Measure Theory, Probability Theory, Stochastic Processes, and Stochastic Analysis

Peter Pfaffelhuber

Version: May 24, 2026

Contents

Part I

Measure Theory

Prelude

Modern probability theory (including statistics) is based on measure theory. This manuscript is part of a course with the aim to introduce measure theory for students with a solid background in mathematics, which aim to dive deeper into probability theory.

In various parts of mathematics, we aim to assign a set some value, and describe it as its content, volume, etc. In probability, this value is the probability of the set. Since this concept of assigning values to sets has the same features in many areas (e.g. if two sets are disjoint, the volume of their union is the sum of the volumes), several areas are dealing with measure theory.

We approach measure theory in several steps. First, in Chapter 1, we have to deal with set systems (i.e. sets of sets), since it turns out that it leads to contradictions if we assign volumes to all sets. Here, we will learn about semi-rings and σ-fields as specific set systems. Second, in Chapter 2, we construct measures on these set systems. We will do so by constructing outer measures (defined on all sets) and restricting them to a σ-field. Third, in Chapter 3, we will be dealing with measurable functions and integrals with respect to measures. In probabilistic terms, these are random variables, and their expectations. Fourth, in Chapter 4, we will study certain subsets of measurable functions (or random variables), known as $\mathcal{L}^p$-spaces. Last, in Chapter 5, we will be dealing with product spaces, which are important for the theory of stochastic processes. Since various notions (Borel sets, compact systems) are from set-theoretic topology, we give a repetition of the relevant terms in Appendix A.
There are various textbooks in measure theory with a focus on probability. The following have guided me as references for the purpose of this manuscript.

- Bogatchev, Vladimir I. Measure Theory. Springer, 2007
- Billingsley, Patrick. Probability and Measure. Wiley, third edition, 1995
- Kallenberg, Olaf. Foundations of Modern Probability Theory. Springer, third edition, 2021
- Klenke, Achim. Probability theory. A comprehensive course. Springer, 2014

The present english version of this manuscript was written based on the German version with the help of DeepL.

1 Set systems

Probability theory formalises the colloquially used word *probable*. This is (in the broadest sense) a property of a possible outcome of an experiment. Fundamental to probability theory is the concept of an *event*, which is intended to describe everything that can happen in the experiment. Events are represented by subsets of an abstract basic space, which is always called Ω. The aim of this section is to assign a probability to as many subsets of Ω as possible. This leads to the concept of a σ-algebra, because these contain exactly the subsets of the base space to which probabilities are then assigned in the next section. In other words, elements of σ-algebras will be events in the above sense. The other set systems introduced in this section will be used to define suitable σ-algebras.

1.1 Semi-rings, rings and σ-fields

The notions in this section are connected as follows: For $\mathcal{C} \subseteq 2^\Omega$, we have

$$\mathcal{C}\ \sigma\text{-field} \quad \Longrightarrow \quad \mathcal{C}\ \text{ring} \quad \Longrightarrow \quad \mathcal{C}\ \text{semi-ring}.$$

Some more properties of the three notions are given in table 1.

Definition 1.1 (Semi-ring, ring, $\boldsymbol{\sigma}$-field)**.** *Let Ω be a set and $\emptyset \neq \mathcal{H}, \mathcal{R}, \mathcal{F} \subseteq 2^\Omega$.*

1. *$\mathcal{H}$ is* $\cap$-stable *(or closed under $\cap$, or a π-system), if ($A, B \in \mathcal{H} \Rightarrow A \cap B \in \mathcal{H}$). It is called* σ-$\cap$-stable *(or closed under σ-$\cap$) if ($A_1, A_2, ... \in \mathcal{H} \Rightarrow \bigcap_{i=n}^\infty A_n \in \mathcal{H}$).*
   It is called $\cup$-stable *(or closed under $\cup$), if ($A, B \in \mathcal{H} \Rightarrow A \cup B \in \mathcal{H}$). It is called* $\sigma - \cup$-stable *(or closed under σ-$\cup$) if ($A_1, A_2, ... \in \mathcal{H} \Rightarrow \bigcup_{i=n}^\infty A_n \in \mathcal{H}$).*
   It is complement-stable *(or closed under complements), if $A \in \mathcal{H} \Rightarrow A^c \in \mathcal{H}$. It is* set-difference-stable *(or closed under set-differences), if ($A, B \in \mathcal{H} \Rightarrow B \setminus A \in \mathcal{H}$).*

2. *$\mathcal{H}$ is a* semi-ring*, if it is (i) closed under $\cap$ and (ii) $\forall A, B \in \mathcal{H} \exists C_1, \dots, C_n \in \mathcal{H}$ with* [1] $B \setminus A = \biguplus_{i=1}^n C_i$.

3. *$\mathcal{R}$ is a* ring*, if it is closed under $\cup$ and set-differences.*

4. *$\mathcal{F}$ is a* σ-field *(or σ-algebra), if $\Omega \in \mathcal{F}$, it is closed under complements and closed under σ-$\cup$. Then, $(\Omega, \mathcal{F})$ is called* measurable space.

Remark 1.2 (Relationships between the collections of sets)**.**

1. Every ring $\mathcal{R}$ is a semi-ring: *For closedness under $\cap$, we write for $A, B \in \mathcal{R}$*

   $$A \cap B = A \setminus (A \setminus B) \in \mathcal{R}.$$

   The second property is trivial.

2. Every σ-field $\mathcal{F}$ is a ring: *We need to understand that $\mathcal{F}$ is closed under set-differences. For this, we write for $A, B \in \mathcal{F}$*

   $$B \setminus A = B \cap A^c = (B^c \cup A)^c.$$

[1] We write $A \uplus B$ for $A \cup B$ if $A \cap B = \emptyset$.

	$\mathcal{C}$ semi-ring	$\mathcal{C}$ ring	$\mathcal{C}$ σ-field
$\mathcal{C}$ is $\cap$-stable	$\bullet$	$\circ$	$\circ$
$\mathcal{C}$ is σ-$\cap$-stable			$\circ$
$\mathcal{C}$ is $\cup$-stable		$\bullet$	$\circ$
$\mathcal{C}$ is σ-$\cup$-stable			$\bullet$
$\mathcal{C}$ is set-difference-stable		$\bullet$	$\circ$
$\mathcal{C}$ is complement-stable			$\bullet$
$B \setminus A = \biguplus_{i=1}^n C_i$	$\bullet$	$\circ$	$\circ$
$\Omega \in \mathcal{C}$			$\bullet$

Table 1: For $\mathcal{C} \subseteq 2^\Omega$, we list all properties for semi-rings, rings and σ-fields. $\bullet$ means that the respective property is a hypothesis in the definition, whereas $\circ$ means that the respective property is a result following from the definition of the collection of subsets.

Example 1.3 (Semi-rings, σ-fields)**.**

1. Semi-open intervals form a semi-ring: *Let $\Omega = \mathbb{R}$. Then,*

$$\mathcal{H} := \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$$

is a semi-ring.
Indeed, if $a_1 \leq b_1, a_1' \leq b_1'$, then[2] $(a_1, b_1] \cap (a_1', b_1'] = (a_1 \vee a_1', b_1 \wedge b_1']$ *and* $(a_1, b_1] \setminus (a_1', b_1'] = (a_1, a_1' \wedge b_1] \uplus (b_1', b_1]$*, where $(a, b] = \emptyset$, falls $a \geq b$.*

2. Examples for σ-fields: *Trivial examples are $\{\emptyset, \Omega\}$ and 2^Ω. (Recall that both are topologies as well; see Definition A.1.)*

   Yet another example will become important in Section 3.1: If $\mathcal{F}'$ is a σ-field on Ω', and $f : \Omega \to \Omega'$. Then,

$$\sigma(f) := \{f^{-1}(A') : A' \in \mathcal{F}'\} \subseteq 2^\Omega \tag{1.1}$$

is a σ-field on Ω.
Indeed: If $A', A_1', A_2', \ldots \in \sigma(f)$, then $(f^{-1}(A'))^c = f^{-1}((A')^c) \in \sigma(f)$ and $\bigcup_{n=1}^\infty f^{-1}(A_n') = f^{-1}\left(\bigcup_{n=1}^\infty A_n'\right) \in \sigma(f)$.

   We will frequently use the so-called Borel σ-field (which is the σ-field generated by a topology; see Definition 1.7.

[2] As usual, we write $x \wedge y := \min(x, y)$ and $x \vee y := \max(x, y)$

1.2 Generators and extensions

On the one hand, we want to use σ-fields as much as possible, since they contain the sets we can assign probabilities to. On the other hand, often only semi-rings can be given constructively. However, we can use the (ring or) σ-field generated by a semi-ring., i.e. the smallest σ-field (or smallest ring) which contains the semi-ring.

Remark 1.4 (Generated set-systems)**.** *First, it is easy to see that the intersection of σ-fields (rings) is a σ-field (ring). (For example, since all σ-fields are closed under $\cup$, and if we take A, B, elements of all σ-fields, then $A \cup B$ is an element of all σ-fields and therefore in the intersection.):*
Let $\mathcal{C} \subseteq 2^\Omega$. Then,

$$\mathcal{R}(\mathcal{C}) := \bigcap \Big\{ \mathcal{R} \supseteq \mathcal{C} : \mathcal{R} \text{ ring} \Big\}$$

is the ring generated from $\mathcal{C}$ *and*

$$\sigma(\mathcal{C}) := \bigcap \Big\{ \mathcal{F} \supseteq \mathcal{C} : \mathcal{F} \ \sigma\text{-field} \Big\}$$

is the σ-field generated from $\mathcal{C}$. *Apparently, $\mathcal{R}(\mathcal{R}(\mathcal{H})) = \mathcal{R}(\mathcal{H})$ and $\sigma(\sigma(\mathcal{H})) = \sigma(\mathcal{H})$.*

The next lemma is shown after Example 1.6.

Lemma 1.5 (Ring generated from a semi-ring)**.** *Let $\mathcal{H}$ be a semi-ring. Then,*

$$\mathcal{R}(\mathcal{H}) = \Big\{ \biguplus_{k=1}^n A_k : A_1, \dots, A_n \in \mathcal{H} \text{ disjoint}, n \in \mathbb{N} \Big\}$$

is the ring generated from $\mathcal{H}$.

Example 1.6 (Ring generated from semi-open intervals)**.** *Let $\mathcal{H}$ be the semi-ring of semi-open intervals from Example 1.3. Then,*

$$\mathcal{R}(\mathcal{H}) = \Big\{ \biguplus_{k=1}^n (a_k, b_k] : a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{Q}, \\ a_k < b_k, k = 1, \dots, n \text{ and } a_k < b_{k+1}, k = 1, \dots, n-1 \Big\}$$

is the ring generated from $\mathcal{H}$.

Proof of Lemma 1.5. It is clear that $\mathcal{R}(\mathcal{H})$ is closed under $\cap$. In order to show that $\mathcal{R}(\mathcal{H})$ is a ring, we start by showing closedness under set-differences. Let $A_1, \dots, A_n \in \mathcal{H}$ and $B_1, \dots, B_m \in \mathcal{H}$ be disjoint, respectively. Then,

$$\Big(\biguplus_{i=1}^n A_i \Big) \setminus \Big(\biguplus_{j=1}^m B_j \Big) = \biguplus_{i=1}^n \bigcap_{j=1}^m A_i \setminus B_j \in \mathcal{R}(\mathcal{H}).$$

In order to show closedness under $\cup$ of $\mathcal{R}(\mathcal{H})$, let $A, B \in \mathcal{R}(\mathcal{H})$. Then, write $A \cup B = (A \cap B) \uplus (A \setminus B) \uplus (B \setminus A) \in \mathcal{R}(\mathcal{H})$, since we already showed closedness under $\cap$ and under set-differences.

Last, note that there is no smaller ring than $\mathcal{R}(\mathcal{H})$, which contains $\mathcal{H}$. Indeed, such a ring would have to be closed under $\cup$, and clearly $\mathcal{R}(\mathcal{H})$ is the minimal set which contains $\mathcal{H}$ and which is closed under $\cup$. $\square$

Definition 1.7 (Borel σ-algebra)**.** *Let $(\Omega, \mathcal{O})$ be a topological space. Then $\mathcal{B}(\Omega) := \sigma(\mathcal{O})$ denotes the* Borel σ-algebra *on Ω. If $\Omega \subseteq \mathbb{R}^d$, we denote by $\mathcal{B}(\Omega)$ the Borel σ-algebra generated by the Euclidean topology on $\mathbb{R}^d$. If $\Omega \subseteq \overline{\mathbb{R}}$, then $\mathcal{B}(\Omega)$ is the Borel σ-algebra generated by the topology from example A.2. Sets in $\mathcal{B}(\Omega)$ are also called* (Borel-)measurable sets.

Lemma 1.8 (Countable base and Borel σ-algebra)**.** *Let $(\Omega, \mathcal{O})$ be a topological space with countable basis $\mathcal{C} \subseteq \mathcal{O}$. Then, $\sigma(\mathcal{O}) = \sigma(\mathcal{C})$.*

Proof. We only need to show that $\mathcal{O} \subseteq \sigma(\mathcal{C})$. However, this is clear since $A \in \mathcal{O}$ can be represented as a countable union of sets from $\mathcal{C}$. See Lemma A.5. □

Lemma 1.9 (Borel σ-algebra is generated by intervals generated)**.** *Let*

$$\begin{aligned}
\mathcal{C}_1 &= \{[-\infty, b] : b \in \mathbb{Q}\} \text{ or} \\
\mathcal{C}_2 &= \{(a, b] : a, b \in \mathbb{Q}, a \leq b\} \\
\mathcal{C}_3 &= \{(a, b) : a, b \in \mathbb{Q}, a \leq b\} \\
\mathcal{C}_4 &= \{[a, b] : a, b \in \mathbb{Q}, a \leq b\}.
\end{aligned}$$

Then $\sigma(\mathcal{C}_i) = \mathcal{B}(\overline{\mathbb{R}})$, $i = 1, \dots, 4$.

Proof. The set system $\mathcal{C}_3$ is a countable basis of Euclidean topology on $\overline{\mathbb{R}}$. So, in this case, the statement follows from Lemma 1.8.

We only show the statement for $\mathcal{C}_1$ and $\mathcal{C}_2$, the statement for $\mathcal{C}_4$ follows analogously. Firstly, $\mathcal{C}_2 := \{A \setminus B : A, B \in \mathcal{C}_1\} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\} \subseteq \sigma(\mathcal{C}_2)$ is the semi-ring generated by $\mathcal{C}_1$ from Example 1.3. Thus $\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2)$ and it is sufficient to show that $\sigma(\mathcal{C}_2) = \mathcal{B}(\overline{\mathbb{R}})$.

Let $\mathcal{O}$ be as in Definition A.1.8 with $\Omega = \mathbb{R}$. We show (i) that $A \in \mathcal{O}$ implies $A \in \sigma(\mathcal{C}_2)$, and (ii) $A \in \mathcal{C}_2$ implies $A \in \sigma(\mathcal{O})$. It then follows that $\mathcal{O} \subseteq \sigma(\mathcal{C}_2) \subseteq \sigma(\mathcal{O})$, i.e. $\sigma(\mathcal{O}) = \sigma(\mathcal{C}_2)$. For (i) let $A \in \mathcal{O}$. We claim

$$A = \bigcup \{(a, b] : [a, b] \subseteq A, a, b \in \mathbb{Q}\}, \tag{1.2}$$

and note that the right-hand side is an element of $\sigma(\mathcal{C}_2)$. Here, '$\supseteq$' is clear. To see '$\subseteq$', we choose $x \in A$. Then, by definition of $\mathcal{O}$, there is a $\varepsilon > 0$ so that $B_\varepsilon(x) \subseteq A$. However, there are also $a, b \in \mathbb{Q}$ with $a \leq b$ and $x \in (a, b] \subseteq B_\varepsilon(x)$. Thus '$\subseteq$' is shown and (i) follows.
For (ii) we proceed similarly; let $A \in \mathcal{C}_2$. Then obviously

$$A = \bigcap_{n=1}^{\infty} \left(a, b + \tfrac{1}{n}\right).$$

Since $\left(a, b + \frac{1}{n}\right) \in \mathcal{O}$, then $A \in \sigma(\mathcal{O})$. □

Example 1.10 (Borel measurable sets)**.** *Of course, all countable intersections and unions of intervals according to Lemma 1.9 in $\mathcal{B}(\overline{\mathbb{R}})$. Let, for example*

$$\begin{aligned}
A_1 &= [0, \tfrac{1}{3}] \cup [\tfrac{2}{3}, 1], \\
A_2 &= [0, \tfrac{1}{9}] \cup [\tfrac{2}{9}, \tfrac{3}{9}] \cup [\tfrac{6}{9}, \tfrac{7}{9}] \cup [\tfrac{8}{9}, 1], \\
A_3 &= [0, \tfrac{1}{27}] \cup [\tfrac{2}{27}, \tfrac{3}{27}] \cup [\tfrac{6}{27}, \tfrac{7}{27}] \cup [\tfrac{8}{27}, \tfrac{9}{27}] \cup [\tfrac{18}{27}, \tfrac{19}{27}] \cup [\tfrac{20}{27}, \tfrac{21}{27}] \cup [\tfrac{24}{27}, \tfrac{25}{27}] \cup [\tfrac{26}{27}, 1], \\
&\dots,
\end{aligned}$$

then $A = \bigcap_{n=1}^\infty$ denotes Cantor's discontinuum. This set is measurable as a countable intersection of finite unions of intervals. In Example 2.27 we will get to know an example of a non-Borel-measurable set.

1.3 Dynkin systems

In measure theory, it is often necessary to show that a certain set system $\mathcal{F}$ is a σ-algebra and contains a semi-ring $\mathcal{H}$. The Dynkin systems discussed in this section are very helpful here. Because of Theorem 1.13 it is sufficient to show that $\mathcal{F}$ is a $\cap$-stable Dynkin system with $\mathcal{H} \subseteq \mathcal{F}$. This is often easier than showing directly that $\mathcal{F}$ is a σ-algebra.

Definition 1.11 (Dynkin system). *1. A set system $\mathcal{D}$ is called* Dynkin system *(on Ω) if (i) $\Omega \in \mathcal{D}$, (ii) it is set-difference-stable for subsets (i.e. $A, B \in \mathcal{D}$ and $A \subseteq B$ imply $B \setminus A \in \mathcal{D}$ and (iii) $A_1, A_2, \ldots \in \mathcal{D}$ and $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ imply $\bigcup_{n=1}^\infty A_n \in \mathcal{D}$.*

2. For $\mathcal{C} \subseteq 2^\Omega$, we set

$$\lambda(\mathcal{C}) := \bigcap \{\mathcal{D} \supseteq \mathcal{C} \text{ Dynkin-system}\}.$$

Example 1.12 (σ-algebras are Dynkin systems). *1.* Every σ-algebra is a Dynkin system: *Let $\mathcal{F}$ be a σ-algebra. Then $A, B \in \mathcal{F}$ imply $A^c \in \mathcal{F}$ and therefore $\Omega = A \cup A^c \in \mathcal{F}$ and $B \setminus A = B \cap A^c \in \mathcal{F}$.*

2. A Dynkin system $\mathcal{D}$ is complement-stable, since

$$A^c = \Omega \setminus A \in \mathcal{D}$$

Theorem 1.13 ($\cap$-stable Dynkin systems). *Let $\mathcal{D}$ be a Dynkin system and $\mathcal{C} \subseteq \mathcal{D}$ be $\cap$-stable. Then $\sigma(\mathcal{C}) \subseteq \mathcal{D}$. In particular, every $\cap$-stable Dynkin system is a σ-algebra.*

Proof. Let $\lambda(\mathcal{C})$ be the Dynkin system generated by $\mathcal{C}$ (see Definition 1.11). So, we find $\lambda(\mathcal{C}) \subseteq \mathcal{D}$. We will show that $\lambda(\mathcal{C})$ is a σ-algebra, because then $\sigma(\mathcal{C}) \subseteq \sigma(\lambda(\mathcal{C})) = \lambda(\mathcal{C}) \subseteq \mathcal{D}$. For showing that $\lambda(\mathcal{C})$ is a σ-algebra, it suffices to show that $\lambda(\mathcal{C})$ is $\cap$-stable. Then, since $\lambda(\mathcal{C})$ is complement-stable, writing $A \cup B = (A^c \cap B^c)^c$, we see that $\lambda(\mathcal{C})$ is $\cup$-stable. Hence, for $A_1, A_2, ... \in \lambda(\mathcal{C})$, we find $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty \bigcup_{i=1}^n A_i \in \lambda(\mathcal{C})$.

So, it remains to show that $A, B \in \lambda(\mathcal{C})$ imply $A \cap B \in \lambda(\mathcal{C})$: If $A, B \in \mathcal{C}$, this is clear due to the $\cap$-stability of $\mathcal{C}$. For $B \in \mathcal{C}$ we set

$$\mathcal{D}_B := \{A \subseteq \Omega : A \cap B \in \lambda(\mathcal{C})\} \supseteq \mathcal{C}.$$

Then $\mathcal{D}_B$ is a Dynkin system since (i) $\Omega \in \mathcal{D}_B$, (ii) for $A, C \subseteq \mathcal{D}_B$ we have $A \cap B, C \cap B \in \lambda(\mathcal{C})$ and if $A \subseteq C$ we find $A \cap B \subseteq C \cap B$, thus $(C \setminus A) \cap B = (C \cap B) \setminus (A \cap B) \in \lambda(\mathcal{C})$ and (iii) for $A_1, A_2, \ldots \in \mathcal{D}_B$ we have $A_1 \cap B, A_2 \cap B, \ldots \in \lambda(\mathcal{C})$ and with $A_1 \subseteq A_2 \subseteq \cdots$ we have $A_1 \cap B \subseteq A_2 \cap B \subseteq \cdots$, thus $\left(\bigcup_{n=1}^\infty A_n\right) \cap B = \left(\bigcup_{n=1}^\infty A_n \cap B\right) \in \lambda(\mathcal{C})$.

Since $\mathcal{C} \subseteq \mathcal{D}_B$ and $\mathcal{D}_B$ is a Dynkin system, we find that $\lambda(\mathcal{C}) \subseteq \mathcal{D}_B$. This means that $A \in \lambda(\mathcal{C})$ and $B \in \mathcal{C}$ imply $A \cap B \in \lambda(\mathcal{C})$. We now set for an $A \in \lambda(\mathcal{C})$

$$\mathcal{B}_A := \{B \subseteq \Omega : A \cap B \in \lambda(\mathcal{C})\}.$$

As above, we show that $\mathcal{B}_A$ is a Dynkin system with $\mathcal{C} \subseteq \mathcal{B}_A$. Therefore, $\lambda(\mathcal{C}) \subseteq \mathcal{B}_A$. In particular, for $A, B \in \lambda(\mathcal{C})$, we find $A \cap B \in \lambda(\mathcal{C})$, i.e. $\lambda(\mathcal{C})$ is $\cap$-stable. This concludes the proof of the first assertion. The second assertion follows from setting $\mathcal{C} := \mathcal{D}$. □

1.4 Compact systems

In topology, compact subsets of an underlying set play an important role; see Appendix A. Here, we introduce an important connection between compact sets and measure theory. The resulting compact systems play an important role in the proof of Theorem 2.10. Here it is shown that the σ-additivity of the set function follows from the additivity of a set function and an approximation property with respect to a compact system.

Definition 1.14 (Compact system). *A $\cap$-stable set system $\mathcal{K}$ is called* compact system *(on Ω) if $\bigcap_{n=1}^{\infty} K_n = \emptyset$ with $K_1, K_2, \ldots \in \mathcal{K}$ implies that there is a $N \in \mathbb{N}$ with $\bigcap_{n=1}^{N} K_n = \emptyset$.*

Example 1.15 (Compact sets). Compact sets form a compact system: *Let (Ω, r) be a metric space and $\mathcal{O}$ the topology generated by r. Then every $\cap$-stable $\mathcal{K} \subseteq \{K \subseteq \Omega : K \text{ compact}\}$ is a compact system.*
Indeed: let $\bigcap_{n=1}^{\infty} K_n = \emptyset$. Then both K_1 and $L_n := K_1 \cap K_n \subseteq K_1$ are closed for $n = 1, 2, \ldots$ according to Lemma A.8 and because of the compactness of K_1 there is an N with $\bigcap_{n=1}^{N} K_n = \emptyset$ according to Proposition A.9.

Lemma 1.16 (Extension of compact systems). *Let $\mathcal{K}$ be a compact system. Then*

$$\mathcal{K}_{\cup} := \Big\{ \bigcup_{i=1}^{n} K_i : K_1, \ldots, K_n \in \mathcal{K}, n \in \mathbb{N} \Big\}$$

is also a compact system.

Proof. It is clear that $\mathcal{K}_{\cup}$ is $\cap$-stable. Let $L_1 = \bigcup_{j=1}^{m_1} K_j^1, L_2 = \bigcup_{j=1}^{m_2} K_j^2, \ldots \in \mathcal{K}_{\cup}$ with $\bigcap_{n=1}^{N} L_n \neq \emptyset$ for all $N \in \mathbb{N}$. We have to show that $\bigcap_{n=1}^{\infty} L_n \neq \emptyset$. For this, we show the following:

> There is a sequence $K_1, K_2, ... \in \mathcal{K}$ with $K_n \subseteq L_n$ for all n, such that for every $N \in \mathbb{N}$ and $k \in \mathbb{N}_0$, we have mat$K_1 \cap \cdots \cap K_N \cap L_{N+1} \cap \cdots \cap L_{N+k} \neq \emptyset$.

The construction of the sequence is recursive. For $N = 1$, note that $\bigcup_{j=1}^{m_1} K_j^1 \cap \bigcap_{i=1}^{k} L_{1+i} \neq \emptyset$ for all $k \in \mathcal{N}_0$. So, there must be $j_1 \in \{1, ..., m_1\}$ with $K_{j_1}^1 \cap \bigcap_{i=1}^{k} L_{1+i} \neq \emptyset$ for all $k \in \mathbb{N}_0$. Set $K_1 := K_{j_1}^1$. (Otherwise, for all $j \in \{1, ..., m_1\}$, there is a k_j^* with $K_{j_1}^1 \cap \bigcap_{i=1}^{k_j^*} L_{1+i} = \emptyset$. Taking $k^* := \max_j k_j^*$ would lead to $\bigcup_{j=1}^{m_1} K_j^1 \cap \bigcap_{i=1}^{k^*} L_{1+i} = \emptyset$, in contradiction to the assumption.) For the recursive step, assume we have constructed the sequence up to K_{N-1}. Write

$$\begin{aligned} &K_1 \cap \cdots \cap K_{N-1} \cap \Big(\bigcup_{j=1}^{m_N} K_j^N \Big) \cap L_{N+1} \cap \cdots \cap L_{N+k} \\ &= \bigcup_{j=1}^{m_N} K_1 \cap \cdots \cap K_{N-1} \cap K_j^N \cap L_{N+1} \cap \cdots \cap L_{N+k} \neq \emptyset. \end{aligned}$$

Again, since this union is not empty, there is a j, so that $K_1 \cap \cdots \cap K_{N-1} \cap K_j^N \cap L_{N+1} \cap \cdots \cap L_{N+k} \neq \emptyset$ for all $k \in \mathbb{N}$. Set $K_N := K_j^N$, and we have constructed the sequence as claimed.

Now we set $k = 0$ in the above construction, and we see that $\bigcap_{n=1}^{N} K_n \neq \emptyset$ for all $N \in \mathbb{N}$. Since $\mathcal{K}$ is a compact system,

$$\emptyset \neq \bigcap_{n=1}^{\infty} K_n \subseteq \bigcap_{n=1}^{\infty} L_n$$

and the assertion is shown. □

2 Set functions

By a set function, we mean a function $m : \mathcal{A} \subseteq 2^\Omega \to \mathbb{R}$. The idea is that $m(A)$ for $A \in \mathcal{A}$ describes the volume of A. Here, *volume* might be an actual volume in space, or something more abstract. In probability theory we think of $m(A)$ as the probability that A occurs. (Mostly, we write $\mathbb{P}$ for the set function.) For any such set function, some requirements seem natural, irrespective of the meaning of *volume*. For example, the empty set (no spatial volume, or an event that never occurs) should be assigned *volume* 0, or m should behave countably additive, see (2.1). In probability theory, Ω consists of all possible outcomes of an experiment, so a natural requirement is $m(\Omega) = 1$. In other words, the probability that there is any outcome of the experiment is 1.

The concept of the probability measure is central to probability theory. As it turns out, measures must be defined on σ-algebras (so usually, $\mathcal{A}$ is a σ-algebra) so that the requirement of countable additivity can be met. In this section we give the most important steps to construct such measures. In *Analysis 3*, the Lebesgue measure was not introduced, which follows along the same lines. However, note that large parts of *Analysis* are dealing with $\Omega \subseteq \mathbb{R}^d$. In probability theory, however, outcomes of experiments might be elements of much larger spaces. When observing a randomly changing quantity (e.g. the position of a particle in space, or a stock price), we might need a probability measure in $\mathcal{C}([0,\infty), \mathbb{R}^d)$ (the set of continuous functions $X : [0,\infty) \to \mathbb{R}$).

2.1 Measures and outer measures

We will now consider functions $\mu : \mathcal{C} \to \overline{\mathbb{R}}_+$ if $\mathcal{C}$ is a semi-ring, ring or σ-algebra. Most important for probability theory is certainly the concept of a *probability measure*, which describes the special case $\mu(\Omega) = 1$.

Definition 2.1 (Measure and outer measure)**.** *For some $\mathcal{F} \subseteq 2^\Omega$, we call any $\mu : \mathcal{F} \to \overline{\mathbb{R}}_+$ a set function.*

1. *The set function μ is called* finitely additive *if for disjoint $A_1, \dots, A_n \in \mathcal{F}$ with $\biguplus_{k=1}^n A_k \in \mathcal{F}$,*

$$\mu\Big(\biguplus_{k=1}^n A_k\Big) = \sum_{k=1}^n \mu(A_k). \tag{2.1}$$

It is called sub-additive *if for (any, not necessarily disjoint) $A_1, \dots, A_n \in \mathcal{F}$ with $\bigcup_{k=1}^n A_k \in \mathcal{F}$,*

$$\mu\Big(\bigcup_{k=1}^n A_k\Big) \le \sum_{k=1}^n \mu(A_k). \tag{2.2}$$

2. *A mapping $\mu : \mathcal{F} \to \mathbb{R}_+$ is called* σ-additive *if* (2.1) *also holds for $n = \infty$. It is called* σ-sub-additive *if* (2.2) *also applies for $n = \infty$. It is called* monotonic *if for any $A, B \in \mathcal{F}$ with $A \subseteq B$ we find $\mu(A) \le \mu(B)$.*

3. *If $\mu(\Omega) < \infty$, then μ is called* finite*. If there is a sequence $\Omega_1, \Omega_2, \dots \in \mathcal{F}$ with $\bigcup_{n=1}^\infty \Omega_n = \Omega$ and $\mu(\Omega_n) < \infty$ for all $n = 1, 2, \dots$, then μ is called* σ-finite.

4. *Let $\mathcal{F}$ be a σ-algebra and $\mu : \mathcal{F} \to \mathbb{R}_+$. If μ is σ-additive, then μ is a* measure *(on $\mathcal{F}$) and $(\Omega, \mathcal{F}, \mu)$ is a* measure space. *If $\mu(\Omega) < \infty$, then μ is called* finite measure *and if $\mu(\Omega) = 1$, then μ is called a* probability measure *or a* probability distribution *or simply a* distribution. *Furthermore, $(\Omega, \mathcal{F}, \mu)$ is then called a* probability space.

5. *Let $(\Omega, \mathcal{O})$ be a topological space and μ a measure on $\mathcal{B}(\mathcal{O})$ (the Borel σ-algebra, see Definition 1.7). Then the smallest closed set F with $\mu(F^c) = 0$ is called the* support of μ[3].

6. *A σ-subadditive, monotone mapping $\mu^* : 2^\Omega \to \mathbb{R}_+$ is called* outer measure *if $\mu^*(\emptyset) = 0$. A set $A \subseteq \Omega$ is called* μ^*-measurable *if*

$$\mu(E) = \mu(E \cap A) + \mu(E \cap A^c) \tag{2.3}$$

for all $E \subseteq \Omega$.

7. *Let $\mathcal{F}$ be $\cap$-stable and $\mathcal{K} \subseteq \mathcal{F}$ a compact system. Then μ is called inner $\mathcal{K}-$regular if for all $A \in \mathcal{K}$*

$$\mu(A) = \sup_{\mathcal{K} \ni K \subseteq A} \mu(K).$$

Example 2.2 (Examples of set functions). 1. *We will often deal with set functions on $\mathcal{H} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$ from Example 1.3. For example, $\mu((a, b]) = b - a$ defines an additive, σ-finite set function on $\mathcal{H}$. We will extend this function uniquely to the Borel σ-algebra $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{H})$ (see lemma 1.9), which will give the Lebesgue measure, see corollary 2.18.*

2. *Another frequently used example is the* Dirac measures. *If $\omega' \in \Omega$, then*

$$\delta_{\omega'} : \begin{cases} 2^\Omega & \to \{0, 1\} \\ A & \mapsto 1_{\{\omega' \in A\}} \end{cases}$$

is a (probability) measure.

3. *If $\mu_i = \delta_{\omega_i}$, $i \in I$, then $\mu := \sum_{i \in I} \delta_{\omega_i}$ is called a* counting measure.

4. *If $\mu_i, i \in I$ are measures on a σ-algebra $\mathcal{F}$. Then for $a_i \in \mathbb{R}_+, i \in I$, $\sum_{i \in I} a_i \mu_i$ is also a measure. Examples of this are well known from the lecture* Elementary probability theory. *There, for example, with $\mathcal{F} = 2^{\mathbb{N}_0}$ and δ_k as in 2.*

$$\mu_{Poi(\gamma)} := \sum_{k=0}^{\infty} e^{-\gamma} \frac{\gamma^k}{k!} \cdot \delta_k$$

the Poisson distribution on $2^{\mathbb{N}_0}$ with parameter γ,

$$\mu_{geo(p)} := \sum_{k=1}^{\infty} (1 - p)^{k-1} p \cdot \delta_k$$

[3] We will see later that this smallest set indeed exists uniquely.

the geometric distribution with success parameter p and

$$\mu_{B(n,p)} := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot \delta_k$$

the binomial distribution $B(n,p)$.

Remark 2.3 (Contents and premeasures)**.** *Finite additive set functions are often called* content*, σ-additive set functions that are not defined on σ-algebras are often called premeasures. The measures defined on a Borel σ-algebra that are regular with respect to the compact sets from the inside are called Radon measures. We will not use these terms.*

Lemma 2.4 (Unions written as disjoint unions)**.** *Let $\mathcal{H}$ be a semi-ring, and $A, A_1, ..., A_n \in \mathcal{H}$. Then, there are $m \in \mathbb{N}$ and $B_1, ..., B_m \in \mathcal{H}$ pairwise disjoint and $A \setminus \bigcup_{i=1}^n A_i = \biguplus_{j=1}^m B_j$.*

Proof. We proceed by induction on n. If $n = 1$, the assertion is true by the definition of a semi-ring. Assume the assertion holds for some n, i.e. there is $m \in \mathbb{N}$ and $B_1, ..., B_m$ with $A \setminus \bigcup_{i=1}^n A_i = \biguplus_{j=1}^m B_j$. Then, we can write $B_j \setminus A_{n+1} = \biguplus_{k=1}^{k_j} C_k^j$ for $C_1^j, ..., C_{k_j}^j \in \mathcal{H}$. Then, write

$$A \setminus \bigcup_{i=1}^{n+1} A_i = \Big(A \setminus \bigcup_{i=1}^{n} A_i\Big) \setminus A_{n+1} = \biguplus_{j=1}^{m} B_j \setminus A_{n+1} = \biguplus_{j=1}^{m} \biguplus_{k=1}^{k_j} C_k^j.$$

This concludes the proof, since the latter disjoint union is over a finite set. $\square$

Lemma 2.5 (Set-functions on semi-rings)**.** *Let $\mathcal{H}$ be a semi-ring and $\mu : \mathcal{H} \to [0, \infty]$ additive. Then, m is monotone and sub-additive. In addition, μ is σ-additive iff it is σ-sub-additive.*

Proof. We start by monotonicity. Let $A, B \in \mathcal{H}$ with $A \subseteq B$ and $C_1, ..., C_k \in \mathcal{H}$ with $B \setminus A = \biguplus_{i=1}^k C_i$. Therefore, we can write $\mu(A) \leq \mu(A) + \sum_{i=1}^k \mu(C_i) = \mu(B)$.

Next, we claim that for $A \in \mathcal{H}$ and $A_1, ..., A_n \in \mathcal{H}$ disjoint with $\biguplus_{I \in \mathcal{I}} A_i \subseteq A$, we have $\sum_{i=1}^n \mu(A_i) \leq m(A)$. For this, write $A \setminus \biguplus_{i=1}^n A_i = \biguplus_{j=1}^m B_j$ as in Lemma 2.4. Then,

$$\mu(A) = \mu\Big(\biguplus_{i=1}^{n} A_i \uplus \biguplus_{j=1}^{m} B_j\Big) = \sum_{i=1}^{n} \mu(A_i) + \sum_{j=1}^{m} \mu(B_j) \geq \sum_{i=1}^{n} \mu(A_i). \tag{2.4}$$

For sub-additivity, let $A_1, ..., A_n \in \mathcal{H}$ with $\bigcup_{i=1}^n A_i \in \mathcal{H}$. We need to show $\mu\Big(\bigcup_{i=1}^n A_i\Big) \leq \sum_{i=1}^n \mu(A_i)$. For $i = 2, ..., n$, we write

$$\bigcup_{i=1}^{n} A_i = \biguplus_{i=1}^{n} \Big(A_i \setminus \bigcup_{j=1}^{i-1} A_j\Big) = \biguplus_{i=1}^{n} \biguplus_{k=1}^{k_i} C_k^i$$

with C_k^i as in Lemma 2.4. So, since $\biguplus_{k=1}^{k_i} C_k^i \subseteq A_i \in \mathcal{H}$,

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} \sum_{k=1}^{k_i} \mu(C_k^i) \leq \sum_{i=1}^{n} \mu(A_i).$$

Now, we show that μ is σ-additive $\iff$ it is σ-sub-additive.

'⇒': Here, just copy the proof of sub-additivity, but using $n = \infty$. For '⇐', let $A_1, A_2, ... \in \mathcal{H}$ be pairwise disjoint with $A = \biguplus_{i=1}^{\infty} A_i \in \mathcal{H}$. Since μ is monotone and for any $n \in \mathbb{N}$, we have $\biguplus_{i=1}^{n} A_i \subseteq A$ (hence $\sum_{i=1}^{n} \mu(A_i) \leq \mu(A)$ by (2.4)),

$$\sum_{i=1}^{\infty} \mu(A_i) = \sup_{n\in\mathbb{N}} \sum_{i=1}^{n} \mu(A_i) \leq \mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

by σ-sub-additivity. So, σ-additivity follows. □

Lemma 2.6 (Extension of set functions on semi-rings)**.** *Let $\mathcal{H}$ be a semi-ring, $\mathcal{R}$ the ring generated by $\mathcal{H}$ from Lemma 1.5 and μ an additive function on $\mathcal{H}$. Define $\widetilde{\mu}$ on $\mathcal{R}$ by*

$$\widetilde{\mu}\Big(\biguplus_{i=1}^{n} A_i\Big) := \sum_{i=1}^{n} \mu(A_i)$$

for $A_1, \ldots, A_n \in \mathcal{H}$ disjoint. Then $\widetilde{\mu}$ is the only additive extension of μ on $\mathcal{R}$ that coincides with μ on $\mathcal{H}$. Moreover, $\widetilde{\mu}$ is σ-additive if and only if μ is σ-additive.

Proof. We only need to show that $\widetilde{\mu}$ is well-defined. All other properties follow by definition of $\widetilde{\mu}$. So, let $A_1, \ldots, A_m, B_1, \ldots, B_n \in \mathcal{H}$ with $\biguplus_{i=1}^{m} A_i = \biguplus_{j=1}^{n} B_j$. Since

$$A_i = \biguplus_{j=1}^{n} A_i \cap B_j, \qquad B_j = \biguplus_{i=1}^{m} A_i \cap B_j,$$

we write using additivity of $\widetilde{\mu}$

$$\sum_{i=1}^{m} \mu(A_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mu(A_i \cap B_j) = \sum_{j=1}^{n} \sum_{i=1}^{m} \mu(A_i \cap B_j) = \sum_{j=1}^{n} \mu(B_j).$$

□

Proposition 2.7 (Inclusion-exclusion principle)**.** *Let μ be an additive set function on a ring $\mathcal{R}$ and I be finite. Then for $A_i \in \mathcal{R}$, $i \in I$, it holds that*

$$\mu\Big(\bigcup_{i\in I} A_i\Big) = \sum_{J\subseteq I} (-1)^{|J|+1} \mu\Big(\bigcap_{j\in J} A_j\Big)$$

In particular, if $I = \{1, 2\}$,

$$\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2) - \mu(A_1 \cap A_2)$$

and if $I = \{1, 2, 3\}$,

$$\begin{aligned}\mu(A_1 \cup A_2 \cup A_3) = \mu(A_1) &+ \mu(A_2) + \mu(A_3) \\ &- \mu(A_1 \cap A_2) - \mu(A_1 \cap A_3) - \mu(A_2 \cap A_3) + \mu(A_1 \cap A_2 \cap A_3).\end{aligned}$$

Proof. We use induction over $|I|$. For $|I| = 2$ the assertion is clear because $A_1 \cup A_2 = A_1 \uplus (A_2 \setminus A_1)$ and $(A_2 \setminus A_1) \uplus (A_1 \cap A_2) = A_2$. Assume it applies to all I with $|I| = n$, and consider some I with $|I| = n+1$. Without loss of gnerality,we write $I = \{1, \dots, n+1\}$. By additivity of μ

$$\begin{aligned}
\mu\Big(\bigcup_{i=1}^{n+1} A_i\Big) &= \mu\Big(\bigcup_{i=1}^{n}(A_i \cup A_{n+1})\Big) \\
&= \sum_{\emptyset \neq J \subseteq \{1,\dots,n\}} (-1)^{|J|+1} \mu\Big(A_{n+1} \cup \bigcap_{j \in J} A_j)\Big) \\
&= \sum_{\emptyset \neq J \subseteq \{1,\dots,n\}} (-1)^{|J|+1} \Big(\mu(A_{n+1}) + \mu\Big(\bigcap_{j \in J} A_j)\Big) - \mu\Big(\bigcap_{j \in J} A_j \cap A_{n+1})\Big)\Big) \\
&= \mu(A_{n+1}) + \sum_{\emptyset \neq J \subseteq \{1,\dots,n\}} (-1)^{|J|+1} \Big(\mu\Big(\bigcap_{j \in J} A_j)\Big) - \mu\Big(\bigcap_{j \in J} A_j \cap A_{n+1})\Big)\Big) \\
&= \sum_{J \subseteq \{1,\dots,n+1\}} (-1)^{|J|+1} \mu\Big(\bigcap_{j \in J} A_j)\Big).
\end{aligned}$$

□

2.2 σ-additivity

The finite additivity of set functions is a requirement that can often be verified. The situation is different with σ-additivity. We will now look at alternative formulations for σ-additivity.

