E$, $(\Omega, \mathcal{A}, \mathcal{F}, P_x)$ is a filtered probability space, X is an E–valued, $\mathcal{F}$–adapted stochastic process defined thereon so that $P_x(X_0 = x) = 1$;

(ii) for every $A \in \mathcal{A}$ the mapping $x \mapsto P_x(A)$ is $\mathcal{E}/\mathcal{B}([0,1])$–measurable;

(iii) for all $s, t \in T$, $x \in E$, $B \in \mathcal{E}$ the *universal Markov property of X*

$$P_x\big(X_{s+t} \in B \mid \mathcal{F}_s\big) = P_{X_s}\big(X_t \in B\big) \tag{8.19}$$

is valid.

8.13 Remark In most of the literature, the universal Markov property as defined above is simply called the *Markov property*. Sometimes the condition $P_x(X_0 = x) = 1$ is not required, and if it holds true, the process is called *normal*.

Notation From now on we shall denote the expectation with respect to P_x, $x \in E$, by $E_x(\cdot)$.

8.14 Lemma $(\Omega, \mathcal{A}, P, \mathcal{F}, X)$ *is a universal Markov process with respect to* $\mathcal{F}$*, if and only if for every* $n \in \mathbb{N}$*, all* $t_1, \dots, t_n \in T$*,* $s \in T$*, all* $f_1, \dots, f_n \in \mathcal{M}_b(E, \mathcal{E})$*, and every* $x \in E$*,*

$$E_x\big(f_1(X_{s+t_1}) \cdots f_n(X_{s+t_n}) \mid \mathcal{F}_s\big) = E_{X_s}\big(f_1(X_{t_1}) \cdots f_n(X_{t_n})\big) \tag{8.20}$$

holds true.

The proof of lemma 8.14 is similar to the proof of lemma 8.3, and can be omitted.

In theorem 8.11 we have shown that associated with a normal Markovian semigroup of kernels on a polish space, one can construct a universal Markov process which has these kernels as transition kernels, namely, we just have to choose the canonically constructed process on the path space. Now we show that also the converse statement is true.

8.15 Theorem *Given a universal Markov process* $(\Omega, \mathcal{A}, P, \mathcal{F}, X)$ *with state space* $(E, \mathcal{E})$*. Then*

$$U_t(x, B) = P_x(X_t \in B), \qquad t \in T,\ x \in E,\ B \in \mathcal{E}, \tag{8.21}$$

defines a normal semigroup of Markovian kernels on $(E, \mathcal{E})$*.*

Proof The hypotheses on $(\Omega, \mathcal{A}, P, \mathcal{F}, X)$ yield that for every $t \in T$, U_t is indeed a Markovian kernel on $(E, \mathcal{E})$. Moreover, for $t = 0$ we find $U_0(x, B) = P_x(X_0 \in B) = 1_B(x) = \varepsilon_x(B)$, so that U_0 is the unit kernel I. It remains to prove the semigroup property of $U = (U_t,\ t \in T)$. Before we do this, we remark that a routine *exercise* along the lines of the construction of the Lebesgue integral gives that equation (8.21) is equivalent to

$$(U_t f)(x) = \int_E f(y)\, U_t(x, dy) = E_x\big(f(X_t)\big) \tag{8.22}$$

for all $x \in E$, $t \in T$, $f \in \mathcal{M}_b(E, \mathcal{E})$. Then with the universal Markov property and (8.22) we get

$$\begin{aligned} U_{s+t}(x, B) &= P_x(X_{s+t} \in B) \\ &= E_x\big(P_x(X_{s+t} \in B \mid \mathcal{F}_s)\big) \\ &= E_x\big(P_{X_s}(X_t \in B)\big) \\ &= \int_E U_s(x, dy)\, P_y(X_t \in B) \\ &= \int_E U_s(x, dy)\, U_t(y, B) \\ &= (U_s \circ U_t)(x, B), \end{aligned}$$

and we are done. □

8.16 Corollary *For a universal Markov process* $(\Omega, \mathcal{A}, P, \mathcal{F}, X)$ *with state space* $(E, \mathcal{E})$ *the relation*

$$E_x\big(f(X_{s+t}) \,\big|\, \mathcal{F}_s\big) = (U_t f)(X_s) \tag{8.23}$$

holds true for all $x \in E$, $f \in \mathcal{M}_b(E, \mathcal{E})$, $s, t \in T$, *where* $U = (U_t,\, t \in T)$ *is the semigroup generated by* X *(cf. theorem 8.15).*

Proof We only need to combine equations (8.19), (8.21), and then to apply a routine argument to go from indicators to general functions f in $\mathcal{M}_b(E, \mathcal{E})$ to make the proof of this statement. □

With a view towards the topic of the next section, namely the *strong* Markov property, we make another reformulation of the Markov property. Let us go back to the path space $(E^T, \mathcal{E}^T)$ and the canonical coordinate process $X_t(\omega) = \omega(t)$, $t \in T$, $\omega \in E^T$. Then on the path space E^T there is a naturally defined family $\theta = (\theta_s,\, s \in T)$ of mappings θ_s which implement a time translation, that is,

$$(\theta_s \omega)(t) = \omega(s + t),$$

Hence we find $X_t \circ \theta_s = X_{s+t}$. We now build this new device into our general framework.

8.17 Definition Suppose that $X = (X_t,\, t \in T)$ is an E–valued stochastic process defined on a probability space $(\Omega, \mathcal{A}, P)$. A family $\theta = (\theta_s,\, s \in T)$ of mappings from Ω into itself so that $X_t \circ \theta_s = X_{s+t}$, $s, t \in T$, is called a family of *shift operators (for X)*.

As before we denote $\mathcal{F}_s^X = \sigma(X_u,\, u \le s)$, $\mathcal{G}_s^X = \sigma(X_u,\, u \ge s)$, and recall that $\mathcal{F}_\infty^X$ is the σ–algebra generated by $\mathcal{F}^X$, or equivalently $\mathcal{F}_\infty^X = \sigma(X_t,\, t \in T)$. From lemma 3.12 we know, that the family of X–cylinders $\mathcal{Z}^X$ is a $\cap$–stable generator of $\mathcal{F}_\infty^X$, and similarly the family $\mathcal{Z}_{\ge s}^X$ of cylinders sets generated by $(X_u,\, u \ge s)$ is a $\cap$–stable generator of $\mathcal{G}_s^X$, $s \in T$. Let $\Lambda_s \in \mathcal{Z}_{\ge s}^X$, then it is of the form

$$\Lambda_s = \{X_{u_1} \in B_1, \dots X_{u_n} \in B_n\}$$

with $n \in \mathbb{N}$, $s \le u_1 \le \cdots \le u_n$, $B_1, \dots, B_n \in \mathcal{E}$. Then

$$\Lambda_s = \theta_s^{-1}(\Lambda),$$

with

$$\Lambda = \{X_{u_1-s} \in B_1, \dots X_{u_n-s} \in B_n\} \in \mathcal{Z}^X.$$

Obviously, for any choice of $s \in T$ we can write every $\Lambda_s \in \mathcal{Z}_{\ge s}^X$ in this way. Thus $\mathcal{Z}_{\ge s}^X = \theta_s^{-1}(\mathcal{Z}^X)$, and consequently for $s \in T$,

$$\mathcal{G}_s^X = \sigma\big(\mathcal{Z}_{\ge s}^X\big) = \sigma\big(\theta_s^{-1}(\mathcal{Z}^X)\big) = \theta_s^{-1}\big(\sigma(\mathcal{Z}^X)\big) = \theta_s^{-1}(\mathcal{F}_\infty^X),$$

where we made use of lemma 3.6.(a). We have proved the following

8.18 Lemma *Suppose that $\theta = (\theta_s,\, s \in T)$ is a family of shift operators on Ω for a stochastic process $X = (X_t,\, t \in T)$. Then for every $s \in T$, θ_s is $\mathcal{G}_s^X/\mathcal{F}_\infty^X$– and therefore in particular $\mathcal{F}_\infty^X/\mathcal{F}_\infty^X$–measurable. If Y is a real-valued, $\mathcal{F}_\infty^X$–measurable random variable, then $Y \circ \theta_s$, $s \in T$, is $\mathcal{G}_s^X$–measurable.*

Now suppose that $(\Omega, \mathcal{A}, P, \mathcal{F}, X)$ is a universal Markov process, and that θ is a family of shift operators for X. Then we can write formula (8.20) in lemma 8.14 in the following way

$$E_x\big(Y \circ \theta_s \mid \mathcal{F}_s\big) = E_{X_s}(Y),$$

where Y is the random variable

$$Y = f_1(X_{t_1}) \cdots f_n(X_{t_n}).$$

A (by now) routine application of the monotone class theorem for functions then yields

8.19 Theorem *Suppose that $(\Omega, \mathcal{A}, P, \mathcal{F}, X, \theta)$ is a universal Markov process with shift θ. Then*

$$E_x(Y \circ \theta_s \mid \mathcal{F}_s) = E_{X_s}(Y) \tag{8.24}$$

for all $s \in T$, and all bounded, $\mathcal{F}_\infty^X$–measurable random variables Y.

8.4 Strong Markov Property and Feller Semigroups

A very powerful idea of the theory of stochastic processes, and for its applications, is to try to establish a Markov property for *random* times. That is, for example, to generalize the universal Markov property (8.24), such that $s \in T$ there is replaced by a stopping time τ. Thus the desired formula would look like

$$E_x(Y \circ \theta_\tau \mid \mathcal{F}_\tau) = E_{X_\tau}(Y), \tag{8.25}$$

and the main purpose of this section is to prove such formulae.

Throughout this section we restrict ourselves to the case where the state space $(E, \mathcal{E})$ is equal to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $d \in \mathbb{N}$. [1]

Clearly, since the left hand side of (8.25) is an $\mathcal{F}_\tau$–measurable random variable, we must have that X_τ is $\mathcal{F}_\tau$–measurable. From exercise 6.21 we know that this is true if $T = \mathbb{N}_0$ and X is $\mathcal{F}$–adapted, which we assume anyway. But for $T = \mathbb{R}_+$ we need some additional hypothesis to guarantee that X is $\mathcal{F}$–progressive. Lemma 6.8 states that this is the case, if the paths of X are continuous from the right — an assumption we shall make throughout below.

As a central ingredient for the purpose of showing a Markov property for a random time, we consider the semigroup $U = (U_t,\, t \in \mathbb{R}_+)$ associated with a Markov process X (see theorem 8.15) in more detail. In what follows, we shall no longer consider U as a semigroup of kernels on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, but rather as a semigroup acting on functions via relation (3.15). To this end, we shall consider two spaces $\mathcal{M}_b(\mathbb{R}^d)$, $C_0(\mathbb{R}^d)$ of functions on $\mathbb{R}^d$: $\mathcal{M}_b(\mathbb{R}^d)$ is the space of all real-valued, bounded, $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R})$–measurable functions on $\mathbb{R}^d$, $C_0(\mathbb{R}^d)$ is the subspace of continuous functions which vanish at infinity. Both function spaces are considered as equipped with the sup-norm, $\|\cdot\|_\infty$, and with this norm, they are Banach spaces (see appendix G).

A priori, the semigroup $U = (U_t,\, t \in \mathbb{R}_+)$ defined by a universal Markov process $X = (X_t,\, t \in \mathbb{R}_+)$ acts on $\mathcal{M}_b(\mathbb{R}^d)$ (cf. (8.22)):

$$(U_t f)(x) = \int_{\mathbb{R}^d} U_t(x, dy)\, f(y) = E_x\big(f(X_t)\big), \; f \in \mathcal{M}_b(\mathbb{R}^d),\, x \in \mathbb{R}^d,\, t \in \mathbb{R}_+. \tag{8.26}$$

Indeed, exercise 3.24 shows that $U_t f$ is measurable. Moreover,

$$|(U_t f)(x)| \le E_x\big(|f(X_t)|\big) \le \|f\|_\infty,$$

so that we obtain

$$\|U_t f\|_\infty \le \|f\|_\infty. \tag{8.27}$$

Therefore, U is a *contraction semigroup* on $\mathcal{M}_b(\mathbb{R}^d)$ — namely, for every $t \in \mathbb{R}_+$, U_t acts as a contraction on the Banach space $\mathcal{M}_b(\mathbb{R}^d)$.

[1] A reasonably general setting would be that E is a locally compact metric space, the general framework one finds in some textbooks is that of a locally compact topological space E with countable topological basis.

8.20 Definition (Feller Semigroup) A contraction semigroup $U = (U_t,\, t \in \mathbb{R}_+)$ acting on $\mathcal{M}_b(\mathbb{R}^d)$ is called a *Feller semigroup*, if the following hold true

(i) $U_0 = \mathrm{id}$,

(ii) for every $t \in \mathbb{R}_+$, U_t maps $C_0(\mathbb{R}^d)$ into itself,

(iii) for every $x \in \mathbb{R}^d$, $f \in C_0(\mathbb{R}^d)$, $(U_t f)(x) \to f(x)$ as $t > 0$ decreases to 0.

8.21 Remarks

(a) Often, a universal Markov process whose associated semigroup is a Feller semigroup is called a *Feller process*.

(b) There are various different definitions for a Feller semigroup in the literature, but most of them are (more or less) equivalent to definition 8.20.

(c) It is well-known (e.g, [15, Theorem 19.6]), that under conditions (i) and (ii), condition (iii) is equivalent to "U is strongly continuous on $(C_0(\mathbb{R}^d), \|\cdot\|_\infty)$", that is, $\|U_t f - f\|_\infty \to 0$ as $t > 0$ decreases to 0.

8.22 Exercise Show that the semigroup associated with a d–dimensional Brownian motion is a Feller semigroup.

8.23 Lemma *Suppose that $X = (X_t,\, t \in \mathbb{R}_+)$ is an $\mathbb{R}^d$–valued universal Markov process relative to a filtration $\mathcal{F} = (\mathcal{F}_t,\, t \in \mathbb{R}_+)$, with exclusively right continuous paths. Assume furthermore that τ is an $\mathcal{F}$–stopping time, and that the semigroup $U = (U_t,\, t \in \mathbb{R}_+)$ generated by X is a Feller semigroup. Then for all $t \in \mathbb{R}_+$, $f \in C_0(\mathbb{R}^d)$, $x \in \mathbb{R}^d$,*

$$E_x\big(f(X_{t+\tau}) \,\big|\, \mathcal{F}_\tau\big) = E_{X_\tau}\big(f(X_t)\big). \tag{8.28}$$

holds true on $\{\tau < +\infty\}$.

Proof Observe that by definition of the semigroup U, equation (8.28) is the same as

$$E_x\big(f(X_{t+\tau}) \,\big|\, \mathcal{F}_\tau\big) = (U_t f)(X_\tau). \tag{8.29}$$

Thus we have to prove that $(U_t f)(X_\tau)$ is a version of the left hand side of (8.28).

Since X is adapted to $\mathcal{F}$, and it has right continuous paths, we know from lemma 6.8 that X is progressively measurable relative to $\mathcal{F}$. Then lemma 6.22 implies that X_τ is $\mathcal{F}_\tau$–measurable. Since the semigroup U maps measurable functions into measurable functions, we conclude that $(U_t f)(X_\tau)$ is $\mathcal{F}_\tau$–measurable. Therefore it remains to show that for every $\Lambda \in \mathcal{F}_\tau$, $\Lambda \subset \{\tau < +\infty\}$, we have

$$E_x\big(f(X_{t+\tau}); \Lambda\big) = E_x\big((U_t f)(X_\tau); \Lambda\big). \tag{8.30}$$

1st step: We suppose in addition, that the range of τ is contained in a finite set $\{r_1, \dots, r_n, +\infty\}$, $n \in \mathbb{N}$. We set $\Lambda_k = \Lambda \cap \{\tau = r_k\}$, $k = 1, \dots, n$. (Note that $\Lambda_k \subset \{\tau < +\infty\}$.) Then

$$\begin{aligned} E_x\big(f(X_{t+\tau}); \Lambda\big) &= \sum_{k=1}^n E_x\big(f(X_{t+\tau}); \Lambda_k\big) \\ &= \sum_{k=1}^n E_x\big(f(X_{t+r_k}); \Lambda_k\big). \end{aligned}$$

For $r \in \mathbb{R}_+$ we find

$$\{\tau = r\} = \bigcap_{m \in \mathbb{N}} \big(\{\tau \le r\} \setminus \{\tau \le r - 1/m\}\big),$$

and

$$\Lambda \cap \{\tau = r\} = \bigcap_{m \in \mathbb{N}} \Big(\big(\Lambda \cap \{\tau \le r\}\big) \setminus \big(\Lambda \cap \{\tau \le r - 1/m\}\big)\Big).$$

By definition of $\mathcal{F}_\tau$, $\Lambda \in \mathcal{F}_\tau$ implies that

$$\begin{gathered} \Lambda \cap \{\tau \le r\} \in \mathcal{F}_r, \\ \Lambda \cap \{\tau \le r - 1/m\} \in \mathcal{F}_{(r-1/m)\vee 0} \subset \mathcal{F}_r. \end{gathered}$$

Hence we obtain $\Lambda \cap \{\tau = r\} \in \mathcal{F}_r$. Now we use the Markov property (8.23) of X to compute in the following way:

$$\begin{aligned} E_x\big(f(X_{t+\tau}); \Lambda\big) &= \sum_{k=1}^n E_x\big(E_x(f(X_{t+r_k}) \,|\, \mathcal{F}_{r_k}); \Lambda_k\big) \\ &= \sum_{k=1}^n E_x\big((U_t f)(X_{r_k}); \Lambda_k\big) \\ &= \sum_{k=1}^n E_x\big((U_t f)(X_\tau); \Lambda_k\big) \\ &= E_x\big((U_t f)(X_\tau); \Lambda\big), \end{aligned}$$

and equation (8.30) is proved under the additional assumption on τ.

2nd step: From exercise 6.14 we know that every $\mathcal{F}$–stopping time can be approximated from above by a sequence $(\tau_n,\, n \in \mathbb{N})$ of $\mathcal{F}$–stopping times which are as in step 1. Take such a sequence, then for every $n \in \mathbb{N}$, and every $\Lambda \in \mathcal{F}_\tau$, $\Lambda \subset \{\tau < +\infty\}$, we get from step 1

$$E_x\big(f(X_{t+\tau_n}); \Lambda\big) = E_x\big((U_t f)(X_{\tau_n}); \Lambda\big). \tag{8.31}$$

As n tends to infinity, the right continuity of the paths of X entails that for every $\omega \in \{\tau < +\infty\}$, $X_{\tau_n}(\omega) \to X_\tau(\omega)$, and $X_{t+\tau_n}(\omega) \to X_{t+\tau}(\omega)$. By the Feller property of U, $U_t f$ belongs to $C_0(\mathbb{R}^d)$, so that the expression under the expectation on the right hand side converges to $(U_t f)(X_\tau)1_\Lambda$. As a function in $C_0(\mathbb{R}^d)$, f is bounded, and U_t is a contraction, so that this expression is bounded by $\|f\|_\infty$, which is an integrable majorant for it. Therefore the dominated convergence theorem implies that the right hand side of (8.31) converges to (8.30) as n tends to infinity. For the left hand side of (8.31) we only need the argument with the dominated convergence theorem to conclude that it converges to the left hand side of (8.30) with $n \to \infty$. $\square$

8.24 Theorem *Suppose that $X = (X_t, t \in \mathbb{R}_+)$ is an $\mathbb{R}^d$–valued universal Markov process relative to a filtration $\mathcal{F} = (\mathcal{F}_t, t \in \mathbb{R}_+)$, with exclusively right continuous paths. Assume furthermore that τ is an $\mathcal{F}$–stopping time, and that the semigroup $U = (U_t, t \in \mathbb{R}_+)$ generated by X is a Feller semigroup. Then for all $t \in \mathbb{R}_+$, $x \in \mathbb{R}^d$, $B \in \mathcal{B}(\mathbb{R}^d)$, $f \in \mathcal{M}_b(\mathbb{R}^d)$, the* strong Markov properties

$$P_x(X_{t+\tau} \in B \,|\, \mathcal{F}_\tau) = P_{X_\tau}(X_t \in B) = U_t(X_\tau, B), \tag{8.32}$$

and

$$E_x\big(f(X_{t+\tau}) \,\big|\, \mathcal{F}_\tau\big) = E_{X_\tau}\big(f(X_t)\big) = (U_t f)(X_\tau). \tag{8.33}$$

hold true on $\{\tau < +\infty\}$.

Proof This theorem follows directly from lemma 8.23 by an approximation argument and an application of the monotone class theorems, theorems 2.4 and 2.7. For convenience of the reader we carry out the approximation argument in detail. Suppose first that $B \subset \mathbb{R}^d$ is a closed, bounded set. We approximate 1_B by functions in $C_0(\mathbb{R}^d)$ as follows. Set

$$\varphi(u) = \begin{cases} 1, & \text{if } t < 0, \\ 1-u, & \text{if } 0 \le u \le 1, \\ 0, & \text{if } u > 1, \end{cases} \tag{8.34}$$

and for $n \in \mathbb{N}$,

$$1_B^n(x) = \varphi\big(nd(x, B)\big), \qquad x \in \mathbb{R}^d, \tag{8.35}$$

where $d(x, B)$ is the distance from x to B, i.e., $d(x, B) = \inf\{d(x, b), b \in B\}$. Then $1_B^n \in C_0(\mathbb{R}^d)$ for all $n \in \mathbb{N}$, and $1_B^n(x) \to 1_B(x)$ for every $x \in \mathbb{R}^d$ as $n \to \infty$. For every $n \in \mathbb{N}$ we get from lemma 8.23

$$E_x\big(1_B^n(X_{t+\tau}) \,\big|\, \mathcal{F}_\tau\big) = E_{X_\tau}\big(1_B^n(X_t)\big).$$

1 is an integrable majorant for $1_B^n(X_{t+\tau})$, and for $1_B^n(X_\tau)$. Therefore the usual dominated convergence theorem, and the dominated convergence theorem for conditional expectations, theorem 5.11.(b), allow to interchange above the limits $n \to \infty$ with the expectations, proving (8.32) for the case that B is closed and bounded. On the other hand, the closed, bounded subsets of $\mathbb{R}^d$ form an $\cap$–stable generator of $\mathcal{B}(\mathbb{R}^d)$, so that we can apply the monotone class theorem, theorem 2.4, to conclude that (8.32) holds for all $B \in \mathcal{B}(\mathbb{R}^d)$. Finally, from (8.32) we get equation (8.33) with the monotone class theorem for functions, theorem 2.7. $\square$

Finally, we consider the case where we have a family $\theta = (\theta_t, t \in \mathbb{R}_+)$ of shift operators for the universal Markov process X: $X_t \circ \theta_s = X_{s+t}$. An easy *exercise* shows that θ_τ defined by $(\theta_\tau)(\omega) = (\theta_{\tau(\omega)})(\omega)$, $\omega \in \{\tau < +\infty\}$, is a measurable mapping from $(\{\tau < +\infty\}, \mathcal{F}_\infty^X \cap \{\tau < +\infty\})$ into $(\Omega, \mathcal{F}_\infty^X)$. Moreover, on $\{\tau < +\infty\}$ we get $X_t \circ \theta_\tau = X_{t+\tau}$. Thus we can write the strong Markov property (8.32) as follows

$$E_x\big((1_B \circ X_t) \circ \theta_\tau \,\big|\, \mathcal{F}_\tau\big) = E_{X_\tau}\big(1_B \circ X_t\big), \qquad \text{on } \{\tau < +\infty\}.$$

Now we can use similar arguments as in the previous sections to obtain the following

8.25 Theorem *Suppose that $X = (X_t, t \in \mathbb{R}_+)$ is an $\mathbb{R}^d$–valued universal Markov process relative to a filtration $\mathcal{F} = (\mathcal{F}_t, t \in \mathbb{R}_+)$, with exclusively right continuous paths, and shift operators $\theta = (\theta_t, t \in \mathbb{R}_+)$. Assume furthermore that τ is an $\mathcal{F}$–stopping time, and that the semigroup $U = (U_t, t \in \mathbb{R}_+)$ generated by X is a Feller semigroup. Then for all $x \in \mathbb{R}^d$, $t \in \mathbb{R}_+$, $Y \in \mathcal{M}_b(\Omega, \mathcal{F}_\infty^X)$,*

$$E_x\big(Y \circ \theta_\tau \,\big|\, \mathcal{F}_\tau\big) = E_{X_\tau}(Y) \tag{8.36}$$

holds true on $\{\tau < +\infty\}$.