Proposition 2.8 (Continuity of from below and from above)**.** *Let $\mathcal{R}$ be a ring and $\mu : \mathcal{R} \to \overline{\mathbb{R}}_+$ be additive. Consider the following properties:*

1. *μ is σ-additive;*
2. *μ is σ-subadditive;*
3. *μ is continuous from below, i.e. for $A, A_1, A_2, \dots \in \mathcal{R}$ and $A_1 \subseteq A_2 \subseteq \dots$ with $A = \bigcup_{n=1}^\infty A_n$ we have $\mu(A) = \lim_{n\to\infty} \mu(A_n)$;*
4. *μ is continuous from above in $\emptyset$, i.e. for $A_1, A_2, \dots \in \mathcal{R}$, $\mu(A_1) < \infty$ and $A_1 \supseteq A_2 \supseteq \dots$ with $\bigcap_{n=1}^\infty A_n = \emptyset$ we have $\lim_{n\to\infty} \mu(A_n) = 0$.*
5. *μ is continuous from above, i.e. for $A, A_1, A_2, \dots \in \mathcal{R}$, $\mu(A_1) < \infty$ and $A_1 \supseteq A_2 \supseteq \dots$ with $A = \bigcap_{n=1}^\infty A_n$ we have $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.*

Then,

$$1. \iff 2. \iff 3. \implies 4. \iff 5.$$

Furthermore, 4. $\implies$ 3. *holds if $\mu(A) < \infty$ for all $A \in \mathcal{R}$.*

Proof. 1.⇔2. follows from Lemma 2.6, since $\mathcal{R}$ is a semi-ring.

1.⇒3.: Let μ be σ-additive and $A, A_1, A_2, \dots \in \mathcal{R}$ as in 3. Then, with $A_0 = \emptyset$,

$$\mu(A) = \sum_{n=1}^\infty \mu(A_n \setminus A_{n-1}) = \lim_{N\to\infty} \sum_{n=1}^N \mu(A_n \setminus A_{n-1}) = \lim_{N\to\infty} \mu(A_N).$$

3.⇒1.: Let $B_1, B_2, \dots \in \mathcal{R}$ be pairwise disjoint and $B = \biguplus_{n=1}^\infty B_n \in \mathcal{R}$. Then, for $A_N = \biguplus_{n=1}^N B_n$,

$$\mu(B) = \lim_{N\to\infty} \mu(A_N) = \sum_{n=1}^\infty \mu(B_n).$$

4.⇒5.: Let $A, A_1, A_2, \dots \in \mathcal{R}$ be as assumed in 5. Further, let $B_n := A_n \setminus A$. Then $B_1, B_2, \dots$ fulfills the conditions of 4., so $\mu(B_n) \xrightarrow{n\to\infty} 0$, i.e. $\mu(A_n) = \mu(B_n) + \mu(A) \xrightarrow{n\to\infty} \mu(A)$.

5.⇒4.: is clear.

3.⇒4.: Let $A_1, A_2, \dots \in \mathcal{R}$ be as assumed in 4. Set $B_n := A_1 \setminus A_n, n \in \mathbb{N}$. Then $B = A_1, B_1, B_2, \dots \in \mathcal{R}$ fulfills the conditions of 3, and thus $\mu(A_1) = \lim_{n\to\infty} \mu(B_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n)$, from which 4. follows.

4.⇒3. if $\mu(A) < \infty$ for all $A \in \mathcal{R}$. Let $A, A_1, A_2, \dots \in \mathcal{R}$ be as assumed in 3. Set $B_n := A \setminus A_n \in \mathcal{R}, n \in \mathbb{N}$. Then $\bigcap_{n=1}^\infty B_n = \emptyset$, i.e. $0 = \lim_{n\to\infty} \mu(B_n) = \mu(A) - \lim_{n\to\infty} \mu(A_n)$, from which 3. follows. Here, the last equality uses the condition that $\mu(A) < \infty$. □

We now want to take a closer look at set functions which are inner regular with respect to a compact system. For measures, inner regularity with respect to the system of all compact sets (which is a compact system due to Example 1.15) is fulfilled on Polish spaces, as the next result shows. This will play an important role in the theory of weak convergence, an important concept in any course on probability theory.

Lemma 2.9. *If $(\Omega, \mathcal{O})$ is Polish and μ is a finite measure on $\mathcal{B}(\mathcal{O})$, then for every $\varepsilon > 0$ there exists a compact set $K \subseteq \Omega$ with $\mu(\Omega \setminus K) < \varepsilon$.*

Proof. First, note that compact sets are closed according to Lemma A.8, so all compact sets are in $\mathcal{B}(\mathcal{O})$ and thus $\mu(\Omega \setminus K)$ is well-defined.

Let $\varepsilon > 0$. Since Ω is separable (see Definition A.1), there is a countable set $\{\omega_1, \omega_2, \dots\} \subseteq \Omega$ which is dense. In particular, for all n, we find $\Omega = \bigcup_{k=1}^\infty B_{1/n}(\omega_k)$. Since μ is continuous from above (Proposition 2.8),

$$0 = \mu\Big(\Omega \setminus \bigcup_{k=1}^\infty B_{1/n}(\omega_k)\Big) = \lim_{N\to\infty} \mu\Big(\Omega \setminus \bigcup_{k=1}^N B_{1/n}(\omega_k)\Big).$$

Thus there is an $N_n \in \mathbb{N}$ with $\mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)\Big) < \varepsilon/2^n$. Now, consider

$$A := \bigcap_{n=1}^\infty \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k).$$

By definition, this set is totally bounded (i.e. for all radii $\varepsilon > 0$. it can be covered by a finite number of balls of radius $\varepsilon > 0$. Hence, according to Lemma A.9, A is relatively compact. Furthermore, (recall that $\overline{A}$ is the closure of A, which is compact according to Proposition A.9),

$$\mu(\Omega \setminus \overline{A}) \le \mu(\Omega \setminus A) \le \mu\Big(\bigcup_{n=1}^\infty \Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k)\Big)\Big) \le \sum_{n=1}^\infty \mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k)\Big) < \varepsilon.$$

This proves the assertion. □

Theorem 2.10 (Inner regular additive set functions are $\boldsymbol{\sigma}$-additive)**.** *Let $\mathcal{H}$ be a semi-ring and $\mu : \mathcal{H} \to \mathbb{R}_+$ finite, finitely additive and inner regular with respect to a compact system $\mathcal{K} \subseteq \mathcal{H}$. Then μ is σ-additive.*

Proof. As in Lemma 2.6, the set function μ can be uniquely extended to the ring $\mathcal{R}(\mathcal{H})$ generated by $\mathcal{H}$ (see Lemma 1.5). Furthermore, according to Lemma 1.16, the system $\mathcal{K}_\cup \subseteq \mathcal{R}(\mathcal{H})$, which consists of unions of sets in $\mathcal{K}$, is also compact. Choose $\varepsilon > 0$ and $A = \bigcup_{i=1}^n A_i \in \mathcal{R}(\mathcal{H})$ with $A_1, \dots, A_n \in \mathcal{H}$, then there are compact sets $K_1, \dots, K_n \in \mathcal{K} \subseteq \mathcal{H}$ with $\mu(A_i) \le \mu(K_i) + \frac{\varepsilon}{n}$ for $i = 1, \dots, n$. This means that the extension of μ to the ring $\mathcal{R}(\mathcal{H})$ is inner regular with respect to $\mathcal{K}_\cup$, since

$$\mu\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n \mu(A_i) \le \Big(\sum_{i=1}^n \mu(K_i)\Big) + \varepsilon = \mu\Big(\bigcup_{i=1}^n K_i\Big) + \varepsilon.$$

This means that μ is $\mathcal{K}_\cup$-regular from the inside. o, without loss of generality, we assume that $\mathcal{H}$ is a ring and $\mathcal{K}$ is $\cup$-stable. We now show that μ is continuous from above in $\emptyset$. This is sufficient according to Proposition 2.8 because of the finiteness of μ on $\mathcal{H}$. Let $A_1, A_2, \dots \in \mathcal{H}$ with $A_1 \supseteq A_2 \supseteq \cdots$ and $\bigcap_{n=1}^\infty A_n = \emptyset$ and $\varepsilon > 0$. Choose $K_1, K_2, \dots \in \mathcal{K}$ with $K_n \subseteq A_n, n \in \mathbb{N}$ and

$$\mu(A_n) \le \mu(K_n) + \varepsilon 2^{-n}.$$

Then, $\bigcap_{n=1}^\infty K_n \subseteq \bigcap_{n=1}^\infty A_n = \emptyset$, which means that there is a $N \in \mathbb{N}$ with $\bigcap_{n=1}^N K_n = \emptyset$ since $\mathcal{K}$ is a compact system. From this,

$$A_N = A_N \cap \Big(\bigcup_{n=1}^N K_n^c\Big) = \bigcup_{n=1}^N A_N \setminus K_n \subseteq \bigcup_{n=1}^N A_n \setminus K_n.$$

Due to the subadditivity and the monotonicity of μ for all $m \ge N$, it follows that

$$\mu(A_m) \le \mu(A_N) \le \sum_{n=1}^N \mu(A_n \setminus K_n) \le \varepsilon \sum_{n=1}^N 2^{-n} \le \varepsilon.$$

This shows the assertion, since $\varepsilon > 0$ was arbitrary. □

2.3 Uniqueness and extension of set functions

Suppose an additive set function $\mu : \mathcal{H} \to \overline{\mathbb{R}}_+$ is given, where $\mathcal{H}$ is a semi-ring. We are concerned with the extension of μ to a measure (i.e. an σ-additive set function) to $\sigma(\mathcal{H})$. The aim is to establish conditions when the measure is already uniquely given by μ. The result is summarised in Theorem 2.16. See also Table 2 for an overview of how the results of previous chapters relate to this.

Proposition 2.11 (Uniqueness of measures)**.** *Let $\mathcal{F}$ be a σ-algebra and $\mu, \nu : \mathcal{F} \to \mathbb{R}_+$ measures. Let $\mathcal{H}$ be a $\cap$-stable set system such that $\sigma(\mathcal{H}) = \mathcal{F}$ and $\mu|_\mathcal{H}, \nu|_\mathcal{H}$ are σ-finite. Then $\mu = \nu$ if and only if $\mu(A) = \nu(A)$ holds for all $A \in \mathcal{H}$.*

Corollary 2.12 (Uniqueness of probability measures)**.** *Let $\mathcal{F}$ be a σ-algebra and $\mu, \nu : \mathcal{F} \to [0,1]$ be probability measures. Let $\mathcal{H}$ be a $\cap$-stable set system with $\sigma(\mathcal{H}) = \mathcal{F}$. Then $\mu = \nu$ if and only if $\mu(A) = \nu(A)$ holds for all $A \in \mathcal{H}$.*

	Lemma 2.5	Theorem 2.10	Theorem 2.16
μ additive	$\circ$	$\circ$	
μ finite		$\circ$	
μ σ-finite			$\circ$
μ defined on semi-ring	$\circ$	$\circ$	$\circ$
μ σ-additive	$\circ/\bullet$	$\bullet$	$\circ$
μ σ-subadditive	$\bullet/\circ$		
μ inner regular wrt a compact system		$\circ$	
μ extends uniquely to $\sigma(\mathcal{H})$			$\bullet$

Table 2: Lemma 2.5 and theorem 2.10 play significant roles in the application of Carathéodory's extension theorem. In the table, the $\circ$'s represent the assumptions of the theorem and $\bullet$ the conclusions. As can easily be seen, Carathéodory's extension theorem applies, for example, if μ is finite and inner regular with respect to a compact system.

Proof. Wlog, $\Omega \in \mathcal{H}$, since $\mu(\Omega) = \nu(\Omega) = 1$. This means that μ and ν are in particular σ-finite and the statement follows from Proposition 2.11. □

Proof of Proposition 2.11. The 'only if' direction is clear. For the 'if' direction, we set for $A \in \mathcal{H}$ with $\mu(A) = \nu(A) < \infty$

$$\mathcal{D}_A := \{B \in \mathcal{F} : \mu(A \cap B) = \nu(A \cap B)\} \supseteq \mathcal{H}.$$

We show that $\mathcal{D}_A$ is a Dynkin system. It is clear that $\Omega \in \mathcal{D}_A$. Furthermore, if $B, C \in \mathcal{D}_A$ and $B \subseteq C$, then $\mu((C \setminus B) \cap A) = \mu(C \cap A) - \mu(B \cap A) = \nu(C \cap A) - \nu(B \cap A) = \nu((C \setminus B) \cap A)$, i.e. $C \setminus B \in \mathcal{D}_A$. If $B_1, B_2, \dots \in \mathcal{D}$ with $B_1 \subseteq B_2 \subseteq B_3 \subseteq \dots \in \mathcal{D}_A$ and $B = \bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$, then because of Proposition 2.8,

$$\mu(B \cap A) = \lim_{n\to\infty} \mu(B_n \cap A) = \lim_{n\to\infty} \nu(B_n \cap A) = \nu(B \cap A),$$

which implies $B \in \mathcal{D}_A$. This means that $\mathcal{D}_A$ is a Dynkin system for all $A \in \mathcal{H}$ with $\mu(A) < \infty$ and thus, due to Theorem 1.13, $\mathcal{F} = \sigma(\mathcal{H}) \subseteq \mathcal{D}_A$. Let $\Omega_1, \Omega_2, \dots \in \mathcal{H}$ with $\Omega_n \uparrow \Omega$ and $\mu(\Omega_n), \nu(\Omega_n) < \infty, n = 1, 2, \dots$ Then for all $n = 1, 2, \dots$ it holds that $\mu(B \cap \Omega_n) = \nu(B \cap \Omega_n)$ for all $B \in \mathcal{F}$. This implies $B \in \mathcal{F}$, since μ and ν are continuous from below, thus

$$\mu(B) = \lim_{n\to\infty} \mu(B \cap \Omega_n) = \lim_{n\to\infty} \nu(B \cap \Omega_n) = \nu(B),$$

i.e. $\mu = \nu$. □

The following theorem explains why the notion of a σ-algebra is so important.

Theorem 2.13 (**$\boldsymbol{\mu^*}$-measurable sets are a $\boldsymbol{\sigma}$-algebra**)**.** *Let μ^* be an outer measure on Ω and $\mathcal{F}^*$ the set of μ^*-measurable sets; recall from* (2.3)*. Then $\mathcal{F}^*$ is a σ-algebra and $\mu := \mu^*|_{\mathcal{F}^*}$ is a measure. Furthermore, $\mathcal{N} := \{N \subseteq \Omega : \mu^*(N) = 0\} \subseteq \mathcal{F}^*$.*

Remark 2.14 (Null-sets and properties almost everywhere)**.** *Sets $N \subseteq \Omega$ with $\mu(N) = 0$ are called* (μ-)null sets*. We further say that $A \subseteq \Omega$ (μ)-*almost everywhere *holds if $A^c \in \mathcal{N}$. If μ is a probability measure, we say* almost surely *instead of 'almost everywhere'.*

Proof of Theorem 2.13. We first show that $\mathcal{F}^*$ is a σ-algebra. It is clear that

$$\mu^*(E) = \mu^*(\emptyset) + \mu^*(E) = \mu^*(E \cap \emptyset) + \mu^*(E \cap \Omega),$$

i.e. $\emptyset \in \mathcal{F}^*$. It is also clear that $A^c \in \mathcal{F}^*$ follows from $A \in \mathcal{F}^*$. Next, let us show that $\mathcal{F}^*$ is $\cap$-stable. For $A, B, E \subseteq \Omega$, note that $(E \cap A \cap B^c) \uplus (E \cap A^c) = E \cap (A \cap B)^c$. So, using the sub-additivity of μ^*, for $A, B \in \mathcal{F}^*$,

$$\begin{aligned}\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) &= \mu^*((E \cap A) \cap B) + \mu^*((E \cap A) \cap B^c) + \mu^*(E \cap A^c) \\ &\geq \mu^*(E \cap (A \cap B)) + \mu^*(E \cap (A \cap B)^c) \geq \mu^*(E),\end{aligned}$$

and we have shown $A \cap B \in \mathcal{F}^*$. Now let $A_1, A_2, \dots \in \mathcal{F}^*$ be disjoint and $B_n = \biguplus_{k=1}^n A_k$ and $B = \bigcup_{n=1}^\infty B_n = \biguplus_{k=1}^\infty A_k$. Since $\mathcal{F}^*$ is $\cap$- and complement stable, it is also $\cup$-stable, so $B_1, B_2, ... \in \mathcal{F}^*$. We further show that $\mu^*(E \cap B_n) = \sum_{k=1}^n \mu^*(E \cap A_k)$ applies to all $E \subseteq \Omega$. For $n = 1$ this is clear, and if it applies to n, it follows that

$$\begin{aligned}\mu^*(E \cap B_{n+1}) &= \mu^*(E \cap B_{n+1} \cap B_n) + \mu^*(E \cap B_{n+1} \cap B_n^c) \\ &= \mu^*(E \cap B_n) + \mu^*(E \cap A_{n+1}) = \sum_{k=1}^{n+1} \mu^*(E \cap A_k).\end{aligned}$$

Therefore, since μ^* is sub-additive and monotone,

$$\mu^*(E \cap B) \leq \sum_{k=1}^\infty \mu^*(E \cap A_k) = \lim_{n\to\infty} \sum_{k=1}^n \mu^*(E \cap A_k) = \lim_{n\to\infty} \mu^*(E \cap B_n) \leq \mu^*(E \cap B),$$

thus

$$\mu^*(E \cap B) = \lim_{n\to\infty} \mu^*(E \cap B_n) = \sum_{k=1}^\infty \mu^*(E \cap A_k). \tag{2.5}$$

Next, we show that $B \in \mathcal{F}^*$, which implies that $\mathcal{F}^*$ is a σ-algebra. For any $E \subseteq \Omega$, since $B_1, B_2, ... \in \mathcal{F}^*$, (2.5) holds,

$$\mu^*(E) = \lim_{n\to\infty} \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B) + \mu^*(E \cap B^c) \geq \mu^*(E).$$

So, $B \in \mathcal{F}^*$ follows. Furthermore, it follows from (2.5) that μ^* is σ-additive, i.e. $\mu = \mu^*|_{\mathcal{F}^*}$ is a measure.

Now let $N \subseteq \Omega$ be such that $\mu^*(N) = 0$ and $E \subseteq \Omega$. Then, due to the monotonicity of μ^*, $\mu^*(E \cap N) = 0$, i.e.

$$\mu^*(E \cap N^c) + \mu^*(E \cap N) \geq \mu^*(E) \geq \mu^*(E \cap N^c) = \mu^*(E \cap N^c) + \mu^*(E \cap N)$$

and therefore $N \in \mathcal{F}^*$. □

Proposition 2.15 (Outer measure generated by finite additive set function)**.** *Let $\mathcal{H}$ be a semi-ring and $\mu : \mathcal{H} \to \mathbb{R}_+$ additive. For $A \subseteq \Omega$ let*

$$\mu^*(A) := \inf_{\mathcal{G}\in\mathcal{U}(A)} \sum_{G\in\mathcal{G}} \mu(G),$$

where

$$\mathcal{U}(A) := \big\{\mathcal{G} \subseteq \mathcal{H} \text{ at most countable, } A \subseteq \bigcup_{G\in\mathcal{G}} G\big\}$$

is the set of at most countable covers of A and $\mu^*(A) = \infty$ if $\mathcal{U}(A) = \emptyset$. Then μ^* is an outer measure.

Proof. The mapping μ^* is monotone (by definition) with $\mu^*(\emptyset) = 0$ (note that $\emptyset \in \mathcal{H}$ and, using finite additivity of $\mu(\emptyset) = \mu(\emptyset) + \mu(\emptyset)$, from which $\mu(\emptyset) = 0$ follows). To check the σ-sub-additivity of μ^*, we choose $A_1, A_2, \dots \subseteq \Omega$. For $n = 1, 2, \dots$ and $\epsilon > 0$ there are sets $G_{nk} \in \mathcal{H}$, $k \in \mathcal{K}_n$ at most countable with

$$A_n \subseteq \bigcup_{k\in\mathcal{K}_n} G_{nk},$$
$$\mu^*(A_n) \geq \sum_{k\in\mathcal{K}_n} \mu(G_{nk}) - \varepsilon 2^{-n}.$$

Since $\bigcup_{n=1}^\infty A_n \subseteq \bigcup_{n=1}^\infty \bigcup_{k\in\mathcal{K}_n} G_{nk}$, and by the monotonicity of μ^* and the definition of μ^*,

$$\mu^*\Big(\bigcup_{n=1}^\infty A_n\Big) \leq \sum_{n=1}^\infty \sum_{k\in\mathcal{K}_n} \mu(G_{nk}) \leq \varepsilon + \sum_{n=1}^\infty \mu^*(A_n).$$

With $\varepsilon \to 0$ the σ-sub-additivity of μ^* follows, i.e. μ^* is an outer measure. □

Theorem 2.16 (Extension of a $\boldsymbol{\sigma}$-additive set function)**.** *Let $\mathcal{H}$ be a semi-ring and $\mu : \mathcal{H} \to \mathbb{R}_+$ σ-finite and σ-additive. Furthermore, let $\widetilde{\mu} = \mu^*|_{\mathcal{F}^*}$ with μ^* from Proposition 2.15 and $\mathcal{F}^*$ from Theorem 2.13. Then $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$ and $\widetilde{\mu}|_{\sigma(\mathcal{H})}$ is the only measure that agrees with μ on $\mathcal{H}$.*

Proof. First we note that μ is both finitely additive and σ-subadditive according to Lemma 2.5. According to Proposition 2.15, μ^* is an outer measure and according to Theorem 2.13, $\mathcal{F}^*$ is a σ-algebra.

Step 1: μ^* coincides with μ on $\mathcal{H}$: Let $H \in \mathcal{H}$. Choose $\mathcal{K}$ at most countable and $H_k \in \mathcal{H}$, $k \in \mathcal{K}$ with $H \subseteq \bigcup_{k\in\mathcal{K}} H_k$ and

$$\mu^*(H) \geq \sum_{k\in\mathcal{K}} \mu(H_k) - \varepsilon.$$

Then, because of $H = \bigcup_{k\in\mathcal{K}} H_k \cap H$ and the σ-sub-additivity of μ

$$\mu^*(H) \leq \mu(H) \leq \sum_{k\in\mathcal{K}} \mu(H_k \cap H) \leq \sum_{k\in\mathcal{K}} \mu(H_k) \leq \mu^*(H) + \varepsilon,$$

where we have used the σ-additivity of μ in the second '$\leq$'. With $\varepsilon \to 0$, we find $\mu^*(H) = \mu(H)$.

Step 2: $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$: Let $E \subseteq \Omega, H \in \mathcal{H}$ and $\varepsilon > 0$. Choose $\mathcal{K}$ at most countable and $H_k \in \mathcal{H}, k \in \mathcal{K}$ with $E \subseteq \bigcup_{k\in\mathcal{K}} H_k$ and $\mu^*(E) \geq \sum_{k\in\mathcal{K}} \mu(H_k) - \varepsilon$. Then, due to σ-additivity of μ

$$\mu^*(E) + \varepsilon \geq \sum_{k\in\mathcal{K}} \mu(H_k) = \sum_{k\in\mathcal{K}} \mu(H_k \cap H) + \sum_{k\in\mathcal{K}} \mu(H_k \cap H^c) \geq \mu^*(E \cap H) + \mu^*(E \cap H^c).$$

With $\varepsilon \to 0$ and the σ-sub-additivity of μ^*, $\mu^*(E) = \mu^*(E \cap H) + \mu^*(E \cap H^c)$, i.e. H is μ^*-measurable and therefore $\mathcal{H} \subseteq \mathcal{F}^*$. Since $\mathcal{F}^*$ is a σ-algebra according to Theorem 2.13, $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$.

Step 3: Uniqueness: According to Theorem 2.13, $\widetilde{\mu}$ is a measure. Since $\widetilde{\mu}$ coincides with μ on $\mathcal{H}$, which in turn coincides with μ^* on $\mathcal{H}$, we find $\widetilde{\mu}|_{\sigma(\mathcal{H})} = \mu^*|_{\sigma(\mathcal{H})}$. Let $\nu : \sigma(\mathcal{H}) \to \mathbb{R}_+$ another measure that is equal to μ on $\mathcal{H}$. Since $\mu = \widetilde{\mu}|_{\mathcal{H}}$ was assumed to be σ-finite, $\nu|_{\mathcal{H}}$ is also σ-finite. With Proposition 2.11 it follows that $\widetilde{\mu} = \nu$ applies to $\sigma(\mathcal{H})$ due to the $\cap$-stability of $\mathcal{H}$.

Now all assertions are proven. □

The above theorem only makes it clear that $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$. The next result shows how sets in $\mathcal{F}^*$ differ from sets in $\sigma(\mathcal{H})$.

Proposition 2.17 (Characterisation of $\boldsymbol{\mathcal{F}^*}$ from Proposition 2.15)**.** *Let $\mathcal{H}$ be a semi-ring, $\mu : \mathcal{H} \to \mathbb{R}_+$ σ-finite and σ-additive, μ^* as in Proposition 2.15 and $\mathcal{F}^*, \mathcal{N}$ as in Theorem 2.13. Then*

$$\mathcal{F}^* = \{A \setminus N : A \in \sigma(\mathcal{H}), N \in \mathcal{N}\}.$$

In particular, the right-hand side is a σ-algebra.

Proof. '$\supseteq$': On the one hand we have $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$ according to theorem 2.16, on the other hand, there is $\mathcal{N} \subseteq \mathcal{F}^*$ from Theorem 2.13. This implies '$\supseteq$', since $\mathcal{F}^*$ is complement stable.

'$\subseteq$': Let $B \in \mathcal{F}^*$. Further, let $\Omega_1, \Omega_2, \dots \in \mathcal{H}$ with $\mu(\Omega_n) < \infty, n = 1, 2, \dots$ and $\Omega = \bigcup_{n=1}^\infty \Omega_n$. Let $\varepsilon_1, \varepsilon_2, \dots > 0$ with $\varepsilon_i \downarrow 0$. For $B_n := B \cap \Omega_n$ and $i = 1, 2, ...$, we choose $\mathcal{K}_{ni}$ at most countable, $A_{nik} \in \mathcal{H}, n \in \mathbb{N}, k \in \mathcal{K}_{ni}$, $B_n \subseteq \bigcup_{k\in\mathcal{K}_{ni}} A_{nik}$ and

$$\mu^*(B_n) \geq \sum_{k\in\mathcal{K}_{ni}} \mu(A_{nik}) - 2^{-n}\varepsilon_i.$$

Clearly, $A_i := \bigcup_{n=1}^\infty \bigcup_{k\in\mathcal{K}_{ni}} A_{nik} \in \sigma(\mathcal{H})$, $B \subseteq A_i$ for all $i = 1, 2, ...$ and $A_i \setminus B = \bigcup_{n=1}^\infty \bigcup_{k\in\mathcal{K}_{ni}} A_{nik} \setminus B_n$. This means that

$$\mu^*(A_i \setminus B) \leq \sum_{n=1}^\infty 2^{-n}\varepsilon_i = \varepsilon_i.$$

Set $A = \bigcap_{i=1}^\infty A_i \in \sigma(\mathcal{H})$. Then $B \subseteq A$, $N := A \setminus B \subseteq A_n \setminus B$ for all $n = 1, 2, ...$ and

$$\mu^*(N) = \mu^*(A \setminus B) \leq \limsup_{i\to\infty} \mu^*(A_i \setminus B) \leq \limsup_{i\to\infty} \varepsilon_i = 0.$$

Thus the assertion follows, since $B = A \setminus N$. □

2.4 Measures on $\mathcal{B}(\mathbb{R})$

From the lecture *Stochastik 1* you already know probability distributions with density. These are measures on $\mathcal{B}(\mathbb{R})$, the Borel σ-algebra on $\mathbb{R}$ (recall from Definition 1.7). We will apply the general theory developed in the last chapters to characterise such measures.

Proposition 2.18 (Lebesgue measure on $\mathbb{R}$)**.** *There is exactly one measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with*

$$\lambda((a, b]) = b - a \tag{2.6}$$

for $a, b \in \mathbb{Q}$ with $a \leq b$.

Proof. Consider $\widetilde{\mathcal{H}} := \{(a, b], [a, b), (a, b), [a, b] : a, b \in \mathbb{Q}, a \leq b\}$, which is a semi-ring with $\sigma(\widetilde{\mathcal{H}}) = \mathcal{B}(\mathbb{R})$. We define $\widetilde{\lambda}$ on $\widetilde{\mathcal{H}}$ by

$$\widetilde{\lambda}((a, b]) = \widetilde{\lambda}([a, b)) = \widetilde{\lambda}((a, b)) = \widetilde{\lambda}([a, b]) = b - a.$$

(Note that $\widetilde{\lambda}$ is the only monotone extension of λ to $\mathcal{H}$.) Then, $\widetilde{\lambda}$ is clearly σ-finite. It is $\mathcal{K} = \{[a, b] : a, b \in \mathbb{Q}, a \leq b\} \subseteq \widetilde{\mathcal{H}}$ a compact system according to Example 1.3. Furthermore, $\widetilde{\lambda}$ is inner $\mathcal{K}$-regular and thus σ-additive according to Theorem 2.10. Hence, Theorem 2.16 gives the only extension of $\widetilde{\lambda}$ to $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R})$. □

Proposition 2.19 (Characterisation of σ-finite measures on $\mathbb{R}$)**.** *A function $\mu : \mathcal{B}(\mathbb{R}) \to \overline{\mathbb{R}}_+$ is a σ-finite measure if and only if there is a non-decreasing and right-continuous function $G : \mathbb{R} \to \mathbb{R}$ with*

$$\mu((a, b]) = G(b) - G(a) \tag{2.7}$$

for $a, b \in \mathbb{Q}$ with $a \leq b$. If $\widetilde{G} : \mathbb{R} \to \mathbb{R}$ is another right-continuous function satisfying (2.7), then $\widetilde{G} = G + c$ for some $c \in \mathbb{R}$.

Corollary 2.20 (Characterisation of probability measures on $\mathbb{R}$)**.** *A function $\mu : \mathcal{B}(\mathbb{R}) \to [0, 1]$ is a probability measure if and only if there is a non-decreasing and right-continuous function $F : \mathbb{R} \to [0, 1]$ with $\lim_{a\to-\infty} F(a) = 0$, $\lim_{b\to\infty} F(b) = 1$ and*

$$\mu((a, b]) = F(b) - F(a) \tag{2.8}$$

for $a, b \in \mathbb{Q}$ with $a \leq b$. In this case, F is uniquely defined by μ.

Proof. The assertion follows directly from Proposition 2.19, since the limit condition uniquely defines c. □

Proof of Proposition 2.19. '⇒': Let μ be a σ-finite measure on $\mathcal{B}(\mathbb{R})$. Define $G(0) := 0$, and $G(x) := \mu((0, x])$ for $x > 0$ and $G(x) := -\mu((x, 0])$ for $x < 0$. Then G is right-continuous, non-decreasing, and (for example for $0 < a < b$) $\mu((a, b]) = \mu((0, b]) - \mu((0, a]) = G(b) - G(a)$.

'⇐': The proof is similar to the proof of Proposition 2.18. Let $\mathcal{H} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$ be the semi-ring of half-open intervals with ends in rational numbers. We show that (2.7) defines a σ-additive set function μ on $\mathcal{H}$. Then, using Theorem 2.16, we see that μ can be uniquely extended to a σ-finite measure on $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R})$. Let $a_1, a_2, \ldots$ be such that

$\bigcup_{n=1}^{\infty}(a_{n+1}, a_n] = (a, b] \in \mathcal{H}$. Without loss of generality, $a_1 \geq a_2 \geq \dots$ Then $b = a_1$ and $a_n \xrightarrow{n\to\infty} a$. Due to the right continuity of G,

$$\mu(a, b] = G(b) - G(a) = G(a_1) - \lim_{N\to\infty} G(a_N) = \sum_{n=1}^{\infty} G(a_n) - G(a_{n+1}) = \sum_{n=1}^{\infty} \mu((a_{n+1}, a_n]),$$

and we have shown the σ-additivity of μ.

Now, let $\widetilde{G}$ be another function for which (2.7) applies. Then for all $a \in \mathbb{R}$,

$$\widetilde{G}(b) = \widetilde{G}(a) + \mu((a, b]) = G(b) + \widetilde{G}(a) - G(a),$$

and the assertion follows with $c = \widetilde{G}(a) - G(a)$. □

Definition 2.21 (Lebesgue measure and distribution functions)**.** *1. The uniquely defined measure λ from Corollary 2.18 is called* one-dimensional Lebesgue measure*.*

2. For a probability measure μ on $\mathcal{B}(\mathbb{R})$, the function F from Corollary 2.20 is called distribution function*.*

Example 2.22 (Some distribution functions)**.** *Let $f : \mathbb{R} \to \mathbb{R}_+$ be piecewise continuous, and*[4] $\int_{-\infty}^{\infty} f(x)dx = 1$*. As known from the lecture* Stochastik 1*, such a function is called a* density*. On the one hand, such density functions define a distribution function by means of*

$$F(x) := \int_{-\infty}^{x} f(a)da.$$

On the other hand, each of these distribution functions defines a probability measure in a unique way due to Corollary 2.20. We will look at distributions with densities in more detail in the Radon-Nikodým theorem (see section 4.4).

As already known,

$$F_{U(0,1)}(x) = \int_{-\infty}^{x} 1_{[0,1]}(a)da = \begin{cases} 0, & x \leq 0, \\ x, & 0 < x \leq 1, \\ 1, & x > 1 \end{cases} \tag{2.9}$$

is the distribution function of the uniform distribution on $[0, 1]$. Further, for $x \geq 0$

$$F_{exp(\lambda)}(x) = \int_{-\infty}^{x} 1_{[0,\infty)}(a)\lambda e^{-\lambda a}da = 1 - e^{-\lambda x} \tag{2.10}$$

is the distribution function of the exponential distribution with parameter λ. Furthermore,

$$F_{N(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\Big(-\frac{(a-\mu)^2}{2\sigma^2}\Big)da =: \Phi(x) \tag{2.11}$$

is the distribution function of the normal distribution $N(\mu, \sigma^2)$ with the expected value μ and the variance σ^2.

[4] We assume here that the Riemann integral $\int_a^b f(x)dx$ is known. (See also definition 3.22.) We will get to know another integral term, the Lebesgue integral, in Chapter 3.

2.5 Image measures

Let μ be a measure on some σ-algebra $\mathcal{F}$. If we transform the base space by means of a function $f : \Omega \to \Omega'$, you can define a measure corresponding to the transformation on Ω', the so-called image measure. Let $\Omega := [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ and $f : u \mapsto -\log u$. We will then see that the image measure of $\mu_{U(0,1)}$ under f is $\mu_{\exp(1)}$. We first recall the situation from example 1.3.2 and define the image measure.

Definition 2.23 (Image measure). *If $(\Omega, \mathcal{F}, \mu)$ is a measure space, $(\Omega', \mathcal{F}')$ is a measurable space and $f : \Omega \to \Omega'$ such that $\sigma(f) \subseteq \mathcal{F}$ for $\sigma(f)$ from* (1.1)*. Then we define a set function $f_*\mu : \mathcal{F}' \to \mathbb{R}_+$ by*

$$f_*\mu(A') := \mu(f^{-1}(A')) = \mu(f \in A'), \qquad A' \in \mathcal{F}'.$$

Here $f_*\mu$ is also called image measure *of μ under f.*

If μ is a probability measure, then $f_*\mu$ is also called distribution *of f (under μ).*

Remark 2.24 (Measurable functions). *If $\sigma(f) \subseteq \mathcal{F}$ as in the definition above, we say that f is measurable (with respect to $\mathcal{F}/\mathcal{F}'$) . This concept will be discussed further in the next section.*

Proposition 2.25 (Image measure is a measure). *Let $(\Omega, \mathcal{F}, \mu)$, $(\Omega', \mathcal{F}')$, $f : \Omega \to \Omega'$ and $f_*\mu$ as in Definition 2.23. Then, $f_*\mu$ is a measure on $\mathcal{F}'$.*

Proof. If $A'_1, A'_2, \dots \in \mathcal{F}'$ are disjoint, then

$$f_*\mu\Big(\biguplus_{n=1}^{\infty} A'_n\Big) = \mu\Big(f^{-1}\Big(\biguplus_{n=1}^{\infty} A'_n\Big)\Big) = \mu\Big(\biguplus_{n=1}^{\infty}(f^{-1}(A'_n)\Big) = \sum_{n=1}^{\infty} \mu(f^{-1}(A'_n)) = \sum_{n=1}^{\infty} f_*\mu(A'_n).$$

This means that $f_*\mu$ is σ-additive and the assertion is shown. □

Example 2.26 (Some transformations).

1. *For $\Omega = [0,1]$, $\{[0,b) : 0 \leq b \leq 1\}$ is a $\cap$-stable generating system of $\mathcal{B}([0,1])$. Let $\mu = \mu_{U(0,1)}$ be the uniform distribution on $[0,1]$ with distribution function $F_{U(0,1)}$ from* (2.9) *and $f : u \mapsto 1-u$. Then $f_*\mu = \mu$, because*

   $$f_*\mu([0,b)) = \mu(f^{-1}([0,b))) = \mu([1-b,1]) = F_{U(0,1)}(1) - F_{U(0,1)}(1-b) = b.$$

   Thus, μ and $f_*\mu$ agree on a $\cap$-stable generator and the statement follows with Proposition 2.11.

2. *Let $\Omega = \mathbb{R}$, $f_y : x \mapsto x+y$ for a $y \in \mathbb{R}$ and λ the Lebesgue measure from Corollary 2.18. Then $(f_y)_*\lambda = \lambda$, because*

   $$(f_y)_*\lambda([a,b]) = \lambda(f_y^{-1}([a,b])) = \lambda([a-y, b-y]) = b-a.$$

   We say that the Lebesgue measure is translation invariant.

3. *Let $\Omega = [0,1], \Omega' = \mathbb{R}_+$, each equipped with Borel's σ-algebra. Further, let $\mu = \mu_{U(0,1)}$ with distribution function $F_{U(0,1)}$ from* (2.9) *and $f : x \mapsto -\frac{1}{\lambda}\log(x)$ for a $\lambda > 0$. Then $f_*\mu = \mu_{\exp(\lambda)}$, where $\mu_{\exp(\lambda)}$ has the distribution function $F_{\exp(\lambda)}$ from* (2.10)*. This is because for $x \geq 0$*

   $$f_*\mu([0,x]) = \mu(f^{-1}([0,x])) = \mu([e^{-\lambda x}, 1]) = 1 - e^{-\lambda x}.$$

Example 2.27 (Example of a non Borel-measurable set (Vitali's set))**.**
So far, there has not yet been an example of a set that is not in $\mathcal{B}(\mathbb{R})$*. We will now construct such a set. It is known as* Vitali's set*. For this purpose, we define an equivalence relation on* $\mathbb{R}$ *by* $x \sim y \iff y - x \in \mathbb{Q}$*. With respect to this equivalence relation,* $\mathbb{R}$ *decomposes into equivalence classes of the form* $\{x + q : q \in \mathbb{Q}\}$*. We select a number from* $[0, 1]$ *from each equivalence class, and put all such numbers into the set* V*. (It should be noted here that this selection is made using the axiom of choice and is therefore not a trivial step). Further, now for* $q \in \mathbb{Q} \cap [-1, 1]$

$$V_q := \{x + q : x \in V\}.$$

Then $[0, 1] \subseteq \biguplus_{q \in \mathbb{Q} \cap [-1,1]} V_q \subseteq [-1, 2]$.

Assume that the quantity V *is measurable. Then the quantities* V_q *would also be measurable and, due to the translation invariance of the Lebesgue measure from Example 2.26.2,* $\lambda(V_q)$ *would not depend on* q*. So let* $\lambda(V_q) =: a \geq 0$*. Furthermore, due to the monotonicity of the Lebesgue measure*

$$1 \leq \sum_{q \in \mathbb{Q} \cap [-1,1]} \lambda(V_q) = \sum_{q \in \mathbb{Q} \cap [-1,1]} a \leq 3.$$

However, this is not possible, neither for $a = 0$ *nor for* $a > 0$*. Because of this contradiction,* $V \notin \mathcal{B}(\mathbb{R})$ *must therefore apply.*

3 Measurable functions and the integral

In this chapter, let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We can now use the measure μ to measure the content of sets of $\mathcal{F}$. The aim of introducing the integral is to weight the elements of Ω differently in such a measurement. This weighting is carried out with a function $f : \Omega \to \mathbb{R}$. Such functions must fulfil the minimal requirement of measurability. The result of this weighting leads to the concept of the integral.

3.1 Measurable functions

We already know what a measurable set (with respect to the σ-algebra $\mathcal{F}$) is, i.e. $A \subseteq \Omega$ is ($\mathcal{F}-$)measurable iff $A \in \mathcal{F}$. We will extend this notion to functions $f : \Omega \to \Omega'$ (for some measurable space $(\Omega', \mathcal{F}')$. Note that for $A \in \mathcal{F}$, there is the indicator function $\omega \mapsto 1_A(\omega)$, which is the simplest form of a measurable function in Definition 3.3. We will call the linear combination of such indicator functions a simple function, which will be measurable as well. These are of particular importance due to Theorem 3.9, which shows that every non-negative measurable function – see below – can be approximated from below (in the sense of pointwise convergence) by simple functions.

Remark 3.1 (Notation)**.** *Let Ω, Ω' be sets, $f : \Omega \to \Omega'$ and I be arbitrary.*

1. *We write $f(A) := \{f(\omega) : \omega \in A\}$ for $A \subseteq \Omega$ and $f^{-1}(A') := \{f^{-1}(\omega') : \omega' \in A'\}$ for $A' \subseteq \Omega'$. We note that the following rules apply to $A', A'_i \subseteq \Omega', i \in I$:*

$$f^{-1}((A')^c) = (f^{-1}(A'))^c, \quad f^{-1}\Big(\bigcap_{i\in I} A'_i\Big) = \bigcap_{i\in I} f^{-1}(A'_i), \quad f^{-1}\Big(\bigcup_{i\in I} A'_i\Big) = \bigcup_{i\in I} f^{-1}(A'_i).$$

   However, some caution is required, since for $A, A_i \subseteq \Omega, i \in I$ only $f(\bigcup_{i\in I} A_n) = \bigcup_{i\in I} f(A_i)$, in general, however, $f(A^c) \neq (f(A))^c$ and $f(\bigcap_{i\in I} A_i) \neq \bigcap_{i\in I} f(A_i)$.

2. *For $\mathcal{C} \subseteq 2^{\Omega'}$ we write analogously*

$$f^{-1}(\mathcal{C}) := \{f^{-1}(A') : A' \in \mathcal{C}\}.$$

Lemma 3.2 (Pre-image of σ-algebras)**.** *Let Ω be a set and $(\Omega', \mathcal{F}')$ a measurable space, $f : \Omega \to \Omega'$ and $\mathcal{C}' \subseteq \mathcal{F}'$ with $\sigma(\mathcal{C}') = \mathcal{F}'$. Then $\sigma(f^{-1}(\mathcal{C}')) = f^{-1}(\sigma(\mathcal{C}'))$. In particular, $f^{-1}(\mathcal{F}')$ is a σ-algebra on Ω.*

Proof. '$\subseteq$': From Remark 3.1, it is clear that $f^{-1}(\sigma(\mathcal{C}'))$ is a σ-algebra. This means that $\sigma(f^{-1}(\mathcal{C}')) \subseteq \sigma(f^{-1}(\sigma(\mathcal{C}'))) = f^{-1}(\sigma(\mathcal{C}'))$.
'$\supseteq$': We define

$$\widetilde{\mathcal{F}}' = \{A' \in \sigma(\mathcal{C}') : f^{-1}(A') \in \sigma(f^{-1}(\mathcal{C}'))\} \subseteq \sigma(\mathcal{C}').$$

Then, again due to Remark 3.1, $\widetilde{\mathcal{F}}'$ is a σ-algebra and $\mathcal{C}' \subseteq \widetilde{\mathcal{F}}' \subseteq \sigma(\mathcal{C}')$. Thus, $\widetilde{\mathcal{F}}' = \sigma(\mathcal{C}')$. For $A' \in \sigma(\mathcal{C}')$, we find $f^{-1}(A') \in \sigma(f^{-1}(\mathcal{C}'))$, which is equivalent to $f^{-1}(\sigma(\mathcal{C}')) \subseteq \sigma(f^{-1}(\mathcal{C}'))$. □

Definition 3.3 (Measurable functions)**.** *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces and $f : \Omega \to \Omega'$.*

1. *The function f is called* $\mathcal{F}/\mathcal{F}'$-measurable *if $f^{-1}(\mathcal{F}') \subseteq \mathcal{F}$. The σ-algebra $f^{-1}(\mathcal{F}')$ (recall from Lemma 3.2 that this is in fact a σ-algebra) is called the* σ-algebra (on Ω) generated by f *and is denoted $\sigma(f)$.*

2. *If $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space and $X : \Omega \to \Omega'$ measurable, then X is called an* Ω'-valued random variable. *The image measure $X_*\mathbf{P}$ from Definition 2.23 is called the* distribution of X.

3. *If $(\Omega', \mathcal{F}') = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, then f is called a* real-valued function. *If f is measurable according to $\mathcal{F}/\mathcal{B}(\overline{\mathbb{R}})$, we say that the function f is (Borel-)measurable.*

4. *If $\Omega' = \overline{\mathbb{R}}$ and $f = 1_A$ for $A \subseteq \Omega$, then f is called* indicator function. *If $f = \sum_{k=1}^{n} c_k 1_{A_k}$ for $c_1, \dots, c_n \in \overline{\mathbb{R}}$ pairwise different and $A_1, \dots, A_n \subseteq \Omega$, then f is called* simple.

Example 3.4. *Let $(\Omega, \mathcal{F})$ be a measurable space.*

1. *The vast majority of functions $f : \Omega \to \overline{\mathbb{R}}$ that one can imagine are (Borel-)measurable. For example, the identity $f : \omega \mapsto \omega$ is measurable, since $f^{-1}(\mathcal{F}) = \mathcal{F}$.*

2. *Let $(\Omega, \mathcal{O})$ and $(\Omega'.\mathcal{O}')$ be topological spaces and $f : \Omega \to \Omega'$ continuous. Then f is $\mathcal{B}(\Omega)/\mathcal{B}(\Omega')$-measurable.*
   Indeed, by continuity we have that $f^{-1}(\mathcal{O}') \subseteq \mathcal{O}$. Therefore, using Lemma 3.2,
$$f^{-1}(\mathcal{B}(\Omega')) = f^{-1}(\sigma(\mathcal{O}')) = \sigma(f^{-1}(\mathcal{O}') \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\Omega).$$

3. *It is important to see that for many measurable functions f it is true that $\sigma(f) \subsetneq \mathcal{F}$, see for example the next example.*

4. *A function $f : \Omega \to \{0,1\}$ is measurable if and only if $f^{-1}(\{1\}) \in \mathcal{F}$. In this case, $\sigma(f) = \{\emptyset, f^{-1}(\{1\}), (f^{-1}(\{1\}))^c, \Omega\}$.*

5. *Let $\mathcal{F} = \mathcal{B}(\mathbb{R})$. To specify a non $\mathcal{F}$-measurable function, you have to make the same effort as to construct a non Borel-measurable set. For example, the function 1_V is not measurable for the Vitali set V from Example 2.27.*

Example 3.5 (Random variables)**.** *Let (E, r) be some metric space (equipped with the Borel σ-algebra).*

1. *Let X be an E-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, and Y an E-valued random variable on $(\Omega', \mathcal{A}, \mathbf{Q})$. If $X_*\mathbf{P} = Y_*\mathbf{Q}$, we say that X and Y are* identically distributed *and write $X \sim Y$. Note, however, since X and Y need not be defined on the same probability space, it does not make sense to write something like $X - Y$. If μ is a measure on $\mathcal{B}(E)$ and $X_*\mathbf{P} = \mu$, we write $X \sim \mu$.*

2. *Let $(X_i)_{i \in I}$ be a family of random variables on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Then, the distribution of $((X_i)_{i \in I})_*\mathbf{P}$ is called the* joint distribution of $(X_i)_{i \in I}$.

Lemma 3.6 (Properties of measurability)**.** *Let $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{F}')$ and $(\Omega'', \mathcal{F}'')$ be measurable spaces.*

1. *If $\mathcal{C}' \subseteq \mathcal{F}'$ with $\mathcal{F}' = \sigma(\mathcal{C}')$, then $f : \Omega \to \Omega'$ is $\mathcal{F}/\mathcal{F}'$-measurable if and only if $f^{-1}(\mathcal{C}') \subseteq \mathcal{F}$.*

2. *If $f : \Omega \to \Omega'$ is $\mathcal{F}/\mathcal{F}'$-measurable and $g : \Omega' \to \Omega''$ is $\mathcal{F}'/\mathcal{F}''$-measurable, then $g \circ f : \Omega \to \Omega''$ is $\mathcal{F}/\mathcal{F}''$-measuarble.*

3. *Let $(\Omega, \mathcal{O})$ and $(\Omega', \mathcal{O}')$ be topological spaces, $f : \Omega \to \Omega'$ continuous and $\mathcal{F} = \sigma(\mathcal{O})$ and $\mathcal{F}' = \sigma(\mathcal{O}')$ the Borel σ-algebras. Then f is $\mathcal{F}/\mathcal{F}'$-measurable.*

4. *A real-valued function f (i.e. $f : \Omega \to \mathbb{R}$) is measurable (with respect to $\mathcal{F}/\mathcal{B}(\mathbb{R})$) if and only if $\{\omega : f(\omega) \le x\} \in \mathcal{F}$ for all $x \in \mathbb{Q}$.*

5. *A simple function $f = \sum_{k=1}^n c_k 1_{A_k}$ with pairwise different $c_1, \dots, c_n \in \overline{\mathbb{R}}$ and $A_1, \dots, A_n \subseteq \Omega$ is measurable if and only if $A_1, \dots, A_n \in \mathcal{F}$.*

Proof. 1. the 'only if' direction is clear. For the 'if' direction, we use Lemma 3.2 and obtain $f^{-1}(\mathcal{F}') = f^{-1}(\sigma(\mathcal{C}')) = \sigma(f^{-1}(\mathcal{C}')) \subseteq \sigma(\mathcal{F}) = \mathcal{F}$. This means that f is $\mathcal{F}/\mathcal{F}'$-measurable.
2. We write directly $(g \circ f)^{-1}(\mathcal{F}'') = f^{-1}(g^{-1}(\mathcal{F}'')) \subseteq f^{-1}(\mathcal{F}') \subseteq \mathcal{F}$, which already shows the assertion.
3. By definition of Borel's σ-algebra, $\mathcal{O}'$ is a generator for $\mathcal{B}(\Omega')$. Since f is continuous (i.e. $f^{-1}(\mathcal{O}') \subseteq \mathcal{O}$), $f^{-1}(\mathcal{O}') \subseteq \mathcal{O} \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\Omega)$ follows. According to 1. f is therefore $\mathcal{B}(\Omega)/\mathcal{B}(\Omega')$-measurable.
4. We use 1 with $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{Q}\}$: If $\Omega' = \mathbb{R}$, then according to Lemma 3.2, $\mathcal{B}(\Omega')$ is generated by $\mathcal{C}$. Therefore, a f is measurable if $f^{-1}(\mathcal{C}') = \{\{\omega : f(\omega) \le x\} : x \in \mathbb{R}\} \subseteq \mathcal{F}$.
5. Let $A := \left(\bigcup_{k=1} A_k\right)^c \in \mathcal{F}$. Then $f^{-1}(\mathcal{B}(\overline{\mathbb{R}})) = \left\{A \cup \bigcup_{j \in J} A_j, \bigcup_{j \in J} A_j : J \subseteq \{1, \dots, n\}\right\}$, from which the assertion follows. □

Lemma 3.7 (Algebraic structure of measurability)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space.*

1. *Let $f, g : \Omega \to \overline{\mathbb{R}}$ be measurable. Then fg, as well as $af + bg$ for $a, b \in \mathbb{R}$ are measurable. In addition, f/g is measurable if $g(\omega) \neq 0$ for all $\omega \in \Omega$.*

2. *Let $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ be measurable. Then,*

$$\sup_{n \in \mathbb{N}} f_n, \quad \inf_{n \in \mathbb{N}} f_n, \quad \limsup_{n \to \infty} f_n, \quad \liminf_{n \to \infty} f_n$$

   are measurable as well. If it exists, $\lim_{n \to \infty} f_n$ is also measurable.

Proof. 1. Consider $\psi : \Omega \to \mathbb{R}^2$, defined by $\psi(\omega) = (f(\omega), g(\omega))$. It is easy to see that ψ is $\mathcal{F}/\mathcal{B}(\mathbb{R}^2)$-measurable. Furthermore, $(x, y) \mapsto ax + by$ and $(x, y) \mapsto xy$ are continuous on $\mathbb{R}$ and $(x, y) \mapsto x/y$ on $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$, i.e. measurable according to Lemma 3.6.3. Thus the assertions according to Lemma 3.6.2 follow.
2. We only show the measurability of $\sup_{n \in \mathbb{N}} f_n$. The other statements then follow using

$$\inf_{n \in \mathbb{N}} f_n = -\sup_{n \in \mathbb{N}}(-f_n), \quad \limsup_{n \to \infty} f_n = \inf_{n \in \mathbb{N}} \sup_{k \ge n} f_k, \quad \liminf_{n \to \infty} f_n = \sup_{n \in \mathbb{N}} \inf_{k \ge n} f_k.$$

We write, for $x \in \mathbb{R}$, according to Lemma 3.6.4

$$\left\{\omega : \sup_{n \in \mathbb{N}} f_n(\omega) \le x\right\} = \bigcap_{n=1}^{\infty} \underbrace{\left\{\omega : f_n(\omega) \le x\right\}}_{\in \mathcal{F}} \in \mathcal{F}$$

and the assertion is shown. □

Corollary 3.8 (Measurability of positive and negative part)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \overline{\mathbb{R}}$. Then f is measurable if and only if $f^+ := f \vee 0$ and $f^- := (-f) \vee 0$ are measurable. Then $|f|$ is also measurable.*

Proof. Note that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Thus the assertion follows from Lemma 3.7.2. □

Theorem 3.9 (Approximation with measurable functions)**.** *Let* $(\Omega, \mathcal{F})$ *be a measurable space and* $f : \Omega \to \overline{\mathbb{R}}_+$ *measurable. Then there is a sequence* $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ *of simple functions with*[5] $f_n \uparrow f$.

Proof. We write for[6] $\omega \in \Omega, n \in \mathbb{N}$

$$f_n(\omega) = n \wedge 2^{-n}[2^n f(\omega)],$$

and note that $f_n \uparrow f$ holds by construction. Furthermore, $\omega \mapsto [2^n f(\omega)]$ is measurable according to Lemma 3.6.4 if f. □

3.2 Definition

The construction of the integral of a function f according to a measure will take place in several steps. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. For the integral of $f : \Omega \to \mathbb{R}$ with respect to μ we use different synonymous notations, namely

$$\mu[f] = \int f d\mu = \int f(\omega)\mu(d\omega). \tag{3.1}$$

The integral is first defined for simple functions $f$and then (see Theorem 3.9) by approximation for general non-negative measurable functions. The integral for (not necessarily non-negative) measurable functions is then defined by the integral of the positive and negative parts; see Definition 3.17.

The application in probability theory is as follows: Recall the notion of a *probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ from Definition 2.1. Here, any measurable $X : \Omega \to \mathbb{R}$ is called a *random variable* (recall from Definition def:measurable). Then, we use the notation

$$\mathbf{E}[X] := \mathbf{P}[X],$$

where $\mathbf{P}[X]$ is defined as in (3.1) and denote this by the *expectation of* X *(with respect to* $\mathbf{P}$*)*.

Definition 3.10 (Integral of simple functions)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* $f = \sum_{k=1}^m c_k 1_{A_k}$ *a simple function with* $c_1, \dots, c_m \geq 0, A_1, \dots, A_m \in \mathcal{F}$. *Then,*

$$\mu[f] := \int f d\mu := \sum_{k=1}^n c_k \mu(A_m)$$

is the integral of f with respect to μ.

Remark 3.11 (Integral is well-defined)**.** *We must make sure that the above integral is well-defined. Let* $f = \sum_{l=1}^n d_l 1_{B_l}$ *be another representation of* f *with* $d_1, \dots, d_n \geq 0$ *and* $B_1, \dots, B_n \in \mathcal{F}$. *Then,*

$$\sum_{k=1}^m c_k \mu(A_k) = \sum_{k=1}^m \sum_{l=1}^n c_k \mu(A_k \cap B_l) = \sum_{k=1}^m \sum_{l=1}^n d_l \mu(A_k \cap B_l) = \sum_{l=1}^n d_l \mu(B_l),$$

so the integral of simple functions is well-defined.

[5] Analogously to '↓', we write for $x, x_1, x_2, \dots \in \mathbb{R}$ that $x_n \uparrow x$ if $x_1 \leq x_2 \leq \dots$ and $x_n \xrightarrow{n\to\infty} x$. For functions $f, f_1, f_2, \dots : \Omega \to \mathbb{R}$, $f_n \uparrow f$ means that $f_n(\omega) \uparrow f(\omega)$ for all $\omega \in \Omega$.

[6] here $[x]$ for $x \in \mathbb{R}$ is the largest integer smaller than x, so $[x] := \sup\{n \in \mathbb{Z} : n \leq x\}$.

Lemma 3.12 (Simple properties)**.** *Let f, g be non-negative, simple functions and $\alpha \geq 0$. Then*[7]

$$\mu[af + bg] = a\mu[f] + b\mu[g], \qquad f \leq g \Rightarrow \mu[f] \leq \mu[g].$$

Proof. Clear. □

Example 3.13 (The integral of indicator functions and the Riemann integral)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $A \in \mathcal{F}$. Then $f = 1_A$ is a simple function and the following applies*

$$\mu[f] = \mu(A)$$

according to definition 3.10. It should be noted that the function $f = 1_A$ no longer has to be piecewise continuous. (Let $A = \mathbb{Q}$ or A be the Cantor continuum considered in Example 1.10). Therefore, it is not clear that the function 1_A is integrable in the sense of Riemann.

Definition 3.14 (The integral of measurable, non-negative functions)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measue space and $f : \Omega \to \overline{\mathbb{R}}_+$ measurable. The integral of f with respect to μ is given by*

$$\mu[f] := \int f(\omega)\mu(d\omega) := \int f d\mu := \sup\{\mu[g] : g \text{ simple, non-negative, } g \leq f\}. \tag{3.2}$$

Remark 3.15 (The integral as an extension)**.** *From Lemma 3.12 it is clear that the definition of $\mu[f]$ for simple, non-negative functions f from Definition 3.10 and Definition 3.14 is the same. The above definition is therefore an extension of $\mu[f]$ to the space of non-negative, measurable functions.*

It is also important to note that, according to Theorem 3.9, each of the functions occurring in Definition 3.14 can be approximated (pointwise) by simple functions. In particular, the supremum is in (3.2) *is over simple functions g which are arbitrarily close to f.*

Proposition 3.16 (Properties of the integral)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g, f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}_+$ measurable. Then the following applies:*

1. *If $f \leq g$, then $\mu[f] \leq \mu[g]$.*
2. *If*

   $$f_n \uparrow f, \quad \text{then} \quad \mu[f_n] \uparrow \mu[f].$$

   We say that the integral obeys monotone convergence*.*
3. *If $a, b \geq 0$, then $\mu[af + bg] = a\mu[f] + b\mu[g]$.*

Proof. 1. is clear from the definition of the integral. 2. From 1., it is clear that $\mu[f_1], \mu[f_2], ...$ is increasing. In particular, $\lim_{n\to\infty} \mu[f_n]$ exists. We need to show $\lim_{n\to\infty} \mu[f_n] \leq \mu[f]$ as well as $\mu[f] \leq \lim_{n\to\infty} \mu[f_n]$. First, since $f_1, f_2, ... \leq f$,

$$\lim_{n\to\infty} \mu[f_n] = \sup_{n\in\mathbb{N}} \mu[f_n] \leq \mu[f].$$

Second, it is sufficient to show that

$$\mu[g] \leq \sup_{n\in\mathbb{N}} \mu[f_n] \tag{3.3}$$

[7]For $f, g : \Omega \to \overline{\mathbb{R}}$, we write $f \leq g$ if $f(\omega) \leq g(\omega)$ holds for all $\omega \in \Omega$.

for all simple functions $g \leq f$. Let $g = \sum_{k=1}^m c_k 1_{A_k} \leq f$ for disjoint sets $A_1, \dots, A_m$ and $c_1, \dots, c_m > 0$. For $\varepsilon > 0$ and $n = 1, 2, \dots$ let $B_n^\varepsilon := \{f_n \geq (1-\varepsilon)g\}$. Since $f_n \uparrow f$ and $g \leq f$, $\bigcup_{n=1}^\infty B_n^\varepsilon = \Omega$ for all $\varepsilon > 0$. Therefore

$$\begin{aligned}\mu[f_n] \geq \mu[(1-\varepsilon)g 1_{B_n^\varepsilon}] &= \sum_{k=1}^m (1-\varepsilon)c_k \mu(A_k \cap B_n^\varepsilon) \\ &\xrightarrow{n\to\infty} \sum_{k=1}^m (1-\varepsilon)c_k \mu(A_k) = (1-\varepsilon)\mu[g].\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, (3.3) follows.
For 3., let $f_1, g_1, f_2, g_2, \dots$ be simple functions with $f_n \uparrow f$ and $g_n \uparrow g$. Then, $af_n + bg_n \uparrow af + bg$ and it follows

$$\mu[af + bg] = \lim_{n\to\infty} \mu[af_n + bg_n] = \lim_{n\to\infty} a\mu[f_n] + b\mu[g_n] = a\mu[f] + b\mu[g]$$

from 2. because of Lemma 3.12. □

We can now define the integral for measurable functions. First, we note that $f^+, f^- \leq |f|$ for any $f : \Omega \to \overline{\mathbb{R}}$. In particular, if f is measurable with $\mu[|f|] < \infty$, then $\mu[f^+], \mu[f^-] < \infty$.

Definition 3.17 (Integral of measurable functions)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \overline{\mathbb{R}}$ measurable. Then f is said to be μ-integrable if $\mu[|f|] < \infty$ and we define*

$$\mu[f] := \int f(\omega)\mu(d\omega) := \int f d\mu := \mu[f^+] - \mu[f^-]. \tag{3.4}$$

We also set

$$\mathcal{L}^1(\mu) := \Big\{ f : \Omega \to \overline{\mathbb{R}} : \mu[|f|] < \infty \Big\}$$

For $A \in \mathcal{F}$ we also write

$$\mu[f, A] := \int_A f d\mu := \mu[f 1_A].$$

Remark 3.18 (Extension of the integral and $\mathcal{L}^p$-spaces)**.**

1. *If at most one of the two terms $\mu[f^+]$ or $\mu[f^-]$ is infinite, we continue to define the integral $\mu[f]$ using* (3.4)*. In other cases, the integral remains undefined.*
2. *The function spaces $\mathcal{L}^p(\mu) := \Big\{ f : \Omega \to \overline{\mathbb{R}} : \mu[|f|^p] < \infty \Big\}$, $p > 0$, will play a special role in Section 4.*

3.3 Properties of the integral

We first establish some properties of the integral. These are, for example, monotonicity and linearity. We will also see that the integral of a function does not change if it is modified on a null-set; see Proposition 3.21.

Proposition 3.19 (Simple properties of the integral)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g \in \mathcal{L}^1(\mu)$. Then, the following holds:*

1. *The integral is monotone, i.e.*

$$f \leq g \text{ almost everywhere} \qquad \Longrightarrow \qquad \mu[f] \leq \mu[g].$$

2. *As a special case of 1., since* $-f, f \leq |f|$,

$$|\mu[f]| \leq \mu[|f|].$$

3. *The integral is linear, so if* $a, b \in \mathbb{R}$, *then* $af + bg \in \mathcal{L}^1(\mu)$ *and*

$$\mu[af + bg] = a\mu[f] + b\mu[g].$$

Proof. All properties follow from Proposition 3.16.1 and 3, and the definition of the integral for measurable functions. □

Proposition 3.20 (Substitution theorem)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space,* $(\Omega', \mathcal{F}')$ *a measurable space,* $f : \Omega \to \Omega'$ *measurable and* $f_*\mu$ *the image measure of* f *from Definition 2.23. Then for* $g \in \mathcal{L}^1(f_*\mu)$ *it is true that* $g \circ f \in \mathcal{L}^1(\mu)$ *and*

$$\mu[g \circ f] = f_*\mu[g].$$

Proof. It is sufficient to show the assertion for simple, non-negative functions g. The general case then follows by means of approximation by simple functions. Let $g = \sum_{k=1}^m c_k 1_{A'_k}$ with $A'_k \in \mathcal{F}'$. Then $g \circ f = \sum_{k=1}^m c_k 1_{f \in A'_k}$ and

$$\mu[g \circ f] = \sum_{k=1}^m c_k \mu(f \in A'_k) = \sum_{k=1}^m c_k f_*\mu(A'_k) = f_*\mu[g].$$

□

Proposition 3.21 (Integrals and properties almost everywhere)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* $f : \Omega \to \overline{\mathbb{R}}_+$ *measurable.*

1. *It is* $f = 0$ *almost everywhere iff* $\mu[f] = 0$.

2. *If* $\mu[f] < \infty$, *then* $f < \infty$ *almost everywhere.*

Proof. 1. Let $N := \{f > 0\} \in \mathcal{F}$. '$\Rightarrow$': Since $\mu(N) = 0$, we find $f \leq \infty \cdot 1_N$, so because of Proposition 3.16.2,

$$0 \leq \mu[f] \leq \mu[\infty, N] = \lim_{n\to\infty} \mu[n, N] = 0.$$

For '$\Leftarrow$', let $N_n := \{f \geq 1/n\}$ and thus $N_n \uparrow N$ and $nf \geq 1_{N_n}$, i.e.

$$0 = \mu[f] \geq \tfrac{1}{n}\mu(N_n).$$

This means that $\mu(N_n) = 0$ and therefore $\mu(N) = \mu(\bigcup_{n=1}^\infty N_n) = 0$ by σ-sub-additivity of μ. For 2., let $A := \{f = \infty\}$. Since $f 1_{f \geq n} \geq n 1_{f \geq n}$,

$$\mu(A) = \mu[1_A] \leq \mu[1_{f \geq n}] \leq \tfrac{1}{n}\mu[f, 1_{f \geq n}] \leq \tfrac{1}{n}\mu[f] \xrightarrow{n\to\infty} 0.$$

This means that $\mu(f = \infty) = 0$, i.e. $f < \infty$ almost everywhere; see Remark 2.14. □

To conclude this section, we show the relationship between the (Lebesgue) integral and the Riemann integral.

Definition 3.22 (Piece-wise constant function and Riemann integral). *Let $f : \mathbb{R} \to \mathbb{R}$ be a piece-wise constant function, i.e.*

$$f(x) = \sum_{j=-\infty}^{\infty} a_j 1_{[x_{j-1},x_j)}(x) \tag{3.5}$$

with $x_{j-1} \le x_j, j \in \mathbb{Z}$, where $a_j \in \mathbb{R}, j \in \mathbb{Z}$. Some $f : [a,b] \to \mathbb{R}$ (with $a < b$) is called Riemann-integrable *if $\lambda[|f|] < \infty$ and there are piece-wise constant functions $f_1^+, f_1^-, f_2^+, f_2^-, \ldots$ with $f_n^- \le f \le f_n^+$ and $\lambda[f_n^+ - f_n^-] \xrightarrow{n\to\infty} 0$. The Riemann integral of f is then defined by $\lambda[f]$. (In particular, the Riemann integral and Lebesgue integral then coincide.*

A function $f : \mathbb{R} \to \mathbb{R}$ is called Riemann-integrable *if $f1_K$ is Riemann-integrable for all compact intervals $K \subseteq \mathbb{R}$ and $\lambda[f1_{[-n,n]}]$ converges. This limit is then the Riemann integral of f with respect to λ.*

Proposition 3.23 (Riemann integrability). *Let $f : [0,\infty) \to \mathbb{R}$ have a discrete set of jump points. Then f is integrable, Riemann-integrable, and*

$$\lambda[f] = \lim_{n\to\infty} \sum_{k=1}^{\infty} f(y_{n,k})(x_{n,k} - x_{n,k-1}) \tag{3.6}$$

for $0 = x_{n,0} \le \ldots \le x_{n,k_n} = t$ with $\max_k |x_{n,k} - x_{n,k-1}| \xrightarrow{n\to\infty} 0$ and any $x_{n,k-1} \le y_{n,k} \le x_{n,k}$.

Proof. It is sufficient to show the assertion for continuous f. Otherwise, f can be broken down into the continuous pieces. It is also sufficient to show the assertion for f with compact support K. Since f is uniformly continuous on K, first choose $\varepsilon_n \downarrow 0$ and $x_{n,0} \le \ldots \le x_{n,k_n}$ such that $K \subseteq [x_{n,0}, x_{n,k_n}]$ and $\max_{x_{n,k-1} \le y < x_{n,k}} |f(x_{n,k-1}) - f(y)| < \varepsilon_n$. Now it is easy to find piecewise constant functions f_n^+ and f_n^- such that $f_n^- \le f \le f_n^+$ and $||f_n^+ - f_n^-|| \le \varepsilon_n$. Integrability and Riemann-integrability follows. The formula (3.6) is valid due to the uniform approximation of the function f. □

Example 3.24 (Differences between Riemann and Lebesgue integral).

1. *We start with a function that is Lebesgue-integrable but not Riemann-integrable. Let $f = 1_{[0,1]\cap\mathbb{Q}}$. Then $1_{[0,1]} \le f^+$ for every piece-wise constant function $f^+ \ge f$ and $f^- \le 0$ for every piece-wise constant function $f^- \le f$. In particular, f is not Riemann-integrable.*

2. *As can be seen from the definition of the Riemann integral, every piece-wise constant function is simple, so every Riemann-integrable function is also Lebesgue-integrable. The situation is different for functions on unbounded domains. Let f be given by $f(t) = \frac{(-1)^{\lceil t\rceil+1}}{\lceil t\rceil}$. Then*

$$\lambda[f1_{[0,2n]}] = \sum_{k=1}^{2n} \frac{(-1)^{k+1}}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \sum_{k=1}^{n} \frac{1}{2k-1} - \frac{1}{2k} = \sum_{k=1}^{n} \frac{1}{(2k-1)2k}$$

and we see that the limit is finite, thus f is Riemann-integrable. However, the following applies

$$\lambda[|f|] = \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

So, according to Definition 3.14, f is not Lebesgue-integrable.

3.4 Convergence results

You may ask whether it is really so important that you can integrate more functions with respect to the Lebesgue integral than with respect to the Riemann integral. After all, most applications involve Riemann-integrable functions. However, there is another advantage of the Lebesgue integral, which we will now discuss. In calculus, the following convergence result for the Riemann integral is frequently given:

Theorem 3.25 (Riemann integral convergence result)**.** *Let $a, b \in \mathbb{R}$ with $a < b$, and $f, f_1, f_2, ... : [a, b] \to \mathbb{R}$ be piecewise continuous. If $f_n \xrightarrow{n\to\infty} f$ uniformly, then (using $\int$ for the Riemann integral)*

$$\int_a^b f_n(x)dx \xrightarrow{n\to\infty} \int_a^b f(x)dx.$$

As you see, this result is concerning the interchange of limits limits and integrals, which requires a uniform convergence in this case. For convergence results with respect to the Lebesgue integral, however, we need much weaker conditions for the exchange of integral and limit. The most important of these are the theorem of monotone convergence and the theorem of dominated convergence.

Theorem 3.26 (Monotone convergence)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $f_1, f_2, \dots \in \mathcal{L}^1(\mu)$ and $f : \Omega \to \overline{\mathbb{R}}$ measurable with $f_n \uparrow f$ almost everywhere. Then,*

$$\lim_{n\to\infty} \mu[f_n] = \mu[f],$$

where both sides can take the value ∞.

Proof. Let $N \in \mathcal{F}$ be such that $\mu(N) = 0$ and $f_n(\omega) \uparrow f(\omega)$ for $\omega \notin N$. Set $g_n := (f_n - f_1)1_{N^c} \geq 0$. This means that $g_n \uparrow (f - f_1)1_{N^c} =: g$ and with Proposition 3.19, Proposition 3.21 and Proposition 3.16.2,

$$\mu[f_n] = \mu[f_1] + \mu[g_n] \xrightarrow{n\to\infty} \mu[f_1] + \mu[g] = \mu[f].$$

□

Theorem 3.27 (Lemma of Fatou)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}_+$ measurable. Then,*

$$\liminf_{n\to\infty} \mu[f_n] \geq \mu[\liminf_{n\to\infty} f_n].$$

Proof. For all $k \geq n$, $f_k \geq \inf_{\ell \geq n} f_\ell$ and thus, for all n,

$$\inf_{k\geq n} \mu[f_k] \geq \mu[\inf_{\ell\geq n} f_\ell]$$

by Proposition 3.16.1 Therefore, with $n \to \infty$

$$\liminf_{n\to\infty} \mu[f_n] = \sup_{n\in\mathbb{N}} \inf_{k\geq n} \mu[f_k] \geq \sup_{n\in\mathbb{N}} \mu[\inf_{k\geq n} f_k] = \mu[\liminf_{n\to\infty} f_n]$$

by monotone convergence, Theorem 3.26, since $\inf_{k\geq n} f_k \uparrow \sup_{n\in\mathbb{N}} \inf_{k\geq n} f_k = \liminf_{n\to\infty} f_n$.
□

Theorem 3.28 (Dominated convergence)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g, f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ measurable with $|f_n| \le g$ almost everywhere, $\lim_{n\to\infty} f_n = f$ almost everywhere, and $g \in \mathcal{L}^1(\mu)$. Then,*

$$\lim_{n\to\infty} \mu[f_n] = \mu[f].$$

Proof. Without loss of generality, $|f_n| \le g$ and $\lim_{n\to\infty} f_n = f$ holds everywhere. (Otherwise, restrict to a set of full measure.) We use Fatou's lemma and $g - f_n, g + f \ge 0$, i.e.

$$\mu[g+f] \le \liminf_{n\to\infty} \mu[g + f_n] = \mu[g] + \liminf_{n\to\infty} \mu[f_n],$$
$$\mu[g-f] \le \liminf_{n\to\infty} \mu[g - f_n] = \mu[g] - \limsup_{n\to\infty} \mu[f_n].$$

After subtracting $\mu[g]$,

$$\mu[f] \le \liminf_{n\to\infty} \mu[f_n] \le \limsup_{n\to\infty} \mu[f_n] \le \mu[f].$$

□

Example 3.29.

1. *Fatou's lemma does not require that any of the f_n is integrable. We now give an example to show that in Fatou's lemma '$<$' rather than '$=$' holds in general. Let λ be the Lebesgue measure and $f_n = 1/n$ (i.e. in particular, f_n constant), $n = 1, 2, \dots$. Then $f_n \downarrow 0$, but*

$$\liminf_{n\to\infty} \mu[f_n] = \infty > 0 = \mu[0] = \mu[\liminf_{n\to\infty} f_n].$$

2. *In the theorem of dominated convergence, the condition that $|f_n| \le g$ and $g \in \mathcal{L}^1(\mu)$ is necessary. For example, let λ be the Lebesgue measure on $[0, 1]$ and $f_n = n \cdot 1_{[0,1/n]}$. Then $\sup_{n\in\mathbb{N}} f_n(x) = \sup\{n : x \le 1/n\} = \left[\frac{1}{x}\right]$[8]. So there is no $g \in \mathcal{L}^1(\lambda)$ with $f_n \le g$. Moreover, $\lim_{n\to\infty} f_n = 0$ almost everywhere (since $\{0\}$ is a null-set) and*

$$\lim_{n\to\infty} \mu[f_n] = 1 \ne 0 = \mu[\lim_{n\to\infty} f_n].$$

The situation is different for $f_n = n \cdot 1_{[0,1/n^2]}$. Here,

$$\sup_{n\in\mathbb{N}} f_n(x) = \sup\{n : x \le 1/n^2\} = \left[\frac{1}{\sqrt{x}}\right] \le \frac{1}{\sqrt{x}} =: g(x).$$

On the one hand, $g \in \mathcal{L}^1(\lambda)$, so dominated convergence applies. On the other hand, $\lim_{n\to\infty} f_n = 0$ almost everywhere and

$$\lim_{n\to\infty} \mu[f_n] = \lim_{n\to\infty} \tfrac{1}{n} = 0 = \mu[0] = \mu[\lim_{n\to\infty} f_n].$$

[8] With $[x] := \sup\{n \in \mathbb{Z} : n \le x\}$ we denote the rounding function.

4 $\mathcal{L}^p$-spaces

Throughout the following section, let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We will now deal with the set of measurable functions $f : \Omega \to \overline{\mathbb{R}}$ satisfying $\mu[|f|^p] < \infty$. We will recognise the resulting function spaces $\mathcal{L}^p(\mu)$ as normed, complete spaces (Proposition 4.8 and Remark 4.4), which also leads to a new concept of convergence. Furthermore, the space $\mathcal{L}^2(\mu)$ will play a special role. It is equipped with a scalar product (namely $\langle f, g\rangle := \mu[fg]$), so general statements are available here, such as the Riesz-Fréchet's theorem (Proposition 4.11). We will use this to characterise σ-finite measures with density by the Radon-Nikodým Theorem (Corollary 4.17).

4.1 Basics

We have already mentioned the spaces $\mathcal{L}^p(\mu)$ in Remark 3.18. By defining the integral in the last section, we can now take a closer look at them. In particular, we show the important Hölder and Minkowski inequalities; see Proposition 4.2. Note that the notation $||.||$ in (4.1) is reminiscent of a norm. As we will discuss in Remark 4.4, it is almost true that $\mathcal{L}^p$, equipped with $||.||_p$ for $p \geq 1$, is a normed space.

Definition 4.1 (**$\mathcal{L}^p(\mu)$**-spaces)**.** *Let $0 < p \leq \infty$. We set*

$$\mathcal{L}^p := \mathcal{L}^p(\mu) := \{f : \Omega \to \overline{\mathbb{R}} \text{ measurable with } ||f||_p < \infty\}$$

for

$$||f||_p := (\mu[|f|^p])^{1/p}, \qquad 0 < p < \infty \tag{4.1}$$

and

$$||f||_\infty := \inf\{K : \mu(|f| > K) = 0\}.$$

On the spaces $\mathcal{L}^p$, $p \geq 1$ we now show a triangle inequality, the Minkowski inequality. It should also be noted that the Hölder inequality in the special case $p = q = 2$ is also called the Cauchy-Schwartz inequality.

Proposition 4.2 (Hölder's and Minkowski's inequality)**.** *Let f, g be measurable.*

1. *Let $0 < p, q, r \leq \infty$ such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Then,*

$$||fg||_r \leq ||f||_p ||g||_q \qquad \text{(Hölder inequality)} \tag{4.2}$$

2. *For $1 \leq p \leq \infty$,*

$$||f + g||_p \leq ||f||_p + ||g||_p. \qquad \text{(Minkowski inequality)} \tag{4.3}$$

Proof. We start with the proof of Hölder's inequality. In the case $p = \infty$ or $q = \infty$, the statement is clear, so let $p, q < \infty$. If either $||f||_p = 0$, $||f||_p = \infty$, $||g||_q = 0$ or $||g||_q = \infty$, the statement is clear as well. Let $f, g \geq 0$ and $0 < ||f|||_p, ||g|||_q < \infty$ and

$$\widetilde{f} := \frac{f}{||f||_p}, \qquad \widetilde{g} = \frac{g}{||g||_q}.$$

Then, we have to show that $||\widetilde{f}\widetilde{g}||_r \le 1$. Due to the convexity of the exponential function

$$(xy)^r = \exp\left(\tfrac{r}{p} p \log x + \tfrac{r}{q} q \log y\right) \le \tfrac{r}{p} x^p + \tfrac{r}{q} y^q,$$

and thus

$$||\widetilde{f}\widetilde{g}||_r^r = \mu[(\widetilde{f}\widetilde{g})^r] \le \tfrac{r}{p}\mu[\widetilde{f}^p] + \tfrac{r}{q}\mu[\widetilde{g}^q] = 1$$

and the assertion follows.