8.5 Reflection Principle for Brownian Motion

In this section we apply the strong Markov property to a standard one dimensional Brownian motion $B = (B_t, t \in \mathbb{R}_+)$ in order to derive the famous reflection principle.

In the sequel let $B = (B_t, t \in \mathbb{R}_+)$ be a one dimensional Brownian motion, given as a universal Markov process, relative to its natural filtration, on a family $\big(\Omega, \mathcal{A}, P = (P_x, x \in \mathbb{R}^d)\big)$ of probability spaces, and with exclusively continuous paths. For example, this could have been constructed on Wiener space as in section 4.3. The only ingredient which is missing there, is the measurability of the mapping $x \mapsto Q_x(A)$, for $A \in \mathcal{C}$ (notation as at the end of section 4.3). But this can proved with the same argument as before theorem 8.8.

We know from exercise 8.22 that the semigroup generated by B has the Feller property. Theorem 8.24 implies that B has the strong Markov property.

We pause here for the following

8.26 Remark Despite all our efforts, this is still not the optimal setting to discuss and apply the strong Markov property for a Brownian motion (or other processes as well). The point is that the natural filtration is to small to allow for a rich class of stopping times. One improves the situation a great deal by moving on to the *right continuous augmentation* of the natural filtration, which results in a very nice class of stopping times. This augmentation can be done by including certain null events into every element of the filtration, but a detailed treatment is beyond the scope of these lectures. The interested reader is referred to the standard literature, e.g., to [16, 28].

Let $x \in \mathbb{R}$, $y > x$. We shall use the strong Markov property of B to derive the law of τ_y, the first hitting time of y, that is the first time the Brownian motion hits the closed interval $[y, +\infty)$: $\tau_y = \inf\{t \geq 0,\ B_t \geq y\}$. We have already shown in example 6.15 that τ_y is indeed a stopping time for the natural filtration of B. The argument uses the *reflection principle*, and this had been already applied before the theory of the strong Markov had been developed. Heuristically, the argument goes as follows: We compute $P_x(\tau_y \leq t)$ for $t \geq 0$ by

$$\begin{aligned} P_x(\tau_y \leq t) &= P_x(\tau_y \leq t, B_t \geq y) + P_x(\tau_y \leq t, B_t < y) \\ &= P_x(B_t \geq y) + P_x(\tau_y \leq t, B_t < y), \end{aligned}$$

because the paths of B are continuous, and therefore the condition $\tau_y \leq t$ becomes trivial if $B_t \geq y$. By the symmetry of the law of the Brownian motion it seems heuristically "obvious", that

$$\begin{aligned} P_x(\tau_y \leq t, B_t < y) &= P_x(\tau_y \leq t, B_t > y) \\ &= P_x(B_t > y) \\ &= P_x(B_t \geq y), \end{aligned}$$

where we used $P_x(B_t = y) = 0$ in the last equality. Here we used the idea, that to every path which hits y at time τ_y and has moved into $[y, +\infty)$ at time t, uniquely corresponds a path, which has been reflected after time τ_y at the horizontal line at y. This path has "the same probability" as the original path, and at time t it is in $(-\infty, y]$, see figure 8.1.

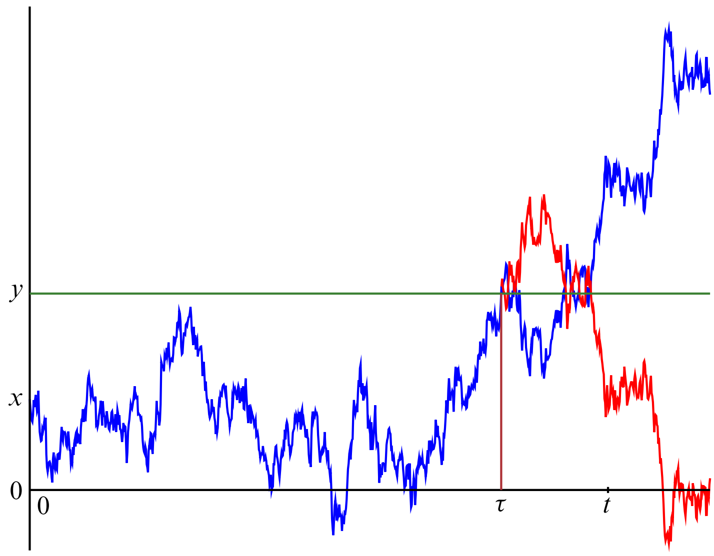

Figure 8.1: first hitting of level 0.35 at time $t = 0.693$

Therefore we get

$$P_x(\tau_y \le t) = 2P_x(B_t \ge y). \tag{8.37}$$

Thus we have derived the distribution function of τ_y, and we are practically done.

To prepare a rigorous proof of equation (8.37), we ask the reader to do the following

8.27 Exercise Let B be a Brownian motion as above, and let τ be a stopping time for $\mathcal{F}^{\boldsymbol{B}}$. Show that $(t,\omega) \mapsto t + \tau(\omega)$ is a measurable mapping from $(\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{A})$ into $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$. (*Hint:* One possibility is to show first that for $s \in \mathbb{R}_+$, $\{(t,\omega),\ t + \tau(\omega) < s\} = \cup_q\big([0, s-q) \times \tau^{-1}([0,q])\big)$, where the union is over all $q \in [0,s) \cap \mathbb{Q}$.) Then use lemma 6.8 to prove that $(t,\omega) \mapsto B_{t+\tau(\omega)}(\omega)$ is measurable from $(\mathbb{R}_+ \times \Omega, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{A})$ into $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

8.28 Lemma (Reflection Principle for Brownian Motion) *For a one dimensional Brownian motion B, $x \in \mathbb{R}$, $y > x$, $t \in \mathbb{R}_+$, equation* (8.37) *holds true.*

Proof $\mathcal{F}^{\boldsymbol{B}} = (\mathcal{F}^{\boldsymbol{B}}_t,\ t \in \mathbb{R}_+)$ denotes the natural filtration of B. For $\lambda > 0$, we compute the following Laplace transform

$$\begin{aligned}
&\int_0^\infty e^{-\lambda t}\, P_x(\tau_y \le t, B_t < y)\, dt \\
&\qquad = E_x\Big(\int_0^\infty e^{-\lambda t}\, 1_{(-\infty,y)}(B_t)\, 1_{[\tau_y,+\infty)}(t)\, dt\Big) \\
&\qquad = E_x\Big(\int_{\tau_y}^\infty e^{-\lambda t}\, 1_{(-\infty,y)}(B_t)\, dt\Big) \\
&\qquad = E_x\Big(e^{-\lambda\tau_y} \int_0^\infty e^{-\lambda t}\, 1_{(-\infty,y)}(B_{\tau_y+t})\, dt\Big) \\
&\qquad = E_x\Big(e^{-\lambda\tau_y} \int_0^\infty e^{-\lambda t}\, E_x\big(1_{(-\infty,y)}(B_{\tau_y+t}) \,|\, \mathcal{F}^{\boldsymbol{B}}_{\tau_y}\big)\, dt\Big)
\end{aligned}$$

where we used Fubini's theorem twice, once for conditional expectations, see theorem E.1, appendix E (for this we needed exercise 8.27). Moreover, we used the fact (see exercise 6.18) that τ_y is $\mathcal{F}^{\boldsymbol{B}}_{\tau_y}$–measurable. The strong Markov property (8.33) for B gives

$$E_x\big(1_{(-\infty,y)}(B_{\tau_y+t}) \,|\, \mathcal{F}^{\boldsymbol{B}}_{\tau_y}\big) = E_{B_{\tau_y}}\big(1_{(-\infty,y)}(B_t)\big) = E_y\big(1_{(-\infty,y)}(B_t)\big) = \frac{1}{2},$$

because by the continuity of the paths of B we get $B_{\tau_y} = y$. Thus we find

$$\begin{aligned}\int_0^\infty e^{-\lambda t}\, P_x(\tau_y \le t, B_t < y)\, dt \\ &= \frac{1}{2\lambda}\, E_x\big(e^{-\lambda \tau_y}\big) \\ &= \frac{1}{2}\, E_x\Big(\int_0^\infty e^{-\lambda t}\, 1_{[0,t]}(\tau_y)\, dt\Big) \\ &= \frac{1}{2}\int_0^\infty e^{-\lambda t}\, P_x(\tau_y \le t)\, dt,\end{aligned}$$

and we used Fubini's theorem again. Since the Laplace transform is injective, see theorem H.4 in appendix H, we get

$$P_x(\tau_y \le t, B_t < y) = \frac{1}{2}\, P_x(\tau_y \le t).$$

Hence,

$$\begin{aligned}P_x(\tau_y \le t) &= P_x(\tau_y \le t, B_t < y) + P_x(\tau_y \le t, B_t \ge y) \\ &= \frac{1}{2}\, P_x(\tau_y \le t) + P_x(B_t \ge y),\end{aligned}$$

and we proved equation (8.37). □

8.29 Corollary *Under P_x, $x \in \mathbb{R}$, τ_y, $y \ne x$, has the density*

$$\varphi_{\tau_y}(t) = \frac{|x-y|}{\sqrt{2\pi t^3}}\, e^{-(x-y)^2/2t}, \qquad t > 0. \tag{8.38}$$

Proof Without loss of generality we consider the case $x < y$. Then τ_y has under P_x the distribution function

$$2\int_y^\infty \frac{1}{\sqrt{2\pi t}}\, e^{-(x-z)^2/2t}\, dz = 2\int_{\frac{y-x}{\sqrt{t}}}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz.$$

Now differentiate with respect to t to obtain the claimed formula. □

To end this section, let us consider the *running maximum* $M = (M_t,\, t \in \mathbb{R}_+)$ of a Brownian motion B starting at $x = 0$:

$$M_t = \sup_{s \le t} B_s. \tag{8.39}$$

8.30 Corollary *The running maximum M_t of a standard Brownian motion at time $t > 0$ has the density*

$$\varphi_{M_t}(y) = 2\frac{1}{\sqrt{2\pi t}}\, e^{-y^2/2t}, \qquad y \ge 0. \tag{8.40}$$

Proof In an *exercise* the reader shows that

$$\{M_t \geq y\} = \{\tau_y \leq t\}.$$

Then lemma 8.28 gives

$$P_0(M_t \geq y) = 2 \int_y^\infty \frac{1}{\sqrt{2\pi t}} e^{-z^2/2t}\, dz.$$

That is, M_t has a distribution function given by

$$F_{M_t}(y) = 1 - 2 \int_y^\infty \frac{1}{\sqrt{2\pi t}} e^{-z^2/2t}\, dz,$$

and differentation with respect to y shows formula (8.40). $\square$

Chapter 9

Fine Structure of Brownian Motion

9.1 Gaussian Processes

In this section we give a characterization of Brownian motion as a Gaussian process, which will serve us well later. We begin with the

9.1 Definition

(a) $\mathcal{N}_1$ denotes the set of normal distributions $N(\mu, \sigma^2)$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with $\mu \in \mathbb{R}$, $\sigma^2 \geq 0$, where $N(\mu, 0)$ is the Dirac measure ε_μ.

(b) $\mathcal{N}_d$, $d \in \mathbb{N}$, denotes the set of probability measures ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ so that for each linear mapping $l : \mathbb{R}^d \to \mathbb{R}$, $l\nu = \nu \circ l^{-1} \in \mathcal{N}_1$. The elements of $\mathcal{N}_d$ are called *Gaussian measures on* $\mathbb{R}^d$.

(c) A random variable X with values in $\mathbb{R}^d$, $d \in \mathbb{N}$, is called a *Gaussian random variable*, if $P_X \in \mathcal{N}_d$. It is called *centered*, if its expectation value is the zero vector in $\mathbb{R}^d$.

(d) A real valued stochastic process $X = (X_t,\ t \in T)$ is called a *Gaussian process*, if all its finite dimensional distributions are given by Gaussian measures. It is called *centered*, if $E(X_t) = 0$ for every $t \in T$.

9.2 Exercise Show that the definition of $\mathcal{N}_1$ as in part (b) with $d = 1$ is consistent with the definition of $\mathcal{N}_1$ in part (a).

9.3 Exercise Let $X = (X_1, \ldots, X_d)$ be a Gaussian random variable with values in $\mathbb{R}^d$. Show that for every $j = 1, \ldots, d$, X_j is normally distributed (possibly degenerate, i.e., with variance zero), and therefore $X_j \in \mathcal{L}^p(\Omega, \mathcal{A}, P)$ for all $p \geq 1$. In particular, all moments of X_j exist.

Let us compute the characteristic function of $l\nu$, $\nu \in \mathcal{N}_d$. $l(y)$, $y \in \mathbb{R}^d$, where l is of the form $l(y) = \sum_{j=1}^d \lambda_j y_j \equiv (\lambda, y)$ for some $\lambda \in \mathbb{R}^d$. By definition of $\mathcal{N}_1$, there exist $\mu_\lambda \in \mathbb{R}$, $\sigma_\lambda^2 \geq 0$, so that $l\nu = N(\mu_\lambda, \sigma_\lambda^2)$:

$$\int_{\mathbb{R}} e^{itx}\, d(l\nu)(x) = e^{it\mu_\lambda - t^2\sigma_\lambda^2/2}, \qquad t \in \mathbb{R}. \tag{9.1}$$

On the other hand, the transformation theorem for Lebesgue integrals gives that the left hand side of equation (9.1) is equal to

$$\int_{\mathbb{R}^d} e^{it(\lambda,y)}\, d\nu(y). \tag{9.2}$$

Denote by $Y = (Y_1, \dots, Y_n)$ a Gaussian random variable with law ν. Then this implies

$$\mu_\lambda = E\Big(\sum_{j=1}^d \lambda_j Y_j\Big) = (\lambda, \mu),$$

where $\mu_j = E(Y_j)$, $j = 1, \dots, d$, and $\mu = (\mu_1, \dots, \mu_d)$. Also we get

$$\begin{aligned}
\sigma_\lambda^2 &= E\Big(\Big(\sum_{j=1}^d \lambda_j (Y_j - \mu_j)\Big)^2\Big) \\
&= \sum_{j,k=1}^d \lambda_j \lambda_k\, E\big((Y_j - \mu_j)(Y_k - \mu_k)\big) \\
&= (\lambda, C\lambda),
\end{aligned}$$

where C is the *covariance matrix* of ν, or Y respectively, with matrix elements

$$C_{jk} = E\big((Y_j - \mu_j)(Y_k - \mu_k)\big), \qquad i, j = 1, \dots, d. \tag{9.3}$$

Clearly, C is symmetric, and the above calculation shows that C is positive definite. Thus choosing above $t = 1$ we get

$$\int_{\mathbb{R}^d} e^{i(\lambda,y)}\, d\nu(y) = e^{i(\lambda,\mu) - 1/2\,(\lambda, C\lambda)}, \qquad \lambda \in \mathbb{R}^d. \tag{9.4}$$

In analogy with the one dimensional case, we call the left hand side of equation (9.4) the *characteristic function of* ν, and have proved the first half of the following

9.4 Theorem *For every Gaussian measure $\nu \in \mathcal{N}_d$, $d \in \mathbb{N}$, there exist a vector $\mu \in \mathbb{R}^d$, and a symmetric, positive definite $d \times d$–matrix C, so that its characteristic function is given by formula* (9.4)*. Conversely, given any $\mu \in \mathbb{R}^d$, and any symmetric, positive definite $d \times d$–matrix C, then the right hand side of equation* (9.4) *defines the characteristic function of a unique Gaussian measure in $\mathcal{N}_d$.*

Proof C is symmetric and positive definite. Therefore there exists an orthogonal $d \times d$–matrix O so that $D = O^t C O$ is a diagonal matrix

$$D = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_d^2 \end{pmatrix},$$

and $\sigma_j^2 \geq 0$ for all $j = 1, \dots, d$. Now let $m = O^t \mu$, and set $\nu_j = N(m_j, \sigma_j^2)$, $j = 1, \dots, d$, and consider the product measure

$$\tilde{\nu} = \nu_1 \otimes \nu_2 \otimes \cdots \otimes \nu_d, \tag{9.5}$$

on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then

$$\int_{\mathbb{R}^d} e^{i(\lambda, y)}\, d\tilde{\nu}(y) = e^{i(\lambda, m) - 1/2\,(\lambda, D\lambda)}, \qquad \lambda \in \mathbb{R}^d.$$

Define $\nu = O\tilde{\nu}$, then we get for any $f \in \mathcal{M}_b(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$

$$\int_{\mathbb{R}^d} f(y)\, d\nu(y) = \int_{\mathbb{R}^d} f\big(Oy\big)\, d\tilde{\nu}(y).$$

In particular, for $f(y) = \exp(i(\lambda, y))$, $\lambda \in \mathbb{R}^d$, we find

$$\begin{aligned} \int_{\mathbb{R}^d} e^{i(\lambda, y)}\, d\nu(y) &= \int_{\mathbb{R}^d} e^{i(\lambda, Oy)}\, d\tilde{\nu}(y) \\ &= \int_{\mathbb{R}^d} e^{i(O^t\lambda, y)}\, d\tilde{\nu}(y) \\ &= e^{i(\lambda, Om) - (\lambda, ODO^t\lambda)} \\ &= e^{i(\lambda, \mu) - (\lambda, C\lambda)}. \end{aligned}$$

The fact that ν is uniquely determined by its characteristic function (e.g., [25] — the proof there for $d = 1$ applies in the same way for general d), ends the proof. □

In the case that C is strictly positive definite, we get that $\sigma_j > 0$ for all $j = 1, \dots, d$, and hence $\tilde{\nu}$ in (9.5) has the density

$$\begin{aligned} &\prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\Big(-\frac{(y_j - m_j)^2}{2\sigma_j^2}\Big) \\ &\qquad = \frac{1}{\sqrt{(2\pi)^d \det(D)}} \exp\Big(-\frac{1}{2}\big(y - m, D^{-1}(y - m)\big)\Big), \qquad y \in \mathbb{R}^d. \end{aligned}$$

Therefore $\nu = O\tilde{\nu}$ has the density (note that $\det(D) = \det(C)$)

$$\frac{1}{\sqrt{(2\pi)^d \det(D)}} \exp\Big(-\frac{1}{2}\big((O^{-1}y - m), D^{-1}(O^{-1}y - m)\big)\Big)$$
$$= \frac{1}{\sqrt{(2\pi)^d \det(D)}} \exp\Big(-\frac{1}{2}\big(O^{-1}(y-\mu), D^{-1}O^{-1}(y-\mu)\big)\Big)$$
$$= \frac{1}{\sqrt{(2\pi)^d \det(C)}} \exp\Big(-\frac{1}{2}\big(y-\mu, C^{-1}(y-\mu)\big)\Big), \tag{9.6}$$

for $y \in \mathbb{R}^d$.

9.5 Corollary *If $\nu \in \mathcal{N}_d$, $d \in \mathbb{N}$, has mean $\mu \in \mathbb{R}^d$, and a strictly positive definite covariance matrix C, then it has the density given by formula* (9.6).

Suppose that X is an $\mathbb{R}^d$–valued Gaussian random variable with law $\nu \in \mathcal{N}_d$, mean $\mu \in \mathbb{R}^d$, and covariance matrix C. Suppose furthermore, that X_i, X_j, i, $j = 1, \ldots, d$, $i \neq j$, are uncorrelated. Then all off-diagonal elements of C are zero. Therefore the above calculation shows that the characteristic function of ν is the product of the characteristic functions of $X_1, \ldots, X_d$. It follows that the random variables $X_1, \ldots, X_d$ are independent:

9.6 Corollary *Suppose that $X_1, \ldots, X_d$, $d \in \mathbb{N}$, are real-valued random variables whose joint distribution is Gaussian. $X_1, \ldots, X_d$ are independent, if and only if they are pairwise uncorrelated.*

9.7 Exercise Let $X = (X_1, X_2)$ be an $\mathbb{R}^2$–valued Gaussian random variable with mean $\mu = (\mu_1, \mu_2)$ and covariance matrix $C = (C_{ij}, i, j = 1, 2)$. Show that $X_1 + X_2$ is normally distributed with mean $\mu_1 + \mu_2$ and variance $V(X_1) + 2\operatorname{Cov}(X_1, X_2) + V(X_2)$.