To prove Minkowski's inequality, we first note that in the cases $p = 1$ and $p = \infty$ the assertion is clear. In the case $1 < p < \infty$, $q = p/(p-1)$ and $r = 1/p + 1/q = 1$ with Hölder's inequality

$$\begin{aligned} ||f+g||_p^p &\le \mu[|f| \cdot |f+g|^{p-1}] + \mu[|g| \cdot |f+g|^{p-1}] \\ &\le ||f||_p \cdot ||(f+g)^{p-1}||_q + ||g||_p \cdot ||(f+g)^{p-1}||_q \\ &= (||f||_p + ||g||_p) \cdot ||f+g||_p^{p-1}, \end{aligned}$$

since $||(f+g)^{p-1}||_q = ||(f+g)^{q(p-1)}||_1^{1/q} = ||(f+g)^p||_1^{(p-1)/p} = ||f+g||_p^{p-1}$. Dividing by $||f+g||_p^{p-1}$ gives the result. □

Proposition 4.3 (Relationship between $\mathcal{L}^r$ and $\mathcal{L}^q$)**.** *Let μ be finite and $1 \le r < q \le \infty$. Then $\mathcal{L}^q(\mu) \subseteq \mathcal{L}^r(\mu)$.*

Proof. The assertion is clear for $q = \infty$. So let $q < \infty$. We use Hölder's inequality. It applies to $f \in \mathcal{L}^q$, since $||1||_p < \infty$ due to the finiteness of μ,

$$||f||_r = ||1 \cdot f||_r \le ||1||_p \cdot ||f||_q < \infty \tag{4.4}$$

for $\frac{1}{p} = \frac{1}{r} - \frac{1}{q} > 0$, from which the assertion immediately follows. □

Remark 4.4 ($\mathcal{L}^p(\mu)$ as a normed space)**.** *For every $p > 0$, we have $||af||_p = |a| \cdot ||f||_p$ for $a \in \mathbb{R}$. Together with Minkowski's inequality (which we have only shown for $1 \le p \le \infty$), this means that $\mathcal{L}^p(\mu)$ is a real vector space. It is crucial to note that the mapping $f \mapsto ||f||_p$ is a pseudo-norm, but not a full norm.*[9] *Indeed, because $||f||_p = 0$ according to Proposition 3.21 only implies that $\mu(f \ne 0) = 0$, but not that $f = 0$, we have $f \ne 0$ with $||f||_p = 0$. In the following, we will therefore identify functions f and g if $f = g$ applies μ almost everywhere. (More precisely, we introduce equivalence classes, where for $f \in \mathcal{L}^p$, the set $\{g \in \mathcal{L}^p : f = g$ almost everywhere$\}$ is the equivalence class of f.) According to the above, $(\{$ equivalence class of $f : f \in \mathcal{L}^p\}, ||\cdot||_p)$ is a normalised space. We will show below that $||\cdot||_p$ is complete (Proposition 4.8), so $(\mathcal{L}^p, ||\cdot||_p)$ is even a Banach space for every $1 \le p \le \infty$. However, we will not make the distinction between $f \in \mathcal{L}^p(\mu)$ and its equivalence class in the sequel.*

Remark 4.5 (Counterexample for $\boldsymbol{\sigma}$-finite $\boldsymbol{\mu}$)**.** *We stress that Proposition 4.3 does not hold if μ is not finite. For example, let λ be the one-dimensional Lebesgue measure and $f : x \mapsto \frac{1}{x} \cdot 1_{x>1}$. Then $f \in \mathcal{L}^2(\lambda)$, but $f \notin \mathcal{L}^1(\lambda)$.*

[9] If V is a real vector space, a mapping $||\cdot|| : V \to \mathbb{R}$ is called *norm* if (i) $||x|| = 0$ iff $x = 0$, (ii) $||a \cdot x|| = |a| \cdot ||x||$ for all $a \in \mathbb{R}$ and $x \in V$, and (iii) $||x+y|| \le ||x|| + ||y||$ for all $x, y \in V$. Then the pair $(V, ||\cdot||)$ is called a normed space. If (i) fails, $||\cdot||$ is called *pseudo-norm*.

4.2 $\mathcal{L}^p$-convergence

We have seen in the theorem of dominated convergence (Theorem 3.28) that for a sequence of functions that converges almost everywhere, their integrals often converge as well. The $\mathcal{L}^p$-convergence considered here now assumes convergence of integrals. We will see that the resulting notion of convergence means that every Cauchy sequence (with respect to $||.||_p$), see Definition 4.1) converges (Proposition 4.8).

Definition 4.6 (Convergence in the $\boldsymbol{p}$-th mean)**.** *A sequence $f_1, f_2, \dots$ in $\mathcal{L}^p(\mu)$* converges to *$f \in \mathcal{L}^p(\mu)$ iff*

$$||f_n - f||_p \xrightarrow{n\to\infty} 0.$$

We then also write $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} f$.

Proposition 4.7 (Convergence in $\mathcal{L}^p$ and in $\mathcal{L}^q$)**.** *Let $\mu(\Omega) < \infty$, $1 \leq r < q \leq \infty$ and $f, f_1, f_2, \dots \in \mathcal{L}^q$. If $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^q} f$, then also $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^r} f$.*

Proof. The assertion is clear for $q = \infty$, so let $q < \infty$. From (4.4) we have $||f-g||_r \leq ||f-g||_q$, from which the assertion already follows. □

Proposition 4.8 (Completeness of $\mathcal{L}^p$)**.** *Let $p \geq 1$ and $f_1, f_2, \dots$ be a Cauchy sequence in $\mathcal{L}^p$. (That is, for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that $||f_n - f_m||_p < \varepsilon$ for all $m, n \geq N$.) Then there is an $f \in \mathcal{L}^p$ with $||f_n - f||_p \xrightarrow{n\to\infty} 0$.*

Proof. Let $\varepsilon_1, \varepsilon_2, \dots$ be summable, e.g. $\varepsilon_n := 2^{-n}$. Since $f_1, f_2, \dots$ is a Cauchy sequence, there is an index n_k for each k with $||f_m - f_n||_p \leq \varepsilon_k$ for all $m, n \geq n_k$. In particular, the following applies

$$\sum_{k=1}^{\infty} ||f_{n_{k+1}} - f_{n_k}||_p \leq \sum_{k=1}^{\infty} \varepsilon_k < \infty.$$

With monotone convergence and Minkowski's inequality,

$$\Big|\Big| \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| \Big|\Big|_p \leq \sum_{k=1}^{\infty} ||f_{n_{k+1}} - f_{n_k}||_p < \infty.$$

In particular $\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty$ almost everywhere, i.e. for almost all $\omega \in \Omega$, the sequence $f_{n_1}(\omega), f_{n_2}(\omega), \dots$ is Cauchy in $\mathbb{R}$. Thus, there is a measurable mapping f with $f_{n_k} \xrightarrow{k\to\infty} f$ almost everywhere. According to Fatou's lemma

$$||f_n - f||_p \leq \liminf_{k\to\infty} ||f_{n_k} - f_n||_p \leq \sup_{m\geq n} ||f_m - f_n||_p \xrightarrow{n\to\infty} 0,$$

i.e. $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} f$. □

4.3 The space $\mathcal{L}^2$

Recall from Remark 4.4, that $\mathcal{L}^p(\mu)$ is in fact a Banach space for all $p \geq 1$. Let us consider the special case $p = 2$. We define a mapping $\mathcal{L}^2 \times \mathcal{L}^2 \to \mathbb{R}$ by

$$\langle f, g \rangle := \mu[fg].$$

Then $\langle\cdot,\cdot\rangle$ is obviously linear, symmetric and positive semi-definite, i.e. a scalar product[10] Consequently, we write $f \perp g$ if and only if $\mu[fg] = 0$. Using

$$||f|| := ||f||_2 = \langle f, f\rangle^{1/2}$$

in this section, $(\mathcal{L}^2, \langle\cdot,\cdot\rangle)$ is therefore a Hilbert space.

Lemma 4.9 (Parallelogram identity)**.** *Let be $f, g \in \mathcal{L}^2$. Then*

$$||f+g||^2 + ||f-g||^2 = 2||f||^2 + 2||g||^2.$$

Proof. From the definition of $||\cdot||$ and the symmetry and bilinearity of $\langle\cdot,\cdot\rangle$,

$$||f+g||^2 + ||f-g||^2 = \langle f+g, f+g\rangle + \langle f-g, f-g\rangle = 2\langle f, f\rangle + 2\langle g, g\rangle = 2||f||^2 + 2||g||^2.$$

□

Proposition 4.10 (Decomposition of $\boldsymbol{f \in \mathcal{L}^2}$)**.** *Let M be a closed, linear subspace of $\mathcal{L}^2$. Then every function $f \in \mathcal{L}^2$ has an almost everywhere unique decomposition $f = g + h$ with $g \in M, h \perp M$.*

Proof. For $f \in \mathcal{L}^2$, we define

$$d_f := \inf_{g \in M}\{||f-g|||\}.$$

Choose $g_1, g_2, \dots$ with $||f - g_n|| \xrightarrow{n\to\infty} d_f$. According to the parallelogram identity

$$4d_f^2 + ||g_m - g_n||^2 \le ||2f - g_m - g_n||^2 + ||g_m - g_n||^2 = 2||f - g_m||^2 + 2||f - g_n||^2 \xrightarrow{m,n\to\infty} 4d_f^2.$$

Thus $||g_m - g_n||^2 \xrightarrow{m,n\to\infty} 0$, i.e. $g_1, g_2, \dots$ is a Cauchy sequence. According to Proposition 4.8, there is some $g \in \mathcal{L}^2$ with $||g_n - g|| \xrightarrow{n\to\infty} 0$. Since M is closed, we find $g \in M$ as well as $||h|| = d_f$ for $h := f - g$. So, for all $t > 0, l \in M$, due to the definition of d_f,

$$d_f^2 \le ||h + tl||^2 = d_f^2 + 2t\langle h, l\rangle + t^2||l||^2.$$

Since this applies to all t, $\langle h, l\rangle = 0$, i.e. $h \perp M$.

To prove uniqueness, let $g' + h'$ be a further decomposition of f. Then, due to the linearity of M, on the one hand $g - g' \in M$, on the other hand, almost everywhere, $g - g' = h - h' \perp M$, i.e. $g - g' \perp g - g'$. This means $||g - g'|| = \langle g - g', g - g'\rangle = 0$, i.e. $g = g'$ almost everywhere. □

Proposition 4.11 (Riesz-Fréchet)**.** *A mapping $F : \mathcal{L}^2 \to \mathbb{R}$ is continuous and linear if and only if there exists some $h \in \mathcal{L}^2$ with*

$$F(f) = \langle f, h\rangle, \qquad f \in \mathcal{L}^2.$$

Then, $h \in \mathcal{L}^2$ is almost everywhere uniquely determined.

[10]If V is a real vector space. Then a mapping is called $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ is a *scalar product* if (i) $\langle x, \alpha y + z\rangle = \alpha\langle x, y\rangle + \langle x, z\rangle$ for all $x, y, z \in V$ and $\alpha \in \mathbb{R}$ (linearity), (ii) $\langle x, y\rangle = \langle y, x\rangle$ (symmetry) and (iii) $\langle x, x\rangle > 0$ for every $x \in V \setminus \{0\}$ (positive definiteness). The norm $||x|| := \langle x, x\rangle^{1/2}$ on V is defined by a scalar product. If $(V, ||\cdot||)$ is complete, then $(V, \langle\cdot,\cdot\rangle)$ is called an *Hilbert space.*

Proof. '⇐': The linearity of $f \mapsto \langle f, h\rangle$ follows from the bilinearity of $\langle\cdot,\cdot\rangle$. The continuity follows from the Cauchy-Schwartz inequality using

$$|\langle |f - f'|, h\rangle| \leq ||f - f'|| \cdot ||h||.$$

'⇒': If $F \equiv 0$, choose $h = 0$. If $F \not\equiv 0$, $M = F^{-1}\{0\}$ is (due to the continuity of F) a closed and (due to the linearity of F) linear subspace of $\mathcal{L}^2$. Choose $f' \in \mathcal{L}^2 \setminus M$ with the (according to Proposition 4.10 almost everywhere unique) orthogonal decomposition $f' = g' + h'$ with $g' \in M$ and $h' \perp M$. Since $f' \notin M$, we have $h' \not\equiv 0$, and $F(h') = F(f') - F(g') = F(f') \neq 0$. We set $h'' = \frac{h'}{F(h')}$, so that $h'' \perp M$ and $F(h'') = 1$ as well as, for all $f \in \mathcal{L}^2$

$$F(f - F(f)h'') = F(f) - F(f)F(h'') = 0.$$

i.e. $f - F(f)h'' \in M$, in particular $\langle F(f)h'', h''\rangle = \langle f, h''\rangle$ and

$$F(f) = \frac{1}{||h''||^2} \cdot \langle F(f)h'', h''\rangle = \frac{1}{||h''||^2} \cdot \langle f, h''\rangle = \langle f, \frac{h''}{||h''||^2}\rangle.$$

Now, the assertion follows with $h := \frac{h''}{||h''||^2}$.

For uniqueness, let $\langle f, h_1 - h_2\rangle = 0$ for all $f \in \mathcal{L}^2$; in particular, with $f = h_1 - h_2$

$$||h_1 - h_2||^2 = \langle h_1 - h_2, h_1 - h_2\rangle = 0,$$

thus $h_1 = h_2$ μ-almost everywhere. □

Remark 4.12 (Generality of the last statements)**.** *Lemma 4.9, as well as the propositions 4.10 and 4.11 also apply if $\mathcal{L}^2$ is replaced by any other Hilbert space.*

4.4 Theorem of Radon-Nikodým

Probability measures with density are already known from the lecture *Elementare probability 1*. This concept is now taken up and embedded in the context of integrals. Let ν be another measure on $\mathcal{F}$. The aim is to specify conditions when the measure ν can be represented by a density. The answer can be found in the Radon-Nikodým theorem (Corollary 4.17). It is a special case of Lebesgue's decomposition theorem, Theorem 4.16. This shows that for every two σ-finite measures μ, ν, the measure ν can be (additively) decomposed into two parts: one absolute continuous with respect to μ and one singular with respect to μ. The absolutely continuous part has a density with respect to μ. First we have to explain all terms.

Definition 4.13 (Absolutely continuous measures)**.**

1. *We say that ν has a* density *f with respect to μ if for all $A \in \mathcal{F}$*

   $$\nu(A) = \mu[f; A].$$

   We then write $f = \frac{d\nu}{d\mu}$ and $\nu = f \cdot \mu$.

2. *The measure ν is called* absolutely continuous with respect to μ *if all μ-zero sets are also ν-zero sets. We then write $\nu \ll \mu$. If both $\nu \ll \mu$ and $\mu \ll \nu$, then μ and ν are called* equivalent.

3. *The measures μ and ν are called* singular *if there is an $A \in \mathcal{F}$ with $\mu(A) = 0$ and $\nu(A^c) = 0$. We then write $\mu \perp \nu$.*

Lemma 4.14 (Chain rule and uniqueness)**.** *Let μ be a measure on $\mathcal{F}$.*

1. *Let ν be a σ-finite measure. If g_1 and g_2 are densities of ν with respect to μ, then $g_1 = g_2$, μ-almost everywhere.*
2. *Let $f : \Omega \to \mathbb{R}_+$ and $g : \Omega \to \mathbb{R}$ be measurable. Then,*

$$(f \cdot \mu)[g] = \mu[fg],$$

if one of the two sides exists.

Proof. 1. Let $\Omega_1, \Omega_2, \dots \in \mathcal{F}$ be such that $\Omega_n \uparrow \Omega$ and $\nu(\Omega_n) < \infty$. Set $A_n := \Omega_n \cap \{g_1 > g_2\}$. Since both g_1 and g_2 are densities of ν with respect to μ,

$$\mu[g_1 - g_2; A_n] = 0.$$

Since only $g_1 > g_2$ is possible on A_n, $g_1 = g_2$ is $1_{A_n}\mu$-almost everywhere. Furthermore,

$$\mu\{g_1 > g_2\} = \mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = 0.$$

Analogously, $\mu\{g_1 < g_2\} = 0$ and thus $g_1 = g_2$ μ-almost everywhere.

2. The statement is clear for $g = 1_A$ with $A \in \mathcal{F}$. This extends step by step to simple functions, positive measurable functions and finally to the general case. □

Example 4.15 (Known densities)**.** 1. *Some density functions are already known from the lecture* Elementary probability 1*. For example, let $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+$ be*

$$f_{N(\mu,\sigma^2)}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2}\Big)$$

and λ is the one-dimensional Lebesgue measure. Then the probability measure $f_{N(\mu,\sigma^2)} \cdot \lambda$ is called normal distribution with expected value μ and variance σ^2*. We can compute for some $X \sim N_{\mu,\sigma^2}$ and $h : x \mapsto (x-\mu)/\sigma$*

$$\mathbf{P}(h(X) \le x) = \mathbf{P}(X \le \mu + x\sigma) = \int_{-\infty}^{\mu+x\sigma} f_{N(\mu,\sigma^2)}(y)dy$$
$$\overset{z=(y-\mu)/\sigma}{=} \int_{-\infty}^{x} f_{N(0,1)}(z)dz,$$

which shows that $(X-\mu)/\sigma \sim N_{(0,1)}$.

For $\gamma \ge 0$, let

$$f_{\exp(\gamma)}(x) := 1_{x\ge 0} \cdot \gamma e^{-\gamma x},$$

the probability measure $f_{\exp(\gamma)} \cdot \lambda$ is called exponential distribution with parameter γ*. For example, you can now use Lemma 4.14 to calculate for some $X \sim \exp(\gamma)$*

$$\mathbf{E}[X] = f_{\exp(\gamma)} \cdot \lambda[id] = \int_0^\infty \gamma e^{-\gamma x} x dx = -e^{-\gamma x} x\Big|_0^\infty + \int_0^\infty e^{-\gamma x} dx = \frac{1}{\gamma}.$$

So, we have computed the expected value of the exponential distribution for the parameter γ.

2. *Of course, there are not only densities with respect to the Lebesgue measure. Let, for example*
$$\mu = \sum_{n=0}^{\infty} \delta_n$$
be the counting measure on $\mathbb{N}_0$ *(see Example 2.2) and* $f : \mathbb{N}_0 \to \mathbb{R}_+$*, given for a* $\gamma \geq 0$ *by*
$$f(k) = e^{-\gamma} \frac{\gamma^k}{k!}.$$
Then $f \cdot \mu$ *is the Poisson distribution for the parameter* γ *on* $2^{\mathbb{N}_0}$ *according to Example 2.2.*

Theorem 4.16 (Lebesgue's decomposition theorem)**.** *Let* μ, ν *be* σ*-finite measures on* $(\Omega, \mathcal{F})$*. Then* ν *can be written uniquely as*
$$\nu = \nu_a + \nu_s \qquad \textit{with} \qquad \nu_a \ll \mu, \nu_s \perp \mu.$$
The measure ν_a *has a density with respect to* μ *that is* μ*-almost everywhere finite.*

Proof. Since μ, ν are σ-finite, we find $\Omega_1, \Omega_2, \dots \in \mathcal{F}$ with $\Omega_n \uparrow \Omega$ and $\nu(\Omega_n), \mu(\Omega_n) < \infty$. In particular, without loss of generality, we can assume that μ, ν are finite measures. With Proposition 4.7. the linear mapping
$$\begin{cases} \mathcal{L}^2(\mu + \nu) & \to \mathbb{R}) \\ f & \mapsto \nu[f] \end{cases}$$
is continuous. According to Proposition 4.11, there is some $h \in \mathcal{L}^2(\mu + \nu)$ with
$$\nu[f] = (\mu + \nu)[fh], \tag{4.5}$$
thus
$$\nu[f(1 - h)] = \mu[fh] \tag{4.6}$$
for each $f \in \mathcal{L}^2(\mu + \nu)$. If one chooses $f = 1_{\{h<0\}}$ in (4.5), we find
$$0 \leq \nu\{h < 0\} = (\mu + \nu)[h; h < 0] \leq 0,$$
i.e. $h \geq 0$ $(\mu + \nu)$-almost everywhere. Similarly, $f = 1_{\{h>1\}}$ can be used to deduce from (4.6) that
$$0 \leq \mu[h; \{h > 1\}] = \nu[1 - h; \{h > 1\} \leq 0,$$
so $h \leq 1$ $(\mu + \nu)$-almost everywhere. Now, let $f \geq 0$ be measurable and $f_1, f_2, \dots \in \mathcal{L}^2(\mu + \nu)$ with $f_n \uparrow f$. With monotone convergence,
$$\nu[f(1 - h)] = \lim_{n\to\infty} \nu[f_n(1 - h)] = \lim_{n\to\infty} \mu[f_n h] = \mu[fh],$$
i.e. (4.6) applies to all measurable $f \geq 0$.

Now let $E := h^{-1}\{1\}$. From (4.6) it follows with $f = 1_E$ that
$$\mu(E) = \mu[h; E] = \nu[1 - h; E] = 0.$$

We define two measures ν_a and ν_s for $A \in \mathcal{F}$ by

$$\nu_a(A) = \nu(A \setminus E), \qquad \nu_s(A) = \nu(A \cap E),$$

so that $\nu = \nu_a + \nu_s$ and $\nu_s \perp \mu$. To show that $\nu_a \ll \mu$ choose $A \in \mathcal{F}$ with $\mu(A) = 0$. This means that after (4.6)

$$\nu[1-h; A \setminus E] = \mu[h; A \setminus E] = 0.$$

Since $h < 1$ on $A \setminus E$, $\nu_a(A) = \nu(A \setminus E) = 0$, i.e. $\nu_a \ll \mu$.

We claim that $g := \frac{h}{1-h} 1_{\Omega \setminus E}$ is the density of ν_a with respect to μ. Indeed, using (4.6),

$$\mu[g; A] = \mu\Big[\frac{h}{1-h}; A \setminus E\Big] = \nu(A \setminus E) = \nu_a(A).$$

To show the uniqueness of the decomposition, let $\nu = \nu_a + \nu_s = \widetilde{\nu}_a + \widetilde{\nu}_s$ for $\nu_a, \widetilde{\nu}_a \ll \mu$, $\nu_s, \widetilde{\nu}_s \perp \mu$. Choose $A, \widetilde{A} \in \mathcal{A}$ with $\nu_s(A) = \mu(A^c) = \widetilde{\nu}_s(\widetilde{A}) = \mu(\widetilde{A}^c) = 0$. Then,

$$\nu_s(A \cap \widetilde{A}) = \widetilde{\nu}_s(A \cap \widetilde{A}) = \nu_a(A^c \cup \widetilde{A}^c) = \widetilde{\nu}_a(A^c \cup \widetilde{A}^c) = 0$$

and therefore

$$\begin{aligned} \nu_a &= 1_{A \cap \widetilde{A}} \cdot \nu_a = 1_{A \cap \widetilde{A}} \cdot \nu = 1_{A \cap \widetilde{A}} \cdot \widetilde{\nu}_a = \widetilde{\nu}_a, \\ \nu_s &= \nu - \nu_a = \nu - \widetilde{\nu}_a = \widetilde{\nu}_s. \end{aligned}$$

□

Corollary 4.17 (Theorem of Radon-Nikodým)**.** *Let μ and ν be σ-finite measures. Then, ν has a density with respect to μ if and only if $\nu \ll \mu$.*

Proof. '⇒': clear.
'⇐': According to Theorem 4.16, there is a unique decomposition $\nu = \nu_a + \nu_s$ with $\nu_a \ll \mu, \nu_s \perp \mu$. Since $\nu \ll \mu$, $\nu_s = 0$ must apply and therefore $\nu = \nu_a$. In particular, the density of ν exists with respect to μ. □

Example 4.18. *In Lebesgue's decomposition Theorem 4.16 and in the Theorem of Radon-Nikodým 4.17, the condition that μ and ν are σ-finite cannot be omitted, as the following example shows:*

Let $(\Omega, \mathcal{F})$ be a measure space with uncountable Ω and

$$\mathcal{F} := \{A : A \text{ or } A^c \text{ countable}\}.$$

Let μ and ν be infinite measures on $(\Omega, \mathcal{F})$, given by

$$\nu(A) := \begin{cases} 0, & A \text{ countable}, \\ \infty, & \text{otherwise}, \end{cases} \qquad \mu(A) := \begin{cases} |A|, & A \text{ finite}, \\ \infty, & \text{otherwise}. \end{cases}$$

Then obviously $\nu \ll \mu$. Assume there is a $\mathcal{F}$-measurable density of ν with respect to μ. Then, for all $\omega \in \Omega$

$$0 = \nu\{\omega\} = \mu[f; \{\omega\}] = f(\omega)\mu(\{\omega\}) = f(\omega).$$

Thus $f = 0$ and $\nu = 0$ would contradict the definition of ν.

5 Product spaces

Let $(\Omega_i)_{i\in I}$ be a family of sets. Then,

$$\Omega := \bigtimes_{i\in I} \Omega_i := \{(\omega_i)_{i\in I} : \omega_i \in \Omega_i\}$$

is the product space of $(\Omega_i)_{i\in I}$. We further define the projections for $H \subseteq J \subseteq I$

$$\pi_H^J : \bigtimes_{i\in J} \Omega_i \to \bigtimes_{i\in H} \Omega_i,$$

as well as $\pi_H := \pi_H^I$ and $\pi_i := \pi_{\{i\}}$, $i \in I$. In this chapter, we will apply all the concepts in the context of measurability. Of particular importance is the theorem on projective limits of probability measures, Theorem 5.24, which will play a fundamental role in the theory of stochastic processes.

5.1 Topology

We start with the definition of a topology on product spaces. In short, this topology is made such that projections are continuous.

Definition 5.1 (Product space and product topology). *If $(\Omega_i, \mathcal{O}_i)_{i\in I}$ is a family of topological spaces, then the topology $\mathcal{O}$, generated by (recall from Definition A.1.7)*[11]

$$\mathcal{C} := \{A_i \times \bigtimes_{j\in I, j\neq i} \Omega_j ; i \in I, A_i \in \mathcal{O}_i\}$$

is called the product topology *on Ω.*

Remark 5.2 (Continuity of projections). *All projections $\pi_i, i \in I$ are continuous with respect to the product topology.*
Indeed, it is

$$\pi_i^{-1}(A_i) = A_i \times \bigtimes_{I\ni j\neq i} \Omega_j \in \mathcal{C} \subseteq \mathcal{O}$$

for $A_i \in \mathcal{O}_i$. The projection is therefore continuous (see Definition A.1.10).

5.2 Semi-rings, rings and σ-algebras

Analogous to topology, the product σ-algebra is just such that projections are measurable functions.

Definition 5.3 (Product-$\boldsymbol{\sigma}$-algebra). *If $(\Omega_i, \mathcal{F}_i)_{i\in I}$ is a family of measurable spaces, the σ-algebra*

$$\bigotimes_{i\in I} \mathcal{F}_i := \sigma(\mathcal{E}), \qquad \mathcal{E} := \{A_i \times \bigtimes_{j\in I, j\neq i} \Omega_j : i \in I, A_i \in \mathcal{F}_i\} \tag{5.1}$$

is called the product-σ-algebra *on $\Omega := \bigtimes_{i\in I} \Omega_i$. If $(\Omega_i, \mathcal{F}_i) = (\Omega, \mathcal{F}), i \in I$, we set $\mathcal{F}^I := \bigotimes_{i\in I} \mathcal{F}$.*

[11] We write $A \subseteq_f B$ if $\subseteq B$ and A is finite.

Remark 5.4 (Measurability of projections). *Analogous to the product topology, the projections π_i are measurable with respect to $\bigotimes_{i\in I}\mathcal{F}_i$. This is because for $A_i \in \mathcal{F}_i$,*

$$\pi_i^{-1}(A_i) = A_i \times \bigtimes_{I\ni j\neq i} \Omega_j \in \bigotimes_{i\in I} \mathcal{F}_i.$$

Lemma 5.5 (Product-σ-algebra for countable products). *Let I be arbitrary and $(\Omega_i, \mathcal{O}_i)_{i\in I}$ a family of topological spaces, and $(\Omega, \mathcal{O})$ the product space, equipped with the product topology from Definition 5.1. Then $\bigotimes_{i\in I}\mathcal{B}(\Omega_i) \subseteq \mathcal{B}(\Omega)$. Moreover, if I is countable and $(\Omega_i, \mathcal{O}_i)_{i\in I}$ a family of separable metric spaces, then $\mathcal{B}(\Omega) = \bigotimes_{i\in I}\mathcal{B}(\Omega_i)$. In particular, $\mathcal{B}(\mathbb{R}^d) = \bigotimes_{i=1}^d \mathcal{B}(\mathbb{R})$.*

Proof. Let $\mathcal{C}$ be as in Definition 5.1, $\mathcal{O}$ the product topology (i.e. $\sigma(\mathcal{O}) = \mathcal{B}(\Omega)$), and $\mathcal{E}$ as in Definition 5.3 with $\mathcal{F}_i$ replaced by $\mathcal{B}(\Omega_i)$. Clearly, $\mathcal{C} \subseteq \mathcal{O}(\mathcal{C})$ as well as $\mathcal{C} \subseteq \mathcal{E}$ by definition. In addition, $\mathcal{E} \subseteq \sigma(\mathcal{C})$ by definition of $\mathcal{B}(\Omega_i)$. This leads to

$$\bigotimes_{i\in I}\mathcal{B}(\Omega_i) = \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{C}) \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\Omega).$$

In case of a countable union of separable spaces, every set in $\mathcal{O}(\mathcal{C})$ is a countable union of sets in $\mathcal{C}$ (see Lemma 1.8), leading to

$$\mathcal{O}(\mathcal{C}) \subseteq \sigma(\mathcal{C}), \qquad \text{so} \qquad \sigma(\mathcal{O}(\mathcal{C})) \subseteq \sigma(\sigma(\mathcal{C})) = \sigma(\mathcal{C}).$$

Hence, all assertions are shown. □

Remark 5.6. *If I is uncountable, by using countable intersections and unions, $\bigotimes_{i\in I}\mathcal{B}(\Omega_i)$ only contains sets which depends on a countable number of coordinates. In contrast, $\sigma(\mathcal{O}(\mathcal{B}))$ contains sets which arise as uncountable intersections of closed sets, which in general depend on an uncountable number of coordinates. This shows that for uncountable product spaces, in general $\bigotimes_{i\in I}\mathcal{B}(\Omega_i) \subsetneq \mathcal{B}(\Omega))$.*

Lemma 5.7 (Products of generators/semi-rings are generators/semi-rings). *Let $(\Omega_i, \mathcal{F}_i)$ be measurable spaces and $\Omega = \bigtimes_{i\in I}\Omega_i$.*

1. *Let I be finite and $\mathcal{H}_i$ a semi-ring with $\sigma(\mathcal{H}_i) = \mathcal{F}_i$. Then*

$$\mathcal{H} := \{\bigtimes_{i\in I} A_i : A_i \in \mathcal{H}_i, i \in I\} \tag{5.2}$$

   is a semi-ring with $\sigma(\mathcal{H}) = \bigotimes_{i\in I}\mathcal{F}_i$.

2. *Let I be arbitrary and $\mathcal{H}_i$ a $\cap$-stable generator of $\mathcal{F}_i$, $i \in I$. Then*

$$\mathcal{H} := \{\bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J}\Omega_i : J \subseteq_f I, A_i \in \mathcal{H}_i, i \in J\}$$

   is a $\cap$-stable generator of $\bigotimes_{i\in I}\mathcal{F}_i$.

Proof. For 1., let $I = \{1, \dots, d\}$ without loss of generality. It is clear that $\mathcal{H}$ is $\cap$-stable. Property (ii) for semi-rings is shown by induction over d. The assertion is clear for $d = 1$, since $\mathcal{H}_1$ is a semi-ring. If it holds to $d-1$, then

$$\begin{aligned}&(A_1 \times \cdots \times A_d) \setminus (B_1 \times \cdots \times B_d)\\ &\quad = \big(A_1 \times \cdots \times A_{d-1} \times (A_d \setminus B_d)\big) \uplus \big((A_1 \times \cdots \times A_{d-1}) \setminus (B_1 \times \cdots \times B_{d-1})\big) \times (A_d \cap B_d)\end{aligned}$$

The first term of the last line can be represented as a disjoint union of sets from $\mathcal{H}$, since $\mathcal{H}_d$ is a semi-ring. The second term can be represented as a disjoint union, since by the induction hypothesis, $(A_1 \times \cdots \times A_{d-1}) \setminus (B_1 \times \cdots \times B_{d-1})$ can be represented as a disjoint union of sets of the form $H_1 \times \cdots \times H_{d-1}$ with $H_i \in \mathcal{H}_i, i = 1, \dots, d-1$.

For 2. it is again clear that $\mathcal{H}$ is $\cap$-stable. From (5.1) it immediately follows that $\mathcal{H} \subseteq \bigotimes_{i\in I} \mathcal{F}_i$, therefore $\sigma(\mathcal{H}) \subseteq \bigotimes_{i\in I} \mathcal{F}_i$. Conversely, it is clear that for $A_i \in \mathcal{F}_i$

$$A_i \times \bigtimes_{j\neq i} \Omega_j \in \sigma\Big(\big\{A_i \times \bigtimes_{j\neq i} \Omega_j : A_i \in \mathcal{H}_i\big\}\Big) \subseteq \sigma(\mathcal{H}),$$

from which $\bigotimes_{i\in I} \mathcal{F}_i \subseteq \sigma(\mathcal{H})$ and thus the assertion follows. □

Corollary 5.8 (Borel's $\boldsymbol{\sigma}$-algebra on $\mathbb{R}^{\boldsymbol{d}}$ is generated by cylinders)**.** *Let $\Omega = \mathbb{R}^d$. For $\underline{a} = (a_1, \dots, a_d), \underline{b} = (b_1, \dots, b_d) \in \mathbb{R}^d$ we set $\underline{a} \le \underline{b}$ if and only if $a_i \le b_i, i = 1, \dots, d$, and with*

$$(\underline{a}, \underline{b}] = (a_1, b_1] \times \cdots \times (a_d, b_d]$$

the half-open cylinder. Then,

$$\mathcal{H} := \{(\underline{a}, \underline{b}] : \underline{a}, \underline{b} \in \mathbb{Q}, \underline{a} \le \underline{b}\}$$

is a semi-ring with $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R}^d)$.

Proof. According to Example 1.3.1 and Lemma 5.7.1, $\mathcal{H}$ is a semi-ring that generates $\bigotimes_{i=1}^d \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d)$; see Lemma 5.5. □

5.3 Measures and integrals

Integrals in multi-dimensional spaces are already known from calculus. We now first define measures on product spaces and the corresponding (multiple) integrals. Fubini's theorem (Theorem 5.13) can then be used to interpret and analyse integrals according to measures on product spaces as multiple integrals. For this purpose, it is necessary that the integrands appearing in the multiple integrals are measurable. This is ensured in Lemma 5.11. In order to be able to define measures on product spaces in sufficient generality, we first need the concept of the transition kernel.

Definition 5.9 (Transition kernel)**.** *Let $(\Omega_i, \mathcal{F}_i), i = 1, 2$ be measurable spaces. A mapping $\kappa : \Omega_1 \times \mathcal{F}_2 \to \mathbb{R}_+$ is called a* transition kernel from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$ *if (i) for all $\omega_1 \in \Omega_1$, the map $\kappa(\omega_1, .)$ is a measure on $\mathcal{F}_2$ and (ii) for all $A_2 \in \mathcal{F}_2$ $\kappa(., A_2)$ is $\mathcal{F}_1$-measurable.*

A transition kernel is called σ-finite if there is a sequence $\Omega_{21}, \Omega_{22}, \dots \in \mathcal{F}_2$ with $\Omega_{2n} \uparrow \Omega_2$ and $\sup_{\omega_1} \kappa(\omega_1, \Omega_{2n}) < \infty$ for all $n = 1, 2, \dots$ It is called stochastic kernel *or* Markov kernel *if for all $\omega_1 \in \Omega_1$ the map $\kappa(\omega_1, .)$ is a probability measure.*

Example 5.10 (Markov chain). *Let $\Omega = \{\omega_1, \dots, \omega_n\}$ be finite and $P = (p_{ij})_{1\leq i,j\leq n}$ with $p_{ij} \in [0,1]$ and $\sum_{j=1}^n p_{ij} = 1$. Then,*

$$\kappa(\omega_i, .) := \sum_{j=1}^n p_{ij} \cdot \delta_{\omega_j}$$

is a Markov kernel from $(\Omega, 2^\Omega)$ to $(\Omega, 2^\Omega)$. Here, P as a stochastic matrix is the transition matrix of a homogeneous, Ω-valued Markov chain.

Lemma 5.11 (Measurability of integrable sections). *Let $(\Omega_i, \mathcal{F}_i), i = 1,2$ be measurable spaces, κ be a σ-finite transition kernel from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$ and $f : \Omega_1 \times \Omega_2 \to \mathbb{R}_+$ to $\mathcal{F}_1 \otimes \mathcal{F}_2$ measurable. Then*

$$\omega_1 \mapsto \kappa(\omega_1, .)[f] := \int \kappa(\omega_1, d\omega_2) f(\omega_1, \omega_2)$$

to $\mathcal{F}_1$-measurable.

Proof. We assume that $\kappa(\omega_1, \Omega_2) < \infty$ for all $\omega_1 \in \Omega_1$. (The general case is then performed using a sequence $\Omega_{11}, \Omega_{12}, \dots \in \mathcal{F}_1$ with $\Omega_{1n} \uparrow \Omega_1$.) Let

$$\mathcal{D} := \{A \in \mathcal{F}_1 \otimes \mathcal{F}_2 : \omega_1 \mapsto \kappa(\omega_1, .)[1_A] \text{ is } \mathcal{F}_1\text{-measurable}\}.$$

Then it is easy to check that $\mathcal{D}$ is a $\cap$-stable Dynkin system. Furthermore, $\mathcal{H} \subseteq \mathcal{D}$, where $\mathcal{H}$ is defined as in (5.2). Thus, according to Theorem 1.13, $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{H}) \subseteq \mathcal{D} \subseteq \mathcal{F}_1 \otimes \mathcal{F}_2$. Therefore, $\omega_1 \mapsto \kappa(\omega_1, .)[1_A]$ is measurable for all $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$ with respect to $\mathcal{F}_1$. This statement can be extended immediately by using a simple function instead of 1_A. By monotonic convergence, it then also follows that $\omega_1 \mapsto \kappa(\omega_1, .)[f]$ is measurable for all measurable, non-negative functions according to $\mathcal{F}_1$. □

Theorem 5.12 (Theorem of Ionescu-Tulcea). *Let $(\Omega_i, \mathcal{F}_i), i = 0, \dots, n$ measurable spaces, μ a σ-finite measure on $\mathcal{F}_0$ and κ_i a σ-finite transition kernel of $\left(\bigtimes_{j=0}^{i-1} \Omega_j, \bigotimes_{j=0}^{i-1} \mathcal{F}_j\right)$ to $(\Omega_i, \mathcal{F}_i)$, $i = 1, \dots, n$. Then there is exactly one σ-finite measure $\mu \bigotimes_{i=1}^n \kappa_i$ on $\left(\bigtimes_{i=0}^n \Omega_i, \bigotimes_{i=0}^n \mathcal{F}_i\right)$ with*

$$\Big(\mu \bigotimes_{i=1}^n \kappa_i\Big)(A_0 \times \dots \times A_n) = \int_{A_0} \mu(d\omega_0) \Big(\int_{A_1} \kappa_1(\omega_0, d\omega_1) \cdots \Big(\int_{A_n} \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) \Big) \cdots \Big). \tag{5.3}$$

Proof. We show the theorem only for $n = 1$, the general case is then done by induction.

The proof is an application of Theorem 2.16. First we establish that according to Lemma 5.7, the set system $\mathcal{H}$ defined in (5.2) is a semi-ring on $\bigtimes_{i=1}^n \Omega_i$. We first show that the given set function is σ-finite on $\mathcal{H}$. Namely, there is $\Omega_{i1}, \Omega_{i2}, \in \mathcal{F}_i$ with $\Omega_{in} \uparrow \Omega_i, i = 0, 1$ with $\mu(\Omega_{0n}) < \infty, \kappa_1(\omega_0, \Omega_{1n}) < \infty, n = 1, 2, \dots, \omega_0 \in \Omega_0$ and $\sup_{\omega_0 \in \Omega_0} \kappa_1(\omega_0, \Omega_{1n}) =: C_n < \infty$. This means that $\mu \otimes \kappa_1(\Omega_{0n} \times \Omega_{1n}) \leq C_n \cdot \mu(\Omega_{0n}) < \infty$ and $\Omega_{0n} \times \Omega_{1n} \uparrow \Omega_0 \times \Omega_1$. This means that $\mu \otimes \kappa_1$ is also σ-finite. If we define $\widetilde{\mu}$ on $\mathcal{H}$ using (5.3), this is therefore a σ-finite set function.

We now show that $\widetilde{\mu}$ is σ-subadditive and finitely additive on $\mathcal{H}$. For $A_1, \dots, A_n \in \mathcal{H}$ and $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{H}$, by σ-subadditivity of $\kappa_1(\omega_0, .)$ for all $\omega_0 \in \Omega_0$

$$\begin{aligned}\widetilde{\mu}(A) &= \int \mu(d\omega_0) \int \kappa_1(\omega_0, d\omega_1) 1_A(\omega_0, \omega_1) \\ &\leq \sum_{n=1}^{\infty} \int \mu(d\omega_0) \int \kappa_1(\omega_0, d\omega_1) 1_{A_n}(\omega_0, \omega_1) = \sum_{n=1}^{\infty} \widetilde{\mu}(A_n).\end{aligned}$$

Similarly, finite additivity is shown. According to Lemma 2.5, $\widetilde{\mu}$ is therefore σ-additive. From Theorem 2.16 it now follows that there is exactly one extension of $\widetilde{\mu}$ to $\sigma(\mathcal{H}) = \bigotimes_{i=1}^{n} \sigma(\mathcal{H}_i)$, which is the one given in the theorem. □

We now deal with the measure defined in Theorem 5.12.

Theorem 5.13 (Fubini's theorem). *Let $(\Omega_i, \mathcal{F}_i)$, μ, κ_i and $\mu \bigotimes_{i=1}^{n} \kappa_i$ be as in Theorem 5.12. Further, let $f : \times_{i=0}^{n} \Omega_i \to \mathbb{R}_+$ be measurable with respect to $\bigotimes_{i=0}^{n} \mathcal{F}_i$. Then,*

$$\int f d\big(\mu \bigotimes_{i=0}^{n} \kappa_i\big) = \int \mu(d\omega_0) \Big(\int \kappa_1(\omega_1, d\omega_2) \cdots \Big(\int \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) f(\omega_0, \dots, \omega_n) \Big) \cdots \Big). \tag{5.4}$$

This equality also applies if $f : \times_{i=0}^{n} \Omega_i \to \mathbb{R}$ is measurable with $\int |f| d\big(\mu \bigotimes_{i=0}^{n} \kappa_i\big) < \infty$.