9.2 Brownian Motion as a Gaussian Process

9.8 Theorem *Suppose that $X = (X_t, t \in \mathbb{R}_+)$ is a real valued stochastic process which a.s. has continuous paths starting at the origin. X is a standard Brownian motion, if and only if X is a centered Gaussian process with covariance $Cov(X_s, X_t) = s \wedge t$.*

Proof Suppose that X is a standard Brownian motion. Then X has independent, stationary increments, and $X_t - X_s$ has the law $N(0, t-s)$. Let $t_0 = 0 < t_1 < t_2 < \cdots < t_n$, $n \in \mathbb{N}$, and set $\Delta t_j = (t_j - t_{j-1})$, $j = 1, \ldots, n$. Furthermore, let f be any bounded, measurable function on $\mathbb{R}^n$. Then

$$E\big(f(X_{t_1}, \ldots, X_{t_n})\big) = E\big(f \circ \varphi(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})\big)$$
$$= \int_{\mathbb{R}^n} (f \circ \varphi)(y) \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi \Delta t_j}} e^{-y_j^2/(2\Delta t_j)}\, dy_j,$$

where φ is the linear mapping from $\mathbb{R}^n$ onto itself defined similarly as in equation (3.41b), with inverse φ^{-1} in (3.45). Moerover, we used that the increments of X are independent, and that $X_t - X_s$, $s < t$, has the law $N(0, t-s)$. Now choose $f(y) = \exp(i(\lambda, y))$, $y \in \mathbb{R}^n$, for $\lambda \in \mathbb{R}^n$. With $X_J = (X_{t_1}, \ldots, X_{t_n})$ we then obtain

$$\begin{aligned} E\big(e^{i(\lambda, X_J)}\big) &= \int_{\mathbb{R}^n} e^{i(\lambda, \varphi(y))} \prod_{j=1}^n \frac{1}{\sqrt{2\pi\Delta t_j}} e^{-y_j^2/(2\Delta t_j)}\, dy_j \\ &= \int_{\mathbb{R}^n} e^{i(\phi^t\lambda, y)} \prod_{j=1}^n \frac{1}{\sqrt{2\pi\Delta t_j}} e^{-y_j^2/(2\Delta t_j)}\, dy_j, \end{aligned}$$

where ϕ is the matrix representation (3.43) of φ. Now the integral is readily computed, and we find

$$E\big(e^{i(\lambda, X_J)}\big) = \exp\Big(-\frac{1}{2}\sum_{j=1}^n \hat{\lambda}_j^2 \Delta t_j\Big),$$

with $\hat{\lambda} = \phi^t\lambda$. Note that $\phi^t_{jk} = 1$ if $j \le k$, and otherwise it is zero. Then

$$\begin{aligned} \sum_{j=1}^n \hat{\lambda}_j^2 \Delta t_j &= \sum_{j=1}^n \Delta t_j \Big(\sum_{k=1}^n \phi^t_{jk}\lambda_k\Big)^2 \\ &= \sum_{j=1}^n \Delta t_j \sum_{k,l=1}^n \phi^t_{jk}\phi^t_{jl}\lambda_k\lambda_l \\ &= \sum_{k,l=1}^n \lambda_k\lambda_l \sum_{j=1}^n \phi^t_{jk}\phi^t_{jl}\Delta t_j \\ &= \sum_{k,l=1}^n \lambda_k\lambda_l \sum_{j=1}^{k\wedge l}(t_j - t_{j-1}) \\ &= \sum_{k,l=1}^n \lambda_k\lambda_l\, t_{k\wedge l} \\ &= \sum_{k,l=1}^n \lambda_k\lambda_l\, t_k \wedge t_l \\ &= (\lambda, C\lambda), \end{aligned}$$

where the matrix C has elements $C_{kl} = t_k \wedge t_l$. Therefore, X is a centered Gaussian process with covariance given by $\mathrm{Cov}(X_s, X_t) = s \wedge t$.

Conversely, suppose that X is a centered Gaussian process with covariance $\mathrm{Cov}(X_s, X_t) = s \wedge t$. Then we obtain for $0 \le v \le u \le s \le t$

$$E\big((X_u - X_v)(X_t - X_s)\big) = 0.$$

Hence the increments of X are pairwise uncorrelated, and by corollary 9.6 X has independent increments. For the variance of $X_t - X_s$, $s < t$, we get $t - s$, so that $X_s - X_t$ has the law $N(0, t - s)$. Thus X is a standard one dimensional Brownian motion. □

With theorem 9.8 we can easily construct new Brownian motions from a given one:

9.9 Theorem *Assume that $B = (B_t, t \in \mathbb{R}_+)$ is a standard one dimensional Brownian motion. Then also the following stochastic processes are standard one dimensional Brownian motions ($t \in \mathbb{R}_+$):*

(i) $B^1_t = B_{t+t_0} - B_{t_0}$, *any fixed* $t_0 \geq 0$;

(ii) $B^2_t = -B_t$;

(iii) $B^3_t = \lambda^{-1/2} B_{\lambda t}$, *any* $\lambda > 0$;

(iv) B^4_t *with* $B^4_0 = 0$, $B^4_t = t B_{1/t}$, $t > 0$.

Proof Clearly, all four processes B^i, $i = 1, \ldots, 4$, are centered Gaussian processes, and easy computations show that all have a covariance $\mathrm{Cov}(B^i_s, B^i_t) = s \wedge t$, $i = 1, \ldots, 4$. Moreover the a.s. continuity of the paths of B obviously extends to B^i, $i = 1, \ldots, 3$. So it remains to show that B^4 has a.s. continuous paths. This is clear for $t > 0$. Since B^4 has the same finite dimensional distributions as a standard one dimensional Brownian motion, by the Kolmogorov–Chentsov–theorem 4.10 B^4 has a modification $B' = (B'_t, t \in \mathbb{R}_+)$ which is a.s. continuous on $\mathbb{R}_+$ (see section 4.3). Moreover, a glance at the proof of the Kolmogorov–Chentsov–theorem tells that this modification can be chosen in such a way that all paths of B^4 and of B' coincide on a dense subset of $\mathbb{R}_+$ (e.g., the dyadic rational numbers). Both processes are a.s. continuous on $(0, +\infty)$, and therefore a.s. their paths are the same on $(0, +\infty)$. But since the paths of B' are a.s. continuous from the right at 0, so must be the paths of B^4. □

9.3 Law of the Iterated Logarithm

In this section we provide a rather detailed analysis of the paths of a standard one dimensional Brownian motion. Basically all results have been found by P. Lévy.

We begin with a simple estimate for the Gaussian error function:

9.10 Lemma *For all $a > 0$*

$$\frac{a}{1 + a^2}\, e^{-a^2/2} \leq \int_a^\infty e^{-x^2/2}\, dx \leq \frac{1}{a^2}\, e^{-a^2/2} \tag{9.7}$$

holds true.

Proof For the upper bound we estimate as follows

$$\begin{aligned}\int_a^\infty e^{-x^2/2}\,dx &\le \frac{1}{a}\int_a^\infty xe^{-x^2/2}\,dx\\ &= \frac{1}{a}\,e^{-a^2/2}.\end{aligned}$$

For the lower bound we make the following computation

$$\begin{aligned}\int_a^\infty \frac{1}{x^2}\,e^{-x^2/2}\,dx &= -\int_a^\infty \Big(\frac{d}{dx}\frac{1}{x}\Big)\,e^{-x^2/2}\,dx\\ &= \frac{1}{a}\,e^{-a^2/2} - \int_a^\infty e^{-x^2/2}\,dx.\end{aligned}$$

Therefore we obtain the formula

$$\int_a^\infty \Big(1+\frac{1}{x^2}\Big)\,e^{-x^2/2}\,dx = \frac{1}{a}\,e^{-a^2/2}.$$

The left hand side of the last equation can be bounded from above by

$$\Big(1+\frac{1}{a^2}\Big)\int_a^\infty e^{-x^2/2}\,dx,$$

so that we get

$$\int_a^\infty e^{-x^2/2}\,dx \ge \Big(1+\frac{1}{a^2}\Big)^{-1}\frac{1}{a}\,e^{-a^2/2} = \frac{a}{1+a^2}\,e^{-a^2/2},$$

which finishes the proof. □

9.11 Remark Somewhat sharper estimates for the Gaussian error function can be found in the papers [17, 23].

From now on we consider a standard one dimensional Brownian motion $B = (B_t,\ t\in\mathbb{R}_+)$, and always take the natural filtration of B.

9.12 Lemma *For all α, $\beta > 0$, $t > 0$ the following estimate holds:*

$$P\Big(\sup_{s\le t}\big(B_s - \frac{\alpha}{2}\,s\big) \ge \beta\Big) \le e^{-\alpha\beta}. \tag{9.8}$$

Proof Since the exponential function is strictly monotone increasing, we get

$$\begin{aligned}P\Big(\sup_{s\le t}\big(B_s - \frac{\alpha}{2}\,s\big)\ge\beta\Big) &= P\Big(e^{\sup_{s\le t}(\alpha B_s - \alpha^2 s/2)} \ge e^{\alpha\beta}\Big)\\ &= P\Big(\sup_{s\le t} e^{\alpha B_s - \alpha^2 s/2} \ge e^{\alpha\beta}\Big)\end{aligned}$$

From example 7.6.(d) we know that $s \mapsto \exp(\alpha B_s - \alpha^2 s/2)$ is a continuous martingale, and so Doob's inequality (7.9), theorem 7.20, implies

$$\begin{aligned} P\Big(\sup_{s\le t}\big(B_s - \frac{\alpha}{2}\,s\big) \ge \beta\Big) &\le e^{-\alpha\beta}\,\sup_{s\le t} E\big(e^{\alpha B_s - \alpha^2 s/2}\big) \\ &= e^{-\alpha\beta}, \end{aligned}$$

because

$$E\big(e^{\alpha B_s - \alpha^2 s/2}\big) = E\big(e^{\alpha B_s - \alpha^2 s/2}\big)\big|_{s=0} = 1,$$

where we used again that $s \mapsto \exp(\alpha B_s - \alpha^2 s/2)$ is a martingale. $\square$

9.13 Theorem (Law of the Iterated Logarithm) *Let $B = (B_t,\, t \in \mathbb{R}_+)$ be a standard one dimensional Brownian motion. Then*

$$P\Big(\limsup_{t\downarrow 0,\, t\in(0,1)} \frac{B_t}{\sqrt{2t\ln\ln 1/t}} = 1\Big) = 1 \tag{9.9}$$

holds true.

Before we begin the proof, let us spell out in detail the statement of theorem 9.13: The set of $\omega \in \Omega$ so that for every sequence $(t_n,\, n \in \mathbb{N})$ in $(0,1)$, which decreases to zero, the largest accumulation point of the sequence

$$\Big(\frac{B_{t_n}(\omega)}{\sqrt{2t_n \ln\ln 1/t_n}},\; n \in \mathbb{N}\Big)$$

is equal to 1, has probability one.

Proof (of theorem 9.13) Set

$$h(t) = \sqrt{2t\ln\ln 1/t}, \qquad t \in (0,1).$$

Choose $\theta, \delta \in (0,1)$, and define

$$\alpha_n = (1+\delta)\theta^{-n} h(\theta^n), \quad \beta_n = \frac{1}{2}\,h(\theta^n), \qquad n \in \mathbb{N}.$$

Observe that

$$\begin{aligned} \alpha_n\beta_n &= (1+\delta)\theta^{-n}\theta^n \ln\ln\theta^{-n} \\ &= (1+\delta)\ln\big(n\ln 1/\theta\big) \\ &= \ln\big(n^{1+\delta}\,(\ln 1/\theta)^{1+\delta}\big), \end{aligned}$$

so that

$$e^{-\alpha_n\beta_n} = n^{-(1+\delta)}\,(\ln 1/\theta)^{-(1+\delta)}.$$

With $t = 1$, lemma 9.12 implies

$$P\Big(\sup_{s\le 1}\big(B_s - \frac{\alpha_n}{2}\, s\big) \ge \beta_n\Big) \le e^{-\alpha_n\beta_n} \le n^{-(1+\delta)}\,(\ln 1/\theta)^{-(1+\delta)}.$$

Since

$$\sum_{n=1}^{\infty} n^{-(1+\delta)} < +\infty,$$

the Borel–Cantelli–Lemma, lemma C.1, appendix C, implies that

$$P\Big(\liminf_n \Big\{\sup_{s\le 1}\big(B_s - \frac{\alpha_n}{2}\, s\big) < \beta_n\Big\}\Big) = 1.$$

This means: For all $\omega \in \Omega$ in an event of probability 1, there exists $n_0(\omega) \in \mathbb{N}$ so that for all $n \ge n_0(\omega)$,

$$\sup_{s\le 1}\big(B_s(\omega) - \frac{\alpha_n}{2}\, s\big) < \beta_n$$

holds true. For such ω, n, and $s \in [0, \theta^{n-1}]$ we then find

$$\begin{aligned} B_s(\omega) &\le \frac{\alpha_n}{2}\, s + \beta_n \\ &\le \frac{\alpha_n}{2}\, \theta^{n-1} + \beta_n \\ &= \frac{1}{2}\,(1+\delta)\theta^{-n}\theta^{n-1} h(\theta^n) + \frac{1}{2}\, h(\theta^n) \\ &= \Big(\frac{1+\delta}{2\theta} + \frac{1}{2}\Big)\, h(\theta^n). \end{aligned}$$

There exists an interval $(0, s_0]$ so that h is monotone increasing on this interval. Indeed, the derivative of $t \mapsto t \ln\ln 1/t$ is equal to $\ln\ln 1/t + (\ln(t))^{-1}$, and for all small enough t this is strictly positive. Thus for all $n \ge n_0(\omega)$ large enough, we get for all $s \in (\theta^n, \theta^{n-1}]$ the estimate

$$B_s(\omega) \le \Big(\frac{1+\delta}{2\theta} + \frac{1}{2}\Big)\, h(s),$$

or

$$\frac{B_s(\omega)}{h(s)} \le \frac{1+\delta}{2\theta} + \frac{1}{2}.$$

Now we let n tend to ∞, so that we obtain for all $\delta, \theta \in (0, 1)$, and all ω in an event of probability 1,

$$\limsup_{s\downarrow 0,\, s\in(0,1)} \frac{B_s(\omega)}{h(s)} \le \frac{1+\delta}{2\theta} + \frac{1}{2}.$$

Finally we let $\delta \to 0$ and $\theta \to 1$, so that we obtain

$$\limsup_{s\downarrow 0,\, s\in(0,1)} \frac{B_s}{h(s)} \le 1, \qquad P\text{–a.s.}$$

Next we prove the corresponding lower bound. To this end, choose $\theta \in (0, 1)$, and consider the events

$$A_n = \big\{B_{\theta^n} - B_{\theta^{n+1}} \geq (1 - \sqrt{\theta})\, h(\theta^n)\big\}, \qquad n \in \mathbb{N}.$$

The independence of the increments of B entail that the sequence $(A_n,\, n \in \mathbb{N})$ is independent. Now

$$\begin{aligned} P(A_n) &= \frac{1}{\sqrt{2\pi(\theta^n - \theta^{n+1})}} \int_{(1-\sqrt{\theta})\, h(\theta^n)}^{\infty} e^{-x^2/(2(\theta^n - \theta^{n+1}))}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{a_n} e^{-x^2/2}\, dx, \end{aligned}$$

with

$$\begin{aligned} a_n &= (1 - \sqrt{\theta})\, h(\theta^n)\, (\theta^n - \theta^{n+1})^{-1/2} \\ &= (1 - \sqrt{\theta})\, \big(2\theta^n \ln\ln \theta^{-n}\big)^{1/2}\, \big(\theta^n(1-\theta)\big)^{-1/2} \\ &= (1 - \sqrt{\theta})\, \sqrt{\frac{2 \ln\ln \theta^{-n}}{1-\theta}}. \end{aligned}$$

The lower bound of the error integral in lemma 9.10 implies

$$\begin{aligned} P(A_n) &\geq \frac{1}{\sqrt{2\pi}}\, \frac{a_n}{1 + a_n^2}\, e^{-a_n^2/2} \\ &= \frac{1}{\sqrt{2\pi}}\, \frac{a_n}{1 + a_n^2} \exp\Big(-\frac{1 - 2\sqrt{\theta} + \theta}{1-\theta}\, \ln\big(n \ln 1/\theta\big)\Big) \\ &= K_\theta\, \frac{a_n}{1 + a_n^2}\, n^{-(1 - 2\sqrt{\theta} + \theta)/(1-\theta)}, \end{aligned}$$

where $K_\theta > 0$ is some numerical constant which only depends on θ. Below we allow for the possibility that K_θ varies from step to step. Note that a_n increases to $+\infty$ with $n \to +\infty$, so that for all n large enough we get $1 + a_n^2 \leq 2a_n^2$. Therefore for all n large enough,

$$\begin{aligned} \frac{a_n}{1 + a_n^2} &\geq \frac{1}{2a_n} \\ &= K_\theta\, \big(\ln(n \ln 1/\theta)\big)^{-1/2} \\ &\geq K_\theta\, \big(\ln n + \ln\ln 1/\theta\big)^{-1} \\ &\geq K_\theta\, \frac{1}{\ln n}, \end{aligned}$$

The estimate of the last step is trivial for $\ln(\ln(\theta^{-1})) \leq 0$. If $\ln(\ln(\theta^{-1})) \geq 0$, we have for all large enough n that $\ln(\ln(\theta^{-1})) \leq \ln(n) \ln(\ln(\theta^{-1}))$, so that we can get the claimed estimate by multiplying K_θ with

$$\frac{1}{1 + \ln(\ln(\theta^{-1}))},$$

which defines the new K_θ. Altogether we have derived for all large enough n the estimate

$$P(A_n) \geq K_\theta \frac{1}{\ln n} n^{-(1-2\sqrt{\theta}+\theta)/(1-\theta)},$$

and we are interested in the convergence properties of a series, of which the right hand side is a generic term. Consider first the term which is a power of n. Since $\theta \in (0,1)$, $\sqrt{\theta} > \theta$, and consequently

$$1 - 2\sqrt{\theta} + \theta < 1 - \theta$$

so that

$$\frac{1 - 2\sqrt{\theta} + \theta}{1 - \theta} < 1.$$

Hence the series

$$\sum_{n=1}^{\infty} n^{-(1-2\sqrt{\theta}+\theta)/(1-\theta)}$$

diverges, and this behaviour is not changed by multiplication of the summands by $(\ln n)^{-1}$. (For example, for any $\varepsilon > 0$, we can estimate $\ln n$ from above by $C_\varepsilon n^\varepsilon$ for some suitable constant C_ε.) Thus we have an independent sequence $(A_n,\ n \in \mathbb{N})$ of events with $\sum_n P(A_n) = +\infty$. The Borel–Cantelli–Lemma implies that

$$P(\limsup_n A_n) = 1.$$

Thus with probability 1 infinitely many events A_n happen, that is, for every ω in an event of probability 1, there are infinitely many $n \in \mathbb{N}$ so that

$$B_{\theta^n}(\omega) \geq B_{\theta^{n+1}} + (1 - \sqrt{\theta})\, h(\theta^n).$$

To get a lower bound on $B_{\theta^{n+1}}$, we use the fact that theorem 9.9 states that also $t \mapsto -B_t$ is a standard Brownian motion, for which we can use the upper bound from the first step of the proof: For all ω in an event of probability 1 there exists an $n_0(\omega)$ so that for all $n \geq n_0(\omega)$,

$$-B_{\theta^{n+1}} \leq 2h(\theta^{n+1})$$

holds. Thus for ω in the intersection of these events of probability 1, and infinitely many $n \in \mathbb{N}$ we have

$$B_{\theta^n}(\omega) \geq (1 - \sqrt{\theta})\, h(\theta^n) - 2h(\theta^{n+1}).$$

Next we claim that for all n large enough

$$h(\theta^{n+1}) \leq 2\sqrt{\theta}\, h(\theta^n).$$

Indeed,

$$\frac{h(\theta^{n+1})}{h(\theta^n)} = \sqrt{\theta} \Big(\frac{\ln n + \ln\ln 1/\theta + \ln(1 + 1/n)}{\ln n + \ln\ln 1/\theta}\Big)^{1/2}$$

and the term in the last bracket decreases to 1 as $n \to \infty$. Thus we obtain that for infinitely many $n \in \mathbb{N}$,

$$B_{\theta^n} \geq (1-\sqrt{\theta})\, h(\theta^n) - 4\sqrt{\theta}\, h(\theta^n) = (1-5\sqrt{\theta})\, h(\theta^n), \qquad P\text{–a.s.},$$

or

$$\frac{B_{\theta^n}}{h(\theta^n)} \geq 1 - 5\sqrt{\theta}, \qquad P\text{–a.s.}$$

Therefore

$$\limsup_{t\downarrow 0,\, t\in(0,1)} \frac{B_t}{h(t)} \geq \limsup_n \frac{B_{\theta^n}}{h(\theta^n)} \geq 1 - 5\sqrt{\theta}, \qquad P\text{–a.s.}$$

Finally, we let θ tend to zero to obtain

$$\limsup_{t\downarrow 0,\, t\in(0,1)} \frac{B_t}{h(t)} \geq 1, \qquad P\text{–a.s.},$$

and the proof is concluded. □

9.14 Corollary *For every standard one dimensional Brownian motion B the following holds true:*

$$P\Big(\liminf_{t\downarrow 0,\, t\in(0,1)} \frac{B_t}{\sqrt{2t \ln\ln 1/t}} = -1\Big) = 1. \tag{9.10}$$

Proof Since with B also $-B$ is a standard one dimensional Brownian motion (cf. thm 9.9), we get from theorem 9.13

$$\begin{aligned} 1 &= P\Big(\limsup_{t\downarrow 0,\, t\in(0,1)} \frac{-B_t}{\sqrt{2t \ln\ln 1/t}} = 1\Big) \\ &= P\Big(\liminf_{t\downarrow 0,\, t\in(0,1)} \frac{B_t}{\sqrt{2t \ln\ln 1/t}} = -1\Big). \end{aligned}$$

□

Since the union of two null sets is a null set, we also have the

9.15 Corollary *For every standard one dimensional Brownian motion B the following holds true:*

$$P\Big(\limsup_{t\downarrow 0,\, t\in(0,1)} \frac{B_t}{\sqrt{2t \ln\ln 1/t}} = 1 \text{ and } \liminf_{t\downarrow 0,\, t\in(0,1)} \frac{B_t}{\sqrt{2t \ln\ln 1/t}} = -1\Big) = 1. \tag{9.11}$$

9.16 Corollary *$t = 0$ is P–a.s. an accumulation point of zeros of the paths of a standard one dimensional Brownian motion.*

Proof Corollary 9.15 states that as t decreases to 0, P–a.s. every path of B becomes infinitely often strictly positive and strictly negative. Since the paths are a.s. continuous, the statement follows from the intermediate value theorem for continuous functions. □

Corollary 9.15 together with theorem 9.9.(i) give

9.17 Corollary *For every standard one dimensional Brownian motion B, and all $s \geq 0$ the following holds true:*

$$P\Big(\limsup_{t\downarrow 0,\, t\in(0,1)} \frac{B_{s+t}-B_s}{\sqrt{2t\ln\ln 1/t}} = 1 \text{ and } \liminf_{t\downarrow 0,\, t\in(0,1)} \frac{B_{s+t}-B_s}{\sqrt{2t\ln\ln 1/t}} = -1\Big) = 1. \quad (9.12)$$

From corollary 9.15 and theorem 9.9.(iv) we obtain

9.18 Corollary *For every standard one dimensional Brownian motion B the following holds true:*

$$P\Big(\limsup_{t\uparrow\infty} \frac{B_t}{\sqrt{2t\ln\ln t}} = 1 \text{ and } \liminf_{t\uparrow\infty} \frac{B_t}{\sqrt{2t\ln\ln t}} = -1\Big) = 1. \quad (9.13)$$

Thus the typical path of a Brownian motion escapes approximately like $\pm\sqrt{t}$ simultaneously to $\pm\infty$ as $t \to +\infty$.

9.19 Corollary (Recurrence) *For a standard one dimensional Brownian motion B and for every $x \in \mathbb{R}$, the set $\{t \geq 0,\ B_t = x\}$ is P–a.s. unbounded.*

Proof Corollary 9.18 states that as $t \to +\infty$, P–a.s. every Brownian path infinitely often passes beyond every negative and every positive number in $\mathbb{R}$. Thus the statement of the corollary follows again from the intermediate value theorem. □

There are many other, curious and interesting path properties of standard one dimensional Brownian motions known, such as (all P–a.s.)

- the paths are nowhere differentiable;
- at no point $t \in \mathbb{R}_+$ a path has a local maximum or a local minimum;
- the level sets $\{t \in \mathbb{R}_+,\ B_t = x\}$, $x \in \mathbb{R}$, are random Cantor sets: they consist solely of accumulation points, are closed, uncountable, have Lebesgue measure zero, and Hausdorff–dimension $1/2$.

The interested reader is referred to the pertinent literature, e.g., [13, 20, 28].

Chapter 10

Introduction to Stochastic Integration

This chapter is intended to give a short introduction to the basic ideas of stochastic integration. As such it has to be quite sketchy, and here and there it will rely on facts which are shown elsewhere and on several "handwaving arguments". For a more thorough treatment the reader is referred to the pertinent literature, such as [7, 12, 16, 20–22, 27, 28, 30].

The importance of the stochastic integral stems from the fact, that it allows to construct (pathwise) many other interesting stochastic processes on the basis of a given one — the latter being very often a Brownian motion, or a stochastic process derived from it via stochastic integration. Here is a simple example: Let $X = (X_t,\, t \in \mathbb{R}_+)$ be the geometric Brownian motion (cf. example 7.6.(d))

$$X_t = \exp\bigl(\lambda B_t - \lambda^2 t/2\bigr), \qquad t \in \mathbb{R}_+, \tag{10.1}$$

where $\lambda \in \mathbb{R}$ is some parameter, and $B = (B_t,\, t \in \mathbb{R}_+)$ is a standard one dimensional Brownian motion. One can show that this process is the unique solution of the following *Itô stochastic integral equation*:

$$X_t = 1 + \lambda \int_0^t X_s\, dB_s. \tag{10.2}$$

As a short hand for (10.2) one often writes

$$dX_t = \lambda X_t\, dB_t, \tag{10.3}$$

the initial condition $X_0 = 1$ being understood. Equations like (10.3) are called *(Itô) stochastic differential equations*, and such equations have to be interpreted in their integrated form, i.e., like in (10.2).

Looking at our example above, the question arises what has been gained? Once the integral in (10.2) has been understood, we may consider more general equations of this type such as

$$X_t = X_0 + \int_0^t b(X_s)\, ds + \int_0^t \sigma(X_s)\, dB_s, \tag{10.4}$$

or in differential form

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dB_t. \tag{10.5}$$

One can show that for appropriate choices of the initial condition X_0, and the coefficient functions b and σ, equation (10.4) has a unique solution, defining a new stochastic process $X = (X_t,\, t \in \mathbb{R}_+)$. Indeed, many properties of X can then be derived from the equation (10.4).

The treatment of stochastic integral or differential equations is far outside the scope of these lectures, and they have been mentioned above only as a motivation for the following sections. There we shall sketch the construction and some of the properties of the Itô integral with respect to Brownian motion B figuring in (10.4).

10.1 Non-Rectifiability of Brownian Paths

The reader has learnt in her analysis course about the integral of a function along a curve, so that she will ask herself why we cannot use that concept for the integation in (10.2). The reason is that the classical method for the construction of the integral

$$\int_0^1 f\big(\gamma(t)\big)\,d\gamma(t) \tag{10.6}$$

of a (smooth, say) function f along a curve γ (in $\mathbb{R}^d$, say) needs in an essential way that γ is *rectifiable*. That is, the total variation of the curve γ,

$$\sup_{\mathcal{T}} \sum_{k=1}^{n} |\gamma(t_k) - \gamma(t_{k-1})|, \tag{10.7}$$

has to be *finite*, where the supremum is over all partitions $\mathcal{T} = \{t_0, t_1, \ldots, t_n\}$ (with $t_0 = 0$, $t_n = 1$) of the interval $[0, 1]$. However, with probability 1 the paths of a Brownian motion are on no subinterval of $\mathbb{R}_+$, however small, rectifiable, as we are going to show now. This means that between any two points of times s, $t \in \mathbb{R}_+$, $s < t$, the piece of a typical path of B between s and t is infinitely long. (It is better to say, that there is no notion of length for Brownian paths.)

Consider the time interval $[0, 1]$, and the sequence of binary partitions

$$\mathcal{T}_n = \left\{\frac{k}{2^n},\ k = 0, 1, \ldots, 2^n\right\}$$

of $[0, 1]$. Let $B = (B_t,\, t \in \mathbb{R}_+)$ be a standard one dimensional Brownian motion, and define

$$L_n^B = \sum_{k=0}^{2^n-1} \big|B_{(k+1)/2^n} - B_{k/2^n}\big|.$$

Observe that L_n^B is pathwise monotone increasing in n, because for any $t_1 < s < t_2$,

$$|B_{t_2} - B_{t_1}| \le |B_s - B_{t_1}| + |B_{t_2} - B_s|.$$

Therefore, for every path of B, $(L_n^B,\ n \in \mathbb{N})$ converges, possibly to $+\infty$. The latter possibility has probability one:

10.1 Lemma $(L_n^B,\ n \in \mathbb{N})$ *increases a.s. to* $+\infty$.

Proof L_∞^B denotes the limit of L_n^B as $n \to +\infty$. The monotone (or the dominated) convergence theorem entails that

$$E\big(e^{-L_\infty^B}\big) = \lim_{n\to\infty} E\big(e^{-L_n^B}\big).$$

From the definition of L_n^B we get

$$\begin{aligned} E\big(e^{-L_n^B}\big) &= E\Big(\exp\big(-\sum_{k=0}^{2^n-1} |B_{(k+1)/2^n} - B_{k/2^n}|\big)\Big) \\ &= E\big(e^{-|B_{1/2^n}|}\big)^{2^n}, \end{aligned}$$

because the increments $B_{(k+1)/2^n} - B_{k/2^n}$, $k = 0, \ldots, 2^n - 1$, are independent, identically distributed, see corollary 3.44. Now we estimate as follows (*exercise*):

$$e^{-x} \le 1 - x + \frac{1}{2}x^2, \qquad x \ge 0.$$

Then

$$E\big(e^{-L_\infty^B}\big) \le \limsup_n E\Big(1 - |B_{1/2^n}| + \frac{1}{2} B_{1/2^n}^2\Big)^{2^n}.$$

$B_{1/2^n}$ having the law $N(0, 1/2^n)$, the last expectation is readily computed, and we get

$$E\big(e^{-L_\infty^B}\big) \le \limsup_n \Big(1 - \sqrt{\frac{2}{\pi}}\, 2^{-n/2} + \frac{1}{2}\, 2^{-n}\Big)^{2^n}.$$

Now, as $n \to +\infty$,

$$\begin{aligned} \ln\bigg(\Big(1 - \sqrt{\frac{2}{\pi}}\, 2^{-n/2} + \frac{1}{2}\, 2^{-n}\Big)^{2^n}\bigg) &= 2^n \ln\Big(1 - \sqrt{\frac{2}{\pi}}\, 2^{-n/2} + \frac{1}{2}\, 2^{-n}\Big) \\ &= 2^n \ln\Big(1 - 2^{-n/2}\Big(\sqrt{\frac{2}{\pi}} + \frac{1}{2}\, 2^{-n/2}\Big)\Big) \\ &\approx -2^n\Big(2^{-n/2}\Big(\sqrt{\frac{2}{\pi}} + \frac{1}{2}\, 2^{-n/2}\Big)\Big) \\ &= -\sqrt{\frac{2}{\pi}}\, 2^{n/2} + \frac{1}{2}, \end{aligned}$$

which tends to $-\infty$. Therefore

$$\lim_n \Big(1 - \sqrt{\frac{2}{\pi}}\, 2^{-n/2} + \frac{1}{2}\, 2^{-n}\Big)^{2^n} = 0,$$

that is, we find

$$E\big(e^{-L_\infty^B}\big) = 0.$$

Hence $\{L_\infty^B < +\infty\}$ has probability zero. $\square$

Statements (i) and (iii) of theorem 9.9 imply

10.2 Theorem *A.s. no path of a one dimensional Brownian motion is rectifiable on a time interval of strictly positive length.*

10.2 Stochastic Integral of Elementary Processes

In this section we construct Itô's stochastic integral for the special class $\mathcal{E}_a$ of elementary, adapted processes X. For simplicity, we shall restrict ourselves to a finite interval for the time parameter domain T, and without loss of generality we can choose $T = [0, 1]$. Throughout we consider a standard one dimensional Brownian motion $B = (B_t,\, t \in [0, 1])$ whose natural filtration will simply be denoted by $\mathcal{F} = (\mathcal{F}_t,\, t \in [0, 1])$.

We will use the following conventions: $\mathcal{I}$ is the set of all finite partitions of $[0, 1]$, such that for $I \in \mathcal{I}$ there exist $n \in \mathbb{N}$, and $t_0, t_1, \ldots, t_n \in [0, 1]$ with $t_0 = 0$, $t_n = 1$, and $t_0 < t_1 < \cdots < t_n$. Then I consists of the intervals $I_1 = [t_0, t_1]$, $I_k = (t_{k-1}, t_k]$, $k = 2, \ldots, n$.

The class $\mathcal{E}_a$ consists of real valued stochastic processes $X = (X_t,\, t \in [0, 1])$ for which there exists $I \in \mathcal{I}$, such that if I is as above, then

$$X_t = \sum_{k=1}^{n} X_{t_{k-1}}\, 1_{I_k}(t) \tag{10.8}$$

where X_{t_k}, $k = 1, \ldots, n-1$, are real valued, bounded random variables, such that X_{t_k} is $\mathcal{F}_{t_k}$–measurable. Furthermore, for simplicity we assume that X starts at some fixed, deterministic point $X_0 \in \mathbb{R}$.

Observe that $X \in \mathcal{E}_a$ is adapted to $\mathcal{F}$, and that is has paths which are continuous from the left (and continuous from the right at $t = 0$). Therefore, by lemma 6.8 $X \in \mathcal{E}_a$ is progressively measurable relative to $\mathcal{F}$.

For an elementary process $X \in \mathcal{E}_a$ of the form (10.8) we define its *Itô integral (with respect to B)* by

$$\int_0^1 X_t\, dB_t = \sum_{k=1}^{n} X_{t_{k-1}}\, (B_{t_k} - B_{t_{k-1}}). \tag{10.9}$$

We leave it as an exercise to the reader to check that this stochastic integral is well-defined, i.e., does not depend on the specific representation of X.

10.3 Remark As it will become clear below, in contrast to, e.g., the construction of the Riemann integal of a function, it is essential that the time t_{k-1}, at which X in (10.9) is evaluated, is at the lower endpoint of the interval I_k.

Let $X \in \mathcal{E}_a$, and suppose that $s, t \in (0, 1]$ $s < t$. Then also $X\, 1_{[0,t]}$, and $X\, 1_{(s,t]}$ belong to $\mathcal{E}_a$. We define

$$\begin{aligned} \int_0^t X_u \, dB_u &= \int_0^1 X_u \, 1_{[0,t]}(u) \, dB_u, \\ \int_s^t X_u \, dB_u &= \int_0^1 X_u \, 1_{(s,t]}(u) \, dB_u. \end{aligned}$$

Observe that by construction for every $X \in \mathcal{E}_a$ and all $t \in [0, 1]$, $\int_0^t X_u \, dB_u$ is $\mathcal{F}_t$–measurable, that is, the process $t \mapsto \int_0^t X_u \, dB_u$ is $\mathcal{F}$–adapted.

The following result is trivial

10.4 Lemma *The stochastic integral is linear on $\mathcal{E}_a$, and additive: If $r, s, t \in [0, 1]$ with $s < r < t$, then for $X \in \mathcal{E}_a$,*

$$\int_s^t X_u \, dB_u = \int_s^r X_u \, dB_u + \int_r^t X_u \, dB_u \tag{10.10}$$

holds true.

If $I \in \mathcal{I}$ is defined as above by $t_1, \ldots, t_{n-1} \in (0, 1)$, we write $\Delta t_k = (t_k - t_{k-1})$, and $\Delta B_k = (B_{t_k} - B_{t_{k-1}})$, $k = 1, \ldots, n$.

The following simple observation is the key point which allows to extend Itô's stochastic integral from elementary adpated processes to a large, interesting class of stochastic processes.

10.5 Lemma (Mini Itô–Isometry) *Suppose that $X \in \mathcal{E}_a$ is as above. Then for all $j, k, l = 1, \ldots, n$, $j \leq k \wedge l$,*

$$E\big(X_{t_{k-1}} \Delta B_k \, X_{t_{l-1}} \Delta B_l \,\big|\, \mathcal{F}_{t_{j-1}}\big) = \delta_{kl} \, E\big(X_{t_{k-1}}^2 \, \Delta t_k \,\big|\, \mathcal{F}_{t_{j-1}}\big) \tag{10.11}$$

and in particular

$$E\big(X_{t_{k-1}} \Delta B_k \, X_{t_{l-1}} \Delta B_l\big) = \delta_{kl} \, E\big(X_{t_{k-1}}^2 \, \Delta t_k\big) \tag{10.12}$$

hold true.

Proof As we had already argued in the proof of lemma 7.5, the independence of the increments of the Brownian motion B entails that for all $s, t \in \mathbb{R}_+$, $s \leq t$, $B_t - B_s$ is independent of $\mathcal{F}_s$, and therefore also of $\mathcal{F}_u$ for every $u \leq s$. First suppose that $k < l$, then we can compute with the properties of the conditional expectation in theorem 5.7.(f), and theorem 5.9.(d), as follows:

$$\begin{aligned} E\big(X_{t_{k-1}} \Delta B_k \, X_{t_{l-1}} \Delta B_l \,\big|\, \mathcal{F}_{t_{j-1}}\big) &= E\big(X_{t_{k-1}} \Delta B_k \, X_{t_{l-1}} E(\Delta B_l \,|\, \mathcal{F}_{t_{l-1}}) \,\big|\, \mathcal{F}_{t_{j-1}}\big) \\ &= 0, \end{aligned}$$

where we used theorem 5.9.(e) to conclude that $E(\Delta B_k \,|\, \mathcal{F}_{t_{k-1}}) = E(\Delta B_k)$, and $E(\Delta B_k) = 0$. Now let $k = l$, then

$$\begin{aligned} E\big(X_{t_{k-1}}^2 \Delta B_k^2 \,\big|\, \mathcal{F}_{t_{j-1}}\big) &= E\big(X_{t_{k-1}}^2 E(\Delta B_k^2 \,|\, \mathcal{F}_{t_{k-1}}) \,\big|\, \mathcal{F}_{t_{j-1}}\big) \\ &= E\big(X_{t_{k-1}}^2 \Delta t_k \,\big|\, \mathcal{F}_{t_{j-1}}\big) \end{aligned}$$

because by a similar reasoning as before $E(\Delta B_k^2 \,|\, \mathcal{F}_{t_{k-1}}) = E(\Delta B_k^2)$, and also $E(\Delta B_k^2) = \Delta t_k$. Thus we have proved equation (10.11), and (10.12) follows by taking expectations on both sides of (10.11). □

10.6 Remark Here is an important remark: It is not really the independence of the increments of the Brownian motion to which the essential relation (10.11) is due. Rather, it is the fact that Brownian motion is a continuous, square integrable martingale. Indeed, if $M = (M_t,\, t \in \mathbb{R}_+)$ is such a martingale, then — with the same arguments as above — one obtains the following generalization of (10.11)

$$E\big(X_{t_{k-1}} \Delta M_k \, X_{t_{l-1}} \Delta M_l \,\big|\, \mathcal{F}_{t_{j-1}}\big) = \delta_{kl} \, E\big(X_{t_{k-1}}^2 \, \Delta A_k \,\big|\, \mathcal{F}_{t_{j-1}}\big), \tag{10.13}$$

where A is the increasing process, which makes $t \mapsto M_t^2 - A_t$ into a martingale. Such a process (with some additonal properties) is uniquely determined via the Doob–Meyer decomposition theorem for the submartingale $t \mapsto M_t^2$. Moreover, we have set $\Delta M_k = M_{t_k} - M_{t_{k-1}}$, and $\Delta A_k = (A_{t_k} - A_{t_{k-1}})$. With this observation it is possible to extend the scope of stochastic integration from Brownian motion as an integrator to continuous, square integrable *local semimartingales*. For details (and more precise statements) the reader is referred to the literature cited at the beginning of this chapter.

10.7 Lemma *Suppose that $X \in \mathcal{E}_a$. Then the stochastic processes defined by ($t \in [0,1]$, $\lambda \in \mathbb{R}$)*

$$I_t = \int_0^t X_u \, dB_u, \tag{10.14}$$

$$J_t = \Big(\int_0^t X_u \, dB_u\Big)^2 - \int_0^t X_u^2 \, du \tag{10.15}$$

$$Z_t = \exp\Big(\lambda \int_0^t X_u \, dB_u - \frac{\lambda^2}{2} \int_0^t X_u^2 \, du\Big) \tag{10.16}$$

are martingales relative to the natural filtration $\mathcal{F}$ of the Brownian motion B.

Proof We only carry out the proof for $J = (J_t,\, t \in [0,1])$, the proofs for I, Z are similar, but easier, and they can be left as an *exercise* to the reader.

Let $s, t \in [0,1]$, $s \leq t$. We have to show that

$$E(J_t \,|\, \mathcal{F}_s) = J_s$$

holds true. Let X be as in equation (10.8). Then we may assume without loss of generality that s and t are among the points $t_0, t_1, \dots, t_n$ defining the partition I. Indeed, if they were not, we can just add them to the list $t_0, t_1, \dots, t_n$, changing the representation of X but not X itself. Thus we suppose from now on that $s = t_k$, $t = t_l$, $k \le l$. Let us denote

$$S_{j,m} = \sum_{i=m}^{j} X_{t_{i-1}} \Delta B_i,$$

then

$$\begin{aligned}\int_0^t X_u \, dB_u &= \int_0^s X_u \, dB_u + \int_s^t X_u \, dB_u \\ &= S_{1,l} + S_{l+1,k}.\end{aligned}$$

We compute

$$\begin{aligned}&E\left(\left(\int_0^t X_u \, dB_u\right)^2 \,\middle|\, \mathcal{F}_s\right) \\ &\qquad = E\big(S_{1,l}^2 \,\big|\, \mathcal{F}_{t_l}\big) + 2E\big(S_{1,l}\, S_{l+1,k} \,\big|\, \mathcal{F}_{t_l}\big) + E\big(S_{l+1,k}^2 \,\big|\, \mathcal{F}_{t_l}\big) \\ &\qquad = S_{1,l}^2 + 2S_{1,l}\, \mathbb{E}\big(S_{l+1,k} \,\big|\, \mathcal{F}_{t_l}\big) + E\big(S_{l+1,k}^2 \,\big|\, \mathcal{F}_{t_l}\big),\end{aligned}$$

because $S_{1,l}$ is $\mathcal{F}_{t_l}$–measurable. For the second term we have

$$\begin{aligned}S_{1,l}\, E\big(S_{l+1,k} \,\big|\, \mathcal{F}_{t_l}\big) &= S_{1,l} \sum_{j=l+1}^{k} E\big(X_{t_{j-1}} \Delta B_j \,\big|\, \mathcal{F}_{t_l}\big) \\ &= S_{1,l} \sum_{j=l+1}^{k} E\big(X_{t_{j-1}} E(\Delta B_j \mid \mathcal{F}_{t_{j-1}}) \,\big|\, \mathcal{F}_{t_l}\big) \\ &= 0,\end{aligned}$$

as in the proof of lemma 10.5. Moreover, with the mini Itô-isometry (10.11) we get

$$\begin{aligned}E\big(S_{l+1,k}^2 \,\big|\, \mathcal{F}_{t_l}\big) &= \sum_{i,j=l+1}^{k} E\big(X_{t_{i-1}} \Delta B_i \, X_{t_{j-1}} \Delta B_j \,\big|\, \mathcal{F}_{t_l}\big) \\ &= \sum_{j=l+1}^{k} E\big(X_{t_{j-1}}^2 \Delta t_j \,\big|\, \mathcal{F}_{t_l}\big) \\ &= E\Big(\int_s^t X_u^2 \, du \,\Big|\, \mathcal{F}_s\Big) \\ &= E\Big(\int_0^t X_u^2 \, du \,\Big|\, \mathcal{F}_s\Big) - \int_0^s X_u^2 \, du.\end{aligned}$$

Altogether we have shown that

$$E\left(\left(\int_0^t X_u\,dB_u\right)^2 \,\Big|\, \mathcal{F}_s\right) = \left(\int_0^s X_u\,dB_u\right)^2 - \int_0^s X_u^2\,du + E\left(\int_0^t X_u^2\,du \,\Big|\, \mathcal{F}_s\right),$$

which proves that J is an $\mathcal{F}$–martingale. □

By taking the expectation of the martingale J we obtain $E(J_t) = E(J_0) = 0$, that is

10.8 Corollary (Itô Isometry for Elementary Processes) *For every $X \in \mathcal{E}_a$ and all $t \in [0, 1]$,*

$$E\left(\left(\int_0^t X_u\,dB_u\right)^2\right) = E\left(\int_0^t X_u^2\,du\right) \tag{10.17}$$

holds true.

10.9 Exercise Derive (10.17) directly from (10.12).

10.3 Mean Square Construction

Suppose that $X \in \mathcal{E}_a$, denote the Itô integral of X temporarily by $I(X)$:

$$I(X) = \int_0^1 X_u\,dB_u,$$

and let us identify $I(X)$ with its P–class in $L^2(P)$. Then the Itô isometry (10.17) states that I is a linear, isometric mapping from $\mathcal{E}_a$ into the Hilbert space $L^2(P)$. Thus we can use a standard result from functional analysis, sometimes called *BLT-theorem* (for "bounded linear transformation"), to extend I to a larger space. For convenience we state this result here together with its simple proof.