Proof. Consider the set function $\widetilde{\mu}$ on $\bigotimes_{i=0}^{n} \mathcal{F}_i$, given by

$$\widetilde{\mu} : A \mapsto \int \mu(d\omega_0) \Big(\int \kappa_1(\omega_1, d\omega_2) \cdots \Big(\int \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) 1_A(\omega_0, \dots, \omega_n) \Big) \cdots \Big).$$

You can see that $\widetilde{\mu}$ corresponds on $\mathcal{H}$ from (5.2) with $\mu \bigotimes_{i=1}^{n} \kappa_i$. Since $\mathcal{H}$ is $\cap$-stable, the equality (5.4) for indicator functions follows due to Proposition 2.11. By means of linearity of the integral, (5.4) is first extended to simple functions and then using monotonicity to any non-negative, measurable function. Note that all occurring integrands are measurable according to Lemma 5.11. □

Corollary 5.14 (Product measures). *Let $\Omega = \times_{i=1}^{n} \Omega_i$ and $\mathcal{H}_i \subseteq 2^{\Omega_i}$ be a semi-ring, $i = 1, \dots, n$, and $\mu_i : \mathcal{H}_i \to \mathbb{R}_+$ σ-finite and, σ-additive, $i = 1, \dots, n$. Then there is exactly one measure $\mu_1 \otimes \cdots \otimes \mu_n$ on $\bigotimes_{i=1}^{n} \sigma(\mathcal{H}_i)$ with*

$$\mu_1 \otimes \cdots \otimes \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n). \tag{5.5}$$

For a measurable function $f : \Omega \to \mathbb{R}_+$, the value of the integral does not depend on the order of integration of the coordinates $\omega, ..., \omega_n$, i.e. for every permutation π on $\{1, \dots, n\}$,

$$\int f d\mu_1 \otimes \cdots \otimes \mu_n = \int \Big(\cdots \Big(\int f(\omega_1, \dots, \omega_n) \mu_{\pi(1)}(d\omega_{\pi_1}) \Big) \cdots \Big) \mu_{\pi(n)}(d\omega_{\pi(n)}).$$

This formula also applies to $f : \Omega \to \mathbb{R}$, if $\int |f| d\mu_1 \otimes \cdots \otimes \mu_n < \infty$.

Proof. The corollary follows directly from Theorem 5.12 and Theorem 5.13 if you set $\kappa_i(\omega_0, \dots, \omega_{i-1}, .) = \mu_i(.)$ for all $\omega_0, \dots, \omega_{i-1}$. □

Definition 5.15 (Finite product measure)**.** *Consider the same situation as in Corollary 5.14. Then, the unique measure* $\mu_1 \otimes \cdots \otimes \mu_n$ *from Corollary 5.14 is called the* product measure of $\mu_1, \ldots, \mu_n$*. We also write*

$$\bigotimes_{i=1}^{n} \mu_i := \mu_1 \otimes \cdots \otimes \mu_n.$$

If $(\Omega_i, \mathcal{H}_i, \mu_i) = (\Omega_0, \mathcal{H}_i, \mu_0)$*,* $i = 1, \ldots, n$*, i.e. all spaces are equal, we also denote it by*

$$\mu_0^{\otimes n} := \mu_1 \otimes \cdots \otimes \mu_n.$$

Example 5.16 (multidimensional Lebesgue measure)**.** 1. *Let* λ *be the one-dimensional Lebesgue measure on* $\mathcal{B}(\mathbb{R})$ *from Proposition 2.18. Then* $\lambda^{\otimes d}$ *is the d-dimensional Lebesgue measure.*

2. *Let* $f : \mathbb{R}^2 \to \mathbb{R}$ *be given by*

$$f(x, y) = \frac{xy}{(x^2 + y^2)^2}.$$

Then, for every $x \in \mathbb{R}$

$$\int \lambda(dy) f(x, y) = 0,$$

since $f(x, .) \in \mathcal{L}^1(\lambda)$ *and* $f(x, y) = -f(x, -y)$*. Therefore, in particular*

$$\int \lambda(dx) \Big(\int \lambda(dy) f(x, y) \Big) = \int \lambda(dy) \Big(\int \lambda(dx) f(x, y) \Big) = 0.$$

However, $|f|$ *is not integrable with respect to* $\lambda^{\otimes 2}$ *because* f *has a non-integrable pole in* $(0, 0)$*. As this example shows, we have to be careful with multiple integrals. In particular, it does not follow from the equality and finiteness of multiple integrals that the integrand is integrable.*

5.4 Convolution of measures

We now consider a simple combination of product dimensions and image measure. To convolve measures μ, ν on $\mathcal{B}(\mathbb{R})$, we first consider the product measure $\mu \otimes \nu$. The image measure under summation is then the convolution of μ, ν. We will later identify this convolution as the distribution of $X + Y$ if X, Y are *independent* random variables with distribution μ and ν, respectively. Sometimes, for example with Poisson distributions and normal distributions, the convolution is again a Poisson or normal distribution.

Definition 5.17 (Convolution of measures)**.** *Let* $\mu_1, \ldots, \mu_n$ *be* σ*-finite measures on* $\mathcal{B}(\mathbb{R})$ *and* $\mu_1 \otimes \cdots \otimes \mu_n$ *their product measure. Further, let* $S(x_1, \ldots, x_n) := x_1 + \cdots + x_n$*. Then the image measure* $S_*(\mu_1 \otimes \cdots \otimes \mu_n)$ *is called the* convolution *of the measures* $\mu_1, \ldots, \mu_n$ *and is denoted by* $\mu_1 * \cdots * \mu_n$ *or* $*_{i=1}^n \mu_i$*.*

Example 5.18 (Convolution of Poisson and geometric distributions)**.** 1. *For* $\gamma_1, \gamma_2 \geq 0$ *let* $\mu_{Poi(\gamma_1)}$ *and* $\mu_{Poi(\gamma_2)}$ *be two Poisson distributions from Example 2.2. We calculate* *the convolution of the two distributions by*

$$\begin{aligned}
\mu_{Poi(\gamma_1)} * \mu_{Poi(\gamma_2)} &= \sum_{m,n} 1_{m+n=k} e^{-(\gamma_1+\gamma_2)} \frac{\gamma_1^m \gamma_2^n}{m!n!} \cdot \delta_k \\
&= \sum_{m=0}^{k} e^{-(\gamma_1+\gamma_2)} \frac{\gamma_1^m \gamma_2^{k-m}}{m!(k-m)} \cdot \delta_k \\
&= e^{-(\gamma_1+\gamma_2)} \frac{(\gamma_1+\gamma_2)^k}{k!} \cdot \delta_k \sum_{m=0}^{k} \binom{k}{m} \frac{\gamma_1^m \gamma_2^{k-m}}{(\gamma_1+\gamma_2)^k} \\
&= \mu_{Poi(\gamma_1+\gamma_2)}.
\end{aligned}$$

2. *The geometric distribution for the parameter $p \in [0,1]$ is as well known from Example 2.2. The convolution of two measures $\mu_{geom(p)}$ is given by*

$$\begin{aligned}
\mu_{geom(p)} * \mu_{geom(p)} &= \sum_{m=2}^{k} (1-p)^{m-1} p (1-p)^{k-m-1} p \cdot \delta_k \\
&= (k-1)(1-p)^{k-2} p^2 \cdot \delta_k.
\end{aligned}$$

This is a negative binomial distribution for the parameters p and 2.

Lemma 5.19 (Convolution of distributions with densities)**.** *Let λ be a measure on $\mathcal{B}(\mathbb{R})$, $\mu = f_\mu \cdot \lambda$ and $\nu = f_\nu \cdot \lambda$ for measurable densities $f_\mu, f_\nu : \mathbb{R} \to \mathbb{R}_+$. Then $\mu * \nu = f_{\mu*\nu} \cdot \lambda$ with*

$$f_{\mu*\nu}(t) = \int f_\mu(s) f_\nu(t-s) \lambda(ds).$$

Proof. The proof is a simple application of Fubini's theorem, Theorem 5.13. □

Example 5.20 (Convolution of normal distributions)**.** *Let $f_{N(\mu_1,\sigma_1^2)}$ and $f_{N(\mu_2,\sigma_2^2)}$ be the density functions of two normal distributions with expected value μ_1, μ_2 and variance σ_1^2 and σ_2^2, respectively. Let further $\mu := \mu_1+\mu_2$ and $\sigma^2 = \sigma_1^2+\sigma_2^2$. Then the density of the convolution is given by*

$$\begin{aligned}
x \mapsto & \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} \int \exp\Big(-\frac{(y-\mu_1)^2}{2\sigma_1^2} - \frac{(x-y-\mu_2)^2}{2\sigma_2^2} \Big) dy \\
\overset{y \to (y-\mu_1)\sigma/(\sigma_1\sigma_2)}{=} & \frac{1}{2\pi\sigma} \int \exp\Big(-\frac{\sigma_2^2 y^2}{2\sigma^2} - \frac{\Big((x-\mu) - y\frac{\sigma_1\sigma_2}{\sigma}\Big)^2}{2\sigma_2^2} \Big) dy \\
= & \frac{1}{2\pi\sigma} \int \exp\Big(- \frac{\sigma_2^2 y^2 + \Big((x-\mu)\frac{\sigma}{\sigma_2} - \sigma_1 y\Big)^2}{2\sigma^2} \Big) dy \\
= & \frac{1}{2\pi\sigma} \int \exp\Big(- \frac{(\sigma y - \frac{\sigma_1}{\sigma_2}(x-\mu))^2}{2\sigma^2} - \frac{(x-\mu)^2\Big(\frac{\sigma^2}{\sigma_2^2} - \frac{\sigma_1^2}{\sigma_2^2}\Big)}{2\sigma^2} \Big) dy \\
= & \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2} \Big).
\end{aligned}$$

So, the convolution is again a normal distribution. This now has expected value μ and variance σ^2.

5.5 Projective families of probability measures

So far we have defined σ-finite measures on finite product spaces. This is not sufficient for the probability theory to be discussed later. To understand this, let us recall the infinite coin toss, which was already considered in the lecture *Elementary probability 1*. Here, we would say that $\Omega = \{\text{head}, \text{tail}\}^{\mathbb{N}}$ and the corresponding probability measure is the product measure $\mathbf{P}^{\otimes\infty}$ of $\mathbf{P} = \frac{1}{2}\delta_{\text{head}} + \frac{1}{2}\delta_{\text{tail}}$. However, this is an infinite (but still countable) product measure whose existence we have not yet shown. More generally, a large part of the lecture *Stochastic Processes* will contain such measures (even on uncountable product spaces). We now give the general construction of probability measures on product measures, which goes back to Kolmogorov (and Daniell). It should be mentioned here that in the resulting theorem of Kolmogorov (theorem 5.24) the assumption is made that Ω is Polish.

Definition 5.21 (Projective limit)**.**

1. *Let $(\Omega, \mathcal{F})$ be a measurable space, I an arbitrary index set and $(\Omega^J, \mathcal{F}^J)_{J\subseteq_f I}$ be a family of measurable product spaces, equipped with the product σ-algebra, as in Definition 5.3. A family of probability measures $(\mathbf{P}_J)_{J\subseteq_f I}$, where $\mathbf{P}_J$ is a probability measure on $\mathcal{F}^J$, is called a projective family if*

$$\mathbf{P}_H = (\pi^J_H)_*\mathbf{P}_J$$

   for all $H \subseteq J \subseteq_f I$. (In other words, projection of coordinates in J to coordinates in H under $\mathbf{P}_J$ leads to $\mathbf{P}_H$.)

2. *If for a projective family $(\mathbf{P}_J)_{J\subseteq_f I}$ of probability measures there exists a probability measure $\mathbf{P}_I$ on $\mathcal{F}^I$ with $\mathbf{P}_J = (\pi_J)_*\mathbf{P}_I$ for all $J \subseteq_f I$, then $\mathbf{P}_I$ is called the projective limit of the projective family. We then write*

$$\mathbf{P}_I = \varprojlim_{J\subseteq_f I} \mathbf{P}_J.$$

Example 5.22 (Projective limits and stochastic processes)**.** *Projective families play a major role in at least two situations.*

1. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and I an infinite index set. In Definition 5.15 we have defined the product measure $\mathbf{P}^{\otimes J}$ on $\mathcal{F}^J$ for each $J \subseteq_f I$. The family $(\mathbf{P}^{\otimes J})_{J\subseteq_f I}$ is projective. If $H \subseteq J \subseteq_f I$, then for $A_i \in \mathcal{F}, i \in H$,*

$$\begin{aligned}
(\pi^J_H)_*\mathbf{P}^{\otimes J}\big(\bigtimes_{i\in H} A_i\big) &= \mathbf{P}^{\otimes J}\big((\pi^J_H)^{-1}\big(\bigtimes_{i\in H} A_i\big)\big) \\
&= \mathbf{P}^{\otimes J}\big(\bigtimes_{i\in H} A_i \times \bigtimes_{i\in J\setminus H} \Omega\big) \\
&= \prod_{i\in H} \mathbf{P}(A_i) \cdot \prod_{i\in J\setminus H} \mathbf{P}(\Omega) \\
&= \prod_{i\in H} \mathbf{P}(A_i) \\
&= \mathbf{P}^{\otimes H}\big(\bigtimes_{i\in H} A_i\big).
\end{aligned}$$

   However, we have not yet shown that the projective limit of $(\mathbf{P}^{\otimes J})_{J\subseteq_f I}$ exists. We would then call this the infinite product measure $\mathbf{P}^{\otimes I}$. (In particular, this would give the probability space for the infinite coin toss from the beginning of this section.)

2. *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space,* I *an arbitrary index set,* $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ *a measurable space and* $X_i : \Omega \to \widetilde{\Omega}, i \in I$ *a random variable (i.e. a function measurable with respect to* $\mathcal{F}/\mathcal{F}_i$*). We will call the family* $\mathcal{X} := (X_i)_{i\in I}$ *a stochastic process. So* $\mathcal{X} : \Omega \to \widetilde{\Omega}^I$ *with* $\mathcal{X}(\omega) = (X_i(\omega))_{i\in I}$*. One can now ask whether the distribution of* $\mathcal{X}$ *(i.e. the image measure* $\mathcal{X}_*\mathbf{P}$*) exists as a distribution on* $\widetilde{\mathcal{F}}^I$*.*

   It should be noted that $\widetilde{\mathbf{P}}_J := ((X_j)_{j\in J})_*\mathbf{P}, J \subseteq_f I$ *is a projective family. If* $H \subseteq J \subseteq_f I$ *and* $\widetilde{A}_i \in \widetilde{\mathcal{F}}, i \in H$*, then*

$$\begin{aligned}(\pi^J_H)_*\widetilde{\mathbf{P}}_J\big(\bigtimes_{j\in H} \widetilde{A}_j\big) &= \widetilde{\mathbf{P}}_J\big((\pi^J_H)^{-1} \bigtimes_{j\in H} \widetilde{A}_j\big)\\ &= \widetilde{\mathbf{P}}_J\big(\bigtimes_{j\in H} \widetilde{A}_j \times \bigtimes_{j\in J\setminus H} \widetilde{\Omega}\big)\\ &= \mathbf{P}\big(X_j \in \widetilde{A}_j, j \in H \text{ and } X_j \in \widetilde{\Omega}, j \in J \setminus H\big)\\ &= \mathbf{P}\big(X_j \in \widetilde{A}_j, j \in H\big)\\ &= \widetilde{\mathbf{P}}_H\big(\bigtimes_{j\in H} \widetilde{A}_j\big).\end{aligned}$$

   As Theorem 5.24 below shows, the distribution $\mathcal{X}_*\mathbf{P}$ *(which is then the projective limit of* $(\widetilde{\mathbf{P}}_J)_{J\subseteq_f I}$*) exists at least if* $\widetilde{\mathcal{F}}$ *is the Borel's* σ*-algebra of a Polish space.*

Remark 5.23 (Uniqueness of the projective limit)**.** *For each projective family* $(\mathbf{P}_J)_{J\subseteq_f I}$ *there is at most one projective limit: If* $\mathbf{P}_I$ *and* $\widetilde{\mathbf{P}}_I$ *are two projective limits, then for*

$$\mathcal{H}' := \Big\{ \bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega_i, A_i \in \mathcal{F}_i, i \in J \subseteq_f I \Big\},$$

we see that $\mathcal{H}'$ *generates* $\mathcal{F}^I$ *(compare with* $\mathcal{H}$ *from Lemma 5.7), and is* $\cap$*-stable. Hence, for* $A = \bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega_i \in \mathcal{H}'$*,*

$$\mathbf{P}_I(A) = \mathbf{P}_J\Big(\bigtimes_{i\in J} A_i\Big) = \widetilde{\mathbf{P}}_J\Big(\bigtimes_{i\in J} A_i\Big) = \widetilde{\mathbf{P}}_I(A).$$

This means that $\mathbf{P}_I$ *and* $\widetilde{\mathbf{P}}_I$ *coincide on the* $\cap$*-stable generator and according to Proposition 2.11,* $\mathbf{P}_I = \widetilde{\mathbf{P}}_I$*. The content of the next theorem is that there is exactly one projective limit for Polish spaces.*

Theorem 5.24 (Existence of processes, Kolmogorov)**.** *Let* $(\Omega, \mathcal{O})$ *be Polish,* $\mathcal{F} = \mathcal{B}(\mathcal{O})$ *and* $(\mathbf{P}_J)_{J\subseteq_f I}$ *a projective family of probability measures on* $\mathcal{F}$*. Then there is the projective limit* $\varprojlim_{J\subseteq_f I} \mathbf{P}_J$*.*

Proof. Let $\mathcal{H}'$ be as in Remark 5.23 and μ be a finite additive set function on $\mathcal{H}'$, defined by the projective family using

$$\mu\Big(\bigtimes_{j\in J} A_j \times \bigtimes_{i\in I\setminus J} \Omega\Big) := \mathbf{P}_J\Big(\bigtimes_{j\in J} A_j\Big).$$

According to Lemma 5.7, $\mathcal{H}$ is a semi-ring and μ is a well-defined content on $\mathcal{H}$. Further,

$$\mathcal{K} := \{\bigtimes_{j\in J} K_j \times \bigtimes_{i\in I\setminus J} \Omega : J \subseteq_f I, K_j \text{ compact}\} \subseteq \mathcal{H}$$

is a compact system.

We now show that μ is inner regular with respect to $\mathcal{K}$. Let $\varepsilon > 0$, $\bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega \in \mathcal{H}$ for $J \subseteq_f I$ and $A_i \in \mathcal{F}, i \in J$. Since $\mathbf{P}_j$ is a measure for $j \in I$, according to Lemma 2.9 there are compact sets $K_j \in \mathcal{F}$ with $K_j \subseteq A_j$ and $\mathbf{P}_j(A_j \setminus K_j) \leq \varepsilon$. This means that

$$\begin{aligned}
\mu\Big(\Big(\bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega\Big)\setminus\Big(\bigtimes_{i\in J} K_i \times \bigtimes_{i\in I\setminus J} \Omega\Big)\Big) &= \mu\Big(\Big((\bigtimes_{i\in J} A_i) \setminus (\bigtimes_{i\in J} K_i)\Big) \times \bigtimes_{i\in I\setminus J} \Omega\Big) \\
&= \mathbf{P}_J\big((\bigtimes_{j\in J} A_j) \setminus (\bigtimes_{j\in J} K_j)\big) \\
&\leq \mathbf{P}_J\Big(\bigcup_{j\in J}(A_j \setminus K_j) \times \bigtimes_{i\neq j} \Omega\Big) \\
&\leq \sum_{j\in J} \mathbf{P}_J\Big((A_j \setminus K_j) \times \bigtimes_{i\neq j} \Omega\Big) \\
&= \sum_{j\in J} \mathbf{P}_j(A_j \setminus K_j) \\
&\leq |J|\varepsilon.
\end{aligned}$$

Since J was finite and $\varepsilon > 0$ was arbitrary, we have shown inner regularity of μ with respect to $\mathcal{K}$. According to Theorem 2.10, μ is σ-additive. Furthermore, $\mu(\Omega^I) = 1$, so μ can be uniquely extended to a measure $\mathbf{P}$ on $\sigma(\mathcal{H}') = \mathcal{F}^I$ according to Theorem 2.16. This must be the projective limit of $(\mathbf{P}_J)_{J\subseteq_f I}$. □

Part II
Probability Theory

Prelude

These are the notes of a lecture which I gave at the University of Freiburg. After some elementary probability and measure theory, this course introduces some main concepts in (measure theoretic) probability theory. As a prerequisite, for measure theoretic contents, we refer to my manuscript *Measure theory for probabilists*. In particular, references to Chapters 1–5 are references to this manuscript.

The following books have guided me as references for the purpose of this manuscript.

- Durrett, Rick. Probability: Theory and Examples, Cambridge Series in Statistical and Probabilistic Mathematics, 2019
- Kallenberg, Olaf. Foundations of Modern Probability Theory. Springer, third edition, 2021
- Klenke, Achim. Probability theory. A comprehensive course. Springer, 2014

Throughout the manuscript, we will use a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ (recall from Definition 2.1). The integral with respect to $\mathbf{P}$ is denoted by $\mathbf{E}[\cdot] := \mathbf{P}[\cdot]$ (recall from Chapter 3). Further, we abbreviate $\mathcal{L}^p := \mathcal{L}^p(\mathbf{P})$ if this does not lead to confusion (recall these spaces from Chapter 4.

Our aim in the present course is to provide the most important probabilistic statements available. Fundamental to this is the concept of the random variable, which we will examine in Chapter 6 (see also Definition 3.3). We will often consider the case of E-valued random variables, where E is a Polish space (see Appendix A in the lecture notes on measure theory). The most influential theorems in probability theory are the strong law of large numbers (LLN, Theorem 8.21) and the central limit theorem (CLT, Theorem10.8). These two theorems are limit statements for random variables, and it is important to note that the type of convergence in both theorems is fundamentally different. While the strong LLN describes an almost sure convergence (refer to Remark 2.14), the CLT is a statement about convergence in distribution (i.e. about the weak convergence of the distributions of the random variables; see Chapter 9). Consequently, one of the tasks will be to understand the relationships between different types of convergence (see Chapter 7 and 9).

The present english version of this manuscript was written based on the German version with the help of DeepL.

6 Random variables

We usually use real-valued random variables $X : \Omega \to \mathbb{R}$ (i.e. Borel-measurable functions, i.e. random variables with values in $\mathbb{R}$, measurable with respect to the Borel σ-algebra in $\mathbb{R}$; recall from Definition 1.7)). We will now recall several concepts from measure theory about random variables and which we will need directly in the following. We will mainly deal with connecting the lecture to measure theory on one side, and *Basic Probability* on the other side.

6.1 Repetition

Recall that we assume throughout that a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is given. In Definition 3.5, we already gave several notions with a relationship to random variables, which we recall for completeness.

Remark 6.1 (Random variables and their distribution)**.** *Let $(\Omega', \mathcal{F}')$ be a measurable space.*

1. *Every $\mathcal{F}/\mathcal{F}'$-measurable function X is called (Ω'-valued)* random variable. *If $(\Omega', \mathcal{F}') = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, it is called* real-valued. *The σ-algebra $\sigma(X) = \{X^{-1}(B) : B \in \mathcal{F}')\}$ is* the σ-algebra generated by X *(see Definition 3.3).*

2. *The probability measure $X_*\mathbf{P}$ on $\mathcal{F}'$ (i.e. the image measure of $\mathbf{P}$ under X; see Section 2.5) is called* distribution of X*. Furthermore, if Y is a random variable and $X_*\mathbf{P} = Y_*\mathbf{P}$ (i.e. $\mathbf{P}(X \in A') = \mathbf{P}(Y \in A')$ for all $A' \in \mathcal{F}'$), then X and Y are identically distributed and we write $X \stackrel{d}{=} Y$. However, this notation should be used with caution, as the equality $X \stackrel{d}{=} Y$ cannot be achieved by equivalence transformations to other statements. (For example $X \stackrel{d}{=} Y$ does in general not imply $X - Y \stackrel{d}{=} 0$).*

3. *For a family $(X_i)_{i\in I}$ of random variables, their* joint distribution *is given by $((X_i)_{i\in I})_*\mathbf{P}$. (This is the image measure under the mapping $(X_i)_{i\in I} : \omega \mapsto (X_i(\omega))_{i\in I}$).*

4. *We will use the following notation:* If X is a random variable with distribution $N(\mu, \sigma^2)$. *(This means means that $X : \Omega \mapsto \mathbb{R}$ is a measurable mapping and $X_*\mathbf{P} = \mu_{N(\mu,\sigma^2)}$; see example 2.22.) Then, we write $X \sim N(\mu, \sigma^2)$. Here, read '$\sim$' as* has distribution.

5. *Let λ be another measure on $\mathcal{F}$ and $f : \Omega \to \mathbb{R}$ with $f \geq 0$ almost everywhere and $\lambda[f] = 1$. Then, X has the density f with respect to μ if and only if $X_*\mathbf{P} = f \cdot \lambda$ (see Definition 4.13). Then, for $A \in \mathcal{F}$,*

$$\mathbf{P}(X \in A) = \mu[f, A].$$

   In this case, for $g : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ that (see Lemma 4.14),

$$\mathbf{E}[g(X)] = (X_*\mathbf{P})[g] = (f \cdot \mu)[g] = \mu[fg],$$

   if the right-hand side exists.

6. *Monotonicity and linearity of the integral means for random variables $X, Y \in \mathcal{L}^1$ and $a, b \in \mathbb{R}$:*

$$\begin{aligned} X \leq Y \text{ almost surely } &\implies \mathbf{E}[X] \leq \mathbf{E}[Y], \\ \mathbf{E}[aX + bY] &= a\mathbf{E}[X] + b\mathbf{E}[Y]. \end{aligned}$$

Furthermore, according to Proposition 3.21,

$$\mathbf{E}[X] < \infty \Longrightarrow \mathbf{P}(X < \infty) = 1.$$

Although we already have a σ-algebra $\mathcal{F}$, in further sections, especially in the introduction of the conditional expectation in Chapter 11, the σ-algebra generated by X will play a special role. Simply put, a real-valued random variable Y is $\sigma(X)$-measurable if and only if $Y = \varphi(X)$ for a Borel-measurable mapping φ. In other words, this means that the value of $Y(\omega)$ is known if you know $X(\omega)$, although you do not know what value ω has assumed. See also Exercise **??**.

Lemma 6.2 (Measurability with respect to $\boldsymbol{\sigma(X)}$)**.** *Let $(\Omega', \mathcal{F}')$ be a measurable space and X a random variable with values in Ω', and $Z : \Omega \to \overline{\mathbb{R}}$. The, Z is $\sigma(X)$-measurable if and only if there is a $\mathcal{F}'/\mathcal{B}(\overline{\mathbb{R}})$-measurable mapping $\varphi : \Omega' \to \overline{\mathbb{R}}$ with $\varphi \circ X = Z$.*

Proof. '$\Longleftarrow$': clear
'$\Longrightarrow$': It suffices to consider the case $Z \geq 0$; otherwise, we write $Z = Z^+ - Z^-$. First, let $Z = 1_A$ for $A \in \sigma(X)$. Then there is an $A' \in \mathcal{F}'$ with $X^{-1}(A') = A$, i.e. $Z = 1_{X^{-1}(A')} = 1_{A'} \circ X$, i.e. $\varphi = 1_{A'}$ fulfills the statement. Due to linearity, the statement is also true for simple functions, i.e. finite linear combinations of indicator functions. In the general case, there are simple functions $Z_1, Z_2, \dots \geq 0$ with $Z_n \uparrow Z$. In addition, there are $\mathcal{F}'$-measurable functions φ_n with $Z_n = \varphi_n \circ X$. Then $\varphi = \sup_n \varphi_n$ is again $\mathcal{F}'$-measurable and, since $Z \geq 0$, and

$$\varphi \circ X = (\sup_n \varphi_n) \circ X = \sup_n (\varphi_n \circ X) = \sup_n Z_n = Z.$$

□

We now briefly repeat the convergence theorems for integrals in the context of of random variables.

Proposition 6.3 (Integral convergence theorems)**.** *Let $X, X_1, X_2, \dots$ be real-valued random variables.*

1. Lemma of Fatou, Theorem 3.27: *If $X_1, X_2, ... \geq 0$, then*

$$\liminf_{n\to\infty} \mathbf{E}[X_n] \geq \mathbf{E}[\liminf_{n\to\infty} X_n].$$

2. Monotone convergence, Theorem 3.26: *If $X_1, X_2, \dots \in \mathcal{L}^1$ and $X_n \uparrow X$ almost surely, then*

$$\lim_{n\to\infty} \mathbf{E}[X_n] = \mathbf{E}[X],$$

where both sides can take the value ∞.

3. Dominated convergence, Theorem 3.28: *Let $X_n \xrightarrow{n\to\infty} X$ almost surely, and Y another real-valued random variable with $|X_1|, |X_2|, \dots \leq Y$ almost surely, and $\mathbf{E}[Y] < \infty$. Then,*

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

We now collect (and re-prove) already known inequalities. They often help to estimate probabilities or expected values. Most of the inequalities are already known from the lecture on *Basic Probability*.

Proposition 6.4 (Markov and Chebyshev inequality). *1. Let X be a random variable with values in $\overline{\mathbb{R}}_+$ and $x \in \mathbb{R}_+$. Then the Markov inequality holds, i.e.,*

$$\mathbf{P}(X \geq x) \leq \frac{\mathbf{E}[X]}{x}.$$

2. If X is a real-valued random variable and $p, x \in \mathbb{R}_+$, then the Chebyshev inequality holds, i.e.

$$\mathbf{P}(|X| \geq x) \leq \frac{\mathbf{E}[|X|^p]}{x^p}.$$

Proof. 1. Since X is non-negative, we find $x \cdot 1_{X \geq x} \leq X$. Thus,

$$x \cdot \mathbf{P}(X \geq x) = \mathbf{E}[x \cdot 1_{X \geq x}] \leq \mathbf{E}[X],$$

and the inequality follows. The inequality in 2. follows from 1. by

$$\mathbf{P}(|X| \geq x) = \mathbf{P}(|X|^p \geq x^p) \leq \frac{\mathbf{E}[|X|^p]}{x^p}.$$

□

The next statement was already given in Proposition 4.2.

Proposition 6.5 (Minkowski and Hölder inequalities). *Let X, Y be real-valued random variables.*

1. *If $0 < p, q, r \leq \infty$ such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Then,*

$$\mathbf{E}[|XY|^r]^{1/r} \leq \mathbf{E}[|X|^p]^{1/p} \cdot \mathbf{E}[|Y|^q]^{1/q} \qquad \textit{(Hölder inequality)} \tag{6.1}$$

Especially, if $p = q = 2$

$$\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^2]^{1/2} \cdot \mathbf{E}[|Y|^2]^{1/2}. \qquad \textit{(Cauchy-Schwartz inequality)} \tag{6.2}$$

2. *For $1 \leq p \leq \infty$,*

$$\mathbf{E}[|X+Y|^p]^{1/p} \leq \mathbf{E}[|X|^p]^{1/p} + \mathbf{E}[|Y|^p]^{1/p}, \quad 1 \leq p \leq \infty \qquad \textit{(Minkowski inequality)} \tag{6.3}$$

Proposition 6.6 (Jensen's inequality). *Let $I \subseteq \mathbb{R}$ be an open interval and $X \in \mathcal{L}^1$ with values in I and $\varphi : I \to \mathbb{R}$ convex.*[12] *Then,*

$$\mathbf{E}[\varphi(X)] \geq \varphi(\mathbf{E}[X]).$$

Proof. Since φ is convex, φ is continuous and

$$t \mapsto \frac{\varphi(tx + (1-t)y) - \varphi(y)}{t(x-y)}$$

[12] A mapping $\varphi : I \to \mathbb{R}$ is convex if $\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$ for all $0 \leq t \leq 1$ and $x, y \in I$.

is monotonically decreasing for $y \leq x$. In particular, for $y \in I$ there exists

$$\lambda(y) := \lim_{x \downarrow y} \frac{\varphi(x) - \varphi(y)}{x - y} = \lim_{t \downarrow 0} \frac{\varphi(tx + (1-t)y) - \varphi(y)}{t(x-y)} \tag{6.4}$$

and

$$\frac{\varphi(x) - \varphi(y)}{x - y} \geq \lambda(y) \quad \Longrightarrow \quad \varphi(y) + \lambda(y)(x-y) \leq \varphi(x) \tag{6.5}$$

for all $x \in I$. (For $y > x$ one argues analogously as above).

Note, since I is an interval, we have $\mathbf{E}[X] \in I$. According to (6.5), for $x \in I$ with $y = \mathbf{E}[X]$

$$\varphi(x) \geq \varphi(\mathbf{E}[X]) + \lambda(\mathbf{E}[X])(x - \mathbf{E}[X])$$

and thus

$$\mathbf{E}[\varphi(X)] \geq \varphi(\mathbf{E}[X]) + \lambda(\mathbf{E}[X])\mathbf{E}[X - \mathbf{E}[X]] = \varphi\big(\mathbf{E}[X]\big). \qquad \square$$

Jensen's inequality can be used to show, for example, that $\mathcal{L}^q \subseteq \mathcal{L}^p$ for $p \leq q$. Alternatively, you can read this property from Proposition 4.3.

Lemma 6.7 ($\mathcal{L}^q$ and $\mathcal{L}^p$)**.** *Let $q > 0$ and $X \in \mathcal{L}^q$ be a reel-valued random variable. Then, for $p \leq q$*

$$\mathbf{E}[|X|^p] \leq \mathbf{E}[|X^q|]^{p/q}.$$

In particular, $\mathcal{L}^q \subseteq \mathcal{L}^p$.

Proof. The mapping $y \mapsto y^{p/q}$ is concave on $\mathbb{R}_+$, so with Jensen's inequality,

$$\mathbf{E}[|X|^p] = \mathbf{E}[(|X|^q)^{p/q}] \leq \mathbf{E}[|X|^q]^{p/q}. \qquad \square$$

6.2 Moments

From the lecture *Basic Probability*, we know terms such as expected value, variance and covariance. When we repeat them now, you will see that all calculation rules that you learned also apply in the measure theoretic sense. The only difference in the formulation is that $\mathbf{E}[\cdot]$ is the integral with respect to a probability measure.

Definition 6.8 (Moments)**.** *Let X, Y be real-valued random variables. Then, if it exists, $\mathbf{E}[X]$ is the* expected value *of X. Furthermore, if it exists, the* variance *of X is given by*

$$\mathbf{V}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2].$$

If it exists, the covariance *of X and Y is given by*

$$\mathbf{COV}[X, Y] := \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])].$$

If $\mathbf{COV}[X, Y] = 0$, we say that X and Y are uncorrelated. Furthermore, $\mathbf{E}[X^p]$ for $p > 0$ is the p-th moment of X and $\mathbf{E}[(X - \mathbf{E}[X])^p]$ is the centered p-th moment of X.

We recall a few properties here.

Proposition 6.9 (Properties of the second moments)**.** *Let $X, Y \in \mathcal{L}^2$ be real-valued random variables. Then,* $\mathbf{V}[X], \mathbf{V}[Y], |\mathbf{COV}[X,Y]| < \infty$ *and*

$$\begin{aligned}\mathbf{V}[X] &= \mathbf{E}[X^2] - (\mathbf{E}[X])^2,\\ \mathbf{COV}[X,Y] &= \mathbf{E}[XY] - \mathbf{E}[X]\cdot\mathbf{E}[Y].\end{aligned}$$

The Cauchy-Schwartz inequality holds, i.e.

$$\mathbf{COV}[X,Y]^2 \le \mathbf{V}[X]\cdot\mathbf{V}[Y].$$

If $X_1, \dots, X_n \in \mathcal{L}^2$, the identity of Bienamyé holds, i.e.

$$\mathbf{V}\Big[\sum_{k=1}^n X_k\Big] = \sum_{k=1}^n \mathbf{V}[X_k] + 2\sum_{1\le k<l\le n} \mathbf{COV}[X_k, X_l].$$

Proof. For the first statement, since $\mathbf{V}[X] = \mathbf{COV}[X,X]$ by definition, it is sufficient to show the second equation. This equation follows from the linearity of the expected value by means of

$$\begin{aligned}\mathbf{COV}[X,Y] &= \mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]\\ &= \mathbf{E}[XY] - \mathbf{E}[\mathbf{E}[X]Y] - \mathbf{E}[X\mathbf{E}[Y]] + \mathbf{E}[X]\mathbf{E}[Y]\\ &= \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y].\end{aligned}$$

The Cauchy-Schwartz inequality follows by applying Proposition 6.5 using the random variables $X-\mathbf{E}[X]$ and $Y-\mathbf{E}[Y]$. In particular, $|\mathbf{COV}[X,Y]| < \infty$. For the equation of Bienamyé let wlog $\mathbf{E}[X_k] = 0$, $k = 1, \dots, n$ (otherwise w use the random variables $X_k - \mathbf{E}[X_k]$). Then,

$$\begin{aligned}\mathbf{V}\Big[\sum_{k=1}^n X_k\Big] &= \mathbf{E}\Big[\Big(\sum_{k=1}^n X_k\Big)^2\Big] = \sum_{k=1}^n\sum_{l=1}^n \mathbf{E}[X_kX_l] = \sum_{k=1}^n \mathbf{E}[X_k^2] + 2\sum_{1\le k<l\le n}\mathbf{E}[X_kX_l]\\ &= \sum_{k=1}^n \mathbf{V}[X_k] + 2\sum_{1\le k<l\le n}\mathbf{COV}[X_kX_l].\end{aligned}$$

□

Proposition 6.10 (Moments of non-negative random variables)**.** *Let X be a random variable with values in $\mathbb{R}_+$. Then,*

$$\mathbf{E}[X^p] = p\int_0^\infty \mathbf{P}(X > t)t^{p-1}dt = p\int_0^\infty \mathbf{P}(X \ge t)t^{p-1}dt.$$

Proof. Note that both, $\mathbf{E}[.]$ and $\int \cdot dt$ are integrals. We use Fubini's theorem in order to be able to change the order of integration,

$$\mathbf{E}[X^p] = p\mathbf{E}\Big[\int_0^X t^{p-1}dt\Big] = p\int_0^\infty \mathbf{E}\Big[1_{X>t}t^{p-1}\Big]dt = p\int_0^\infty \mathbf{P}(X>t)t^{p-1}dt.$$

The proof of the second equation is analogous. □

6.3 Characteristic functions and Laplace transforms

We now introduce expected values of certain functions of random variables. This results in the characteristic function (of the distribution of real-valued random variables) and the Laplace transform (of the distribution of of non-negative random variables). (Both are covered in some courses on Basic Probability Theory, but not in all.) These functions are useful since they allow the calculation of moments (see Proposition 6.14). In addition, later in Proposition 9.25, we will show that these functions uniquely determine the underlying measure.

Definition 6.11 (Characteristic function and Laplace transform)**.**

1. *The* characteristic function *of an $\mathbb{R}^d$-valued random variable X is given by*

$$\psi_X := \psi_{X_*\mathbf{P}} := \begin{cases} \mathbb{R}^d & \to \mathbb{C}, \\ t & \mapsto \mathbf{E}[e^{itX}] := \mathbf{E}[\cos(tX)] + i\mathbf{E}[\sin(tX)], \end{cases}$$

where $tx := \langle t, x\rangle$ is the scalar product in $\mathbb{R}^d$.

2. *The* Laplace transform *of X is given by*

$$\mathscr{L}_X := \mathscr{L}_{X_*\mathbf{P}} := \begin{cases} \mathbb{R}^d & \to \mathbb{R}, \\ t & \mapsto \mathbf{E}[e^{-tX}], \end{cases}$$

provided the integral on the right-hand side exists. (The Laplace transform is used frequently for probability measures on $\mathbb{R}^d_+$.)

Proposition 6.12 (Properties of characteristic functions)**.** *Let X, Y be random variables with values in $\mathbb{R}^d$ and characteristic functions ψ_X, ψ_Y. Then,*

1. *$|\psi_X(t)| \leq 1$ for each $t \in \mathbb{R}^d$ and $\psi_X(0) = 1$.*
2. *ψ_X is uniformly continuous.*
3. *$\psi_{aX+b}(t) = \psi_X(at)e^{ibt}$ for all $a \in \mathbb{R}, b \in \mathbb{R}^d$.*

Proof. 1. is clear. For uniform continuity, we have the bound

$$\begin{aligned} |e^{ihx} - 1| &= \sqrt{|\cos(hx) + i\sin(hx) - 1|^2} = \sqrt{(\cos(hx) - 1)^2 + \sin(hx)^2} \\ &= \sqrt{2(1 - \cos(hx))} = 2|\sin(hx/2)| \leq |hx| \wedge 2. \end{aligned}$$

From this, 2. follows because of

$$\begin{aligned} \sup_{t\in\mathbb{R}^d} |\psi_X(t+h) - \psi_X(t)| &= \sup_{t\in\mathbb{R}^d} |\mathbf{E}[e^{i(t+h)X} - e^{itX}]| = \sup_{t\in\mathbb{R}^d} |\mathbf{E}[e^{itX}(e^{ihX} - 1)]| \\ &\leq \mathbf{E}[|e^{ihX} - 1|] \leq \mathbf{E}[|hX| \wedge 2] \xrightarrow{h\to 0} 0. \end{aligned}$$

For 3. we calculate using linearity

$$\mathbf{E}[e^{it(aX+b)}] = e^{itb}\mathbf{E}[e^{i(at)X}] = e^{itb}\psi_X(at).$$

□

Example 6.13 (Examples of characteristic functions functions)**.** *1. The characteristic function of $X \sim B(n,p)$ is given by*

$$\psi_{B(n,p)}(t) = (1-p+pe^{it})^n.$$

Indeed: By definition,

$$\mathbf{E}[e^{itX}] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{itk} = (1-p+pe^{it})^n.$$

2. *The characteristic function of $X \sim Poi(\gamma)$ is given by*

$$\psi_{Poi(\gamma)} = e^{\gamma(e^{it}-1)}.$$

Indeed: We calculate

$$\psi_{Poi(\gamma)} = e^{-\gamma} \sum_{n=0}^{\infty} \frac{\gamma^n e^{itn}}{n!} = e^{\gamma(e^{it}-1)}.$$

3. *The characteristic function of $X \sim N(\mu, \sigma^2)$ is given by*

$$\psi_{N(\mu,\sigma^2)}(t) = e^{it\mu} e^{-\sigma^2 t^2/2}.$$

Indeed: We use that $X \sim \sigma Z + \mu$ for $Z \sim N(0,1)$. According to Proposition 6.12.2, it is sufficient to compute this assertion for $\mu = 0, \sigma^2 = 1$. For this case, using partial integration,

$$\frac{d}{dt}\psi_{N(0,1)}(t) = \frac{i}{\sqrt{2\pi}} \int x e^{-x^2/2} e^{itx} dx = -\frac{i}{\sqrt{2\pi}} \int e^{-x^2/2} it e^{itx} dx = -t\psi_{N(0,1)}(t).$$

This differential equation with $\psi_{N(0,1)}(0) = 1$ has the unique solution $\psi_{N(0,1)}(t) = e^{-t^2/2}$.