10.10 Theorem (BLT) *Suppose that $(E_i, \|\cdot\|_i)$ are normed vector spaces, $(E_2, \|\cdot\|_2)$ complete. Assume furthermore, that T is a bounded linear transformation from D_1 into E_2, where D_1 is dense in E_1. Then T can be extended from D_1 to E_1 without increasing the bound, the extension being unique.*

Proof Let $u \in E_1$, and let $(u_n,\ n \in \mathbb{N})$ be a sequence which converges to u with respect to $\|\cdot\|_1$. Then $(T(u_n),\ n \in \mathbb{N})$ is a Cauchy sequence in $(E_2, \|\cdot\|_2)$:

$$\begin{aligned}\|T(u_n) - T(u_m)\|_2 &= \|T(u_n - u_m)\|_2\\ &\le \|T\|\|u_n - u_m\|_1\\ &\to 0,\end{aligned}$$

as $n, m \to \infty$, where $\|T\|$ denotes the bound of T. Since E_2 is complete relative to $\|\cdot\|_2$, there exists a unique $w \in E_2$, so that $T(u_n) \to w$, $n \to \infty$. We set $T(u) = w$, thereby extend T to E_1. It is easy to check that this extension is well-defined (i.e., does not depend on the choice of the approximating sequence $(u_n,\ n \in \mathbb{N})$), hence unique, and that the extended map admits the same bound. □

We want to apply this theorem with a choice where E_1 is an appropriate subspace of $L^2(P \otimes \lambda)$, λ denoting the Lebegue measure on $([0,1], \mathcal{B}([0,1))$, and $E_2 = L^2(P)$. On the other hand, we prefer to work with random variables instead of with equivalence classes of random variables. This is a standard hassle in such a context, because the norm $\| \cdot \|_2$ of $L^2(P)$ is not a norm on $\mathcal{L}^2(P)$, but only a seminorm. Since however the differences between $L^2(P)$, $L^2(P \otimes \lambda)$, and $\mathcal{L}^2(P)$, respectively $\mathcal{L}^2(P \otimes \lambda)$, are well-understood, we shall brush this problem under the rug, and pretend that we can apply theorem 10.10 directly to $\mathcal{L}^2(P)$, and $\mathcal{L}^2(P \otimes \lambda)$.

Thus, if we can find any space of stochastic processes, so that $\mathcal{E}_a$ is dense in this space with respect to the seminorm of $\mathcal{L}^2(P \otimes \lambda)$, then we can extend the stochastic integral to this space by theorem 10.10. Obviously this space must be a subspace of $\mathcal{L}^2(P \otimes \lambda)$. We let $\mathcal{L}^2_\pi(P \otimes \lambda)$ denote the space of stochastic processes $X = (X_t,\, t \in [0,1])$ which are progressively measurable (see definition 6.4) relative to the natural filtration $\mathcal{F}$ of B, and such that $(\omega, t) \mapsto X_t(\omega)$ belongs to $\mathcal{L}^2(P \otimes \lambda)$. We understand $\mathcal{L}^2_\pi(P \otimes \lambda)$ as being equipped with the seminorm of $\mathcal{L}^2(P \otimes \lambda)$.

10.11 Lemma *$\mathcal{E}_a$ is dense in $\mathcal{L}^2_\pi(P \otimes \lambda)$.*

Sketch of the Proof Let X be in $\mathcal{L}^2_\pi(P \otimes \lambda)$. We must show that X can be approximated in L^2–sense by processes in $\mathcal{E}_a$. This is done in a sequence of steps, which we sketch below. For complete details, in particular of the smoothing step 2 below, we refer the interested reader to, e.g., [14].

Step 1: We approximate X by the sequence $(X^n,\, n \in \mathbb{N})$ of bounded, progressively measurable processes X^n defined by $X^n = (X \wedge n) \vee (-n)$. Indeed, the dominated convergence theorem immediately shows that $X^n \to X$ in $\mathcal{L}^2(P \otimes \lambda)$. That this cutting procedure does not destroy the progressivity of X is obvious. Thus from now on we may assume in addition that the X we start with is bounded.

Step 2: Suppose that X is a progressively measurable stochastic process which is bounded, say by $M > 0$: $|X_t(\omega)| \le M$, $t \in [0,1]$, $\omega \in \Omega$. We want to approximate X relative to the seminorm of $\mathcal{L}^2(P \otimes \lambda)$ by a sequence $(X^n,\, n \in \mathbb{N})$ of bounded, adapted processes which have continuous paths. X^n can be defined by

$$X^n_t(\omega) = n \int_{(t-1/n)\vee 0}^{t} X_s(\omega)\, ds, \qquad t \in [0,1],\ \omega \in \Omega.$$

Since $(\omega, t) \mapsto X_t(\omega)$ is measurable and bounded, the Lebesgue integral on the right hand side is indeed continuous in t. Moreover, the fact that the mapping $(\omega, t) \mapsto X_t(\omega)$ is progressively measurable entails (with the first statement of the theorem of Fubini–Tonelli) that X^n is an adapted process. (Actually, it then follows from lemma 6.8 that X^n is progressively measurable, too — but we will not use this here.) Also it is obvious that for every $n \in \mathbb{N}$, X^n is bounded by M. With results of classical functional analysis (e.g., [32], see also [14, Appendix D]) one can show that X^n converges indeed in $\mathcal{L}^2(P \otimes \lambda)$ to X.

Step 3: It remains to prove that an adapted, bounded stochastic process X with continuous paths can be approximated in $\mathcal{L}^2(P \otimes \lambda)$ by a sequence $(X^n,\, n \in \mathbb{N})$ of processes in $\mathcal{E}_a$. To this end, define $X^n = X_{k/2^n}$ on the interval $(k/2^n, (k+1)/2^n]$, $k = 1, \ldots, 2^n - 1$, and $X^n = X_0$ on $[0, 1/2^n]$. Then it is clear that $X^n \in \mathcal{E}_a$. Moreover, X^n converges pointwise to X as n tends to infinity, while $|X^n| \leq |X|$ for all $n \in \mathbb{N}$. Therefore the dominated convergence theorem gives that X^n converges in $\mathcal{L}^2(P \otimes \lambda)$ to X. □

10.12 Remark In most of the literature one finds that instead of $\mathcal{L}^2_\pi(P \otimes \lambda)$ the larger space of $P \otimes \lambda$–square integrable, adapted processes X is employed. Then there is a small technical problem with step 2 in the preceding proof: It is not clear whether the smoothing procedure retains the property of adaptedness. One remedy is to prove first that every adapted square integrable process has a progressive modification, and then to use this modification as above. That such a modification exists has been shown in [8], a variant of the proof in [8] can be found in [14]. On the other hand, if X is adapted and has left or right continuous paths, then we know from lemma 6.8 that X is already progressive, so that in this case — which will be the typical situation — the above mentioned differences in approach do not matter.

10.13 Definition For every $X \in \mathcal{L}^2_\pi(P \otimes \lambda)$ its *Itô integral* is defined as the (P–a.s.) unique extension of the Itô integral from $\mathcal{E}_a$ to $\mathcal{L}^2_\pi(P \otimes \lambda)$, and it is denoted by

$$\int_0^1 X_u \, dB_u. \tag{10.18}$$

10.14 Remark Based on the martingale property of the process $Z = (Z_t,\ t \in [0,1])$ in (10.16) (with $X \in \mathcal{E}_a$), McKean derives in [20] a powerful inequality which allows a *pathwise* construction of the stochastic integral. Also, it is proved there that for $X \in \mathcal{L}^2_\pi(P \otimes \lambda)$ the stochastic process

$$t \mapsto \int_0^t X_u \, dB_u$$

has a.s. continuous paths.

As for elementary processes X as integrands, we define for $X \in \mathcal{L}^2_\pi(P \otimes \lambda)$, s, $t \in (0,1]$, $s < t$

$$\int_0^t X_u \, dB_u = \int_0^1 X_u \, 1_{[0,t]}(u) \, dB_u, \tag{10.19}$$

$$\int_s^t X_u \, dB_u = \int_0^1 X_u \, 1_{(s,t]}(u) \, dB_u. \tag{10.20}$$

As a byproduct of our construction of the stochastic integral with the help of theorem 10.10, we immediately get

10.15 Theorem *The Itô integral is linear and additive on $\mathcal{L}^2_\pi(P \otimes \lambda)$. Moreover, for every $X \in \mathcal{L}^2_\pi(P \otimes \lambda)$, the* Itô isometry

$$E\left(\Big(\int_0^1 X_t\,dB_t\Big)^2\right) = E\Big(\int_0^1 X_t^2\,dt\Big) = \int_0^1 E(X_t^2)\,dt \tag{10.21}$$

holds.

Consider the stochastic integrals (10.19) as depending on t, that is, as stochastic processes. *A priori*, we only know that each of these stochastic integrals is a random variable, i.e., $\mathcal{A}/\mathcal{B}(\mathbb{R})$–measurable. But for every $t \in [0, t]$, and for every sequence $(X^n, n \in \mathbb{N})$ of elementary processes in $\mathcal{E}_a$ which converge in $\mathcal{L}^2(P \otimes \lambda)$ to X we know that

$$\int_0^1 X_u^n\, 1_{[0,t]}(u)\,dB_u,$$

is $\mathcal{F}_t/\mathcal{B}(\mathbb{R})$–measurable. Thus we consider the pre-Hilbert space $\mathcal{L}^2(\Omega, \mathcal{F}_t, P)$ as embedded in $\mathcal{L}^2(\Omega, \mathcal{A}, P)$. Then the stochastic integrals of $X^n 1_{[0,t]}$ form a Cauchy sequence in $\mathcal{L}^2(\Omega, \mathcal{F}_t, P)$, and therefore (by the Riesz–Fischer–theorem) converge to an $\mathcal{F}_t/\mathcal{B}(\mathbb{R})$–measurable random variable. Since limits in $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ can only differ by a function which is P–a.s. null, we find that the limit we constructed previously, and the limit in $\mathcal{L}^2(\Omega, \mathcal{F}_t, P)$ we just discussed, can only differ on a P–null set. That is, if we consider the process

$$t \mapsto I_t = \int_0^t X_u\,dB_u,$$

then our argument shows that this process has a modification which is $\mathcal{F}$–adapted. Henceforth we shall always consider this modification. But then it is easy to show the following

10.16 Theorem *For every $X \in \mathcal{L}^2(P \otimes \lambda)$ the following two processes are martingales relative to $\mathcal{F}$:*

$$I_t = \int_0^t X_u\,dB_u,$$
$$J_t = \Big(\int_0^t X_u\,dB_u\Big)^2 - \int_0^t X_u^2\,du$$

The *proof* is left as an *exercise* to the interested reader.

10.4 Itô's Formula — A Sketch

Itô's formula is an indispensable tool for computations with stochastic integrals and stochastic differential equations, and as such it is at the basis of a vast number of results in modern stochastic analysis. Even though this formula is relatively simple, and it is easy to understand and to apply, its proof in its full form involves a large amount of technicalities which cannot be treated here. Thus we only give a very rough sketch of it, for the full theory the interested reader is referred to the literature cited at the beginning of this chapter.

The basic observation is that the Brownian increment $B_t - B_s$, $s, t \in \mathbb{R}_+$, $s < t$, having the law $N(0, t-s)$ is of the order of magnitude $\sqrt{t-s}$ for small $t - s$. For example, $E((B_t - B_s)^2) = t - s$. The combination of this observation with Taylor's formula produces the Itô formula, and hence the Itô calculus. In the following we make this observation a bit more precise. To this end, assume that $Y = (Y_t,\ t \in [0,1])$ is a "nice" adapted [1] stochastic process, for example, bounded with continuous paths. Furthermore, suppose that $0 \le s < t \le 1$, and let $u_0, u_1, \dots, u_n \in [s,t]$ define a partition of $[s,t]$ with $u_0 = s < u_1 < \cdots < u_n = t$. Then

$$\sum_{k=1}^{n} Y_{u_{k-1}} (B_{u_k} - B_{u_{k-1}})^2 \to \int_s^t Y_u \, du \tag{10.22}$$

in $\mathcal{L}^2(P)$, as the mesh $\Delta = \max_k (u_k - u_{k-1})$ of the partition tends to zero, as we shall prove now. Write

$$\begin{aligned}
&\sum_{k=1}^{n} Y_{u_{k-1}} (B_{u_k} - B_{u_{k-1}})^2 \\
&\quad = \sum_{k=1}^{n} Y_{u_{k-1}} (u_k - u_{k-1}) + \sum_{k=1}^{n} Y_{u_{k-1}} \big((B_{u_k} - B_{u_{k-1}})^2 - (u_k - u_{k-1}) \big).
\end{aligned}$$

The first term on the right hand side converges pathwise, i.e., pointwise on Ω, to the Riemann integral of Y over $[s,t]$. Since we assume that Y is bounded, the dominated convergence theorem implies its convergence to $\int_s^t Y_u \, du$ in $\mathcal{L}^2(P)$. Thus to prove our claim, we have to show that the second term on the right hand side vanishes in $\mathcal{L}^2(P)$ with $\Delta \to 0$. Therefore we have to estimate

$$E\left(\Big(\sum_{k=1}^{n} Y_{u_{k-1}} \big((B_{u_k} - B_{u_{k-1}})^2 - (u_k - u_{k-1}) \big) \Big)^2 \right).$$

[1] Throughtout we mean here by "adapted" that the process is adapted to the natural filtration generated by the underlying Brownian motion B.

Let us consider a single cross term, i.e., for $l < k$ consider

$$
\begin{aligned}
E\Big(&Y_{u_{l-1}}Y_{u_{k-1}}\big((B_{u_l}-B_{u_{l-1}})^2-(u_l-u_{l-1})\big)(B_{u_k}-B_{u_{k-1}})^2-(u_k-u_{k-1})\Big)\\
&= E\Big(Y_{u_{l-1}}Y_{u_{k-1}}\big((B_{u_l}-B_{u_{l-1}})^2-(u_l-u_{l-1})\big)\\
&\qquad\qquad \times E\big((B_{u_k}-B_{u_{k-1}})^2-(u_k-u_{k-1}) \,\big|\, \mathcal{F}_{u_{k-1}}\big)\Big)\\
&= E\Big(Y_{u_{l-1}}Y_{u_{k-1}}\big((B_{u_l}-B_{u_{l-1}})^2-(u_l-u_{l-1})\big)\\
&\qquad\qquad \times E\big((B_{u_k}-B_{u_{k-1}})^2-(u_k-u_{k-1})\big)\Big)\\
&= 0,
\end{aligned}
$$

where we used the independence of $(B_{u_k}-B_{u_{k-1}})$ of $\mathcal{F}_{u_{k-1}}$, and the above mentioned fact that $E((B_{u_k}-B_{u_{k-1}})^2) = u_k - u_{k-1}$. Thus

$$
\begin{aligned}
E\bigg(&\Big(\sum_{k=1}^{n} Y_{u_{k-1}}\big((B_{u_k}-B_{u_{k-1}})^2-(u_k-u_{k-1})\big)\Big)^2\bigg)\\
&= \sum_{k=1}^{n} E\Big(Y_{u_{k-1}}^2\big((B_{u_k}-B_{u_{k-1}})^2-(u_k-u_{k-1})\big)^2\Big)\\
&= \sum_{k=1}^{n} E\Big(Y_{u_{k-1}}^2 E\Big(\big((B_{u_k}-B_{u_{k-1}})^2-(u_k-u_{k-1})\big)^2 \,\Big|\, \mathcal{F}_{u_{k-1}}\Big)\Big)\\
&= \sum_{k=1}^{n} E\Big(Y_{u_{k-1}}^2 E\Big(\big((B_{u_k}-B_{u_{k-1}})^2-(u_k-u_{k-1})\big)^2\Big)\Big),
\end{aligned}
$$

by the same argument as before. The inner expectation is readily computed:

$$
E\Big(\big((B_{u_k}-B_{u_{k-1}})^2-(u_k-u_{k-1})\big)^2\Big) = 2(u_k-u_{k-1})^2,
$$

and so we get

$$
\begin{aligned}
E\bigg(\Big(\sum_{k=1}^{n} Y_{u_{k-1}}\big((B_{u_k}-B_{u_{k-1}})^2-(u_k-u_{k-1})\big)\Big)^2\bigg)&\\
&= 2\sum_{k=1}^{n} E(Y_{u_{k-1}}^2)(u_k-u_{k-1})^2\\
&\le 2\Delta\sum_{k=1}^{n} E(Y_{u_{k-1}}^2)(u_k-u_{k-1}).
\end{aligned}
$$

By our assumptions on Y, the last sum converges to the Riemann integral

$$
\int_s^t E(Y_u^2)\,du
$$

as $\Delta \to 0$, and therefore the whole term vanishes as $\Delta \to 0$, as claimed.

We pause for the following remark. What really made the above argument work is the fact that $u \mapsto B_u^2 - u$ is a martingale. Therefore, if M is a continuous martingale in $\mathcal{L}^4(P)$, and if — according to the Doob–Meyer–decomposition — A is the predictable increasing process which makes $u \mapsto M_u^2 - A_u$ a martingale, then (with some care) we can replace everywhere above ΔB by ΔM and Δu by ΔA, and get an analogous, more general result.

Now suppose that g is a smooth, bounded function on $\mathbb{R}$, $s, t \in [0, 1]$, $s < t$. Let $u_0 = s < u_1 < \cdots < u_n = t$, Δ be as above. Write

$$g(B_t) - g(B_s) = \sum_{k=1}^{n} \big(g(B_{u_k}) - g(B_{u_{k-1}})\big).$$

We want to make a Taylor approximation of $g(B_{u_k})$ at $B_{u_{k-1}}$ under the sum. Then (10.22) tells us that we have at least terms of second order in $(B_{u_k} - B_{u_{k-1}})$ into account, and a similar computation to the one which used to prove this convergence shows that terms of higher order do not contribute in the limit. Thus we compute

$$\begin{aligned} g(B_{u_k}) - g(B_{u_{k-1}}) &= g'(B_{u_{k-1}})(B_{u_k} - B_{u_{k-1}}) \\ &\quad + \frac{1}{2} g''(B_{u_{k-1}})(B_{u_k} - B_{u_{k-1}})^2 + O\big((B_{u_k} - B_{u_{k-1}})^3\big). \end{aligned}$$

We insert this above, and — taking (10.22) into account — obtain in the limit $\Delta \to 0$ the formula

$$g(B_t) = g(B_{u_s}) + \int_s^t g'(B_u)\, dB_u + \frac{1}{2} \int_s^t g''(B_u)\, du. \tag{10.23}$$

Clearly, in the same way we can handle the case where g depends explicitly on time, too. Then we obtain the simplest form of *Itô's formula*:

$$\begin{aligned} g(t, B_t) = g(s, B_{u_s}) &+ \int_s^t \dot g(u, B_u)\, du \\ &+ \int_s^t g'(u, B_u)\, dB_u + \frac{1}{2} \int_s^t g''(u, B_u)\, du, \end{aligned} \tag{10.24}$$

where $\dot g(t, x)$ denotes the partial derivative of g with respect to the time variable t.

It is quite handy and common to write Itô's formula (10.24) in *differential form* as follows

$$dg(t, B_t) = \dot g(t, B_t)\, dt + g'(t, B_t)\, dB_t + \frac{1}{2} g''(t, B_t)\, dt. \tag{10.25}$$

Let us work out two examples. [2] First let us choose $g(x) = 1/2\, x^2$, $x \in \mathbb{R}$. This

[2]Actually, both examples make use of functions g which do *not* satisfy the boundedness condition which we have used above to derive Itô's formula. As a matter of fact, one can show that the formula holds for a much larger class of functions by using elaborate approximation techniques. In the sketch of Itô calculus presented here we have to avoid to go further into such details.

case can be treated with the simpler formula (10.23), and we obtain (with $s = 0$)

$$\int_0^t B_u \, dB_u = \frac{1}{2}\left(B_t^2 - t\right).$$

This formula is strikingly different from its counterpart for standard Riemann integrals (due to the second order term in (10.23)). Next let choose $g(t, x) = \exp(\lambda x - \lambda^2 t/2)$, $t \in \mathbb{R}_+$, $x \in \mathbb{R}$, where λ is a real parameter. Then Itô's formula (10.24) gives

$$g(t, B_t) = 1 + \lambda \int_0^t g(u, B_u) \, dB_u,$$

or in differential form

$$dg(t, B_t) = \lambda g(t, B_t) \, dB_t,$$

showing that $g(t, B_t)$ plays the role of the exponential function in Itô calculus.

We end this section by pointing out that Itô's formula has very far going generalizations, which can be found in the literature quoted above.

Appendix A

Kolmogorov Extension Theorem

The aim of this appendix is to prove Kolmogorov's extension theorem, theorem 3.16. Throughout we use the notation developed in chapter 3, and in particular in section 3.3.

First we need to collect some properties of polish spaces. The interested reader is referred to, e.g., [1, Section 4.3] or [2, § 26] for details, proofs and further references.

A topological space $(E, \mathcal{O})$ is called *polish*, if it is metrizable, first countable (which for a metrizable space is equivalent to separable), and complete for a metric defining the topology $\mathcal{O}$. Obviously, a polish space is a Hausdorff space (i.e., points can be separated by open sets). In particular, compact sets in a polish space are closed.

Thus every complete, separable metric space is a polish space. Obvious examples are $\mathbb{R}^d$, $\mathbb{C}^d$, $d \in \mathbb{N}$, with any of their standard metrics, and arbitrary closed subsets of these spaces (with the corresponding induced metric).

Let J be any finite index set, and consider the Cartesian product E^J, equipped with the product topology $\mathcal{O}^J$. Then also $(E^J, \mathcal{O}^J)$ is polish.

The Borel–σ–algebra of any topological space $(E, \mathcal{O})$ is by definition the σ–algebra generated by $\mathcal{O}$: $\sigma(\mathcal{O})$. The importance of polish spaces for measure and integration theory, and therefore in particular for probability theory, stems from the following fact (e.g., [1, Theorem 4.3.8] or [2, Lemma 26.2]). Suppose that $(E, \mathcal{O})$ is polish, and denote $\mathcal{E} = \sigma(\mathcal{O})$. Assume furthermore that μ is a finite measure on $(E, \mathcal{E})$. Then it is *inner regular*. This means that for every $B \in \mathcal{E}$

$$\mu(B) = \sup\{\mu(K),\ K \subset B,\ K \text{ compact}\} \tag{A.1}$$

holds true. (Note that $\mu(K)$ for compact K makes sense, because as a compact set K is closed (s.a.), and therefore belongs to $\mathcal{E}$.) That is, for the computation of $\mu(B)$, $B \in \mathcal{E}$, we can approximate B by compact sets $K \subset B$, and thereby use the special topological properties of compact sets for this purpose. The reader will see a particular instance of this in the proof of Kolmogorov's extension theorem below.