4. *The Laplace transform of $X \sim \exp(\gamma)$ is given by*

$$\mathscr{L}_{\exp(\gamma)}(t) = \frac{\gamma}{\gamma + t}.$$

Indeed: A straight-forward calculation reveals

$$\mathbf{E}[e^{-tX}] = \int_0^\infty \gamma e^{-\gamma x} e^{-tx} dx = \frac{\gamma}{\gamma + t}.$$

Characteristic functions and Laplace transforms are a simple tool to calculate the moments of random variables.

Proposition 6.14 (Characteristic function and moments)**.** *Let X be a real-valued random variable.*

1. *If $X \in \mathcal{L}^p$, then ψ_X is p-times continuously differentiable and for $k = 0, \dots, p$,*

$$\psi_X^{(k)}(t) = \mathbf{E}[(iX)^k e^{itX}].$$

In particular, $\psi_X^{(k)}(0) = i^k \mathbf{E}[X^k]$.

2. In particular, if $X \in \mathcal{L}^2$, then

$$\psi_X(t) = 1 + it\mathbf{E}[X] - \frac{t^2}{2}\mathbf{E}[X^2] + \varepsilon(t)t^2$$

as $\varepsilon(t) \xrightarrow{t\to 0} 0$.

Proof. 1. With $|X|^p$ also $|X|^p \vee 1$ is integrable. Thus, since $|X|^k \leq |X|^p \vee 1$, all $|X|^k$ can be dominated by an integrable random variable and the right-hand side exists. Since the statement is obviously true for $k = 0$, we assume assume that it is valid for $k < n$. Then

$$|\frac{d^{k+1}}{dt^{k+1}} e^{itX}| = \lim_{h\to 0} \Big| \frac{(iX)^k e^{i(t+h)X} - (iX)^k e^{itX}}{h} \Big| \leq \big| X^k \frac{e^{ihX} - 1}{h} \big| \leq \big| X^{k+1} \big|.$$

Due to dominated convergence, the derivative and integral can be interchanged and it follows

$$\psi_X^{(k+1)}(t) = \mathbf{E}\Big[\frac{d}{dt}(iX)^k e^{itX}\Big] = \mathbf{E}[(iX)^{k+1} e^{itX}].$$

The continuity of the derivative also follows with dominated convergence.
2. For the estimation, we need the Taylor expansion of ψ_X with remainder term. We have

$$e^{itX} = 1 + itX - \frac{t^2 X^2}{2}(\cos(\theta_1 tX) + i\sin(\theta_2 tX))$$

with random numbers θ_1, θ_2, so that $|\theta_1|, |\theta_2| \leq 1$. Therefore we get

$$\psi_X(t) = 1 + it\mathbf{E}[X] - \frac{t^2}{2}\mathbf{E}[X^2] + \varepsilon(t)t^2$$

with $2\varepsilon(t) = \mathbf{E}[X^2(1 - \cos(\theta_1 tX) + i\sin(\theta_2 tX))] \xrightarrow{t\to 0} 0$ from dominated convergence. □

Example 6.15 (Moments of the exponential and normal distribution). *1. Let $X \sim \exp(\gamma)$. We know that $\mathscr{L}_{\exp(\gamma)}(t) = \gamma/(\gamma+t)$ from Example 6.13.4. From this and the last Proposition,*

$$\mathbf{E}[X^n] = (-1)^n \frac{d^n}{dt^n}\mathbf{E}[e^{-tX}]\big|_{t=0} = (-1)^n \frac{d^n}{dt^n}\frac{\gamma}{\gamma+t}\Big|_{t=0} = \frac{n!\gamma}{(\gamma+t)^{n+1}}\Big|_{t=0} = \frac{n!}{\gamma^n}.$$

2. For $X \sim N(\mu, \sigma^2)$, we already know $\psi_{N(\mu,\sigma^2)}(t) = e^{it\mu - \sigma^2 t^2/2}$. For small t we develop this with

$$\psi_{N(\mu,\sigma^2)}(t) = 1 + it\mu - \sigma^2 t^2/2 - \mu^2 t^2/2 + \varepsilon(t)t^2$$

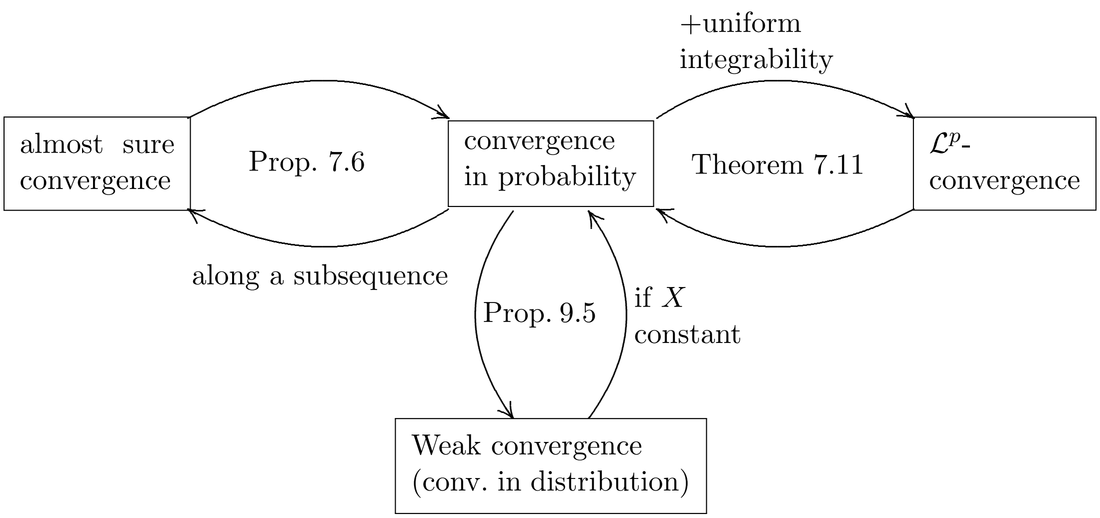
with $\varepsilon(t) \xrightarrow{t\to 0} 0$. From this one reads by means of Proposition 6.14.2 that

$$\mathbf{E}[X] = \mu, \qquad \mathbf{V}[X] = \mathbf{E}[X^2] - \mu^2 = \sigma^2.$$

7 Almost sure, stochastic and $\mathcal{L}^p$-convergence

It is already known from lectures on *Analysis* that there are different types of convergence, such as uniform and pointwise convergence. We will now discuss the most important types of convergence with respect to random variables. (For definitions, see below.)

In our course on measure theory, we have already seen almost sure convergence. In addition, we will discuss convergence in probability and the $\mathcal{L}^p$-convergence (see also Section 4). In Section 9, we will learn about convergence in distribution (which is the same as the weak convergence of the distributions of random variables). The following diagram summarizes all types of convergence:



7.1 Definition and examples

Let's start with some definitions.

Definition 7.1 (Almost sure convergence and convergence in probability)**.** *Let $X, X_1, X_2, \dots$ be random variables with values in a metric space (E, r).*

1. *If*
$$\mathbf{P}(\lim_{n\to\infty} r(X_n, X) = 0) = 1,$$
we say that the sequence $X_1, X_2, \dots$ converges almost surely to X and write $X_n \xrightarrow{n\to\infty}_{as} X$.

2. *If, for all $\varepsilon > 0$,*
$$\lim_{n\to\infty} \mathbf{P}(r(X_n, X) > \varepsilon) = 0,$$
we say that the sequence $X_1, X_2, \dots$ converges to X in probability (or stochastically) and write $X_n \xrightarrow{n\to\infty}_{p} X$.

3. *If the random variables are real-valued and for some $p > 0$*
$$\lim_{n\to\infty} \mathbf{E}[|X_n - X|^p] = 0,$$
we say that the sequence $X_1, X_2, \dots$ converges in $\mathcal{L}^p$ (or in the p-th mean) to X and also write $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$.

Remark 7.2 (Properties of $\mathcal{L}^p$ convergence)**.** *From section 4 we already know a lot about the $\mathcal{L}^p$-convergence. For example, if $X, X_1, X_2, \dots$ is such that $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^q} X$ and $p < q$, then $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$ according to Proposition 4.7. In addition, the spaces $\mathcal{L}^p$ are complete according to Proposition 4.8. So, if for all $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that for all $m, n \geq n$*

$$\mathbf{E}[|X_n - X_m|^p] < \varepsilon,$$

then there is a random variable $X \in \mathcal{L}^p$ with $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$.

Example 7.3 (Counterexamples)**.** *If we look at the diagram at the beginning of the chapter, we can see that convergence in probability follows from almost sure convergence, but not vice versa. Furthermore, convergence in probability follows from $\mathcal{L}^1$ convergence, but not even almost sure convergence implies $\mathcal{L}^1$ convergence. We first give two examples for these two cases.*

1. *Convergence in probability does not imply almost sure convergence: Let U be a $[0,1]$ uniformly distributed random variable. Further we set*

$$\begin{aligned} &A_1 = [0, \tfrac{1}{2}], && A_2 = [\tfrac{1}{2}, 1], \\ &A_3 = [0, \tfrac{1}{4}], && A_4 = [\tfrac{1}{4}, \tfrac{2}{4}], && A_5 = [\tfrac{2}{4}, \tfrac{3}{4}], && A_6 = [\tfrac{3}{4}, 1], \\ &\dots \end{aligned}$$

   and $X_n := 1_{U \in A_n}$. Then it is clear for $0 < \varepsilon < 1$

$$\lim_{n\to\infty} \mathbf{P}(|X_n| > \varepsilon) = \lim_{n\to\infty} \mathbf{P}(U \in A_n) = 0,$$

   i.e. $X_n \xrightarrow{n\to\infty}_p 0$, but for each $n \in \mathbb{N}$ there is an $m > n$ with $X_m = 1$. Therefore, $X_1, X_2, \dots$ does not converge almost surely to 0.

2. *Almost sure convergence does not imply $\mathcal{L}^1$ convergence: Let U again be a uniformly distributed random variable on $[0,1]$ random variable. Further, $B_n = [0, \frac{1}{n}]$ and $Y_n = n \cdot 1_{U \in B_n}$. Then $Y_n \xrightarrow{n\to\infty}_{fs} Y = \infty \cdot 1_{U=0}$, so $Y = 0$ is almost sure. On the other hand*

$$\mathbf{E}[Y_n] = n \cdot \mathbf{P}(U \in A_n) = 1,$$

   so $Y_1, Y_2, \dots$ does not converge to 0 in $\mathcal{L}^1$.

Lemma 7.4 (Limit in probability is (almost surely) unique)**.** *Let $X, Y, X_1, X_2, \dots$ be random variables with values in a metric space space (E, r) and $X_n \xrightarrow{n\to\infty}_p X$, as well as $X_n \xrightarrow{n\to\infty}_p Y$. Then $X = Y$ almost surely.*

Proof. For all $\varepsilon > 0$,

$$\begin{aligned} \mathbf{P}(r(X,Y) > 2\varepsilon) &\leq \mathbf{P}(r(X_n, X) > \varepsilon \text{ or } r(X_n, Y) > \varepsilon) \\ &\leq \mathbf{P}(r(X_n, X) > \varepsilon) + \mathbf{P}(r(X_n, Y) > \varepsilon) \xrightarrow{n\to\infty} 0. \end{aligned}$$

Therefore,

$$\mathbf{P}(X \neq Y) = \mathbf{P}\Big(\bigcup_{k=1}^\infty \Big\{r(X,Y) > 1/k\Big\}\Big) \leq \sum_{k=1}^\infty \mathbf{P}(r(X,Y) > 1/k) = 0,$$

and the statement follows. ◻

7.2 Almost sure convergence and convergence in probability

We now show a result that relates almost sure convergence and convergence in probability.

Lemma 7.5 (Characterization of convergence in probability)**.** *Let $X, X_1, X_2, \dots$ be random variables with values in a metric space (E, r). Then,*

$$X_n \xrightarrow{n\to\infty}_p X \qquad \Longleftrightarrow \qquad \mathbf{E}[r(X_n, X) \wedge 1] \xrightarrow{n\to\infty} 0. \tag{7.1}$$

Proof. If $X_n \xrightarrow{n\to\infty}_p X$, then for all $\varepsilon > 0$,

$$\begin{aligned}\lim_{n\to\infty} \mathbf{E}[r(X_n, X) \wedge 1] &= \lim_{n\to\infty} \mathbf{E}[r(X_n, X) \wedge 1, r(X_n, X) \leq \varepsilon] + \mathbf{E}[r(X_n, X) \wedge 1, r(X_n, X) > \varepsilon] \\ &\leq \lim_{n\to\infty} \big(\varepsilon + \mathbf{P}(r(X_n, X) > \varepsilon)\big) = \varepsilon,\end{aligned}$$

which shows the right-hand side since $\varepsilon > 0$ was arbitrary. However, if the right-hand side applies, the Chebyshev inequality for $0 < \varepsilon \leq 1$ implies that

$$\mathbf{P}(r(X_n, X) > \varepsilon) \leq \frac{\mathbf{E}[r(X_n, X) \wedge 1]}{\varepsilon} \xrightarrow{n\to\infty} 0. \qquad \square$$

Proposition 7.6 (Convergence in probability and almost sure convergence)**.** *Let $X, X_1, X_2, \dots$ be random variables with values in a metric space (E, r). Then, the following are equivalent:*

1. $X_n \xrightarrow{n\to\infty}_p X$
2. *For each sequence $(n_k)_{k=1,2,\dots}$ there is a subsequence $(n_{k_\ell})_{\ell=1,2,\dots}$ with $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty}_{as} X$.*

In particular,

$$X_n \xrightarrow{n\to\infty}_{fs} X \quad \longrightarrow \quad X_n \xrightarrow{n\to\infty}_p X.$$

Proof. 1. $\to$ 2.: Because of (7.1), for each sequence $(n_k)_{k=1,2,\dots}$, there is a subsequence $(n_{k_\ell})_{\ell=1,2,\dots}$ such that (using monotone convergence in the first equality)

$$\mathbf{E}\Big[\sum_{\ell=1}^\infty (r(X_{n_{k_\ell}}, X) \wedge 1)\Big] = \sum_{\ell=1}^\infty \mathbf{E}[r(X_{n_{k_\ell}}, X) \wedge 1] < \infty.$$

This means that

$$1 = \mathbf{P}\Big(\sum_{\ell=1}^\infty (r(X_{n_{k_\ell}}, X) \wedge 1) < \infty\Big) \leq \mathbf{P}\big(\limsup_{\ell\to\infty} r(X_{n_{k_\ell}}, X) = 0\big) \leq 1,$$

i.e. $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty}_{fs} X$.
2. $\to$ 1.: Let's assume that 1. is not valid. (We must show that 2. cannot hold.) Because of (7.1), there is $\varepsilon > 0$ and a subsequence $(n_k)_{k=1,2,\dots}$, so that $\lim_{k\to\infty} \mathbf{E}[r(X_{n_k}, X) \wedge 1] > \varepsilon$. Assuming that there is a subsequence $(n_{k_\ell})_{\ell=1,2,\dots}$ such that $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty}_{as} X$. Then also

$$\lim_{\ell\to\infty} \mathbf{E}[r(X_{n_{k_\ell}}, X) \wedge 1] = \mathbf{E}\big[\lim_{\ell\to\infty} r(X_{n_{k_\ell}}, X) \wedge 1\big] = 0$$

due to dominated convergence, i.e. a contradiction. So we have found a sequence $(n_k)_{k=1,2,\dots}$ for which there is no further subsequence $(n_{k_\ell})_{\ell=1,2,\dots}$ with $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty}_{as} X$, and we have shown that 2. does not hold. $\square$

7.3 Convergence in probability and $\mathcal{L}^p$-convergence

In Example 7.3 we had already seen that almost sure convergence (as well as convergence in probability) does not imply $\mathcal{L}^1$-convergence. This is not surprising, since the theorem of dominated convergence states that a sequence $X_1, X_2, \ldots$, which converges almost surely to X and has an integrable dominating random variable converges in $\mathcal{L}^1$ to X. If the almost sure convergence implies the $\mathcal{L}^1$ convergence, one would not need to make the requirement of an integrable dominating random variable. In the following, we want to find conditions of the integrable dominating random variable in order to suffice for $\mathcal{L}^1$ convergence. See theorem 7.11 and Corollary 7.12. The concept of uniform integrability is central to this, see Definition 7.7.

Definition 7.7 (Uniform integrability)**.** *A family $(X_i)_{i\in I}$ is called* uniformly integrable *if*

$$\inf_K \sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K] = 0.$$

Example 7.8 (Uniform integrability)**.**

1. *Let $Y \in \mathcal{L}^1$ and $(X_i)_{i\in I}$ with $\sup_i |X_i| \le |Y|$. Then $(X_i)_{i\in I}$ is uniformly integrable since*

$$\sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K] \le \mathbf{E}[|Y|; |Y| > K] \xrightarrow{K\to\infty} 0$$

   by dominated convergence. In particular, every $Y \in \mathcal{L}^1$ is uniformly integrable.

2. *Every finite family $(X_i)_{i=1,\ldots,n}$ with $X_1, \ldots, X_n \in \mathcal{L}^1$ is uniformly integrable, because* $\sup_{1\le i\le n} |X_i| \in \mathcal{L}^1$ *and therefore, 1. applies with $Y = \sup_{1\le i\le n} |X_i|$.*

3. *Let us consider the example 7.3.2 Here, for $n > K$*

$$\mathbf{E}[|Y_n|; |Y_n| > K] = \mathbf{E}[Y_n] = 1.$$

   In particular, $(Y_n)_{n=1,2,\ldots}$ is not uniformly integrable.

4. *Let $p > 1$. Then $(X_i)_{i\in I}$ with $X_i \in \mathcal{L}^p, i \in I$ is uniformly integrable if $\sup_{i\in I} ||X_i||_p < \infty$. This is because $K^{p-1}|X_i|1_{|X_i|>K} \le |X_i|^p$, therefore*

$$\sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K] \le \sup_{i\in I} \frac{\mathbf{E}[|X_i|^p]}{K^{p-1}} \xrightarrow{K\to\infty} 0.$$

   .

Lemma 7.9 (Characterization of uniform integrability)**.** *Let $(X_i)_{i\in I}$ be a family of random variables. Then, the following are equivalent:*

1. *$(X_i)_{i\in I}$ is uniformly integrable.*

2. *It holds*

$$\sup_{i\in I} \mathbf{E}[|X_i|] < \infty \quad \text{and} \quad \lim_{\varepsilon\to 0} \sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] = 0.$$

   .

3. *It holds*

$$\lim_{K\to\infty} \sup_{i\in I} \mathbf{E}[(|X_i| - K)^+] = 0.$$

4. *There exists a function* $f : \mathbb{R}_+ \to \mathbb{R}_+$ *such that* $\frac{f(x)}{x} \xrightarrow{x\to\infty} \infty$ *and* $\sup_{i\in I} \mathbf{E}[f(|X_i|)] < \infty$.

If one of the four statements is true, the function f in 4. can be chosen to be monotonically increasing and convex.

Proof. '1. → 2.': Let $\delta > 0$ be given and $K = K_\delta$ such that $\sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K] \leq \delta$. Then, for $A \in \mathcal{F}$,

$$\mathbf{E}[|X_i|; A] = \mathbf{E}[|X_i|; A \cap \{|X_i| > K\}] + \mathbf{E}[|X_i|; A \cap \{|X_i| \leq K\}] \leq \delta + K \cdot \mathbf{P}(A).$$

In particular,

$$\sup_{i\in I} \mathbf{E}[|X_i|] = \sup_{i\in I} \mathbf{E}[|X_i|; \Omega] \leq \delta + K < \infty$$

and

$$\sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] \leq \delta + K\varepsilon \xrightarrow{\varepsilon\to 0} \delta.$$

Since $\delta > 0$ was arbitrary,

$$\lim_{\varepsilon\to 0} \sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] = 0.$$

'2. ⇒ 3.': First, we note that $(|X_i| - K)^+ \leq |X_i| 1_{|X_i|\geq K}$. Let $\varepsilon > 0$. Choose $K = K_\varepsilon$ large enough so that – using the Markov inequality –

$$\sup_{i\in I} \mathbf{P}(|X_i| > K) \leq \sup_{i\in I} \frac{\mathbf{E}[|X_i|]}{K} < \varepsilon.$$

This means that 3. follows from

$$\begin{aligned}\lim_{K\to\infty} \sup_{i\in I} \mathbf{E}[(|X_i| - K)^+] &= \lim_{\varepsilon\to 0} \sup_{i\in I} \mathbf{E}[(|X_i| - K_\varepsilon)^+] \leq \lim_{\varepsilon\to 0} \sup_{i\in I} \mathbf{E}[|X_i|; |X_i| > K_\varepsilon] \\ &\leq \lim_{\varepsilon\to 0} \sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] = 0.\end{aligned}$$

'3. ⇒ 4.': There is a sequence $K_1, K_2, \ldots$ with $K_n \uparrow \infty$ and $\sup_{i\in I} \mathbf{E}[(|X_i| - K_n)^+] \leq 2^{-n}$. We set

$$f(x) := \sum_{n=1}^{\infty} (x - K_n)^+.$$

Then f is monotonically increasing and convex as a sum of convex functions. Furthermore, for $x \geq 2K_n$,

$$\frac{f(x)}{x} \geq \sum_{k=1}^{n} \Big(1 - \frac{K_k}{x}\Big) \geq \frac{n}{2},$$

thus $\frac{f(x)}{x} \xrightarrow{x\to\infty} \infty$. Because of monotone convergence,

$$\mathbf{E}[f(|X_i|)] = \sum_{n=1}^{\infty} \mathbf{E}[(|X_i| - K_n)^+] \leq \sum_{n=1}^{\infty} 2^{-n} = 1.$$

'4. ⇒ 1.': Set $a_K := \inf_{x\geq K} \frac{f(x)}{x}$, so that $a_K \xrightarrow{K\to\infty} \infty$. Thus,

$$\sup_{i\in I} \mathbf{E}[|X_i|; |X_i| \geq K] \leq \frac{1}{a_K} \sup_{i\in I} \mathbf{E}[f(|X_i|); |X_i| \geq K] \leq \frac{1}{a_K} \sup_{i\in I} \mathbf{E}[f(|X_i|)] \xrightarrow{K\to\infty} 0. \qquad \square$$

Example 7.10 (Differences and uniform integrability)**.** *For $X \in \mathcal{L}^1$, $(X_i)_{i\in I}$ is uniformly integrable iff $(X_i - X)_{i\in I}$ is uniformly integrable.*

To see this, let $(X_i)_{i\in I}$ be uniformly integrable. According to Example 7.8.2, X is uniformly integrable. Furthermore,

$$\sup_{i\in I} \mathbf{E}[|X_i - X|] \leq \mathbf{E}[|X|] + \sup_{i\in I} \mathbf{E}[|X_i|] < \infty$$

and

$$\sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i - X|; A] \leq \sup_{A:\mathbf{P}(A)<\varepsilon} \sup_{i\in I} \mathbf{E}[|X_i|; A] + \sup_{A:\mathbf{P}(A)<\varepsilon} \mathbf{E}[|X|; A] \xrightarrow{\varepsilon\to 0} 0,$$

i.e. according to Lemma 7.9, $(X_i - X)_{i\in I}$ is uniformly integrable. The inverse follows analogously.

Theorem 7.11 (Convergence in probability and convergence in the p-th mean)**.** *Let $X_1, X_2, \ldots$ be a sequence in $\mathcal{L}^p$ with $1 \leq p < \infty$. The following statements are equivalent:*

1. *There is a measurable function $X \in \mathcal{L}^p$ with $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$.*
2. *The family $(|X_i|^p)_{i=1,2,\ldots}$ is uniformly integrable and there is a measurable function X with $X_n \xrightarrow{n\to\infty}_p X$.*

If 1. or 2. applies, then the limits coincide almost surely.

Proof. 1.→ 2.: First, due to Chebyshev's inequality for every $\varepsilon > 0$

$$\mathbf{P}(|X_n - X| > \varepsilon) \leq \frac{\mathbf{E}[|X_n - X|^p]}{\varepsilon^p} = \frac{||X_n - X||_p^p}{\varepsilon^p} \xrightarrow{n\to\infty} 0,$$

i.e. convergence in probability applies. For the proof of uniform integrability, we use Lemma 7.9. Let $\varepsilon > 0$ and $N = N_\varepsilon$ such that $||X_n - X||_p < \varepsilon$ for $n \geq N$. Then for $A \in \mathcal{F}$, with Minkowski's inequality,

$$\begin{aligned}\sup_{n\in\mathbb{N}} (\mathbf{E}[|X_n|^p; A])^{1/p} &= \sup_{n\in\mathbb{N}} ||X_n 1_A||_p \\ &\leq \sup_{n<N} ||X_n 1_A||_p + \sup_{n\geq N} ||(X_n - X)1_A||_p + ||X 1_A||_p \\ &\leq \sup_{n<N} (\mathbf{E}[|X_n|^p; A]) + \varepsilon + (\mathbf{E}[|X|^p; A]).\end{aligned}$$

Using $A = \Omega$, we find $\sup_{n\in\mathbb{N}}(\mathbf{E}[|X_n|^p]) < \infty$. Moreover, since N is finite, we find

$$\lim_{\delta\to 0} \sup_{A:\mathbf{P}(A)<\delta} \sup_{n\in\mathbb{N}} \mathbf{E}[|X_n|^p; A] \leq \varepsilon^p.$$

Because $\varepsilon > 0$ was arbitrary, the assertion follows.
2.→ 1.: Since $X_n \xrightarrow{n\to\infty}_p X$, according to proposition 7.6 there is a subsequence $n_1, n_2, \ldots$ with $X_{n_k} \xrightarrow{k\to\infty} X$ almost surely. With Fatou's Lemma,

$$\mathbf{E}[|X|^p] = \mathbf{E}[\liminf_{k\to\infty} |X_{n_k}|^p] \leq \sup_{n\in\mathbb{N}} \mathbf{E}[|X_n|^p] < \infty$$

because of Lemma 7.9. In particular, $X \in \mathcal{L}^p$. Just like in Example 7.10, $\{|X_n - X|^p : n \in \mathbb{N}\}$ is uniformly integrable. For every $\delta > 0$, due to convergence in probability,

$$\mathbf{P}(|X_n - X| > \delta) \xrightarrow{n\to\infty} 0.$$

From lemma 7.9 now follows with dominated convergence

$$\lim_{n\to\infty} \mathbf{E}[|X_n - X|^p] = \lim_{n\to\infty} \mathbf{E}[|X_n - X|^p; |X_n - X| > \delta] + \mathbf{E}[|X_n - X|^p; |X_n - X| \le \delta] \le \delta^p.$$

Since $\delta > 0$ was arbitrary, $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$ follows. □

Corollary 7.12 ($\mathcal{L}^p$-convergence and uniform integrability)**.** *Let $1 \le p < \infty$ and $X_1, X_2, \dots \in \mathcal{L}^p$ and X be measurable with $X_n \xrightarrow{n\to\infty}_p X$. Then, the following are equivalent:*

1. *$X_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} X$,*
2. *$||X_n||_p \xrightarrow{n\to\infty} ||X||_p$,*
3. *The family $(|X_n|^p)_{n=1,2,\dots}$ is uniformly integrable.*

Proof. The equivalence 1. ⇔ 3. is clear from Theorem 7.11.
1. ⇒ 2.: follows from Minkowski's inequality with

$$\big|||X_n||_p - ||X||_p\big| \le ||X_n - X||_p \xrightarrow{n\to\infty} 0.$$

2. → 3.: For fixed K, we write

$$\mathbf{E}[|X_n|^p; |X_n| > K] \le \mathbf{E}[|X_n|^p - (|X_n| \wedge (K - |X_n|)^+)^p] \xrightarrow{n\to\infty} \mathbf{E}[|X|^p - (|X| \wedge (K - |X|)^+)^p].$$

Convergence follows from $\mathbf{E}[|X_n|^p] \xrightarrow{n\to\infty} \mathbf{E}[|X|^p]$, and $(|X_n| \wedge (K - X_n|)^+)^p \xrightarrow{n\to\infty}_{\mathcal{L}^1} |X| \wedge (K - X)^+)^p$, since the convergence according to Proposition 7.6 is in probability, and $((|X_n| \wedge (K - |X_n|)^+)^p)_{n=1,2,\dots}$ is bounded, in particular uniformly integrable. Since $\mathbf{E}[|X|^p - (|X| \wedge (K - |X|)^+)^p] \xrightarrow{K\to\infty} 0$ after dominated convergence, $(|X_n|^p)_{n=1,2,\dots}$ is uniformly integrable. □

8 Independence and the strong law

With our knowledge on probability measures and σ-algebras we now shed light on the concept of independence. In particular, in this chapter we will prove the strong law of large numbers, see Theorem 8.21. On the way, we prove the Borel-Cantelli lemma (Theorem 8.8) and Kolmogorov's 0-1 law (Theorem 8.15).

8.1 Definition and simple properties

Already in the lecture *Basic Probability*, independent random variables were considered. The intuitive idea of independence is often correct, but should sometimes be treated with caution.

Definition 8.1 (Independence)**.** 1. *A family of sets $(A_i)_{i\in I}$ with $A_i \in \mathcal{F}$ is called* independent *if*

$$\mathbf{P}\Big(\bigcap_{j\in J} A_j\Big) = \prod_{j\in J} \mathbf{P}(A_j) \tag{8.1}$$

for all $J \subseteq_f I$.[13]

[13] Recall that we write $J \subseteq_f I$ iff $J \subseteq I$ and J is finite.

2. *A family* $(\mathcal{C}_i)_{i\in I}$ *of set systems* $\mathcal{C}_i \subseteq \mathcal{F}$ *is called* independent *if* (8.1) *holds for all* $J \subseteq_f I$ *and* $A_j \in \mathcal{C}_j, j \in J$.

3. *A family of random variables* $(X_i)_{i\in I}$ *is called independent if* $(\sigma(X_i))_{i\in I}$ *is independent.*

We first deal with the question if there are probability spaces with an arbitrary number of independent random variables. Here we benefit from our knowledge of product measures.

Proposition 8.2 (Independence and product measures)**.** *A family* $(X_i)_{i\in I}$ *of random variables is independent iff for each* $J \subseteq_f I$

$$((X_i)_{i\in J})_*\mathbf{P} = \bigotimes_{i\in J}(X_i)_*\mathbf{P},$$

i.e. the joint distribution of each finite subfamily is the product distribution of the individual distributions.

Proof. By definition, the family $(X_i)_{i\in I}$ is independent if and only if for each $J \subseteq_f I$ and $A_i \in \mathcal{F}, i \in J$,

$$\mathbf{P}(X_i \in A_i, i \in J) = \prod_{i\in J}\mathbf{P}(X_i \in A_i).$$

The assertion now follows from the fact that $\mathbf{P}(X_i \in A_i) = (X_i)_*\mathbf{P}(A_i)$ (see Definition 2.23) and $\mathbf{P}(X_i \in A_i, i \in J) = ((X_i)_{i\in J})_*\mathbf{P}\big(\times_{i\in J} A_i\big)$ (see Corollary 5.14). □

Corollary 8.3 (Existence of uncountably many independent random variables)**.** *Let E be a Polish space and I an arbitrary index set. Let $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i)$ be probability spaces and X_i an E-valued random variable, $i \in I$. Then there is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a family $(Y_i)_{i\in I}$ E-valued, independent random variable with $Y_i \stackrel{d}{=} X_i$.*

Proof. It should be noted that $(((X_i)_{i\in J})_* \bigotimes_{i\in J}\mathbf{P}_i)_{J\subseteq_f I}$ is a projective family of probability measures on $(E, \mathcal{B}(E))$. Using Theorem 5.24 we find the projective limit $\mathbf{P}_I$. This is a probability measure on $(E^I, (\mathcal{B}(E))^I)$. Furthermore, with $\pi_i : E^I \to E$, the i-th projection, $(\pi_i)_*\mathbf{P}_I = (X_i)_*\mathbf{P}_i$, i.e. $\pi_i \stackrel{d}{=} X_i$. □

Lemma 8.4 (Functions of independent random variables)**.** *Let $(\Omega_i', \mathcal{F}_i'), (\Omega_i'', \mathcal{F}_i'')$, $i \in I$, measurable spaces. Let $(X_i)_{i\in I}$ be a family of independent random variables, $X_i : \Omega \to \Omega_i'$, and $\varphi_i : \Omega_i' \to \Omega_i''$ measurable, $i \in I$. Then the family $(\varphi_i(X_i))_{i\in I}$ is independent.*

Proof. According to Lemma 6.2, the random variable $\varphi_i(X_i)$ is measurable according to $\sigma(X_i)$, $i \in I$, i.e. $\sigma(\varphi_i(X_i)) \subseteq \sigma(X_i)$. Since $(\sigma(X_i))_{i\in I}$ is an independent family by assumption, the assertion follows from the definition of independence. □

Proposition 8.5 (Independent and Uncorrelated)**.** *Let $X, Y \in \mathcal{L}^1$ be independent, real-valued random variables. Then $XY \in \mathcal{L}^1$ and*

$$\mathbf{E}[XY] = \mathbf{E}[X] \cdot \mathbf{E}[Y].$$

Proof. The assertion is clear if X and Y are indicator functions. Then, note that if the assertion applies to the pairs (X_i, Y_j), $i, j = 1, \dots, n$, then it also for $\sum_{i=1}^n X_i$ and $\sum_{j=1}^n Y_j$: Indeed, due to the linearity of the expected value,

$$\mathbf{E}\Big[\sum_{i=1}^n X_i \cdot \sum_{j=1}^n Y_j\Big] = \sum_{i=1}^n \sum_{j=1}^n \mathbf{E}[X_i Y_j] = \sum_{i=1}^n \sum_{j=1}^n \mathbf{E}[X_i]\mathbf{E}[Y_j] = \mathbf{E}\Big[\sum_{i=1}^n X_i\Big] \cdot \mathbf{E}\Big[\sum_{j=1}^n Y_j\Big].$$

So, since the assertion applies to indicator functions, it is also valid for simple functions, and thus with monotonic convergence also for non-negative measurable functions. The general case follows with the decomposition $X = X^+ - X^-$ and $Y = Y^+ - Y^-$. □

Example 8.6 (Uncorrelated, non-independent random variables)**.** *Let U be a random variable uniformly distributed on $[0, 1]$, $X = \cos(2\pi U)$ and $Y = \sin(2\pi U)$. Then $\mathbf{E}[X] = \mathbf{E}[Y] = 0$ and*

$$\mathbf{E}[XY] = \int_0^1 \cos(2\pi u)\sin(2\pi u)du = \tfrac{1}{2}\int_0^1 \sin(4\pi u)du = 0$$

and thus X, Y are are uncorrelated. However, $\{|X| < \varepsilon, |Y| < \varepsilon\} = \emptyset$ for $\varepsilon > 0$ is small enough and thus $\mathbf{P}(X^{-1}(-\varepsilon, \varepsilon), Y^{-1}(-\varepsilon, \varepsilon)) = 0 < \mathbf{P}(X^{-1}(-\varepsilon, \varepsilon)) \cdot \mathbf{P}(Y^{-1}(-\varepsilon, \varepsilon))$. This means that X and Y are not independent.

If there is a probability space and (countably) many events, you can ask yourself how many of these events will likely occur. The Borel-Cantelli lemma gives a sharp criterion for the occurrence of only finitely many events.

Definition 8.7 (Limsup of sets)**.** *For $A_1, A_2, \dots \in \mathcal{F}$,*

$$\limsup_{n\to\infty} A_n := \bigcap_{n\geq 1} \bigcup_{m\geq n} A_m$$

is the event infinitely many of the A_n occur*.*

Theorem 8.8 (Borel-Cantelli lemma)**.** *1. Let $A_1, A_2, ... \in \mathcal{F}$. Then,*

$$\sum_{n=1}^\infty \mathbf{P}(A_n) < \infty \Longrightarrow \mathbf{P}(\limsup_{n\to\infty} A_n) = 0.$$

2. If $A_1, A_2, \dots$ are independent,

$$\sum_{n=1}^\infty \mathbf{P}(A_n) = \infty \Longrightarrow \mathbf{P}(\limsup_{n\to\infty} A_n) = 1.$$

Proof. We start with 1. Because of the continuity of $\mathbf{P}$ from above (see Proposition 2.8),

$$\mathbf{P}(\limsup_{n\to\infty} A_n) = \lim_{n\to\infty} \mathbf{P}\Big(\bigcup_{m\geq n} A_m\Big) \leq \lim_{n\to\infty} \sum_{m=n}^\infty \mathbf{P}(A_m) = 0$$

by assumption. For 2. we use that $\log(1-x) \leq -x$ for $x \in [0,1]$. From this and the continuity of $\mathbf{P}$ from below and the independence of $(A_n)_{n=1,2,\dots}$,

$$\begin{aligned}\mathbf{P}((\limsup_{n\to\infty} A_n)^c) &= \mathbf{P}\Big(\bigcup_{n=1}^{\infty}\bigcap_{m\geq n} A_m^c\Big) \\ &= \lim_{n\to\infty}\mathbf{P}\Big(\bigcap_{m=n}^{\infty} A_m^c\Big) \\ &= \lim_{n\to\infty}\prod_{m=n}^{\infty}(1-\mathbf{P}(A_m)) \\ &= \lim_{n\to\infty}\exp\Big(\sum_{m=n}^{\infty}\log(1-\mathbf{P}(A_m))\Big) \\ &\leq \lim_{n\to\infty}\exp\Big(-\sum_{m=n}^{\infty}\mathbf{P}(A_m)\Big) \\ &= 0,\end{aligned}$$

and the assertion follows. □

Example 8.9 (Infinite coin toss and geometric distributions)**.**

1. *We consider an infinite coin toss. This means that we have a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *and independent random variables* $X_1, X_2, \dots$ *with values in* $\{heads, tails\}$. *The coin toss is fair, i.e.* $\mathbf{P}(X_n = head) = 1/2$. *We consider the events* $A_n = \{X_n = head\}$. *Since*

$$\sum_{n=1}^{\infty}\mathbf{P}(A_n) = \sum_{n=1}^{\infty}\tfrac{1}{2} = \infty$$

and the family $(A_n)_{n\in\mathbb{N}}$ *is independent, it follows from the Borel-Cantelli lemma that almost surely infinitely often* head *occurs.*

2. *We consider the same situation as in 1, but the events* $B_n := \{X_1 = Kopf\}$. *It is clear that the family* $(B_n)_{n\in\mathbb{N}}$ *is not independent. (For example* $\mathbf{P}(B_1 \cap B_2) = \mathbf{P}(B_1) = 1/2 \neq \frac{1}{4} = \mathbf{P}(B_1)\cdot\mathbf{P}(B_2)$.*) Just like in 1.* $\sum_{n=1}^{\infty}\mathbf{P}(B_n) = \infty$. *It is also clear that* $\mathbf{P}(\limsup_{n\to\infty} B_n) = \frac{1}{2}$. *It follows from this, that in the Borel-Cantelli lemma the condition of independence in 2. does not apply.*

3. *Let* $X_1, X_2, \dots$ *be geometrically distributed with the success parameter* p. *We consider the events* $A_n := \{X_n \geq n\}$ *and ask ourselves whether an infinite number of these events can occur. Since*

$$\sum_{n=1}^{\infty}\mathbf{P}(A_n) = \sum_{n=1}^{\infty}\mathbf{P}(X_n \geq n) = \sum_{n=1}^{\infty}(1-p)^{n-1} = \tfrac{1}{p} < \infty.$$

Therefore, almost surely only a finite number of the events $\{X_n \geq n\}$ *occur.*

8.2 Kolmogorov's 0-1 law

The Borel-Cantelli lemma is already a statement about when an event that depends on an infinite number of events is occur almost surely. We will now examine this situation further.

Proposition 8.10 (Independence of generated $\boldsymbol{\sigma}$-algebras)**.** *Let $(\mathcal{C}_i)_{i\in I}$ be a family of independent, $\cap$-stable set systems. Then, $(\sigma(\mathcal{C}_i))_{i\in I}$ is also an independent family.*

Proof. Let $J = \{i_1, \dots, i_n\} \subseteq_f I$ and (wlog) $n > 1$. Then, (8.1) holds for any $A_{i_1}, \dots, A_{i_n}$ with $A_{i_k} \in \mathcal{C}_{i_k}, k = 1, \dots, n$. We keep $A_{i_2}, \dots, A_{i_n}$ fixed and define

$$\mathcal{D} := \{A_{i_1} \in \mathcal{F} : (8.1) \text{ holds}\}.$$

We will now show that $\mathcal{D}$ is a Dynkin system. Namely, if $A \subseteq B \in \mathcal{D}$, then $B \setminus A \in \mathcal{D}$, because

$$\begin{aligned}\mathbf{P}\Big((B \setminus A) \cap \bigcap_{k=2}^{n} A_{i_k}\Big) &= \mathbf{P}\Big(B \cap \bigcap_{k=2}^{n} A_{i_k}\Big) - \mathbf{P}\Big(A \cap \bigcap_{k=2}^{n} A_{i_k}\Big) \\ &= (\mathbf{P}(B) - \mathbf{P}(A)) \cdot \prod_{k=2}^{n} \mathbf{P}(A_{i_k}) \\ &= \mathbf{P}(B \setminus A) \cdot \prod_{k=2}^{n} \mathbf{P}(A_{i_k}).\end{aligned}$$

Furthermore, if $A_1, A_2, \dots \in \mathcal{D}$ with $A_1 \subseteq A_2 \subseteq A_2 \dots$, then due to the continuity of $\mathbf{P}$ from below,

$$\begin{aligned}\mathbf{P}\Big(\Big(\bigcup_{j=1}^{\infty} A_j\Big) \cap \bigcap_{k=2}^{n} A_{i_k}\Big) &= \sup_{j\in\mathbb{N}} \mathbf{P}\Big(A_j \cap \bigcap_{k=2}^{n} A_{i_k}\Big) \\ &= \sup_{j\in\mathbb{N}} \mathbf{P}(A_j) \cdot \prod_{k=2}^{n} \mathbf{P}(A_{i_k}) \\ &= \mathbf{P}\Big(\bigcup_{j=1}^{\infty} A_j\Big) \cdot \prod_{k=2}^{n} \mathbf{P}(A_{i_k}).\end{aligned}$$

Since $\mathcal{C}_{i_1}$ is $\cap$-stable and $\mathcal{C}_{i_1} \subseteq \mathcal{D}$, $\sigma(\mathcal{C}_{i_1}) \subseteq \mathcal{D}$ according to theorem 1.13. In particular, (8.1) applies for $A_{i_1} \in \sigma(\mathcal{C}_{i_1}), A_{i_2} \in \mathcal{C}_{i_2}, \dots, A_{i_n} \in \mathcal{C}_{i_n}$. Iterating the above procedure for $k = 2, \dots, n$, you get the statement. □

Corollary 8.11 (Independence of indicator functions)**.** *A family of sets $(A_i)_{i\in I}$ is independent if and only if the family of random variables $(1_{A_i})_{i\in I}$ is independent. In particular,*

$$\mathbf{P}\Big(\bigcap_{j\in J} B_j\Big) = \prod_{j\in J} \mathbf{P}(B_j)$$

for $J \subseteq_f I$, $B_j \in \{A_j, A_j^c\}, j \in J$.

Proof. For $i \in I$ let $\mathcal{C}_i = \{A_i\}$. Then $\sigma(1_{A_i}) = \{\emptyset, A_i, A_i^c, \Omega\} = \sigma(\mathcal{C}_i)$. Since $\mathcal{C}_i$ is trivially cut-stable, the statement follows from Proposition 8.10. □

Corollary 8.12 (Grouping). *Let $(\mathcal{F}_i)_{i\in I}$ be a family of independent σ-algebras. Further, let $\mathcal{I}$ be a partition of I, i.e. $\mathcal{I} = \{I_k, k \in K\}$ with $\biguplus_{k\in K} I_k = I$, so the I_k are disjoint and their union is I. Then, $(\sigma(\mathcal{F}_i : i \in I_k))_{k\in K}$ is also an independent system.*

Proof. The set system $\mathcal{C}_k := \{\bigcap_{i\in J_k} A_i : J_k \subseteq_f I_k, A_i \in \mathcal{F}_i\}$ is $\cap$-stable and $\sigma(\mathcal{C}_k) = \sigma(\mathcal{F}_i : i \in I_k)$, $k \in K$. Since, according to the assumption, the family $(\mathcal{C}_k)_{k\in K}$ is independent, the assertion follows from Proposition 8.10. □

We now come to the main statement of this section, Kolmogorov's 0-1 law. For this we introduce a certain σ-algebra, the terminal σ-algebra.

Definition 8.13 (Terminal and trivial σ-algebras).

1. *Let $\mathcal{F}_1, \mathcal{F}_2, ... \subseteq \mathcal{F}$ be a sequence of σ-algebras. Then*
$$\mathcal{T}(\mathcal{F}_1, \mathcal{F}_2, \dots) = \bigcap_{n\geq 1} \sigma\Big(\bigcup_{m>n} \mathcal{F}_m\Big)$$
the σ-algebra of terminal events *of $\mathcal{F}_1, \mathcal{F}_2, \dots$*

2. *A σ-algebra $\widetilde{\mathcal{F}} \subseteq \mathcal{F}$ is called* **P***-trivial if $\mathbf{P}(A) \in \{0, 1\}$ for all $A \in \widetilde{\mathcal{F}}$.*

Lemma 8.14 (Trivial σ-algebras).

1. *A σ-algebra $\widetilde{\mathcal{F}}$ is* **P***-trivial if and only if $\widetilde{\mathcal{F}}$ is independent of itself.*

2. *Let $\widetilde{\mathcal{F}}$ be a* **P***-trivial σ-algebra and X a $\widetilde{\mathcal{F}}$-measurable random variable with values in a separable metric space E. Then X is constant, almost surely.*

Proof. 1. Let $\widetilde{\mathcal{F}}$ be **P**-trivial and $A, B \in \widetilde{\mathcal{F}}$. Then $\mathbf{P}(A \cap B) = \mathbf{P}(A) \wedge \mathbf{P}(B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$, therefore $\widetilde{\mathcal{F}}$ is independent of itself. If on the other hand, $\widetilde{\mathcal{F}}$ is independent of itself and $A \in \widetilde{\mathcal{F}}$, then $\mathbf{P}(A) = \mathbf{P}(A \cap A) = (\mathbf{P}(A))^2$, i.e. $\mathbf{P}(A) \in \{0, 1\}$.
2. For $n \in \mathbb{N}$, let $(B_{nj})_{j=1,2,...}$ be a countable covering of E with balls of radius $1/n$. Since $\widetilde{\mathcal{F}}$ is a **P**-trivial σ-algebra then $\mathbf{P}(X \in B_{nj}) \in \{0, 1\}$ applies to all n, j. For $n \in \mathbb{N}$ let $J_n := \{j \in \mathbb{N} : \mathbf{P}(X \in B_{nj}) = 1\} \neq \emptyset$. Thus, due to the continuity from above, $\mathbf{P}\Big(X \in \bigcap_{n=1}^\infty \bigcap_{j\in J_n} B_{nj}\Big) = 1$. Since $\bigcap_{n=1}^\infty \bigcap_{j\in J_n} B_{nj}$ has at most one element, the assertion follows. □

Under independence, the σ-algebra of terminal events is particularly simple.

Theorem 8.15 (Kolmogorov's 0-1 law). *Let $\mathcal{F}_1, \mathcal{F}_2, \dots \subseteq \mathcal{F}$ be a sequence of independent σ-algebras. Then $\mathcal{T} := \mathcal{T}(\mathcal{F}_1, \mathcal{F}_2, \dots)$* **P***-trivial.*

Proof. Let $\mathcal{T}_n := \sigma\Big(\bigcup_{m>n} \mathcal{F}_m\Big)$, $n = 1, 2, \dots$. According to Corollary 8.12, $(\mathcal{F}_1, \dots, \mathcal{F}_n, \mathcal{T}_n)$ are independent, $n = 1, 2, \dots$ This means that $(\mathcal{F}_1, \dots, \mathcal{F}_n, \mathcal{T})$ are also independent, $n = 1, 2, \dots$ and thus also $(\mathcal{T}, \mathcal{F}_1, \mathcal{F}_2, \dots)$. Again with Corollary 8.12, it follows that $(\mathcal{T}_0, \mathcal{T})$ are independent and, since $\mathcal{T} \subseteq \mathcal{T}_0$ it also follows that $\mathcal{T}$ is independent of itself. Therefore, the assertion follows from Lemma 8.14. □

8.3 Sums of independent random variables

Many important theorems in probability theory deal with independent random variables. In this lecture, these are in particular the Strong Law of Large Numbers (Theorem 8.21) and the Central Limit Theorem (Theorem 10.8). We present here important tools for analyzing sums of independent random variables. The first is the connection with the convolution of probability measures (see section 5.4).

Proposition 8.16 (Convolution is distribution of the independent sum)**.** *Let $X_1, \dots, X_n$ be independent, real-valued random variables. Then,*

$$(X_1 + \cdots + X_n)_*\mathbf{P} = (X_1)_*\mathbf{P} * \cdots * (X_n)_*\mathbf{P}.$$

Further, for the characteristic functions

$$\psi_{X_1+\cdots+X_n} = \psi_{X_1} \cdots \psi_{X_n}$$

and, if $X_1, \dots, X_n$ assume values in $\mathbb{R}_+$,

$$\mathscr{L}_{X_1+\cdots+X_n} = \mathscr{L}_{X_1} \cdots \mathscr{L}_{X_n}.$$

Proof. First of all, according to Proposition 8.2 $((X_1, \dots, X_n))_*\mathbf{P} = (X_1)_*\mathbf{P} \otimes \cdots \otimes (X_n)_*\mathbf{P}$. Thus, the first assertion already follows from Definition 5.17 of the convolution of measures. The further assertions follow from Proposition 8.5, since for example

$$\begin{aligned}\psi_{X_1+\cdots+X_n}(t) &= \mathbf{E}[e^{it(X_1+\cdots+X_n)}] = \mathbf{E}\big[e^{itX_1} \cdots e^{itX_n}\big] \\ &= \mathbf{E}[e^{itX_1}] \cdots \mathbf{E}[e^{itX_n}] = \psi_{X_1}(t) \cdots \psi_{X_n}(t).\end{aligned}$$ □

Kolmogorov's 0-1 law provides a very simple statement as to when sums of independent random variables are almost sure to converge.

Proposition 8.17 (Convergence of sums of independent random variables)**.** *Let $X_1, X_2, \dots$ be independent random variables and $S_n := X_1 + \cdots + X_n$.*

1. *Then,*

$$\mathbf{P}(\omega : S_n(\omega) \text{ converges for } n \to \infty) \in \{0, 1\}$$

2. *Further,*

$$\mathbf{P}(\omega : S_n(\omega)/n \text{ converges for } n \to \infty) \in \{0, 1\}.$$

   If $\mathbf{P}(S_n/n$ converges$) = 1$, the limit value is almost surely constant.

Proof. Set $\mathcal{F}_i := \sigma(X_i)$, $i = 1, 2, \dots$ This means that the family $(\mathcal{F}_i)_{i=1,2,\dots}$ is independent. The set $\{\omega : S_n(\omega) \text{ converges for } n \to \infty\}$ is measurable with respect to $\mathcal{T}(\mathcal{F}_1, \mathcal{F}_2, \dots)$ and thus the first statement from Theorem 8.15 follows. In the same way it follows that $\mathbf{P}(S_n/n \text{ converges}) \in \{0, 1\}$. Let $S = \lim_{n\to\infty} S_n(n)/n$. Thus, for all $m = 1, 2, \dots,$

$$S = \lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \lim_{n\to\infty} \frac{X_m + \cdots + X_n}{n},$$

so S is measurable wrt $\sigma\Big(\bigcup_{k\geq m} \mathcal{F}_k\Big)$. This means that S is also $\mathcal{T}$-measurable and therefore almost surely constant according to Theorem 8.15 and Lemma 8.14. □

Proposition 8.18 (Maximum inequality of Kolmogorov)**.** *Let $X_1, X_2, \dots \in \mathcal{L}^2$ be independent random variables. Then, for $K > 0$,*

$$\mathbf{P}\Big(\sup_{n\in\mathbb{N}}\Big|\sum_{k=1}^n X_k - \mathbf{E}[X_k]\Big| > K\Big) \le \frac{\sum_{n=1}^\infty \mathbf{V}(X_n)}{K^2}.$$

Proof. Wlog, let $\mathbf{E}[X_k] = 0, k = 1, 2, \dots$. We further set $S_n = X_1 + \dots + X_n$ and $T := \inf\{n : |S_n| > K\}$. Then, $\mathbf{P}(\sup_n |S_n| > K) = \mathbf{P}(T < \infty)$. Because of Corollary 8.12, $S_k \cdot 1_{T=k}$ and $S_n - S_k$ are independent for $k \le n$. Therefore

$$\begin{aligned}\sum_{k=1}^n \mathbf{E}[X_k^2] = \mathbf{E}[S_n^2] &\ge \sum_{k=1}^n \mathbf{E}[S_n^2, T = k]\\ &= \sum_{k=1}^n \mathbf{E}[S_k^2 + (S_n - S_k + 2S_k)(S_n - S_k), T = k]\\ &\ge \sum_{k=1}^n \mathbf{E}[S_k^2, T = k] + 2\mathbf{E}[S_k(S_n - S_k), T = k]\\ &= \sum_{k=1}^n \mathbf{E}[S_k^2, T = k] \ge K^2\mathbf{P}(T \le n)\end{aligned}$$

Now follows the assertion with $n \to \infty$. □

Theorem 8.19 (Convergence criterion for series)**.** *Let $X_1, X_2, \dots \in \mathcal{L}^2$ be independent random variables with $\sum_{n=1}^\infty \mathbf{V}[X_n] < \infty$. Then, $\sum_{k=1}^n X_k - \mathbf{E}[X_k]$ converges almost surely.*

Proof. Again, let $\mathbf{E}[X_k] = 0, k = 1, 2, \dots$ and we write $S_n = X_1 + \dots + X_n$. For $\varepsilon > 0$, according to Proposition 8.18,

$$\lim_{k\to\infty} \mathbf{P}(\sup_{n\ge k} |S_n - S_k| > \varepsilon) \le \lim_{k\to\infty} \frac{\sum_{n=k+1}^\infty \mathbf{E}[X_n^2]}{\varepsilon^2} = 0.$$

Therefore, $\sup_{n\ge k} |S_n - S_k| \xrightarrow{k\to\infty}_p 0$. So, by Proposition 7.6, there is a subsequence $k_1, k_2, \dots$ with $\sup_{n\ge k_i} |S_n - S_{k_i}| \xrightarrow{i\to\infty}_{fs} 0$. However, since $(\sup_{n\ge k} |S_n - S_k|)_{k=1,2,\dots}$ is decreasing, $\sup_{n\ge k} |S_n - S_k| \xrightarrow{k\to\infty}_{fs} 0$ applies. This means, however, that $(S_n)_{n=1,2,\dots}$ converges. □

8.4 The Strong Law of Large Numbers

In the lecture *Basic Probability*, we already proved the weak law of large numbers: if $X_1, X_2, \dots \in \mathcal{L}^2$ are identically distributed and uncorrelated, then, for $\varepsilon > 0$

$$\mathbf{P}\Big(\frac{1}{n}\Big|\sum_{k=1}^n (X_k - \mathbf{E}[X_k])\Big| > \varepsilon\Big) \le \frac{1}{\varepsilon^2}\mathbf{V}\Big[\frac{1}{n}\sum_{k=1}^n X_k\Big] = \frac{1}{\varepsilon^2 n^2}\sum_{k=1}^n \mathbf{V}[X_k] = \frac{\mathbf{V}[X_1]}{\varepsilon^2 n} \xrightarrow{n\to\infty} 0.$$

As we now know, this means in other terms,

$$\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{n\to\infty}_p \mathbf{E}[X_0].$$

We now want to improve this statement in two directions. On the one hand, we want to replace convergence in probability by almost sure convergence, and on the other hand only assume the existence of first moments (but not the existence of second moments). First, however, we define what exactly what we mean when we say that a sequence of random variables follows a law of large numbers.

Definition 8.20 (Law of large numbers). *Let $X_1, X_2, \dots \in \mathcal{L}^1$ be a sequence of real-valued random variables. We say that the sequence follows the weak law of large numbers if*

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbf{E}[X_k]) \xrightarrow{n\to\infty}_p 0.$$

The sequence satisfies the strong law of large numbers if

$$\frac{1}{n}\sum_{k=1}^{n}(X_k - \mathbf{E}[X_k]) \xrightarrow{n\to\infty}_{fs} 0.$$

Theorem 8.21 (Strong law for independent random variables). *A sequence $X_1, X_2, \dots \in \mathcal{L}^1$ of independent and identically distributed random variables satisfies the strong law of large numbers, i.e.*

$$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{n\to\infty}_{fs} \mathbf{E}[X_1].$$

Remark 8.22 (Weak law of large numbers). *Since convergence in probability is implied by almost sure convergence (see Proposition 7.6), the sequence $X_1, X_2, \dots$ from the theorem also satisfies the weak law of large numbers. Furthermore, the sequence $X_1^+, X_2^+, \dots$ also satisfies the strong law and $\mathbf{E}[\frac{1}{n}(X_1^+ + \cdots X_n^+)] = \mathbf{E}[X_1^+]$. This means that the sequence $(\frac{1}{n}(X_1^+ + \cdots + X_n^+))_{n=1,2,\dots}$ is uniformly according to Corollary 7.12. In the same way, the sequence of partial sums of the negative parts is uniformly integrable. It follows from Theorem 7.11 that $\frac{1}{n}(X_1 + \cdots + X_n) \xrightarrow{n\to\infty}_{\mathcal{L}^1} \mathbf{E}[X_1]$.*

Remark 8.23 (Finite fourth and second moments). *The difficulty in proving the strong law is that only may be used that $X_1 \in \mathcal{L}^1$. The proof is significantly easier if we use $X_1 \in \mathcal{L}^4$ or $X_1 \in \mathcal{L}^2$. We start with these two proofs and write $S_n := X_1 + \cdots + X_n$.*

1. *The case $X_1 \in \mathcal{L}^4$: Here you can get by without further aids: From the linearity of the expected value, it is clear that $\mathbf{E}[S_n/n] = \mathbf{E}[X_1]$. Wlog, let $\mathbf{E}[X_1] = 0$, otherwise you go to the random variables $X_1 - \mathbf{E}[X_1], X_2 - \mathbf{E}[X_2], \dots \in \mathcal{L}^4$. First we calculate with the help of the independence of $(X_k)_{k=1,2,\dots}$*

$$\mathbf{E}[S_n^4] = \sum_{k=1}^{n}\mathbf{E}[X_k^4] + 3\sum_{\substack{k,l=1\\k\neq l}}^{\infty}\mathbf{E}[X_k^2 X_l^2] \leq (n + 6n^2)\mathbf{E}[X_1^4]$$

because of the Cauchy-Schwartz inequality. From this,

$$\mathbf{E}\Big[\sum_{n=1}^{\infty}\Big(\frac{S_n}{n}\Big)^4\Big] \leq \sum_{n=1}^{\infty}\frac{n + 6n^2}{n^4}\mathbf{E}[X_1^4] < \infty.$$

Therefore, $\sum_{n=1}^{\infty}\Big(\frac{S_n}{n}\Big)^4 < \infty$ applies is almost sure, in particular $\frac{S_n}{n} \xrightarrow{n\to\infty}_{fs} 0$.

2. *The case $X_1 \in \mathcal{L}^2$: Here the convergence criterion for series, theorem 8.19 is of crucial help. We also need the following result:*

Lemma 8.24 (Kronecker Lemma)**.** *Let $x_1, x_2, \dots \in \mathbb{R}$, $y_1, y_2, \dots \in \mathbb{R}$ be monotone with $y_n \uparrow \infty$ and $\sum_{n=1}^\infty x_n/y_n < \infty$. Then, $\sum_{k=1}^n x_k/y_n \xrightarrow{n\to\infty} 0$.*

Proof. Let $z_0 = 0, z_n := \sum_{k=1}^n x_k/y_k$. Then $z_n \xrightarrow{n\to\infty} z_\infty < \infty$ and $x_k = y_k(z_k - z_{k-1})$. We write with $y_0 = 0$

$$\begin{aligned}\frac{\sum_{k=1}^n x_k}{y_n} &= \frac{1}{y_n}\sum_{k=1}^n y_k(z_k - z_{k-1}) = z_n + \frac{1}{y_n}\Big(\sum_{k=0}^{n-1} y_k z_k - \sum_{k=1}^n y_k z_{k-1}\Big) \\ &= z_n - \frac{1}{y_n}\Big(\sum_{k=1}^n y_k z_{k-1} - y_{k-1}z_{k-1}\Big) \\ &\xrightarrow{n\to\infty} z_\infty - z_\infty \cdot \lim_{n\to\infty}\frac{1}{y_n}\sum_{k=1}^n y_k - y_{k-1} = 0.\end{aligned}$$

□

Back to the proof of the strong law in the case $X_1 \in \mathcal{L}^2$. Wlog, let $\mathbf{E}[X_1] = 0$. Consider the sequence $X_1/1, X_2/2, \dots$ Because $\sum_{n=1}^\infty \mathbf{V}[X_n/n] = \mathbf{V}[X_1]\sum_{n=1}^\infty 1/n^2$ applies according to Theorem 8.19 that $\sum_{k=1}^n X_k/k$ almost surely converges. With Lemma 8.24 it follows that $S_n/n \xrightarrow{n\to\infty}_{fs} 0$.

Proof of theorem 8.21 if $X_1 \in \mathcal{L}^1$. It is sufficient to consider the case of non-negative random variables. In the general case, note that $X_1^+, X_2^+, \dots \in \mathcal{L}^1$ and $X_1^-, X_2^-, \dots \in \mathcal{L}^1$ fulfill the conditions of the theorem, and from $(X_1^+ + \dots + X_n^+)/n \xrightarrow{n\to\infty}_{fs} \mathbf{E}[X_1^+]$ and $(X_1^- + \dots + X_n^-)/n \xrightarrow{n\to\infty}_{fs} \mathbf{E}[X_1^-]$ the statement follows due to linearity of the expectation.

For $S_n = X_1 + \dots + X_n$ we will show that

$$\mathbf{E}[\limsup_{n\to\infty} S_n/n] \le \mathbf{E}[X_1]. \tag{8.2}$$

If this is true, then firstly

$$\begin{aligned}\mathbf{E}[\liminf_{n\to\infty} S_n/n] &\ge \mathbf{E}[\liminf_{n\to\infty}(X_1 \wedge k + \dots + X_n \wedge k)/n] \\ &= k - \mathbf{E}[\limsup_{n\to\infty}((k - X_1)^+ + \dots + (k - X_n)^+)/n] \\ &\ge \mathbf{E}[k - (k - X_1)^+] \xrightarrow{k\to\infty} \mathbf{E}[X_1].\end{aligned}$$

Secondly, then $\mathbf{E}[\limsup_{n\to\infty} S_n/n - \liminf_{n\to\infty} S_n/n] = 0$, i.e. $\limsup_{n\to\infty} S_n/n = \liminf_{n\to\infty} S_n/n = 0$ almost surely, since both $\liminf_{n\to\infty} S_n/n$ as well as $\limsup_{n\to\infty} S_n/n$ are terminal functions, and thus according to Theorem 8.15 and Lemma 8.14 are almost surely constant. Furthermore,

$$\liminf_{n\to\infty} S_n/n = \mathbf{E}[\liminf_{n\to\infty} S_n/n] \ge \mathbf{E}[X_1] \ge \mathbf{E}[\limsup_{n\to\infty} S_n/n] = \limsup_{n\to\infty} S_n/n,$$

from which the assertion follows.

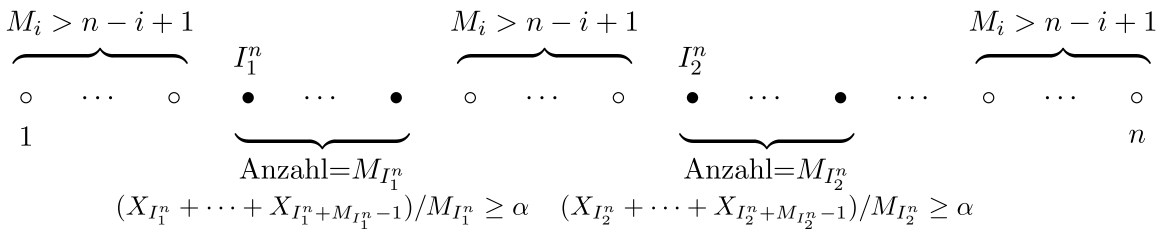


Figure 1: Illustration of M_i, I_j^n, introduced below (8.3). The size L_n is the number of contiguous areas of $\bullet$'s.

It therefore remains to show (8.2). Wlog let $\mathbf{E}[X_1] > 0$, otherwise $X_k = 0$ is almost sure, $k = 1, 2, \dots$ and the statement is trivial. For this we will use

$$0 < \alpha < \mathbf{E}[\limsup_{n\to\infty} S_n/n] \qquad \Longrightarrow \qquad \alpha \leq \mathbf{E}[X_1] \tag{8.3}$$

can be proved. According to the assumption, for $i = 0, 1, 2, \dots$

$$\alpha < \mathbf{E}[\limsup_{n\to\infty} S_n/n] = \limsup_{n\to\infty} S_n/n = \limsup_{n\to\infty} (X_{i+1} + \cdots X_{i+n})/n.$$

Thus,

$$M_i := \inf\{n \in \mathbb{N} : (X_i + \cdots + X_{i+n-1})/n \geq \alpha\}$$

is finite, almost surely, $i = 1, 2, \dots$ The M_i's are identically distributed. We define recursively for $n = 1, 2, \dots$ (see also Figure 1) $I_1^n = 0$ and for $j = 0, 1, 2, \dots$ (with $M_0 := 0$)

$$I_{j+1}^n := \inf\{i \in \mathbb{N} : i \geq I_j^n + M_{I_j}^n, M_i \leq n - i + 1\}$$

with $\inf \emptyset = \infty$ and $L_n := \sup\{n \in \mathbb{N}_0 : I_j^n < \infty\}$. This means that for $1 \leq j \leq L_n$, $I_j^n + M_{I_j^n} \leq n$, i.e. $(X_{I_j^n} + \cdots + X_{I_j^n + M_{I_j^n} - 1})/M_{I_j^n} \geq \alpha$. We now use this by means of

$$\begin{aligned}
\mathbf{E}[X_1] &= \mathbf{E}[(X_1 + \cdots + X_n)/n] \\
&\geq \tfrac{1}{n}\mathbf{E}\Big[\sum_{j=1}^{L_n} M_{I_j^n} \cdot (X_{I_j^n} + \cdots + X_{I_j^n + M_{I_j^n} - 1})/M_{I_j^n}\Big] \\
&\geq \tfrac{\alpha}{n}\mathbf{E}\Big[\sum_{j=1}^{L_n} M_{I_j^n}\Big] = \alpha - \tfrac{\alpha}{n}\mathbf{E}\Big[n - \sum_{j=1}^{L_n} M_{I_j^n}\Big] \\
&\geq \alpha - \tfrac{\alpha}{n}\mathbf{E}\Big[\sum_{i=1}^{n} 1_{M_i > n-i+1}\Big] \\
&= \alpha\Big(1 - \tfrac{1}{n}\sum_{i=1}^{n} \mathbf{P}(M_i > i)\Big) \xrightarrow{n\to\infty} \alpha,
\end{aligned}$$

since $\big(\frac{1}{n}\sum_{i=1}^n \mathbf{P}(M_i > i)\big)_{n=1,2,\dots}$ as Cesàro-Limes of $(\mathbf{P}(M_i > i))_{i=1,2,\dots}$ because of the identity of the distributions of the M_i's converges to 0. Thus (8.3) is shown and the assertion is proven. $\square$

We now give a simple application of the strong law. It often happens in statistics that a large number of independent, identically distributed, real-valued random variables must be studied. The Glivenko-Cantelli theorem (Theorem 8.26) states that the empirical distribution of the random variables almost surely converges to the underlying distribution.

Definition 8.25 (Empirical distribution)**.** *Let $X_1, X_2, \dots$ be random variables. For $n = 1, 2, \dots$ the distribution is called (random) probability distribution*

$$\widehat{\mu}_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}$$

the empirical distribution *of $X_1, \dots, X_n$. If the random variables are real-valued, then in addition*

$$\widehat{F}_n(x) := \frac{1}{n}\sum_{k=1}^{n} 1_{X_k \le x},$$

the empirical distribution function *of $X_1, \dots, X_n$.*

Theorem 8.26 (Glivenko-Cantelli Theorem)**.** *Let $X_1, X_2, \dots$ be independent, real-valued random variables with identical distribution with distribution function F. Then,*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} widehasF_n(x) - F(x)| \xrightarrow{n\to\infty}_{fs} 0.$$

Proof. For $x \in \mathbb{R}$ and $n = 1, 2, \dots$ let $Y_n(x) := 1_{X_n \le x}$ and $Z_n(x) := 1_{X_n < x}$. According to Theorem 8.21, for each $x \in \mathbb{R}$

$$\begin{aligned}
\widehat{F}_n(x) &= \frac{1}{n}\sum_{k=1}^{n} Y_k(x) \xrightarrow{n\to\infty}_{fs} \mathbf{E}[Y_1(x)] = \mathbf{P}(X_1 \le x) = F(x),\\
\widehat{F}_n(x-) &= \frac{1}{n}\sum_{k=1}^{n} Z_k(x) \xrightarrow{n\to\infty}_{fs} \mathbf{E}[Z_1(x)] = \mathbf{P}(X_1 < x) = F(x-).
\end{aligned}$$

We must show that these limits hold uniformly for all $x \in \mathbb{R}$. For $N = 1, 2, \dots$ and $j = 0, \dots, N$ we set

$$x_j^N := \inf\{x \in \mathbb{R} : F(x) \ge j/N\}$$

and

$$R_n^N := \max_{j=1,\dots,N-1} \big(|\widehat{F}_n(x_j^N) - F(x_j^N)| + |\widehat{F}_n(x_j^N-) - F(x_j^N-)|\big).$$

For $N = 1, 2, \dots$, therefore, $R_n^N \xrightarrow{n\to\infty}_{fs} 0$. Furthermore, for $x \in (x_{j-1}^N, x_j^N)$

$$\begin{aligned}
\widehat{F}_n(x) &\le \widehat{F}_n(x_j^N) \le \widehat{F}_n(x_j^N) + R_n^N \le F(x) + R_n^N + \tfrac{1}{N},\\
\widehat{F}_n(x) &\ge \widehat{F}_n(x_{j-1}^N) \ge F(x_{j-1}^N) - R_n^N \ge F(x) - R_n^N - \tfrac{1}{N},
\end{aligned}$$

thus, for each $N = 1, 2, \dots$

$$\sup_{x\in\mathbb{R}} widehatF_n(x) - F(x)| \le \tfrac{1}{N} + R_n^N \xrightarrow{n\to\infty}_{fs} \tfrac{1}{N}.$$

Since the left-hand side does not depend on N, the assertion follows with $N \to \infty$. □

9 Weak convergence

For measurable spaces, we have often used the Borel σ-algebra, i.e. the σ-algebra that is generated by a topology. In this section we will often assume that the topological space is Polish, i.e. separable and metrizable by a complete metric; recall from Definition A.1 in the manuscript on measure theory. To save us some work, we will assume throughout that (E, r) is a metric space and sometimes we will assume that it is complete and separable.

For a measurable mapping $f : E \to \mathbb{R}$ and a measure μ on $\mathcal{B}(E)$ (the Borel's σ-algebra of E) we will use throughout this and the next chapter the notation

$$\mu[f] := \int f d\mu.$$

9.1 Definition and simple properties

So far, we have dealt with different types of convergence of random variables. The convergence in distribution of random variables is the same as the weak convergence of the distributions of random variables. For the motivation behind the following definitions, let us recall a fact: in a metric space (E, r) we have $x_n \xrightarrow{n\to\infty} x$ if and only if $f(x_n) \xrightarrow{n\to\infty} f(x)$ for all continuous functions on E (i.e. $f \in \mathcal{C}(E, \mathbb{R})$).

Definition 9.1 (Weak convergence and convergence in distribution)**.**

1. *We denote by $\mathcal{P}(E)$ the set of probability measures on $\mathcal{B}(E)$ and with $\mathcal{P}_{\leq 1}(E)$ the set of finite measures μ on $\mathcal{B}(E)$ with $\mu(E) \leq 1$. Further, $\mathcal{C}_b(E)$ is the set of the real-valued, bounded, continuous functions on E and $\mathcal{C}_c(E) \subseteq \mathcal{C}_b(E)$ is the set of the real-valued, bounded continuous functions on E with compact support.*

2. *A sequence $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$ converges* weakly *to $\mathbf{P} \in \mathcal{P}(E)$, if*

$$\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mathbf{P}[f] \tag{9.1}$$

   for all $f \in \mathcal{C}_b(E)$. We then write

$$\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}.$$

3. *Let $\mu_1, \mu_2, \dots \in \mathcal{P}_{\leq 1}$ and μ be a measure on E. If* (9.1) *only applies to all $f \in \mathcal{C}_c(E)$, we say that μ_n converges* vagely *to μ. We then write*

$$\mu_n \underset{v}{\xRightarrow{n\to\infty}} \mu.$$

4. *Let $X, X_1, X_2, \dots$ be random variables on probability spaces $(\Omega, \mathcal{A}, \mathbf{P})$, $(\Omega_1, \mathcal{A}_1, \mathbf{P}_1)$, $(\Omega_2, \mathcal{A}_2, \mathbf{P}_2), \dots$ with values in E. Then, $X_1, X_2, \dots$ converges* in distribution *to X if $(X_n)_*\mathbf{P}_n \xRightarrow{n\to\infty} X_*\mathbf{P}$. We then write*

$$X_n \xRightarrow{n\to\infty} X.$$

Remark 9.2. *1. Note that for random variables $X, X_1, X_2, \dots$ with values in E, we have $X_n \xrightarrow{n\to\infty} X$ if*

$$\mathbf{P}[f(X_n)] \xrightarrow{n\to\infty} \mathbf{P}[f(X)]$$

for all $f \in \mathcal{C}_b(E)$. Many of the following results can therefore be formulated in two ways: either by means of probability distributions, or by means of random variables. The connection here is always that the statement about the probability distributions is also a statement about the distributions of the random variables.

2. The weak limit of probability measures must again be a probability measure, since $1 \in \mathcal{C}_b(E)$. The vague limit of probability measures does not necessarily have to be a probability measure, since $1 \notin \mathcal{C}_c(E)$ if E is not compact; see also Example 9.3.1. After all, the vague limit is in $\mathcal{P}_{\leq 1}(E)$, as Lemma 9.12 shows.

3. We already know the almost sury convergence, the convergence in probability, and the convergence in $\mathcal{L}^p$ of random variables $X_1, X_2, \dots$ to X. The difference to convergence in distribution is that the latter does not require that the random variables are defined on the same probability space.

4. By Definition 9.1, the topology of weak convergence on $\mathcal{P}(E)$ is the weakest (i.e. the smallest) topology for which $\mathbf{P} \mapsto \mathbf{P}[f]$ for all $f \in \mathcal{C}_b(E)$ is continuous.

Example 9.3. *1. Let $x, x_1, x_2, \dots \in \mathbb{R}$ with $x_n \xrightarrow{n\to\infty} x$ and $\mathbf{P} = \delta_x, \mathbf{P}_1 = \delta_{x_1}, \mathbf{P}_2 = \delta_{x_2}, \dots$ Then, $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$, since*

$$\mathbf{P}_n[f] = f(x_n) \xrightarrow{n\to\infty} f(x) = \mathbf{P}[f]$$

for all $f \in \mathcal{C}_b(\mathbb{R})$.

If the sequence $x_1, x_2, \dots$ diverges, for example $x_n = n$, then $\mathbf{P}_n \xrightarrow{n\to\infty}_v 0$ (this is the 0-measure on $\mathcal{B}(\mathbb{R})$), since

$$\mathbf{P}_n[f] = f(x_n) \xrightarrow{n\to\infty} 0 = 0[f]$$

for all $f \in \mathcal{C}_c(\mathbb{R})$. However, weak convergence does not hold, since $\mathbf{P}_n[1] = 1 \neq 0 = 0[1]$.

2. Let $X, X_1, X_2, \dots$ be identically distributed. Then $X_n \xRightarrow{n\to\infty} X$, but in general the convergence is neither almost sure, nor in probability nor in $\mathcal{L}^p$ for any $p > 0$.

3. As we will see, the Central Limit Theorem (Theorem 10.8), is a result about convergence in distribution. In its simplest form, the theorem of deMoivre-Laplace (see also Remark 9.8 and Example 9.34), it states: let $p \in (0,1)$, $X_n \sim B(n,p), n = 1, 2, \dots$ and $X \sim N(0,1)$. Then,

$$\frac{X_n - np}{\sqrt{np(1-p)}} \xRightarrow{n\to\infty} X.$$

4. Similarly, the Poisson approximation of the binomial distribution is a statement about convergence in distribution (see the course in Basic Probability *and Theorem 10.5): let $X_n \sim B(n, p_n), n = 1, 2, \dots$ with $n \cdot p_n \xrightarrow{n\to\infty} \lambda$ and $X \sim Poi(\lambda)$. Then,*

$$X_n \xrightarrow{n\to\infty} X.$$

Lemma 9.4 (Uniqueness of the weak limit)**.** *Let* $\mathbf{P}, \mathbf{Q}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$ *with* $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$ *and* $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{Q}$*. Then* $\mathbf{P} = \mathbf{Q}$*.*

Proof. According to Proposition 2.11 it suffices to show that $\mathbf{P}(A) = \mathbf{Q}(A)$ for all closed $A \subseteq E$. (The set of all closed sets is a $\cap$-stable generator of $\mathcal{B}(E)$.) So let $A \subseteq E$ be closed. We set

$$r(x, A) := \inf_{y\in A} r(x, y)$$

and

$$f_m(x) \mapsto (1 - m \cdot r(x, A))^+.$$

for $m = 1, 2, \dots$. Then $f_m \xrightarrow{m\to\infty} 1_A$, since A is is closed. Then, using cominated convergence,

$$\mathbf{P}(A) = \lim_{m\to\infty} \mathbf{P}[f_m] = \lim_{m\to\infty} \lim_{n\to\infty} \mathbf{P}_n[f_m] = \lim_{m\to\infty} \lim_{m\to\infty} \mathbf{Q}[f_m] = \mathbf{Q}(A)$$

and the assertion follows. □

Recall the initial figure of Chapter 7. A sequence of random variables can converge almost surely, in probability, in $\mathcal{L}^p$ or in distribution. Convergence in distribution is the weakest of these terms in the following sense.

Proposition 9.5 (Convergence in probability and in distribution)**.** *Let* $X, X_1, X_2, \dots$ *be random variables with values in* E*. If* $X_n \xrightarrow{n\to\infty}_p X$*, then* $X_n \xRightarrow{n\to\infty} X$*. If* X *is constant, the inversion also applies.*

Proof. Let $X_n \xrightarrow{n\to\infty}_p X$. Suppose that there is an $f \in \mathcal{C}_b(E)$ such that $\lim_{n\to\infty} \mathbf{P}[f(X_n)] \neq \mathbf{P}[f(X)]$. Then there is a subsequence $(n_k)_{k=1,2,\dots}$ and a $\varepsilon > 0$ with

$$\lim_{k\to\infty} |\mathbf{P}[f(X_{n_k})] - \mathbf{P}[f(X)]| > \varepsilon. \tag{9.2}$$

Because of $X_n \xrightarrow{n\to\infty}_p X$ and Proposition 7.6 there is a subsequence $(n_{k_\ell})_{\ell=1,2,\dots}$ such that $X_{n_{k_\ell}} \xrightarrow{\ell\to\infty} X$ almost surely. By dominated convergence, this would imply

$$\lim_{\ell\to\infty} \mathbf{P}[f(X_{n_{k_\ell}})] = \mathbf{P}[f(X)]$$

in contradiction to (9.2).

For the inverse, let $X = s \in E$. Note that $x \mapsto r(x, s) \wedge 1$ is a bounded, continuous function and therefore

$$\mathbf{P}[r(X_n, s) \wedge 1] \xrightarrow{n\to\infty} \mathbf{P}[r(X, s) \wedge 1] = 0.$$

Thus, $X_n \xrightarrow{n\to\infty}_p X$ holds because of (7.1). □

Theorem 9.6 (Portmanteau theorem)**.** *Let* $X, X_1, X_2, \dots$ *be random variables with values in* E*. The following conditions are equivalent:*

(i) $X_n \xRightarrow{n\to\infty} X$

(ii) $\mathbf{P}[f(X_n)] \xrightarrow{n\to\infty} \mathbf{P}[f(X)]$ *for all bounded, Lipschitz-continuous functions* f*.*

(iii) $\liminf_{n\to\infty} \mathbf{P}(X_n \in G) \geq \mathbf{P}(X \in G)$ *for all open* $G \subseteq E$.

(iv) $\limsup_{n\to\infty} \mathbf{P}(X_n \in F) \leq \mathbf{P}(X \in F)$ *for all completed* $F \subseteq E$.

(v) $\lim_{n\to\infty} \mathbf{P}(X_n \in B) = \mathbf{P}(X \in B)$ *for all* $B \in \mathcal{B}(E)$ *with*[14] $\mathbf{P}(X \in \partial B) = 0$.