Let T be an arbitrary index set, $\mathcal{T}_0$ denotes the family of all finite subsets of T. Recall that for $J \subset T$, π_T^J denotes the projection of E^T onto E^J. $\mathcal{E}^T$ is the σ–algebra on E^T generated by the one dimensional projections $\pi_T^{\{t\}}$, $t \in T$. Equivalent descriptions of $\mathcal{E}^T$ are given in lemma 3.12 (with a slightly different notation). In particular, it has been proved there that $\mathcal{E}^T$ is generated by the family $\mathcal{Z}^T$ of cylinder sets. Recall that $Z \in \mathcal{Z}^T$, if there exists $J \in \mathcal{T}_0$, and $B_1, \ldots, B_n \in \mathcal{E}$, where n is the number of elements in J, so that $Z = (\pi_T^J)^{-1}(B_1 \times \cdots \times B_n)$. Also it has been shown in section 3.2 that for every $J \subset T$, π_T^J is $\mathcal{E}^T/\mathcal{E}^J$–measurable.

It will be useful to consider a larger family $\mathcal{C}^T$ than $\mathcal{Z}^T$, namely all subsets C of E^T which are of the form $C = (\pi_T^J)^{-1}(B)$ with $B \in \mathcal{E}^J$, $J \in \mathcal{T}_0$. That is, instead of just considering pre-images of Cartesian products under the finite dimensional projections, we take pre-images of the full σ–algebras $\mathcal{E}^J$ into account. Obviously $\mathcal{Z}^T \subset \mathcal{C}^T \subset \mathcal{E}^T$, and therefore also $\mathcal{C}^T$ generates $\mathcal{E}^T$. In an *exercise* the reader will check that $\mathcal{C}^T$ is an algebra over E^T, that is, it contains E^T, and is stable under intersections and unions. For simplicity, we shall also call the elements of $\mathcal{C}^T$ *cylinder sets*.

We suppose that we are given a family of probability spaces $\big((E^J, \mathcal{E}^J, P_J),\ J \in \mathcal{T}_0\big)$ which is projective. Thus whenever $I, J \in \mathcal{T}_0$ with $I \subset J$, then $P_I = \pi_J^I P_J$. Our aim is to show the unique existence of a probability measure P on $(E^T, \mathcal{E}^T)$ so that for every $J \in \mathcal{T}_0$, $\pi_T^J P = P_J$.

Clearly, by the given data, the measure on $(E^T, \mathcal{E}^T)$ we aim at is already determined on the algebra $\mathcal{C}^T$: For $C \in \mathcal{C}^T$ we (have to) define

$$Q(C) = P_J(B), \quad \text{for } C = (\pi_T^J)^{-1}(B),\ B \in \mathcal{E}^J,\ J \in \mathcal{T}_0. \tag{A.2}$$

By the projectivity of the system $(P_J,\ J \in \mathcal{T}_0)$, Q is a well-defined set function on $\mathcal{C}^T$, as an *exercise* shows. Another *exercise* proves that Q is finitely additive on $\mathcal{C}^T$. Moreover, $Q(E^T) = 1$ is obvious, so that the finite additivity of Q entails also its subtractivity: If $C_1 \subset C_2$, $C_i \in \mathcal{C}^T$, $i = 1, 2$, then $Q(C_2 \setminus C_1) = Q(C_2) - Q(C_1)$.

Thus our aim is to prove that Q extends from $\mathcal{C}^T$ to $\mathcal{E}^T$ to a unique probability measure P. If there is such an extension it has to be unique by theorem 2.5, because $\mathcal{C}^T$ is $\cap$–stable. The existence of P follows from Carathéodory's theorem (e.g., [2, Satz 5.1], [1, 1.3.10]), once we have established that Q is *countably* additive on $\mathcal{C}^T$. In order to show this, it is enough to prove that Q is continuous from above at $\emptyset$ because Q is finite (e.g., [1, Theorem 1.2.8] or [2, Satz 3.2]). That is, we have to show that if $(C_n,\ n \in \mathbb{N})$ is a sequence in $\mathcal{C}^T$ which decreases to $\emptyset$, then $\lim_n Q(C_n) = 0$. Equivalently, we shall prove the following: If $(C_n,\ n \in \mathbb{N})$ is a decreasing sequence in $\mathcal{C}^T$ so that $Q(C_n) \geq \varepsilon > 0$ for all $n \in \mathbb{N}$, then $\cap_n C_n \neq \emptyset$. This is the content of the following two lemmas.

A.1 Lemma *Assume that $(C_n,\ n \in \mathbb{N})$ is a decreasing sequence in $\mathcal{C}^T$ so that for every $n \in \mathbb{N}$, $Q(C_n) \geq \varepsilon > 0$. Then there exists a decreasing sequence $(B_n,\ n \in \mathbb{N})$ of non-empty sets $B_n \in \mathcal{C}^T$ with the following property: For every $n \in \mathbb{N}$ there exist $J_n \in \mathcal{T}_0$ as well as a compact set $K_n \subset E^{J_n}$ so that $B_n \subset (\pi_T^{J_n})^{-1}(K_n) \subset C_n$.*

Proof Let $n \in \mathbb{N}$, and consider C_n. Then there exists $J_n \in \mathcal{T}_0$ and $A_n \in \mathcal{E}^{J_n}$ so that $C_n = (\pi_T^{J_n})^{-1}(A_n)$. By construction of Q (cf. (A.2)) we obtain

$$\varepsilon \leq Q(C_n) = P_{J_n}(A_n).$$

P_{J_n} is inner regular on $\mathcal{E}^{J_n}$, and therefore there exists a compact set $K_n \subset E^{J_n}$ so that $K_n \subset A_n$, and

$$P_{J_n}(A_n) - P_{J_n}(K_n) \le 2^{-(n+1)}\varepsilon.$$

Set $\tilde{B}_n = (\pi_T^{J_n})^{-1}(K_n)$. Then we obtain $C_n \supset \tilde{B}_n$, and

$$Q(C_n \setminus \tilde{B}_n) = Q(C_n) - Q(\tilde{B}_n) \le 2^{-(n+1)}\varepsilon.$$

Now define

$$B_n = \tilde{B}_1 \cap \cdots \cap \tilde{B}_n, \qquad n \in \mathbb{N}.$$

Then $(B_n, n \in \mathbb{N})$ is a decreasing sequence of subsets of E^T, and for each $n \in \mathbb{N}$, $B_n \subset \tilde{B}_n = (\pi_T^{J_n})^{-1}(K_n) \subset C_n$. We claim that $Q(B_n) \ge \varepsilon/2$. Indeed, from the subtractivity of Q and $C_n \subset C_j$ for $j \le n$, we get

$$\begin{aligned}
Q(C_n) - Q(B_n) &= Q(C_n \setminus B_n) \\
&= Q\Big(C_n \setminus (\bigcap_{j=1}^n \tilde{B}_j)\Big) \\
&= Q\Big(\bigcup_{j=1}^n C_n \setminus \tilde{B}_j\Big) \\
&\le Q\Big(\bigcup_{j=1}^n C_j \setminus \tilde{B}_j\Big) \\
&\le \sum_{j=1}^n Q(C_j \setminus \tilde{B}_j) \\
&\le \varepsilon \sum_{j=1}^n 2^{-(j+1)} \\
&\le \frac{\varepsilon}{2}.
\end{aligned}$$

In particular, for each $n \in \mathbb{N}$, B_n is non-empty. $\square$

A.2 Lemma *Suppose that $(B_n, n \in \mathbb{N})$ is a decreasing sequence of non-empty sets in $\mathcal{E}^T$, such that for every $n \in \mathbb{N}$ there exist $J_n \in \mathcal{T}_0$ and a compact set K_n in E^{J_n} with $B_n \subset (\pi_T^{J_n})^{-1}(K_n)$. Then $\cap_n (\pi_T^{J_n})^{-1}(K_n)$ is non-empty.*

Proof Basically, the proof is an application of the diagonal sequence principle. Since every B_n, $n \in \mathbb{N}$, is non-empty, we can find a sequence $(\omega_n, n \in \mathbb{N})$ in E^T with $\omega_n \in B_n$ for every $n \in \mathbb{N}$. $(B_n, n \in \mathbb{N})$ is decreasing, so that for all $m, n \in \mathbb{N}$ with $n \ge m$, we have $\omega_n \in B_m$. Set $J = \cup_n J_n$, and notice that J is at most countable. Actually we shall assume that J is countable — the case that J is finite makes the argument below even easier. Choose $t \in J$, say with $t \in J_m$, $m \in \mathbb{N}$. Then for all $n \in \mathbb{N}$, $n \geq m$,

$$\begin{aligned}\omega_n(t) &= \pi_T^{\{t\}}(\omega_n) = \pi_{J_m}^{\{t\}} \circ \pi_T^{J_m}(\omega_n)\\ &\in \pi_{J_m}^{\{t\}} \circ \pi_T^{J_m}(B_m) \subset \pi_{J_m}^{\{t\}}(K_m).\end{aligned}$$

By definition of the product topology of E^{J_m}, the projection $\pi_{J_m}^{\{t\}}$ is continuous from E^{J_m} onto E, so that the set $\pi_{J_m}^{\{t\}}(K_m)$ is compact. The topology of E being metrizable, this implies that every sequence in $\pi_{J_m}^{\{t\}}(K_m)$ has a subsequence which converges in $\pi_{J_m}^{\{t\}}(K_m)$. Since almost all (i.e., all but finitely many) elements of $(\omega_n(t),\ n \in \mathbb{N})$ belong to this compact set, we have shown the following: For every $t \in J$, every subsequence of $(\omega_n(t),\ n \in \mathbb{N})$ has a further subsequence which converges in E. By the diagonal sequence principle there exists a subsequence $(\omega_{n'})$ and an element $\eta \in E^J$, such that for every $t \in J$, $\omega_{n'}(t) \to \eta(t)$ as $n' \to \infty$. Consider again $t \in J$ such that $t \in J_m$, $m \in \mathbb{N}$. Above we have seen that then $\eta(t) \in \pi_{J_m}^{\{t\}}(K_m)$. Thus if $J_m = \{s_1, \ldots, s_{r_m}\}$, then $(\eta(s_1), \ldots, \eta(s_{r_m})) \in K_m$. Fix an arbitrary element $\xi_0 \in E$. For $t \notin J$ define $\xi(t) = \xi_0$, and if $t \in J$ set $\xi(t) = \eta(t)$. Thus we obtain an element $\xi \in E^T$. Then for every $m \in \mathbb{N}$,

$$\pi_T^{J_m}(\xi) = \big(\eta(s_1), \ldots, \eta(s_{r_m})\big) \in K_m.$$

That is, for every $m \in \mathbb{N}$,

$$\xi \in (\pi_T^{J_m})^{-1}(K_m)$$

holds true. Hence the intersection $\cap_m (\pi_T^{J_m})^{-1}(K_m)$ is non-empty. $\square$

Putting the two lemmas together we find that for the given sequence $(C_n,\ n \in \mathbb{N})$ there exists the sequence $(K_n,\ n \in \mathbb{N})$ such that

$$\bigcap_{n \in \mathbb{N}} C_n \supset \Big(\bigcap_{n \in \mathbb{N}} (\pi_T^{J_n})^{-1}(K_n)\Big) \neq \emptyset,$$

and therefore theorem 3.16 is proved.

Appendix B

Construction of a Poisson Process

Throughout we assume that $\lambda > 0$, and that we are given a probability space $(\Omega, \mathcal{A}, P)$ on which there is defined an independent sequence $(H_n,\, n \in \mathbb{N})$ of exponentially distributed random variables H_n, $n \in \mathbb{N}$, all having rate λ. (For example, the probability space could be constructed as the product space of $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), \mu_\lambda)$ over $\mathbb{N}$, where μ_λ is the exponential law of parameter λ, and H_n is the projection onto the n–th coordinate of $\mathbb{R}_+^{\mathbb{N}}$.)

Define a sequence $J = (J_n,\, n \in \mathbb{N}_0)$ of *jump times* by $J_0 = 0$, and $J_n = J_{n-1} + H_n$, $n \in \mathbb{N}$. Thus, $(J_n,\, n \in \mathbb{N}_0)$ has independent, stationary increments, and the law of $J_{n+m} - J_n$, $m \in \mathbb{N}$, $n \in \mathbb{N}_0$, is the *Erlang* or *Gamma* distribution $\mathcal{G}(m, \lambda)$ with parameters m and λ. For example, one can see this by computing the Laplace transform of the law of $J_{n+m} - J_n$ at $\mu > 0$:

$$E\big(e^{-\mu(J_{n+m}-J_n)}\big) = E\big(e^{-\mu H_1}\big)^m = \Big(\frac{\lambda}{\lambda+\mu}\Big)^m, \tag{B.1}$$

while the Laplace transform of $\mathcal{G}(m, \lambda)$ at μ is given by

$$\int_0^\infty e^{-\mu t}\, \frac{\lambda}{\Gamma(m)}\, (\lambda t)^{m-1}\, e^{-\lambda t}\, dt = \Big(\frac{\lambda}{\lambda+\mu}\Big)^m. \tag{B.2}$$

Therefore the injectivity of the Laplace transform proves the above claim.

Define a stochastic process $X = (X_t,\, t \in \mathbb{R}_+)$ with values in $\mathbb{N}_0$ as follows: $X_0 = 0$, and for $t > 0$ set

$$X_t = n \quad \text{on} \quad J_n \le t < J_{n+1}, \qquad n \in \mathbb{N}_0.$$

By construction, X has right continuous, increasing paths. We claim that X is a Poisson process with rate λ. That is, that X has independent, stationary increments, and that for $0 \le s < t$ the law of $X_t - X_s$ is the Poisson distribution $\mathcal{P}\big(\lambda(t-s)\big)$ with parameter $\lambda(t-s)$.

First we show the following relation for $m, n \in \mathbb{N}_0$, $0 \le s < t$:

$$P\big(X_t = m+n, X_s = n \mid \mathcal{J}_n\big) = P\big(X_{t-s} = m\big)\, P\big(X_s = n \mid \mathcal{J}_n\big), \tag{B.3}$$

where $\mathcal{J} = (\mathcal{J}_n,\, n \in \mathbb{N}_0)$ is the filtration generated by J: $\mathcal{J}_n = \sigma(J_0, J_1, \dots, J_n)$, $n \in \mathbb{N}$. We do this by calculating the Laplace transform of both sides of (B.3) in t and s. (Some details of the calculations below are omitted, the reader is asked to fill them in as an *exercise*.) For the left hand side we get for $\mu_1, \mu_2 > 0$

$$\begin{aligned}
L(\mu_1,\mu_2) &= \int_0^\infty \int_0^\infty e^{-\mu_1 s - \mu_2 t}\, 1_{\{s<t\}}\, P\big(X_t = n+m, X_s = n \,|\, \mathcal{J}_n\big)\, dt\, ds \\
&= E\Big(\int_0^\infty \int_0^\infty e^{-\mu_1 s - \mu_2 t}\, 1_{\{s<t\}}\, 1_{\{J_{n+m} \le t < J_{n+m+1}\}} \\
&\qquad\qquad\qquad \times 1_{\{J_n \le s < J_{n+1}\}}\, dt\, ds \,\Big|\, \mathcal{J}_n\Big) \\
&= \frac{1}{\mu_2} E\Big(\int_{J_n}^{J_{n+1}} e^{-\mu_1 s}\big(e^{-\mu_2 (s \vee J_{n+m})} - e^{-\mu_2 J_{n+m+1}}\big)\, ds \,\Big|\, \mathcal{J}_n\Big),
\end{aligned}$$

and we used Fubini's theorem for conditional expectations, theorem E.1. For $m = 0$ we get $s \vee J_{n+m} = s$, and

$$\begin{aligned}
L(\mu_1,\mu_2) &= \frac{1}{\mu_2}\frac{1}{\mu_1+\mu_2} E\big(e^{-(\mu_1+\mu_2)J_n} - e^{-(\mu_1+\mu_2)J_{n+1}} \,|\, \mathcal{J}_n\big) \\
&\qquad - \frac{1}{\mu_1\mu_2} E\Big(\big(e^{-\mu_1 J_n} - e^{-\mu_1 J_{n+1}}\big) e^{-\mu_2 J_{n+1}} \,\Big|\, \mathcal{J}_n\Big) \\
&= \frac{1}{\lambda+\mu_2}\frac{1}{\lambda+\mu_1+\mu_2} e^{-(\mu_1+\mu_2)J_n},
\end{aligned}$$

where we made use of the independence of $J_{n+1} - J_n$ of $\mathcal{J}_n$ and of (B.1). (Note that $\mathcal{J}_n$ and the σ–algebra generated by the increments of J up to index n coincide — see, e.g., lemma 6.10.) For $m \ge 1$ we find in a similar way

$$\begin{aligned}
L(\mu_1,\mu_2) &= \frac{1}{\mu_1\mu_2} e^{-(\mu_1+\mu_2)J_n} E\Big(\big(1 - e^{-\mu_1(J_{n+1}-J_n)}\big) \\
&\qquad \times e^{-\mu_2(J_{n+1}-J_n)}\, e^{-\mu_2(J_{n+m}-J_{n+1})}\big(1 - e^{-\mu_2(J_{n+m+1}-J_{n+m})}\big)\Big) \\
&= \frac{1}{\mu_1\mu_2} e^{-(\mu_1+\mu_2)J_n} E\big(e^{-\mu_2(J_{n+1}-J_n)} - e^{-(\mu_1+\mu_2)(J_{n+1}-J_n)}\big) \\
&\qquad \times E\big(e^{-\mu_2(J_{n+m}-J_{n+1})}\big)\, E\big(1 - e^{-\mu_2(J_{n+m+1}-J_{n+m})}\big) \\
&= \frac{\lambda^m}{(\lambda+\mu_2)^{m+1}}\frac{1}{\lambda+\mu_1+\mu_2} e^{-(\mu_1+\mu_2)J_n}.
\end{aligned}$$

Thus we have for all $m \in \mathbb{N}_0$

$$L(\mu_1,\mu_2) = \frac{\lambda^m}{(\lambda+\mu_2)^{m+1}}\frac{1}{\lambda+\mu_1+\mu_2} e^{-(\mu_1+\mu_2)J_n}. \tag{B.4}$$

For the right hand side of (B.3) we find

$$\begin{aligned}
R(\mu_1,\mu_2) &= \int_0^\infty \int_0^\infty e^{-\mu_1 s - \mu_2 t}\, 1_{\{s<t\}}\, P\big(X_{t-s} = m\big)\, P\big(X_s = n \,|\, \mathcal{J}_n\big)\, dt\, ds \\
&= \int_0^\infty e^{-(\mu_1+\mu_2)s} P\big(X_s = n \,|\, \mathcal{J}_n\big)\, ds \int_0^\infty e^{-\mu_2 t} P\big(X_t = m\big)\, dt.
\end{aligned}$$

We compute the last integral first. It is equal to

$$\begin{aligned}
E\Big(\int_0^\infty e^{-\mu_2 t}\, 1_{\{J_m \le t < J_{m+1}\}}\, dt\Big) &= \frac{1}{\mu_2}\, E\big(e^{-\mu_2 J_m} - e^{-\mu_2 J_{m+1}}\big) \\
&= \frac{1}{\mu_2}\, E\big(e^{-\mu_2 J_m}\big) E\big(1 - e^{-\mu_2 (J_{m+1} - J_m)}\big) \\
&= \frac{\lambda^m}{(\lambda + \mu_2)^{m+1}},
\end{aligned}$$

where we used once again the independence of the increments of J and relation (B.1). From (B.2) we can read off that

$$\int_0^\infty e^{-\mu_2 t}\, \frac{(\lambda t)^m}{m!}\, e^{-\lambda t}\, dt = \frac{\lambda^m}{(\lambda + \mu_2)^{m+1}}.$$

Thus with the injectivity of the Laplace transform we can conclude that for every $t > 0$ the law of X_t is the Poisson law $\mathcal{P}(\lambda t)$. Moreover we obtain

$$\begin{aligned}
R(\mu_1, \mu_2) &= \frac{\lambda^m}{(\lambda + \mu_2)^{m+1}}\, \frac{1}{\mu_1 + \mu_2}\, E\big(e^{-(\mu_1+\mu_2) J_n} - e^{(\mu_1+\mu_2) J_{n+1}} \mid \mathcal{J}_n\big) \\
&= \frac{\lambda^m}{(\lambda + \mu_2)^{m+1}}\, \frac{1}{\mu_1 + \mu_2}\, e^{-(\mu_1+\mu_2) J_n} \Big(1 - E\big(e^{(\mu_1+\mu_2)(J_{n+1} - J_n)}\big)\Big) \\
&= \frac{\lambda^m}{(\lambda + \mu_2)^{m+1}}\, \frac{1}{\lambda + \mu_1 + \mu_2}\, e^{-(\mu_1+\mu_2) J_n} \\
&= L(\mu_1, \mu_2).
\end{aligned}$$

Thus formula (B.3) is proved.

We rewrite (B.3) as

$$P\big(X_t - X_s = m, X_s = n \mid \mathcal{J}_n\big) = P\big(X_{t-s} = m\big)\, P\big(X_s = n \mid \mathcal{J}_n\big), \tag{B.5}$$

and take expectations on both sides to obtain

$$P\big(X_t - X_s = m, X_s = n\big) = P\big(X_{t-s} = m\big)\, P\big(X_s = n\big).$$

Summation over $n \in \mathbb{N}_0$ yields

$$P\big(X_t - X_s = m\big) = P\big(X_{t-s} = m\big),$$

which shows that the increments of X are stationarily distributed, and that $(X_t - X_s) \sim \mathcal{P}(\lambda(t-s))$, $0 \le s < t$. Moreover, we found

$$P\big(X_t - X_s = m \mid X_s = n\big) = P\big(X_t - X_s = m\big). \tag{B.6}$$

Let $r \in N$, $0 \le s_1 < s_2 < \cdots < s_r < s$, $n_1, \ldots, n_r \in \mathbb{N}$, and consider the event

$$A = \{X_{s_j} = n_j,\ j = 1, \ldots, r\}$$

on $\{X_s = n\}$. Then we have $J_{n+1} > s$, and therefore A is completely determined by $J_1, \ldots, J_n$. With (B.5) we can compute as follows

$$\begin{aligned} P\big(X_t - X_s = m, &X_s = n, X_{s_r} = n_r, \ldots, X_{s_1} = n_1\big) \\ &= E\Big(P\big(X_t - X_s = m, X_s = n, X_{s_r} = n_r, \ldots, X_{s_1} = n_1 \mid \mathcal{J}_n\big)\Big) \\ &= E\Big(P\big(X_t - X_s = m, X_s = n \mid \mathcal{J}_n\big); X_{s_r} = n_r, \ldots, X_{s_1} = n_1\Big) \\ &= P\big(X_t - X_s = m\big)\, P\big(X_s = n, X_{s_r} = n_r, \ldots, X_{s_1} = n_1\big). \end{aligned}$$

Hence we have proved that for all $s, t \in \mathbb{R}_+$, $s < t$, $X_t - X_s$ is independent of $\mathcal{F}_s^X$. It is now an easy exercise to show that this implies that X has independent increments. Altogether we have proved that X is a Poisson process of rate λ.

Rewrite the result of the last computation as follows

$$\begin{aligned} P\big(X_t = n + m, &X_s = n, X_{s_r} = n_r, \ldots, X_{s_1} = n_1\big) \\ &= P\big(X_t - X_s = m\big)\, P\big(X_s = n, X_{s_r} = n_r, \ldots, X_{s_1} = n_1\big), \end{aligned}$$

which is the same as

$$P\big(X_t = n + m \mid X_s = n, X_{s_r} = n_r, \ldots, X_{s_1} = n_1\big) = P\big(X_t - X_s = m\big).$$

By relation (B.6), we may write this as

$$\begin{aligned} P\big(X_t = n + m \mid X_s = n, X_{s_r} = n_r, \ldots, X_{s_1} = n_1\big) &= P\big(X_t - X_s = m \mid X_s = n\big) \\ &= P\big(X_t = n + m \mid X_s = n\big). \end{aligned}$$

Thus we also obtain the (simple) Markov property of the stochastic process X.

Appendix C

Borel–Cantelli Lemma

C.1 Lemma *Suppose that* $(A_n,\, n \in \mathbb{N})$ *is a sequence of events.*

(a) *If* $\sum_n P(A_n) < +\infty$ *then* $P(\limsup_n A_n) = 0$.

(b) *If* $(A_n,\, n \in \mathbb{N})$ *is an independent sequence, and* $\sum_n P(A_n) = +\infty$ *then* $P(\limsup_n A_n) = 1$.

Proof
(a): By definition

$$\limsup_n A_n = \lim_{n\to\infty} \bigcup_{k\geq n} A_k,$$

and therefore by the continuity of P,

$$\begin{aligned} P(\limsup_n A_n) &= \lim_{n\to\infty} P\Big(\bigcup_{k\geq n} A_k\Big) \\ &\leq \limsup_n \sum_{k=n}^{\infty} P(A_k). \end{aligned}$$

But the last series is the Cauchy remainder term of the convergent series $\sum_n P(A_n)$, and therefore vanishes in the limit $n \to \infty$.

(b): We equivalently show that $P(\liminf_n \complement A_n) = 0$. Since

$$\liminf_n \complement A_n = \lim_{n\to\infty} \bigcap_{k\geq n} \complement A_k,$$

we get

$$
\begin{aligned}
P(\liminf_n \complement A_n) &= \lim_{n\to\infty} P\Big(\bigcap_{k\geq n} \complement A_k\Big) \\
&= \lim_{n\to\infty} P\Big(\lim_{N\to\infty} \bigcap_{k\geq n}^{N} \complement A_k\Big) \\
&= \lim_{n\to\infty} \lim_{N\to\infty} P\Big(\bigcap_{k=n}^{N} \complement A_k\Big) \\
&= \lim_{n\to\infty} \lim_{N\to\infty} \prod_{k=n}^{N} P\big(\complement A_k\big) \\
&= \lim_{n\to\infty} \lim_{N\to\infty} \prod_{k=n}^{N} \big(1 - P(A_k)\big),
\end{aligned}
$$

where we used the independence of $\complement A_n, \ldots, \complement A_N$. Now we apply $1 - x \leq \exp(-x)$ for $x \geq 0$ to obtain

$$
\begin{aligned}
P(\liminf_n \complement A_n) &\leq \limsup_{n\to\infty} \limsup_{N\to\infty} \prod_{k=n}^{N} \exp\big(-P(A_k)\big) \\
&\leq \limsup_{n\to\infty} \limsup_{N\to\infty} \exp\Big(-\sum_{k=n}^{N} P(A_k)\Big) \\
&= 0,
\end{aligned}
$$

because by hypothesis the series in the argument of the exponential function diverges as N tends to ∞. $\square$

Appendix D

Hilbert Spaces and Riesz' Theorem

$\mathbb{K}$ denotes one of the fields $\mathbb{R}$ or $\mathbb{C}$.

D.1 Definition A vector space H over $\mathbb{K}$, equipped with an inner product $(\cdot\,,\,\cdot)$, which is complete with respect to the metric induced by the inner product, is called a *Hilbert space*.

D.2 Remark In the case that $\mathbb{K} = \mathbb{C}$ there are different conventions about the sesquilinear inner product (i.e., $\overline{(u, v)} = (v, u)$, $u, v \in H$) in the literature. We make the convention that it is anti-linear in the first, and linear in the second argument.

D.3 Example Let $(M, \mathcal{M}, \mu)$ be a measure space, and denote the set of μ–square integrable, $\mathbb{K}$–valued functions on M by $\mathcal{L}^2(\mu)$. It follows from Schwarz' inequality that $\mathcal{L}^2(\mu)$ is a $\mathbb{K}$–vector space. Equip $\mathcal{L}^2(\mu)$ with the sesquilinear form

$$(f, g) = \int_M \overline{f} g \, d\mu.$$

Let $\mathcal{N}$ denote the subspace of those $f \in \mathcal{L}^2(\mu)$ so that $(f, f) = 0$, and let $L^2(\mu)$ denote the quotient space $\mathcal{L}^2(\mu)/\mathcal{N}$. It is easy to check that $\mathcal{N}$ consists of the set of $\mathcal{M}$–measurable functions on M which are μ–a.e. equal to 0. For $f \in \mathcal{L}^2(\mu)$ let $[f]$ denotes its class in $L^2(\mu)$. That is, $f, g \in [f]$ if and only if $f - g$ is μ–a.e. equal to 0. Extend the sesquilinear form $(\cdot\,,\,\cdot)$ to $L^2(\mu)$ by setting

$$\big([f], [g]\big)_{L^2(\mu)} = (f, g).$$

The reader will check in an *exercise* that the left hand side is well-defined, and that $(\cdot\,,\,\cdot)_{L^2(\mu)}$ is indeed an inner product on $L^2(\mu)$. Moreover, the Riesz–Fischer–theorem implies that $L^2(\mu)$ is complete (since so is $\mathcal{L}^2(\mu)$ under the degenerate metric defined by $(\cdot\,,\,\cdot)$). Thus $L^2(\mu)$ is a Hilbert space. $\diamond$

Throughout these lectures we shall consider a Hilbert space H as equipped with the metric topology defined by its inner product.

D.4 Definition The *dual* H^* of a Hilbert space H over $\mathbb{K}$ is the space of all linear continuous mappings from H to $\mathbb{K}$.

H^* is equipped with the following norm

$$\|T\|_{H^*} = \sup_{u \in H, u \neq 0} \frac{|T(u)|}{\|u\|_H}, \qquad T \in H^*,$$

where $\|\cdot\|_H$ is the norm defined by the inner product, i.e., $\|u\|_H = \sqrt{(u,u)}$, $u \in H$.

Fix an element $v \in H$. Then it is clear that the mapping

$$\begin{aligned} v : H &\to \mathbb{K} \\ u &\mapsto (v, u) \end{aligned} \tag{D.1}$$

defines an element in H^* of H^*–norm equal to $\|v\|_H$. Thus we see that the mapping (D.1) defines an isometric embedding of H into H^*. The representation theorem of Riesz for Hilbert spaces states that also the converse is true:

D.5 Theorem (Riesz' Representation Theorem) *For every Hilbert space H there exists an isometric bijection between H and its dual H^*. In particular, for every $T \in H^*$ there exists a unique element $v \in H$ so that $T(u) = (v, u)$ for every $u \in H$, and $\|y\|_H = \|T\|_{H^*}$.*

Appendix E

Fubini's Theorem for Conditional Expectations

Throughout this appendix we assume that $(\Omega, \mathcal{A}, P)$ is a probability space, and that $(M, \mathcal{M}, \mu)$ is a σ–finite measure space. Furthermore, X is an $(\mathcal{M} \otimes \mathcal{A})/\mathcal{B}(\mathbb{R})$–measurable mapping from $M \times \Omega$ into $\mathbb{R}$ which is non-negative, $\mu \otimes P$–integrable, respectively. For $m \in M$, $E(X(m) \,|\, \mathcal{A}_0)$ denotes any real valued random variable which is a conditional expectation of the random variable $X(m)$ with respect to $\mathcal{A}_0$.

E.1 Theorem *The family* $\bigl(E(X(m) \,|\, \mathcal{A}_0),\, m \in M\bigr)$ *has a modification* $\bigl(Y(m),\, m \in M\bigr)$ *(i.e.,* $P\bigl(Y(m) = E(X(m) \,|\, \mathcal{A}_0)\bigr) = 1$ *for all* $m \in M$*), possibly with values in* $\overline{\mathbb{R}}$*, which is* $(\mathcal{M} \otimes \mathcal{A}_0)/\mathcal{B}(\overline{\mathbb{R}})$*–measurable,* $\mu \otimes P$*–a.e. non-negative,* $\mu \otimes P$*–integrable, respectively, and such that*

$$\int_M Y(m)\, d\mu(m) = E\Bigl(\int_M X(m)\, d\mu(m) \Bigm| \mathcal{A}_0\Bigr) \tag{E.1}$$

holds true.

Proof Suppose first that such a modification Y of $E(X \,|\, \mathcal{A}_0)$ exists, and assume in addition that X is non-negative. Then for every $m \in M$ almost surely $E(X(m) \,|\, \mathcal{A}_0) \geq 0$, and therefore also P–a.s. $Y(m) \geq 0$. Actually, without loss of generality we may assume that Y is pointwise non-negative: Otherwise replace Y by $|Y|$, then $|Y|$ is $\mathcal{M} \otimes \mathcal{A}_0$–measurable, and a modification of $\bigl(E(X(m) \,|\, \mathcal{A}_0),\, m \in M\bigr)$, too. Let $A \in \mathcal{A}_0$, then Fubini's theorem gives

$$\begin{aligned}
\int_A \Bigl(\int_M Y(m)\, d\mu(m)\Bigr)\, dP &= \int_M \Bigl(\int_A Y(m)\, dP\Bigr)\, d\mu(m) \\
&= \int_M \Bigl(\int_A E\bigl(X(m) \,|\, \mathcal{A}_0\bigr)\, dP\Bigr)\, d\mu(m) \\
&= \int_M \Bigl(\int_A X(m)\, dP\Bigr)\, d\mu(m) \\
&= \int_A \Bigl(\int_M X(m)\, d\mu(m)\Bigr)\, dP.
\end{aligned}$$

Here we used the fact that P–a.s. $Y(m) = E(X(m) \mid \mathcal{A}_0)$ for second equality, for the third the property of the conditional expectation, and for the last again Fubini's theorem. Thus in the case that $X \geq 0$ we have proved that formula (E.1) holds. Moreover, this calculation shows that the $\mu \otimes P$–integrability of X entails the one of Y. Therefore, if X is not necessarily non-negative but $\mu \otimes P$–integrable, then (E.1) also holds true. It remains to show the existence of a modification Y of $E(X \mid \mathcal{A}_0)$ which is $\mathcal{M} \otimes \mathcal{A}_0$–measurable.

Consider first the case that X is of the form $X(m, \omega) = 1_{B \times D}(m, \omega) = 1_B(m)\, 1_D(\omega)$ with $B \in \mathcal{M}$, $D \in \mathcal{A}$. Then

$$E(X(m) \mid \mathcal{A}_0)(\omega) = 1_B(m) P(D \mid \mathcal{A}_0)(\omega), \qquad P\text{–a.s.},$$

and we may set $Y(m, \omega) = 1_B(m) P(D \mid \mathcal{A}_0)$, where $P(D \mid \mathcal{A}_0)$ is any version of this conditional probability. Indeed, this Y is $\mathcal{M} \otimes \mathcal{A}_0$–measurable: If $x \in \mathbb{R}$, then

$$\{Y \leq x\} = \begin{cases} B \times \big\{P(D \mid \mathcal{A}_0) \leq x\big\}, & \text{if } x < 0, \\ \big(\complement B \times \Omega\big) \cup \big(B \times \big\{P(D \mid \mathcal{A}_0) \leq x\big\}\big), & \text{if } x \geq 0. \end{cases}$$

(NB: If $x < 0$, $\big\{P(D \mid \mathcal{A}_0) \leq x\big\}$ is in general a P–null set in $\mathcal{A}_0$ which does not need to be the empty set.) Clearly, the sets on the right hand side belong to $\mathcal{M} \otimes \mathcal{A}_0$.

In order to pass to $X = 1_C$ with general $C \in \mathcal{M} \otimes \mathcal{A}$, we use the monotone class theorem. To this end, let $\mathcal{D}$ denote the family of sets in $\mathcal{M} \otimes \mathcal{A}$ which are such that $\big(E(X(m) \mid \mathcal{A}_0),\, m \in M\big)$ has a modification which is $\mathcal{M} \otimes \mathcal{A}_0$–measurable. Above we have shown that $\mathcal{D} \supset \mathcal{M} \times \mathcal{A}$, the latter being a $\cap$–stable generator of $\mathcal{M} \otimes \mathcal{A}$. We prove that $\mathcal{D}$ is a Dynkin system. The fact that $M \times \Omega$ belongs to $\mathcal{D}$ is obvious. If $C_i \in \mathcal{D}$, $i = 1, 2$, with $C_1 \subset C_2$, then $1_{C_2 \setminus C_1} = 1_{C_2} - 1_{C_1}$ so that

$$E(1_{C_2 \setminus C_1} \mid \mathcal{A}_0) = E(1_{C_2} \mid \mathcal{A}_0) - E(1_{C_1} \mid \mathcal{A}_0), \qquad P\text{–a.s.}$$

Let Y_i denote the modification of $E(1_{C_i} \mid \mathcal{A}_0)$, $i = 1, 2$. Then $Y_2 - Y_1$ is a modification of $E(1_{C_2 \setminus C_1} \mid \mathcal{A}_0)$ which is $\mathcal{M} \otimes \mathcal{A}_0$–measurable. Thus $C_2 \setminus C_1 \in \mathcal{D}$. Next let $(C_n,\, n \in \mathbb{N})$ be a sequence in $\mathcal{D}$ which increases to C, and denote an $\mathcal{M} \otimes \mathcal{A}_0$–measurable modification of $E(1_{C_n} \mid \mathcal{A}_0)$ by Y_n, $n \in \mathbb{N}$. The monotone convergence theorem for conditional expectations, theorem 5.11.(a), entails that for every $m \in M$,

$$E(1_C(m) \mid \mathcal{A}_0) = \sup_n E(1_{C_n}(m) \mid \mathcal{A}_0), \qquad P\text{–a.s.}$$

For every $m \in M$, and every $n \in \mathbb{N}$ there exists a P–null set $N_{m,n}$ in $\mathcal{A}_0$, so that for all $\omega \in \complement N_{m,n}$

$$E(1_{C_n}(m) \mid \mathcal{A}_0)(\omega) = Y_n(m, \omega)$$

holds true. Set $N_m = \cup_n N_{m,n}$ which is a P–null set in $\mathcal{A}_0$, and such that for all $\omega \in \complement N_m$, and all $n \in \mathbb{N}$ the last equation is valid. Therefore we get

$$E(1_C(m) \mid \mathcal{A}_0) = \sup_n Y_n(m), \qquad P\text{–a.s.},$$

for all $m \in M$. Thus we may choose $Y = \sup_n Y_n$, which is an $\mathcal{M} \otimes \mathcal{A}_0$–measurable modification of $E(1_C \mid \mathcal{A}_0)$. Hence $C \in \mathcal{D}$, and therefore the monotone class theorem 2.4 gives that $\mathcal{D} = \mathcal{M} \otimes \mathcal{A}$.

Clearly, we now get by (P–a.s.) linearity of the conditional expectation that every $\mathcal{M} \otimes \mathcal{A}$–elementary mapping X (i.e., finite linear combinations of indicators in $\mathcal{M} \otimes \mathcal{A}$ with non-negative coefficients) admits a modification Y as in the statement of the theorem. If X is non-negative and $\mathcal{M} \otimes \mathcal{A}$–measurable, then there exists a sequence $(X_n,\, n \in \mathbb{N})$ of $\mathcal{M} \otimes \mathcal{A}$–elementary mappings X_n, each of which admits a modification Y_n as above, and which is such that X_n increases to X. Thus we may use the same argument as above for the sequence $(1_{C_n},\, n \in \mathbb{N})$ to conclude that $E(X \mid \mathcal{A}_0)$ has an $\mathcal{M} \otimes \mathcal{A}_0$–measurable modification given by $Y = \sup_n Y_n$. Finally, for general $\mathcal{M} \otimes \mathcal{A}$–measurable X we consider its decomposition $X = X^+ - X^-$ into positive and negative parts with $\mathcal{M} \otimes \mathcal{A}_0$–measurable modifications $Y^\pm$ of $E(X^\pm \mid \mathcal{A}_0)$, so that $Y^+ - Y^-$ is a modification of $E(X \mid \mathcal{A}_0)$ with the desired properties. $\square$

After all these precautions it will now be safe to state the above theorem and formula (E.1) in the somewhat sloppy form

$$\int_M E\big(X(m) \mid \mathcal{A}_0\big)\, d\mu(m) = E\Big(\int_M X(m)\, d\mu(m) \,\Big|\, \mathcal{A}_0\Big), \tag{E.2}$$

where the preceding arguments boil down to the statement that one has to choose appropriate versions of the conditional expectations $E\big(X(m) \mid \mathcal{A}_0\big)$, $m \in M$, which, however, exist.

Appendix F

Doob's Factorization Lemma

F.1 Lemma *Suppose that* Ω *is a set,* $(\Omega', \mathcal{A}')$ *a measurable space,* f *a real valued function on* Ω*, and* T *a mapping from* Ω *to* Ω'*.* f *is* $\sigma(T)$*–measurable, if and only if there exists a measurable function* g *on* Ω'*, so that* $f = g \circ T$*.*

Sketch of the Proof "$\Rightarrow$" Assume first that $f = 1_A$ with $A \in \sigma(T)$. Then there exists $A' \in \mathcal{A}'$ so that $A = T^{-1}(A')$. Set $g = 1_{A'}$. Then $g \circ T = 1_{T^{-1}(A')} = 1_A = f$. Now go through the usual routine: Prove the statement for f being a positive linear combination of indicators, then for positive measurable f, and finally — by decomposition in positive and negative parts — for general f.

"$\Leftarrow$" This is trivial. □

F.2 Remarks For a detailed proof the interested reader can consult, e.g., [2, p. 71f]. The statement of the factorization lemma is readily generalized to functions f taking values in the extended real line.

Appendix G

Some Banach Spaces of Functions

$\mathbb{K}$ denotes $\mathbb{R}$ or $\mathbb{C}$.

G.1 Definition A *Banach space* is a vector space B over $\mathbb{K}$, equipped with a norm $\|\cdot\|$, such that B is complete with respect to $\|\cdot\|$.

For notational simplicty, consider henceforth only the case $\mathbb{K} = \mathbb{R}$, the case of complex valued functions will only require minor modifications.

Suppose that $(E, \mathcal{E})$ is a measurable space. $\mathcal{M}_b(E, \mathcal{E})$ denotes the real vector space of all real-valued, bounded, $\mathcal{E}/\mathcal{B}(\mathbb{R})$–measurable functions on E. $\mathcal{M}_b(E, \mathcal{E})$ is equipped with the sup-norm $\|\cdot\|_\infty$, i.e., for $f \in \mathcal{M}_b(E, \mathcal{E})$,

$$\|f\|_\infty = \sup_{x \in E} |f(x)|.$$

In order to show that $(\mathcal{M}_b(E, \mathcal{E}), \|\cdot\|_\infty)$ indeed is a Banach space, we have to prove its completeness. Assume that $(f_n,\ n \in \mathbb{N})$ is a Cauchy sequence with respect to $\|\cdot\|_\infty$. Then for every $x \in E$, $(f_n(x),\ n \in \mathbb{N})$ is a real Cauchy sequence, because for $m, n \in \mathbb{N}$,

$$|f_n(x) - f_m(x)| \le \|f_n - f_m\|_\infty \underset{m,n\to\infty}{\longrightarrow} 0.$$

The completeness of $\mathbb{R}$ entails that for every $x \in E$ there exists $f(x) \in \mathbb{R}$, so that $f_n(x) \to f(x)$ as $n \to \infty$. As a pointwise limit of measurable functions, $f : x \mapsto f(x)$ is measurable. We show that f is bounded: There exists $n_0 \in \mathbb{N}$ such that for all $n \ge n_0$, $\|f_n - f_{n_0}\|_\infty \le 1$. Thus we get for all $n \ge n_0$, $\|f_n\|_\infty \le 1 + \|f_{n_0}\|_\infty$, and as consequence, for all $n \in \mathbb{N}$,

$$\|f_n\|_\infty \le 1 + \sum_{k=1}^{n_0} \|f_k\|_\infty \equiv \rho.$$

In other words, $(f_n,\, n \in \mathbb{N})$ is uniformly bounded by ρ. Then we get for every $x \in E$,

$$\begin{aligned}|f(x)| &= \lim_{n\to\infty} |f_n(x)| \\ &\le \limsup_n \|f_n\|_\infty \\ &\le \rho.\end{aligned}$$

Hence $\|f\|_\infty \le \rho$, and we have shown that $f \in \mathcal{M}_b(E, \mathcal{E})$. Finally we prove that $\|f_n - f\|_\infty \to 0$ with $n \to \infty$: For all $n \in \mathbb{N}$, all $m \in \mathbb{N}$, $m \ge n$, and every $x \in E$ we have

$$|f_n(x) - f_m(x)| \le \sup_{m\ge n} |f_n(x) - f_m(x)|,$$

and since $f_m(x) \to f(x)$, as $m \to \infty$, also

$$\lim_{m\to\infty} |f_n(x) - f_m(x)| \le \sup_{m\ge n} |f_n(x) - f_m(x)|.$$

Now we can estimate as follows:

$$\begin{aligned}\sup_{x\in E} |f_n(x) - f(x)| &= \sup_{x\in E} \lim_{m\to\infty} |f_n(x) - f_m(x)| \\ &\le \sup_{x\in E} \sup_{m\ge n} |f_n(x) - f_m(x)| \\ &= \sup_{m\ge n} \sup_{x\in E} |f_n(x) - f_m(x)| \\ &= \sup_{m\ge n} \|f_n - f_m\|_\infty,\end{aligned}$$

because suprema may be interchanged (*exercise!*). Let $\varepsilon > 0$ be given, then there exists $n_0 \in \mathbb{N}$ so that for all $n, m \ge n_0$, $\|f_n - f_m\|_\infty < \varepsilon$. Thus if $n \ge n_0$ we find from above

$$\|f_n - f\|_\infty \le \sup_{m\ge n_0} \|f_n - f_m\|_\infty < \varepsilon,$$

and we are done.

Next we assume that $(E, \mathcal{T})$ is a topological space, and that $\mathcal{E} = \sigma(\mathcal{T})$. Define the subspace $C_b(E, \mathcal{T})$ of $\mathcal{M}_b(E, \mathcal{E})$ of functions in $\mathcal{M}_b(E, \mathcal{E})$ which are continuous, which we shall abbreviate by $C_b(E)$. Also $C_b(E)$ is equipped with the sup-norm $\|\cdot\|_\infty$. Then $(C_b(E), \|\cdot\|_\infty)$ is a Banach space, too. To see this let $(f_n,\, n \in \mathbb{N})$ be a Cauchy sequence in $C_b(E)$ with respect to $\|\cdot\|_\infty$. We already know that this sequence converges in the sup-norm to some $f \in \mathcal{M}_b(E, \mathcal{E})$, and so we only have to prove that f is continuous. But as a uniform limit of continuous functions, f is continuous. Explicitly: Suppose that $x \in E$, and let $\varepsilon > 0$ be given. Then there exists $n \in \mathbb{N}$ so that $\|f - f_n\|_\infty < \varepsilon/3$. On the other hand, f_n is a continuous function so that there exists a neighborhood [1] U of x so that for all $y \in U$ we have

[1] A *neighborhood* of a point x in a topological space E, is a subset of E which contains an open set containing x.

$|f_n(x) - f_n(y)| < \varepsilon/3$. Thus for all $y \in U$ we get

$$|f(x) - f(y)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| < \varepsilon.$$

Hence f is continous, and therefore $C_b(E)$ complete with respect to $\|\cdot\|_\infty$.

Finally we consider the case where $(E, \mathcal{T})$ is a locally compact topological space. That is, every point x in E has a compact neighborhood. Obviously, every compact topological space is locally compact. Examples of locally compact spaces which are not compact are $\mathbb{R}^d$, $d \in \mathbb{N}$, and $\mathbb{R}_+$. A real valued function f on E is said to *vanish at infinity*, if for every $\varepsilon > 0$ the set $\{x \in E, |f(x)| \ge \varepsilon\}$ is compact — in other words, if for every $\varepsilon > 0$ there exists a compact subset K of E so that $|f(x)| < \varepsilon$ for all $x \in \complement K$.

$C_0(E)$ is defined as the space of continuous real valued functions on E which vanish at infinity. In an easy *exercise* the reader will check that $C_0(E)$ is a subspace of $C_b(E)$. Endowed with the sup-norm $C_0(E)$ is a Banach space. To see this, let $(f_n, n \in \mathbb{N})$ be a Cauchy sequence in $C_0(E)$. Since this is also a Cauchy sequence in $C_b(E)$ we already know from above that $f_n \to f$ uniformly as $n \to \infty$ with $f \in C_b(E)$. Thus we only have to prove that f vanishes at infinity. Given $\varepsilon > 0$, we can find $n \in \mathbb{N}$ so that $\|f_n - f\|_\infty < \varepsilon/2$, and a compact set $K_n \subset E$ so that $|f_n(x)| < \varepsilon/2$ for all x in the complement of K_n. But then for all $x \in \complement K_n$,

$$\begin{aligned} |f(x)| &\le |f(x) - f_n(x)| + |f_n(x)| \\ &\le \|f - f_n\|_\infty + |f_n(x)| \\ &< \varepsilon, \end{aligned}$$

so that f vanishes at infinity.

Appendix H

On the Laplace Transform

In this appendix we prove the two most basic and well-known results on the (real) Laplace transform of a measurable function defined on $\mathbb{R}_+$.

$\mathcal{F}_L$ denotes the space of real valued, measurable functions f on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ which are such that for every $\lambda > 0$, $x \mapsto \exp(-\lambda x) f(x)$, $x \in \mathbb{R}_+$, is integrable with respect to the Lebesgue measure. Then for $\lambda > 0$,

$$\mathcal{L} f(\lambda) = \int_0^\infty e^{-\lambda x} f(x)\, dx$$

is called the *Laplace transform of* f *at* λ. Thus, for $f \in \mathcal{F}_L$, $\mathcal{L} f$ is a well-defined real valued function on $(0, +\infty)$.

Suppose that f, g are two real valued, measurable functions on $(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$, such that for every $x \in \mathbb{R}_+$ the mapping $y \mapsto f(x-y)g(y)$, $y \in \mathbb{R}_+$, is integrable. Define the *convolution* $f * g$ by

$$f * g(x) = \int_0^x f(x-y)g(y)\, dy, \qquad x \in \mathbb{R}_+. \tag{H.1}$$

The first result we prove is that the Laplace transform of $f * g$ is given by the product of their Laplace transforms:

H.1 Theorem *Suppose that* $f, g \in \mathcal{F}_L$. *Then also* $f * g \in \mathcal{F}_L$, *and*

$$\mathcal{L}(f * g)(\lambda) = \mathcal{L} f(\lambda)\, \mathcal{L} g(\lambda), \qquad \lambda > 0, \tag{H.2}$$

holds true.

Proof Suppose first in addition that f, g are non-negative. Then we can compute with the theorem of Fubini–Tonelli as follows ($\lambda > 0$):

$$\begin{aligned}
&\int_0^\infty e^{-\lambda x}\Big(\int_0^x f(x-y)g(y)\,dy\Big)\,dx \\
&\qquad = \int_0^\infty \Big(\int_0^\infty e^{-\lambda(x-y)}\,1_{\mathbb{R}_+}(x-y)f(x-y)\,e^{-\lambda y}g(y)\,dy\Big)\,dx \\
&\qquad = \int_0^\infty e^{-\lambda y}g(y)\Big(\int_0^\infty e^{-\lambda(x-y)}\,1_{\mathbb{R}_+}(x-y)f(x-y)\,dx\Big)\,dy \\
&\qquad = \Big(\int_0^\infty e^{-\lambda y}g(y)\,dy\Big)\Big(\int_0^\infty e^{-\lambda x}f(x)\,dx\Big) \\
&\qquad = \mathcal{L}f(\lambda)\,\mathcal{L}g(\lambda) < +\infty.
\end{aligned}$$

On one hand this shows that $f, g \in \mathcal{F}_L$ implies that $f * g \in \mathcal{F}_L$, and on the other hand that the application of the theorem of Fubini–Tonelli is also justified when we drop the assumption that f and g are non-negative. Thus repeating the calculation above for the general case, we obtain the statement of the theorem. □

Next we prove that $\mathcal{L}$ is injective from $\mathcal{F}_L$ to $\mathcal{L}\mathcal{F}_L$ in the sense, that if $\mathcal{L}f = \mathcal{L}g$ for $f, g \in \mathcal{F}_L$, then $f = g$ a.e. (Thus, if F_L denotes the usual quotient space of $\mathcal{F}_L$ in which functions which only differ on Lebesgue null sets form an equivalence class, then $\mathcal{L}$ is injective on F_L in the usual sense.) A very slick and elegant proof of this result, based on the classical approximation theorem of Weierstraß, can be found, e.g., in [11, Chapter I, Lemma 1.1]. Here we give a rather elementary, detailed proof based on the full Stone–Weierstraß theorem, which we quote in the following form from [19, Chapter I, Theorem 4e]:

H.2 Theorem (Stone–Weierstraß) *Let E be a compact topological space, and let $\mathcal{A}$ be an algebra of real valued functions on E which separates points. Then the closure of $\mathcal{A}$ with respect to the sup-norm is either the algebra $C(E)$ of all real valued, continuous functions on E, or else it is the algebra of all real valued, continuous functions which vanish at one single point p_∞ in E.*

We apply this in the following way. We choose E as the one point compactification of $\mathbb{R}_+$: $E = \mathbb{R}_+ \cup \{+\infty\}$ (i.e., we choose $p_\infty = +\infty$ above), with a topology defined as the collection of all open sets in $\mathbb{R}_+$ together with all sets of the form $O \cup \{+\infty\}$, where O is open in $\mathbb{R}_+$ and has a compact complement. Then the space $C_0(\mathbb{R}_+)$ of all continuous functions which vanish at infinity (see appendix G) coincides with the functions $f \in C(E)$ so that $f(+\infty) = 0$ (*exercise!*). For $\lambda > 0$, define $e_\lambda(x) = \exp(-\lambda x)$, $x \in \mathbb{R}_+$, and extend these functions to E by setting $e_\lambda(+\infty) = 0$. Then $e_\lambda \in C(E)$. $\mathcal{A}$ denotes the algebra generated by $\{e_\lambda, \lambda > 0\}$. Thus, by the functional relation of the exponential function, f is an element in $\mathcal{A}$, if and only if f is of the form

$$f = \sum_{k=1}^n a_k\, e_{\lambda_k},$$

with $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{R}$, $\lambda_1, \dots, \lambda_n > 0$. Clearly $\mathcal{A}$ separates points, since already every e_λ, $\lambda > 0$, is doing this: If $x, y \in E$ are different, then $e_\lambda(x) \neq e_\lambda(y)$ for any choice of $\lambda > 0$. Therefore we infer from H.2 the following

H.3 Corollary *The set of exponential functions $\{e_\lambda, \lambda > 0\}$ is total in $C_0(\mathbb{R}_+)$ with respect to the sup-norm, that is, its linear hull is uniformly dense in $C_0(\mathbb{R}_+)$.*

H.4 Theorem (Injectivity of the Laplace Transform) *If for $f, g \in \mathcal{F}_L$, and all $\lambda > 0$, $\mathcal{L}f(\lambda) = \mathcal{L}g(\lambda)$, then $f = g$ a.e. If in addition f and g are continuous, then $f = g$.*

Proof The second statement is trivial, so we only have to prove that if $f \in \mathcal{F}_L$ is such that $\mathcal{L}f(\lambda) = 0$ for all $\lambda > 0$, then $f = 0$ a.e.

In a first step we reduce to the case where $f \in \mathcal{L}^1(\mathbb{R}_+)$. Indeed suppose that we have proved the claim for all integrable functions g. Given $f \in \mathcal{F}_L$, set $g(x) = \exp(-x)f(x)$, $x \in \mathbb{R}_+$. Then $g \in \mathcal{L}^1(\mathbb{R}_+)$. Moreover, by hypothesis $\mathcal{L}g(\lambda) = \mathcal{L}f(\lambda + 1) = 0$ for all $\lambda > 0$. Therefore $g = 0$ a.e., and hence also $f = 0$ a.e.

Now assume that $f \in \mathcal{L}^1(\mathbb{R}_+)$ is such that $\mathcal{L}f(\lambda) = 0$ for all $\lambda > 0$. Then by linearity of the integral we get

$$\int_0^\infty g(x) f(x)\, dx = 0, \tag{H.3}$$

for all g in the algebra $\mathcal{A}$ spanned by the exponential functions e_λ, $\lambda > 0$. By corollary H.3, $\mathcal{A}$ is dense with respect to $\|\cdot\|_\infty$ in the space $C_0(\mathbb{R}_+)$. Thus for every $g \in C_0(\mathbb{R}_+)$ there exists a sequence $(g_n, n \in \mathbb{N})$ in $\mathcal{A}$ so that $\|g_n - g\|_\infty \to 0$ with $n \to \infty$. Then

$$\begin{aligned}\Big|\int_0^\infty g(x) f(x)\, dx\Big| &\leq \int_0^\infty |g_n(x) - g(x)||f(x)|\, dx \\ &\leq \|g_n - g\|_\infty\, \|f\|_{\mathcal{L}^1(\mathbb{R}_+)}\end{aligned}$$

and this converges to zero as n tends to infinity. Therefore we get (H.3) for all functions $g \in C_0(\mathbb{R}_+)$.

Next let B be a closed, bounded subset of $\mathbb{R}_+$. For $n \in \mathbb{N}$ let $1^n_B(x)$, $x \in \mathbb{R}_+$, denote the smoothed indicator function of B defined in equation (8.35). By construction, 1^n_B belongs to $C_0(\mathbb{R}_+)$, and thus we find

$$\int_0^\infty 1^n_B(x) f(x)\, dx = 0$$

for all $n \in \mathbb{N}$. Since $f \in \mathcal{L}^1(\mathbb{R}_+)$, $|f|$ is an integrable majorant of $x \mapsto 1^n_B(x) f(x)$, uniform in $n \in \mathbb{N}$, and because 1^n_B converges pointwise to 1_B as n tends to infinity, we get from the dominated convergence theorem that

$$\int_0^\infty 1_B(x) f(x)\, dx = 0$$

We let $\mathcal{H}$ denote the set of all bounded, measurable functions g which are such that equation (H.3) holds true. We show that $\mathcal{H}$ satisfies all conditions for an application of theorem 2.7. Obviously, $\mathcal{H}$ is a real vector space. $\mathcal{H}$ contains the constant function 1 because for an integrable function f, as a consequence of the dominated convergence theorem, $\lambda \mapsto \mathcal{L}f(\lambda)$ extends continuously from $(0, +\infty)$ to $\mathbb{R}_+$. Thus we get $\mathcal{L}f(0) = 0$, i.e., $1 \in \mathcal{H}$. If $(g_n, n \in \mathbb{N})$ is a sequence of non-negative functions in $\mathcal{H}$ which pointwise increases to a bounded function g, then it follows again from the dominated convergence theorem that $g \in \mathcal{H}$. Moreover, by the argument above, $\mathcal{H}$ contains 1_B for all bounded, closed subsets $\mathbb{R}_+$. Since these form a $\cap$–stable generator of $\mathcal{B}(\mathbb{R}_+)$, theorem 2.7 implies that $\mathcal{H}$ contains all bounded, measurable functions on $\mathbb{R}_+$. In particular, we may choose for g in equation (H.3) the function $f\, 1_{\{|f|\le n\}}$, $n \in \mathbb{N}$, and obtain

$$\int_0^\infty f(x)^2\, 1_{\{|f|\le n\}}(x)\, dx = 0.$$

The monotone convergence theorem allows us to let n tend to infinity under the last integral, so that we find

$$\int_0^\infty f(x)^2\, dx = 0.$$

Clearly, this implies that $f = 0$ a.e. □

Bibliography

[1] R. B. Ash, *Real Analysis and Probability*, Academic Press, New York and London, 1972.

[2] H. Bauer, *Maß- und Integrationstheorie*, de Gruyter, Berlin, New York, 1990.

[3] ______, *Wahrscheinlichkeitstheorie*, de Gruyter, Berlin, New York, 2002.

[4] L. Breiman, *Probability*, Addison-Wesley, Reading, Massachusetts, 1968.

[5] K. L. Chung, *Lectures from Markov Processes to Brownian Motion*, Springer, New York, Heidelberg, Berlin, 1982.

[6] ______, *A Course in Probability Theory*, third ed., Academic Press, New York, London, 2001.

[7] T. Deck, *Der Itô–Kalkül*, Springer, Berlin, Heidelberg, New York, 2006.

[8] C. Dellacherie and P. A. Meyer, *Probabilities and Potential*, North-Holland, Amsterdam, New York,Oxford, 1978.

[9] J. L. Doob, *Probability*, Wiley, New York, 1953.

[10] ______, *Classical Potential Theory and Its Probabilistic Counterpart*, Springer, Berlin, Heidelberg, New York, 1984.

[11] E. B. Dynkin, *Markov Processes*, vol. 1, Springer, Berlin, Heidelberg, New York, 1965.

[12] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, 2nd ed., North Holland, Amsterdam, Oxford, New York, 1989.

[13] K. Itô and H. P. McKean Jr., *Diffusion Processes and their Sample Paths*, 2nd ed., Springer, Berlin, Heidelberg, New York, 1974.

[14] S. Kaden and J. Potthoff, *Progressive Stochastic Processes and an Application to the Itô Integral*, Stoch. Proc. Appl. **22** (2004), 834–865.

[15] O. Kallenberg, *Foundations of Modern Probability*, Springer, Berlin, Heidelberg, New York, 2002.

[16] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer, Berlin, Heidelberg, New York, 1991.

[17] Y. Komatu, *Elementary inequalities for Mill's ratio*, Rep. Statist. Appl. Res. Un. Jap. Sci. Engrs. **4** (1955), 69–70.

[18] M. Loève, *Probability Theory*, 4th ed., vol. II, Springer, New York, Heidelberg, Berlin, 1978.

[19] L. H. Loomis, *An Introduction to Abstract Harmonic Analysis*, Van Nostrand, London, 1953.

[20] H. P. McKean Jr., *Stochastic Integrals*, Academic Press, New York and London, 1969.

[21] M. Metivier and J. Pellaumail, *Stochastic Integration*, Academic Press, New York, 1980.

[22] B. Øksendal, *Stochastic Differential Equations*, 5th ed., Springer-Verlag, Berlin, Heidelberg, New York, 2000.

[23] H. O. Pollak, *A remark on "Elementary inequalities for Mill's ratio" by Yûsaku Komatu*, Rep. Statist. Appl. Res. Un. Jap. Sci. Engrs. **4** (1956), 110.

[24] J. Potthoff, *Stochastic Simulation*, 2012, Lecture Notes, Mannheim University, http://math5.math.uni-mannheim.de/.

[25] ———, *Wahrscheinlichkeitstheorie I*, 2012, Lecture Notes, Mannheim University, http://math5.math.uni-mannheim.de/.

[26] ———, *Einführung in die Wahrscheinlichkeitstheorie*, 2014, Lecture Notes, Mannheim University, http://math5.math.uni-mannheim.de/.

[27] P. E. Protter, *Stochastic Integration and Differential Equations*, 2nd ed., Springer, Berlin, Heidelberg, New York, 2005.

[28] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, Springer, Berlin, Heidelberg, New York, 1999.

[29] J. Ville, *Étude Critique de la Notion de Collectif*, Gauthier-Villars, Paris, 1939.

[30] H. von Weizsäcker and G. Winkler, *Stochastic Integrals*, Vieweg, Braunschweig, 1990.

[31] D. Williams, *Diffusions, Markov Processes, and Martingales*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, 1979.

[32] A. Zygmund, *Trigonometric Series*, Cambridge University Press, Cambridge, 1988.

Index

σ–cylinders, 13
π–system, 4

Borel–σ–algebra, 7
Borel–Cantelli Lemma, 131
Brownian motion, 30
  gemometric, 62

canonical construction, 23
canonical coordinate process, 16
conditional expectation
  definition, 42
  dominated convergence theorem, 48
  Fubini's theorem, 49, 135
  Jensen's inequality, 50
  monotone convergence theorem, 48
conditional probability, 45
consistency condition, 17
covariance matrix, 96
cylinder set, 12

d-system, 4
Daniell–Kolmogorov
  theorem, 18
Doob's downcrossing inequality, 72
Doob–decomposition, 64
Doob–inequality, 70
Doob–Meyer–decomposition, 63
downcrossing, 71

entry time, 55

filtered probability space, 52
filtration, 52

Gauß kernel, 19
Gaussian
  measure, 95
  process, 95
  random variable, 95

Hölder continuity, 32

increments
  independent, 28
  stationary, 28
inner regular, 123
Itô formula, 121
Itô integral, 111, 117
Itô isometry, 112, 118

kernel, 19
  Brown, 26
  composition, 21
  Poisson, 27
  semigroup, 21
  translation invariant, 26
Kolmogorov
  extension theorem, 17, 123
Kolmogorov–Chentsov
  theorem, 37

Lévy process, 30
law of the iterated logarithm, 102

Markov process, 76
Markov property, 76
Markovian kernel, 19
martingale, 60
martingale convergence theorem, 74
martingale transform, 65
monotone class theorem, 5
  for functions, 6

natural filtration, 54

optional sampling theorem, 66
optional stopping theorem, 65

path space, 11
Poisson process, 30, 127
polish space, 17, 123
projective family, 17
projective limit, 17

random walk, 8
reflection principle, 92
running maximum, 93

semigroup, 21
  Brown, 24
  contraction, 86
  convolution, 27
  Feller, 87
  normal, 21
  Ornstein–Uhlenbeck, 25
  Poisson, 25
stochastic differential equation, 108
stochastic integral equation, 108
stochastic process
  adapted to a filtration, 52
  canonical version, 18
  continuous in probability, 30
  definition, 7
  equivalence, 10
  Feller, 87
  finite dimensional distribution, 10
  indistinguishable, 10
  measurable, 53
  modification, 10
  predictable, 63
  progressively measurable, 53
  stopped, 58
  version, 10
stopping time, 55
strong Markov property, 89
submartingale, 60
supermartingale, 60

transition kernel, 19

unit kernel, 19
universal Markov process, 82
universal Markov property, 83

Wiener
  measure, 41
  process, 41
  space, 41

Notation

$(\Omega, \mathcal{A})$	measurable space
$(\Omega, \mathcal{A}, P)$	probability space
$(E, \mathcal{E})$	measurable space, state space
$(E, \mathcal{O})$	topological space
$(E, \mathcal{O})$	topological space
$\mathcal{A}_0$	σ–subalgebra of $\mathcal{A}$
$\mathcal{F}$	filtration
$\mathcal{F}_\infty$	$\sigma(\mathcal{F}_t,\ t \in T)$
$\mathcal{F}_\tau$	σ–algebra defined by a stopping time τ
$\mathcal{G}_s^X$	σ–algebra of the future of X after time s
$\mathcal{H}$	real vector space of functions
$\mathcal{L}^0((\Omega, \mathcal{A}, P); (E, \mathcal{E}))$	space of E–valued random variables
$\mathcal{L}^p(\Omega, \mathcal{A}, P)$	space of real valued, p–fold integrable random variables
$\mathcal{M}_b(C, \mathcal{C})$	space of bounded, $\mathcal{C}$–measurable functions on C
Δ	ideal point separate from E
Ω	non-empty set, sample space
$\sigma(T_i,\ i \in I)$	σ–algebra generated by a family $(T_i,\ i \in I)$ of mappings T_i
$\mu \otimes \nu$	product measure of the (σ–finite) measures μ and ν
$\mathbb{N}_0$	$\mathbb{N} \cup \{0\}$
$\overline{T}$	$T \cup \{+\infty\}$

$\overline{\mathbb{R}}_+$	$\mathbb{R}_+ \cup \{+\infty\}$
π_I^J	projection in path space
$\mathbb{R}_+$	$[0, +\infty)$
τ	stopping time
$\theta = (\theta_t,\, t \in T)$	family of shift operators
φ	linear bijection from $\mathbb{R}^{n+1}$ onto itself
$E(R; A)$	$E\big(R\, 1_A\big)$
$E(X \mid \mathcal{A}_0)$	conditional expecation of X given $\mathcal{A}_0$
E^T, E^I	path space
$g \otimes f$	tensor product of the functions g and f
$P(A \mid \mathcal{A}_0)$	conditional probability of A given $\mathcal{A}_0$
T	time parameter domain
$X_{t\wedge\tau}$	process stopped at τ