Proof. $(i) \to (ii)$: clear.
$(ii) \Rightarrow (iv)$: Let $F \subseteq E$ be closed and $f_1, f_2, \dots$ Lipschitz-continuous such that $f_k \downarrow 1_F$. (For example, one chooses $\varepsilon_k \downarrow 0$ and $f_k(x) = (1 - \frac{1}{\varepsilon_k} r(x, F))^+$, where $r(x, F) := \inf_{y \in F} r(x, y)$.) This means that

$$\limsup_{n\to\infty} \mathbf{P}(X_n \in F) \leq \inf_{k=1,2,\dots} \limsup_{n\to\infty} \mathbf{P}[f_k(X_n)] = \inf_{k=1,2,\dots} \mathbf{P}[f_k(X)] = \mathbf{P}(X \in F).$$

$(iii) \iff (iv)$: That is clear. For $(iii) \Rightarrow (iv)$, set $F := E \setminus G$ and for $(iv) \Rightarrow (iii)$, set $G := E \setminus F$.
$(iii) \Rightarrow (i)$: Let $f \geq 0$ be continuous. By Proposition 6.10 and Fatou's lemma,

$$\begin{aligned} \mathbf{P}[f(X)] &= \int_0^\infty \mathbf{P}(f(X) > t)dt \leq \int_0^\infty \liminf_{n\to\infty} \mathbf{P}(f(X_n) > t)dt \\ &\leq \liminf_{n\to\infty} \int_0^\infty \mathbf{P}(f(X_n) > t)dt = \liminf_{n\to\infty} \mathbf{P}[f(X_n)]. \end{aligned}$$

For $-c < f < c$, since $-f + c \geq is0$,

$$\begin{aligned} \limsup_{n\to\infty} \mathbf{P}[f(X_n)] &= c - \liminf_{n\to\infty} \mathbf{P}[-f(X_n) + c] \leq c - \mathbf{P}[-f(X) + c] = \mathbf{P}[f(X)] \\ &\leq \liminf_{n\to\infty} \mathbf{P}[f(X_n)], \end{aligned}$$

thus $\mathbf{P}[f(X_n)] \xrightarrow{n\to\infty} \mathbf{P}[f(X)]$.
$(iii), (iv) \to (v)$ For $B \in \mathcal{B}(E)$,

$$\mathbf{P}(X \in B^\circ) \leq \liminf_{n\to\infty} \mathbf{P}(X_n \in B^\circ) \leq \limsup_{n\to\infty} \mathbf{P}(X_n \in \overline{B}) \leq \mathbf{P}(X \in \overline{B}).$$

Given $\mathbf{P}(X \in \partial B) = \mathbf{P}(X \in \overline{B}) - \mathbf{P}(X \in B^\circ) = 0$, therefore $\mathbf{P}(X_n \in B) \xrightarrow{n\to\infty} \mathbf{P}(X \in B)$.
$(v) \to (iv)$: Assume (v) is true and $F \subseteq E$ is closed. We write $F^\varepsilon := \{x \in E : r(x, F) \leq \varepsilon\}$ for $\varepsilon > 0$. The sets $\partial F^\varepsilon \subseteq \{x : r(x, F) = \varepsilon\}$ are disjoint, so

$$\mathbf{P}(X \in \partial F^\varepsilon) = 0 \tag{9.3}$$

for Lebesgue-almost every ε. Let $\varepsilon_1, \varepsilon_2, \dots$ denote a sequence with $\varepsilon_k \downarrow 0$ such that (9.3) holds for all $\varepsilon_1, \varepsilon_2, \dots$. This means that

$$\limsup_{n\to\infty} \mathbf{P}(X_n \in F) \leq \inf_{k=1,2,\dots} \limsup_{n\to\infty} \mathbf{P}(X_n \in F^{\varepsilon_k}) = \inf_{k=1,2,\dots} \mathbf{P}(X \in F^{\varepsilon_k}) = \mathbf{P}(X \in F).$$

□

[14] For the closure $\overline{B}$ and the interior B° denote here $\partial B := \overline{B} \setminus B^\circ$ the edge of B.

Corollary 9.7 (Convergence of distribution functions)**.** *Let* $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$ *with distribution functions* $F, F_1, F_2, \dots$ *Then* $\mathbf{P}_n \xrightarrow{n\to\infty} \mathbf{P}$ *exactly if* $F_n(x) \xrightarrow{n\to\infty} F(x)$ *for all continuity points* x *of* F.

Proof. '$\Rightarrow$': If x is a continuity point of F, then $\mathbf{P}(\partial(-\infty; x]) = \mathbf{P}(\{x\}) = 0$. This means that – according to Theorem 9.6 (direction $(i) \Rightarrow (v)$) – that

$$F_n(x) = \mathbf{P}_n((-\infty; x]) \xrightarrow{n\to\infty} \mathbf{P}((-\infty; x]) = F(x).$$

'$\Leftarrow$': According to Theorem 9.6 (direction $(ii) \Rightarrow (i)$), it suffices to show that $\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mathbf{P}[f]$ for all bounded, Lipschitz functions f. Wlog, we assume that $|f| \le 1$ and f has Lipschitz constant 1. For $\varepsilon > 0$ choose $N \in \mathbb{N}$ and continuity points $y_0 < \dots < y_N$ of F, so that $F(y_0) < \varepsilon$, $F(y_N) > 1 - \varepsilon$ and $y_i - y_{i-1} < \varepsilon$ for $i = 1, \dots, N$. Then $F_n(y_i) \xrightarrow{n\to\infty} F(y_i)$ and

$$f \le 1_{(-\infty, y_0]} + 1_{(y_N, \infty)} + \sum_{i=1}^{N-1} (f(y_i) + \varepsilon) 1_{(y_i, y_{i+1}]},$$

as well as

$$\begin{aligned}\limsup_{n\to\infty} \mathbf{P}_n[f] &\le \limsup_{n\to\infty} F_n(y_0) + 1 - F_n(y_N) + \sum_{i=1}^{N} (f(y_i) + \varepsilon)(F_n(y_i) - F_n(y_{i-1}) \\ &\le 3\varepsilon + \sum_{i=1}^{N} f(y_i)(F(y_i) - F(y_{i-1})) \le 4\varepsilon + \mathbf{P}[f].\end{aligned}$$

With $\varepsilon \to 0$ and by replacing f with $1 - f$, we find $\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mathbf{P}[f]$. □

Remark 9.8 (The Theorem of deMoivre-Laplace)**.** *In Example 9.3 we claimed that deMoivre-Laplace's Theorem makes a statement about weak convergence. The Theorem states that for* $B(n, p)$*-distributed random variables* X_n, $n = 1, 2, \dots$,

$$\mathbf{P}\Big(\frac{X_n - np}{\sqrt{np(1-p)}} \le x \Big) \xrightarrow{n\to\infty} \Phi(x),$$

where Φ *is the distribution function of the standard normal distribution. As Corollary 9.7 shows, this means exactly the convergence in distribution to a standard normal distribution.*

Corollary 9.9 (Slutzky's Theorem)**.** *Let* $X, X_1, X_2, \dots, Y_1, Y_2, \dots$ *be random variables with values in* E. *If* $X_n \xRightarrow{n\to\infty} X$ *and* $r(X_n, Y_n) \xrightarrow{n\to\infty}_p 0$, *then* $Y_n \xrightarrow{n\to\infty} X$.

Proof. Let $f : E \to \mathbb{R}$ be bounded and Lipschitz-continuous with Lipschitz constant L. Then,

$$|f(x) - f(y)| \le L \cdot r(x, y) \wedge (2||f||_\infty)$$

for all $x, y \in E$. From this,

$$\limsup_{n\to\infty} \mathbf{E}[f(X_n) - f(Y_n)] \le \limsup_{n\to\infty} \mathbf{E}[L \cdot r(X_n, Y_n) \wedge (2||f||_\infty)] = 0$$

according to Lemma 7.5. Thus,

$$\limsup_{n\to\infty} \big|\mathbf{E}[f(Y_n)] - \mathbf{E}[f(X)]\big| \le \limsup_{n\to\infty} \big|\mathbf{E}[f(Y_n)] - \mathbf{E}[f(X_n)]\big| + \big|\mathbf{E}[f(X_n)] - \mathbf{E}[f(X)]\big| = 0,$$

and the claimed convergence follows with Theorem 9.6. □

Theorem 9.10 (Continuous mapping theorem)**.** *Let E be separable, (E', r') another metric space and $\varphi : E \to E'$ measurable and $U_\varphi \subseteq E$ the set of discontinuity points of φ.*

1. *If $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$ and $\mathbf{P}(U_\varphi) = 0$ and $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$, then $\varphi_*\mathbf{P}_n \xRightarrow{n\to\infty} \varphi_*\mathbf{P}$.*

2. *If $X, X_1, X_2, \dots$ are random variables with values in E and $\mathbf{P}(X \in U_\varphi) = 0$ and $X_n \xRightarrow{n\to\infty} X$, then also $\varphi(X_n) \xRightarrow{n\to\infty} \varphi(X)$.*

Proof. First, we note that 2. is an application of 1. if one sets $\mathbf{P}_n = (X_n)_*\mathbf{P}$. The set U_φ is Borel-measurable, since

$$U_\varphi^{\delta,\varepsilon} = \{x \in E : \exists y, z \in B_\delta(x), r'(\varphi(y), \varphi(z)) > \varepsilon\}$$

is Borel-measurable (here the separability of E is included) and

$$U_\varphi = \bigcup_{n=1}^\infty \bigcap_{k=1}^\infty U_\varphi^{1/k,1/n}.$$

Let $G \subseteq E'$ be open and $x \in \varphi^{-1}(G) \cap U_\varphi^c$. Since φ is continuous in x, there is a $\delta > 0$ with $\varphi(y) \in G$ (i.e. $y \in \varphi^{-1}(G)$) for all y with $r(x, y) < \delta$. Therefore, $\varphi^{-1}(G) \cap U_\varphi^c \subseteq (\varphi^{-1}(G))^\circ$. This follows with Theorem 9.6 (direction $(i) \Rightarrow (iii)$)

$$\begin{aligned}\varphi_*\mathbf{P}(G) &= \mathbf{P}(\varphi^{-1}(G)) = \mathbf{P}(\varphi^{-1}(G) \cap U_\varphi^c) \le \mathbf{P}((\varphi^{-1}(G))^\circ) \\ &\le \liminf_{n\to\infty} \mathbf{P}_n((\varphi^{-1}(G))^\circ) \le \liminf_{n\to\infty} \mathbf{P}_n(\varphi^{-1}(G)) = \liminf_{n\to\infty} \varphi_*\mathbf{P}_n(G).\end{aligned}$$

Again due to theorem 9.6 (direction $(iii) \Rightarrow (i)$), this implies $\varphi_*\mathbf{P}_n \xRightarrow{n\to\infty} \varphi_*\mathbf{P}$. □

Apart from the vague convergence, convergence in distribution is the weakest form of convergence. However, there is a connection with almost sure convergence, as the following theorem shows.

Theorem 9.11 (Weak and almost sure convergence, Skorohod)**.** *Let $X, X_1, X_2, \dots$ be random variables with values in a complete and separable space (E, r). Then, $X_n \xRightarrow{n\to\infty} X$ holds if and only if there is a probability space on which random variables $Y, Y_1, Y_2, \dots$ are defined with $Y_n \xrightarrow{n\to\infty}_{fs} Y$ and $Y \stackrel{d}{=} X, Y_1 \stackrel{d}{=} X_1, Y_2 \stackrel{d}{=} X_2, \dots$*

Proof. '$\Leftarrow$': This is clear, since almost sure convergence implies weak convergence (see Proposition 9.5).
'$\Rightarrow$': We extend the probability space on which X is defined, and we set $Y = X$. Let $E = \{1, \dots, m\}$ be finite, U be uniformly distributed on $[0, 1]$ and independent of Y, and $W_1, W_2, \dots$ independent with

$$\mathbf{P}(W_n = k) = \frac{\mathbf{P}(X_n = k) - \mathbf{P}(X = k) \wedge \mathbf{P}(X_n = k)}{1 - \sum_{l=1}^m \mathbf{P}(X = l) \wedge \mathbf{P}(X_n = l)}.$$

We set $Y_n = k$ if either

$$X = k \text{ and } U \le \frac{\mathbf{P}(X_n = k)}{\mathbf{P}(X = k)}$$

or

$$X = l \text{ and } U > \frac{\mathbf{P}(X_n = l)}{\mathbf{P}(X = l)} \text{ and } W_n = k.$$

Then

$$\begin{aligned}
\mathbf{P}(Y_n = k) &= \mathbf{P}(X = k) \cdot \frac{\mathbf{P}(X_n = k)}{\mathbf{P}(X = k)} \wedge 1 \\
&\quad + \sum_{l=1}^{m} \mathbf{P}(X = l) \cdot \Big(1 - \frac{\mathbf{P}(X_n = l)}{\mathbf{P}(X = l)}\Big)^{+} \frac{\mathbf{P}(X_n = k) - \mathbf{P}(X = k) \wedge \mathbf{P}(X_n = k)}{1 - \sum_{l'=1}^{m} \mathbf{P}(X = l') \wedge \mathbf{P}(X_n = l')} \\
&= \mathbf{P}(X_n = k) \wedge \mathbf{P}(X = k) \\
&\qquad + \sum_{l=1}^{m} (\mathbf{P}(X = l) - \mathbf{P}(X_n = l) \wedge \mathbf{P}(X = l)) \frac{\mathbf{P}(X_n = k) - \mathbf{P}(X = k) \wedge \mathbf{P}(X_n = k)}{1 - \sum_{l'=1}^{m} \mathbf{P}(X = l') \wedge \mathbf{P}(X_n = l')} \\
&= \mathbf{P}(X_n = k).
\end{aligned}$$

Thus $Y_n \stackrel{d}{=} X_n$. Since according to the condition $\mathbf{P}(X_n = k) \xrightarrow{n\to\infty} \mathbf{P}(X = k)$, the almost sure convergence follows.

For general E, let $p = 1, 2, ...$ and choose a partition of E in sets $B_1, B_2, \ldots$ in E with $\mathbf{P}(Y \in \partial B_k) = 0$ and diameter at most 2^{-p}. Choose m large enough, so that $\mathbf{P}(Y \notin B_0) < 2^{-p}$ with $B_0 := E \setminus \bigcup_{k \le m} B_k$. For $k = 1, 2, \ldots$, define random variables $\widetilde{Z}, \widetilde{Z}_1, \widetilde{Z}_2, \ldots$ such that $\widetilde{Z} = k$ exactly when $Y \in B_k$ and $\widetilde{Z}_n = k$ if $Y_n \in B_k$. Then $\widetilde{Z}_n \xRightarrow{n\to\infty} \widetilde{Z}$. Since $\widetilde{Z}, \widetilde{Z}_1, \ldots$ only takes values in a finite set, we can use random variables $Z, Z_1, Z_2, \ldots$ with $Z_n \xrightarrow{n\to\infty}_{fs} Z$. Furthermore, let $W_{n,k}$ be random variables with distribution $\mathbf{P}[X_n \in .|X_n \in B_k]$ and $\widetilde{Y}_{n,p} = \sum_k W_{n,k} 1_{Z_n = k}$, so that $\widetilde{Y}_{n,p} \stackrel{d}{=} X_n$ for all n. It is now clear

$$\Big\{ r(\widetilde{Y}_{n,p}, Y) > 2^{-p} \Big\} \subseteq \{Z_n \neq Z\} \cup \{Y \in B_0\}.$$

Since $Z_n \xrightarrow{n\to\infty}_{fs} Z$ and $\mathbf{P}\{Y \in B_0\} < 2^{-p}$, for each p there are numbers $n_1 < n_2 < \ldots$ with

$$\mathbf{P}\Big(\bigcup_{n \ge n_p} \{ r(\widetilde{Y}_{n,p}, Y) > 2^{-p} \} \Big) < 2^{-p}$$

for all p. With the Borel-Cantelli lemma we get

$$\sup_{n \ge n_p} r(\widetilde{Y}_{n,p}, Y) \le 2^{-p}$$

for almost all p. We therefore define $Y_n := \widetilde{Y}_{n,p}$ for $n_p \le n < n_{p+1}$ and note that $X_n \stackrel{d}{=} Y_n \xrightarrow{n\to\infty}_{fs} Y$. □

9.2 Prohorov' Theorem

In this section, we first examine the concept of vague convergence. We will restrict ourselves to the space $E = \mathbb{R}$. (Most of the statements shown here are still valid in locally compact spaces). It is already clear that weak convergence of distributions implies vague convergence (since all continous functions with compact support are bounded), and that the weak convergence is equivalent to the convergence of the distribution functions (Corollary 9.7). The main result here is the theorem of Helly (Theorem 9.13), which states that every sequence of probability measures has a vaguely convergent subsequence.

We then examine the question when a sequence of probability measures also has weakly convergent subsequence. This leads us to the notion of tightness of probability measures and Prohorov's theorem (Theorem 9.19).

As we have already seen in Remark 9.2.1, it can be that the vague limit measure of probability measures is not a probability measure. However, the following result shows that the limit measure has total mass at most 1.

Lemma 9.12 (Mass loss at vague convergence). *Let $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$ and μ a measure on $\mathcal{B}(\mathbb{R})$ with $\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mu[f], f \in \mathcal{C}_c(\mathbb{R})$, then $\mu \in \mathcal{P}_{\leq 1}(\mathbb{R})$ applies.*

Proof. Let $f_1, f_2, \dots \in \mathcal{C}_c(\mathbb{R})$ with $f_k \uparrow 1$. Then with monotonic convergence

$$\mu(\mathbb{R}) = \sup_{k\in\mathbb{N}} \mu[f_k] = \sup_{k\in\mathbb{N}} \limsup_{n\to\infty} \mathbf{P}_n[f_k] \leq 1.$$

. □

Theorem 9.13 (Helly's theorem). *Let $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$. Then there is a subsequence $(n_k)_{k=1,2,\dots}$ and a $\mu \in \mathcal{P}_{\leq 1}(\mathbb{R})$ with $\mathbf{P}_{n_k} \xRightarrow[v]{k\to\infty} \mu$.*

Proof. Let $F_1, F_2, \dots$ be the distribution functions of $\mathbf{P}_1, \mathbf{P}_2, \dots$ Further, let $(x_1, x_2, \dots)$ be a count of $\mathbb{Q}$. Since $[0,1]$ is compact, for each sequence there is $(F_n(x_i))_{n=1,2,\dots}$ a convergent subsequence. By means of a diagonal argument, there is a sequence $(n_k)_{k=1,2,\dots}$ such that $(F_{n_k}(x_i))_{k=1,2,\dots}$ for all i against a limit $G(x_i)$ converges to $\mathbb{Q}$. We define

$$F(x) := \inf\{G(r) : r \in \mathbb{Q}, r > x\}.$$

Since all F_n and therefore G have non-negative increments, the same applies to F. From the definition of F and the monotonicity of G, it also follows that F is right-continuous. According to Proposition 2.19, there is a measure μ on $\mathbb{R}$ with $\mu((x,y]) = F(y) - F(x)$ for all $x, y \in \mathbb{R}, x \leq y$. It remains to show that $\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mu[f]$ for all $f \in \mathcal{C}_c(\mathbb{R})$. Wlog we can assume that $f \geq is0$.

It is $F_n(x) \xrightarrow{n\to\infty} F(x)$ at all continuity points x of F by construction. There is a countable set $D \subseteq \mathbb{R}$ such that F is continuous on D^c is continuous. This means that $\mathbf{P}_n(U) \xrightarrow{n\to\infty} \mu(U)$ for all finite unions U of intervals with vertices in D^c. Now let $B \subseteq \mathbb{R}$ be open and bounded. Let $U_1, U_2, \dots$ and $V_1, V_2, \dots$ be sequences of finite unions of open intervals with vertices in C such that such that $U_k \uparrow B$, $V_k \downarrow \overline{B}$. Then,

$$\begin{aligned}\mu(B) &= \lim_{k\to\infty} \mu(U_k) = \lim_{k\to\infty} \liminf_{n\to\infty} \mathbf{P}_n(U_k) \leq \liminf_{n\to\infty} \mathbf{P}_n(B)\\ &\leq \limsup_{n\to\infty} \mathbf{P}_n(B) \leq \lim_{k\to\infty} \limsup_{n\to\infty} \mathbf{P}_n(V_k) = \lim_{k\to\infty} \mu(V_k) = \mu(\overline{B}).\end{aligned}$$

Since $\mu(f = t) > 0$ for at most countably many t, and since $\mathbf{P}_n(f > t) \leq 1_{t \geq ||f||}$, it follows with dominated convergence

$$\begin{aligned}\mu[f] &= \int_0^\infty \mu(f > t)dt \leq \int_0^\infty \liminf_{n\to\infty} \mathbf{P}_n(f > t)dt = \liminf_{n\to\infty} \int_0^\infty \mathbf{P}_n(f > t)dt = \liminf_{n\to\infty} \mathbf{E}_n[f]\\ &\leq \limsup_{n\to\infty} \mathbf{P}_n[f] = \limsup_{n\to\infty} \int_0^\infty \mathbf{P}_n(f > t)dt = \int_0^\infty \limsup_{n\to\infty} \mathbf{P}_n(f > t)dt \leq \int_0^\infty \mu(f \geq t)\\ &= \mu[f].\end{aligned}$$

□

We now return to the case of a general metric space (E, r). To show the existence of accumulation points in the sense of weak convergence, it must be ensured that limit measures are again limit measures are again probability measures. In particular no mass is lost at the boundary crossing as in the case of vague convergence (see Lemma 9.12). Here, the concept of *tightness* is central.

Definition 9.14 (Tightness). *Let $\mathcal{K}$ be the system of all compact sets in E. A family $(\mathbf{P}_i)_{i\in I}$ in $\mathcal{P}(E)$ is* tight, *if*

$$\sup_{K\in\mathcal{K}} \inf_{i\in I} \mathbf{P}_i(K) = 1.$$

A family $(X_i)_{i\in I}$ of E-valued random variables is tight *if $((X_i)_*\mathbf{P})_{i\in I}$ is tight, i.e.*

$$\sup_{K\in\mathcal{K}} \inf_{i\in I} \mathbf{P}(X_i \in K) = 1.$$

Remark 9.15 (Equivalent formulations).

1. *The definition of the tightness of a family $(\mathbf{P}_i)_{i\in I}$ in $\mathcal{P}(E)$ is equivalent to the following condition: for all $\varepsilon > 0$ there exists $K \subseteq E$ compact with $\inf_{i\in I} \mathbf{P}_i(K) \geq 1 - \varepsilon$.*

2. *If $E = \mathbb{R}^d$, a family $(\mathbf{P}_i)_{i\in I}$ is tight if and only if*

$$\sup_{r>0} \inf_{i\in I} \mathbf{P}_i(B_r(0)) = 1,$$

where $B_r(0)$ is the sphere around 0 with radius r.

3. *In Lemma 2.9 we have shown that $\mathbf{P} \in \mathcal{P}(E)$ is tight if (E, r) is complete and is separable. It also follows that every finite family of probability measures on the Borel's σ-algebra of a Polish space is tight.*

4. *Further, a countable family $(\mathbf{P}_i)_{i=1,2,\dots}$ is of probability measures on a Polish space (E, r) is tight if and only if*

$$\sup_{K\in\mathcal{K}} \liminf_{i=1,2,\dots} \mathbf{P}_i(K) = 1.$$

Proof. '⇒': This is clear, since $\liminf_{i=1,2,\dots} \mathbf{P}_i(K) \geq \inf_{i=1,2,\dots} \mathbf{P}_i(K) = 1$.

'⇐': Let $\varepsilon > 0$ and K such that $\liminf_{i=1,2,\dots} \mathbf{P}_i(K) \geq 1 - \varepsilon/2$. Choose N such that $\inf_{i=N+1,N+2,\dots} \mathbf{P}_i(K) \geq 1 - \varepsilon$ and $K_1, \dots, K_N$ compact such that $\mathbf{P}_i(K_i) \geq 1 - \varepsilon$ for $i = 1, \dots, N$. Since $\widetilde{K} = K \cup K_1 \cup \dots \cup K_N$ is compact and $\inf_{i=1,2,\dots} \mathbf{P}_i(\widetilde{K}) \geq 1 - \varepsilon$ the tightness of $(\mathbf{P}_i)_{i=1,2,\dots}$ follows. □

Example 9.16 (Tight sets of probability measures).

1. *If E is compact, every family of probability measures on $\mathcal{B}(E)$ is tight.*

2. *A family $(X_i)_{i\in I}$ of real-valued random variables with*

$$\sup_{i\in I} \mathbf{P}[|X_i|] < \infty,$$

is tight. This is because

$$\inf_{r>0} \sup_{i\in I} \mathbf{P}(|X_i| \geq r) \leq \inf_{r>0} \sup_{i\in I} \frac{\mathbf{P}[|X_i|]}{r} = 0.$$

.

3. *The family $(\delta_n)_{n=1,2,\dots}$, where δ_n is the Dirac measure on n, is not tight.*

Lemma 9.17 (Vague convergence and tightness)**.** *Let $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$ and $\mu \in \mathcal{P}_{\leq 1}(\mathbb{R})$ with*

$$\mathbf{P}_n \xRightarrow[v]{n\to\infty} \mu.$$

Then

$$\mu(\mathbb{R}) = 1 \qquad \Longleftrightarrow \qquad (\mathbf{P}_n)_{n=1,2,\dots} \text{ is tight}$$

. In this case, $\mathbf{P}_n \xRightarrow{n\to\infty} \mu$.

Proof. For $r > 0$ choose a $g_r \in \mathcal{C}_c(\mathbb{R})$, $1_{B_r(0)} \leq g_r \leq 1_{B_{r+1}(0)}$. Then $(\mathbf{P}_n)_{n=1,2,\dots}$ is tight if and only if

$$\sup_{r>0} \liminf_{n\to\infty} \mathbf{P}_n[g_r] = 1.$$

'$\Rightarrow$': Since μ is continuous from below, we find

$$1 = \sup_{r>0} \mu(B_r(0)) \leq \sup_{r>0} \mu[g_r] = \sup_{r>0} \liminf_{n\to\infty} \mathbf{P}_n[g_r] \leq 1.$$

'$\Leftarrow$': Let $(\mathbf{P}_n)_{n=1,2,\dots}$ be be tight. Then, from Lemma 9.12,

$$1 \geq \mu(\mathbb{R}) = \sup_{r>0} \mu(B_r(0)) = \sup_{r>0} \mu[g_r] = \sup_{r>0} \liminf_{n\to\infty} \mathbf{P}_n[g_r] = 1.$$

It remains to show the weak convergence. Assuming that $(\mathbf{P}_n)_{n=1,2,\dots}$ is tight and $f \in \mathcal{C}_b(\mathbb{R})$. Then,

$$\begin{aligned} \limsup_{n\to\infty} |\mathbf{P}_n[f] - \mu[f]| &\leq \inf_{r>0} \limsup_{n\to\infty} \big(|\mathbf{P}_n[f - fg_r]| + |\mathbf{P}_n[fg_r]| - \mu[fg_r]| + |\mu[f - fg_r]|\big) \\ &\leq ||f|| \inf_{r>0} \limsup_{n\to\infty} \mathbf{P}_n(B_r(0)^c) + \inf_{r>0} \mu[B_r(0)^c] = 0, \end{aligned}$$

and $\mathbf{P}_n \xRightarrow{n\to\infty} \mu$ follows. □

Corollary 9.18 (Weak convergence and tightness)**.** *Let $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R})$. If $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$, then $(\mathbf{P}_n)_{n\in\mathbb{N}}$ is tight.*

Proof. Since weak convergence of $\mathbf{P}_1, \mathbf{P}_2, \dots$ to $\mathbf{P}$ implies vague convergence, for $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots$ the conditions of Lemma 9.17 and $\mathbf{P}(\mathbb{R}) = 1$ are satisfied. Therefore, $(\mathbf{P}_n)_{n\in\mathbb{N}}$ is tight. □

To determine the weak convergence of probability measures Theorem 9.6 is helpful. We now turn to the question whether a sequence of probability measures can have an accumulation point. This means that there is a subsequence that converges weakly to a probability measure.

Theorem 9.19 (Prohorov's theorem)**.** *Let (E, r) be complete and separable and $(\mathbf{P}_i)_{i\in I}$ a family in $\mathcal{P}(E)$. The following are equivalent:*

1. *The family $(\mathbf{P}_i)_{i\in I}$ is relatively compact with respect to the topology of weak convergence, i.e. every sequence in $(\mathbf{P}_i)_{i\in I}$ has a weakly convergent subsequence.*

2. For every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ and $x_1, \dots, x_N \in E$, so that

$$\inf_{i \in I} \mathbf{P}_i\Big(\bigcup_{k=1}^{N} B_\varepsilon(x_k) \Big) \geq 1 - \varepsilon.$$

3. The family $(\mathbf{P}_i)_{i \in I}$ is tight.

Proof. Let $x_1, x_2, \dots$ be a dense subsequence in E (which exists since (E, r) is separable).
1. $\Rightarrow$ 2.: Suppose 2. is not true. Then there is $\varepsilon > 0$ and for each $N = 1, 2, \dots$ a $\mathbf{P}_{i_N}$ with $\mathbf{P}_{i_N}\Big(\bigcup_{k=1}^{N} B_\varepsilon(x_k) \Big) \leq 1 - \varepsilon$. By relative compactness, there would then be some subsequence $(\mathbf{P}_{i_M})_{M=1,2,\dots}$ which is weakly convergent to some $\mathbf{P} \in \mathcal{P}(E)$. Thus, because of Theorem 9.6 ($(i) \Rightarrow (iii)$) we find that

$$1 = \mathbf{P}(E) = \sup_{N \in \mathbb{N}} \mathbf{P}\Big(\bigcup_{k=1}^{N} B_\varepsilon(x_i) \Big) \leq \sup_{N \in \mathbb{N}} \liminf_{M \to \infty} \mathbf{P}_{i_M}\Big(\bigcup_{k=1}^{N} B_\varepsilon(x_i) \Big) \leq 1 - \varepsilon,$$

thus a contradiction.
2. $\Rightarrow$ 3.: Let $\varepsilon > 0$. For $j = 1, 2, \dots$ we choose $x_{j1}, \dots, x_{jN_j}$ such that

$$\inf_{i \in I} \mathbf{P}_i\Big(\bigcup_{k=1}^{N_j} B_{\varepsilon 2^{-j}}(x_{jk}) \Big) > 1 - \varepsilon 2^{-j}.$$

We further set

$$K := \bigcap_{j=1}^{\infty} \bigcup_{k=1}^{N_j} B_{\varepsilon 2^{-j}}(x_{jk}).$$

Then $K \subseteq E$ is totally bounded by construction, according to Proposition A.9 therefore relatively compact, so $\overline{K}$ is compact. Furthermore

$$\sup_{i \in I} \mathbf{P}_i(\overline{K}^c) \leq \sup_{i \in I} \sum_{j=1}^{\infty} \mathbf{P}_i\Big(\bigcap_{k=1}^{N_j} (B_{\varepsilon 2^{-j}}(x_{jk}))^c \Big) \leq \varepsilon.$$

Thus the family $(\mathbf{P}_i)_{i \in I}$ is tight.
3. $\Rightarrow$ 1.: Let $\mathbf{P}_1, \mathbf{P}_2, \dots$ be a sequence in the family of the family $(\mathbf{P}_i)_{i \in I}$. The aim is to find a convergent subsequence. For this purpose, we choose compact sets $K_1 \subseteq K_2 \subseteq \cdots \subseteq E$ with $\inf_{n=1,2,\dots} \mathbf{P}_n(K_j) \geq 1 - 1/j$. Further, we choose the system of compact sets

$$\mathcal{K} := \Big\{ \bigcup_{k=1}^{N} K_{j_k} \cap \overline{B_{\varepsilon_k}(x_k)} : N, j_k \in \mathbb{N}, \varepsilon_k \in \mathbb{Q}^+ \Big\}.$$

Since $\mathcal{K}$ is countable, we can use a diagonal argument in order to create a subsequence $\mathbf{P}_{n_1}, \mathbf{P}_{n_2}, \dots$ from $\mathbf{P}_1, \mathbf{P}_2, \dots$ so that $\mathbf{P}_{n_k}(A)$ converges for all $A \in \mathcal{K}$. Define the set function μ on $\mathcal{K}$ by

$$\mu(A) = \lim_{k \to \infty} \mathbf{P}_{n_k}(A), \qquad A \in \mathcal{K}.$$

Our goal is to construct a probability measure $\mathbf{P}$, such that, for all open sets B,

$$\mathbf{P}(B) = \sup_{\mathcal{K} \ni A \subseteq B} \mu(A). \tag{9.4}$$

Indeed, if we find such a $\mathbf{P}$, we can write for B open

$$\mathbf{P}(B) = \sup_{\mathcal{K}\ni A\subseteq B} \lim_{k\to\infty} \mathbf{P}_{n_k}(A) \leq \liminf_{k\to\infty} \mathbf{P}_{n_k}(B),$$

and $\mathbf{P}_{n_k} \xrightarrow{k\to\infty} \mathbf{P}$ follows by Theorem 9.6. In order to find $\mathbf{P}$, we are going to construct an outer measure γ, and show that the open sets are γ-measurable. Then, $\mathbf{P}$ can be defined via γ on the σ-algebra of all measurbale sets; see Lemma 6.2.

We first extend μ to all open sets (giving rise to β below), and directly construct γ by setting

$$\gamma(C) := \inf_{B\supseteq C \text{ open}} \beta(B), \qquad \beta(B) =: \sup_{\mathcal{K}\ni K\subseteq B} \mu(K).$$

So, β is defined on all open sets, and, by construction, β is monotone, additive, sub-additive, and $\gamma = \beta$ on all open sets.
We claim that

$$\gamma \text{ is an outer measure and all closed sets are } \gamma\text{-measurable.} \tag{9.5}$$

(Recall that C is measurable with respect to the outer measure γ, if $\gamma(S) \geq \gamma(S\cap C)+\gamma(S\cap C^c)$ for all $S \subseteq E$; see Definition 2.1.6 and sub-additivity of γ). Then, we write for B open $\mathbf{P}(B) = \gamma(B) = \beta(B) = \sup_{K\ni A\subseteq B} \mu(A)$, i.e. (9.4) follows.
In order to show (9.5), we proceed in steps:
Step 1: *If $F \subseteq B \cap K$ is closed, with B open and $K \in \mathcal{K}$, then there is $K' \in \mathcal{K}$ with $F \subseteq K' \subseteq B$.*
For each $x \in F$, choose $\varepsilon(x) \in \mathbb{Q}$ such that $B_{\varepsilon(x)}(x) \subseteq B$. Since $(B_{\varepsilon(x)}(x))_{x\in F}$ is an open cover of $F \cap K$, which is compact, there must be a finite subcover, i.e. some $F = F \cap K \subseteq \bigcup_{n=1}^N \overline{B_{\varepsilon(x_n)}(x_n)} \cap K \subseteq B$. We can now read off the required K'.
Step 2: *β is σ-sub-additive on the open sets.*
For finite sub-additivity, let B_1, B_2 be open, and $\mathcal{K} \ni K \subseteq B_1 \cup B_2$. Define

$$F_1 := \{x \in K : r(x, B_1^c) \geq r(x, B_2^c\}, \quad F_2 := \{x \in K : r(x, B_2^c) \geq r(x, B_1^c)\}.$$

Note that $F_1 \subseteq B_1$: Indeed, if $x \in F_1 \subseteq K \subseteq B_1 \cup B_2$ and $x \in B_2 \setminus B_1$, then $0 = r(x, B_1^c) < r(x, B_2^c)$ since B_2^c is closed, which is a contradiction. Analogously, $F_2 \subseteq B_2$.
So, for $i = 1, 2$, we find $F_i \subseteq B_i \cap K$, and we find $K_i \in \mathcal{K}$ with $F_i \subseteq K_i \subseteq B_i$ with Step 1. So, note that $F_1 \cup F_2 = K$, and we can write

$$\mu(K) \leq \mu(K_1 \cup K_2) \leq \mu(K_1) + \mu(K_2) \leq \beta(B_1) + \beta(B_2).$$

Finite sub-additivity follows by taking the supremum over $\mathcal{K} \ni K \subseteq B_1 \cup B_2$ on the left hand side. For σ-sub-additivity, take $\mathcal{K} \ni K \subseteq \bigcup_{n=1}^\infty B_n$. Since K is compact, choose n_0 such that $K \subseteq \bigcup_{n=1}^{n_0} B_n$ and write

$$\mu(K) \leq \beta\Big(\bigcup_{n=1}^{n_0} B_n\Big) \leq \sum_{n=1}^{n_0} \beta(B_n) \leq \sum_{n=1}^{\infty} \beta(B_n).$$

Then, σ-sub-additivity by taking the supremum over $\mathcal{K} \ni K \subseteq \bigcup_{n=1}^\infty B_n$ on the left hand side.
Step 3: *γ is an outer measure.*
Since $\gamma(\emptyset) = 0$ and γ is monotone by construction, it remains to show σ-sub-additivity. If

$\S_1, S_2, ... \subseteq E$, let $\varepsilon > 0$ and choose $B_1 \subseteq S_1, B_2 \subseteq S_2, ...$ open with $\beta(B_n) < \gamma(S_n) + \varepsilon/2^n$. Then, using Step 2,

$$\gamma\Big(\bigcup_{n=1}^\infty S_n\Big) \leq \beta\Big(\bigcup_{n=1}^\infty S_n\Big) \leq \sum_{n=1}^\infty \beta(B_n) \leq \varepsilon + \sum_{n=1}^\infty \gamma(S_n).$$

The assertion follows by letting $\varepsilon \downarrow 0$.
Step 4: Closed sets are γ-measurable.
It suffices to show

$$\beta(B) \geq \gamma(F \cap B) + \gamma(F^c \cap B)$$

for F closed and B open. Once this is shown, consider an arbitrary S and $B \supseteq S$ open. Then, $\beta(B) \geq \gamma(F \cap B) + \gamma(F^c \cap B) \geq \gamma(F \cap S) + \gamma(F^c \cap S)$ by monotonicity og γ. From here, the assertion follows by taking $\inf_{B \subseteq S \text{ open}}$ on the left hand side.
So, let F be closed and B be open and $\varepsilon > 0$. Choose $K_1, K_2 \in \mathcal{K}$ with $K_1 \subseteq F^c \cap B$ and $K_2 \subseteq K_1^c \cap B$ (in particular, K_1, K_2 are disjoint) with $\mu(K_1) > \beta(F^c \cap B) - \varepsilon$ and $\mu(K_2) > \beta(K_1^c \cap B) - \varepsilon$. Then, since $\beta(K_1^c \cap B) \geq \gamma(F \cap B)$

$$\beta(B) \geq \mu(K_1 \cup K_2) = \mu(K_1) + \mu(K_2) > \gamma(F^c \cap B) + \gamma(K_1^c \cap B) - 2\varepsilon.$$

By letting $\varepsilon \to 0$, this concludes the proof, i.e. (iii)⇒(i) is shown. □

9.3 Separating classes of functions

Now we will introduce separating classes of functions. In particular, this will shed some light on the usefulness of characteristic functions and Laplace transforms of distributions (see Definition 6.11). These are based on two specific classes of functions that are separating.

Definition 9.20 (Classes of functions separating points and separating function classes)**.**

1. *A function class $\mathcal{M} \subseteq \mathcal{C}(E)$ is said to separate points in E if for all $x, y \in E$ with $x \neq y$ there exists an $f \in \mathcal{M}$ with $f(x) \neq f(y)$.*

2. *A class of functions $\mathcal{M} \subseteq \mathcal{C}(E)$ is called separating in $\mathcal{P}(E)$ if from $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(E)$ and*

$$\mathbf{P}[f] = \mathbf{Q}[f] \text{ for all } f \in \mathcal{M}$$

implies that $\mathbf{P} = \mathbf{Q}$.

Example 9.21. *1. The class of functions $\mathcal{M} := \mathcal{C}_b(E)$ is both, separating points and separating. Namely, if $x \neq y$, then $z \mapsto r(x,z) \wedge 1$ is a bounded, continuous function that separates x and y. Furthermore, if $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(E)$ and $\mathbf{P} \neq \mathbf{Q}$, then there is an open ball A with $\mathbf{P}(A) \neq \mathbf{Q}(A)$. Let $f_1, f_2, \ldots$ be a sequence in $\mathcal{C}_b(E)$ with $f_n \uparrow 1_A$. If $\mathbf{P}[f_n] = \mathbf{Q}[f_n]$ for all $n = 1, 2, \ldots$, then it would also*

$$\mathbf{P}(A) = \lim_{n\to\infty} \mathbf{P}[f_n] = \lim_{n\to\infty} \mathbf{Q}[f_n] = \mathbf{Q}(A)$$

in contradiction to the assumption.

2. *The class of functions $\{x \mapsto cx : c \in \mathbb{R}\}$ of all linear functions separates points, but is not separating.*

The next result requires the Stone-Weierstrass theorem, which we repeat first.

Definition 9.22 (Algebra). *A set system $\mathcal{M} \subseteq \mathcal{C}(E)$ is called an algebra, if $1 \in \mathcal{M}$, and if $\alpha, \beta \in \mathbb{R}$ and it contains f, g it also contains $\alpha f + \beta g$, as well as fg.*

Theorem 9.23 (Stone-Weierstrass). *Let (E, r) be compact and $\mathcal{M} \subseteq C_b(E)$ an algebra separating points. Then, $\mathcal{M}$ is dense in $\mathcal{C}_b(E)$ with respect to the supremum norm.*

Proof. See some lecture on *Analysis*. □

Theorem 9.24 (Algebras separating points and separating algebras).
Let (E, r) be complete and separable. If $\mathcal{M} \subseteq \mathcal{C}_b(E)$ separates points and is such that $f, g \in \mathcal{M}$ implies $fg \in \mathcal{M}$. Then $\mathcal{M}$ is is separating.

Proof. Let $\mathbf{P}, \mathbf{Q} \in \mathcal{P}(E)$. Without restriction, $1 \in \mathcal{M}$, since $\mathbf{P}[1] = \mathbf{Q}[1]$ always holds. Thus $\mathcal{M}$ is wlog an algebra. Let $\varepsilon > 0$ and K be compact such that $\mathbf{P}(K) > 1-\varepsilon$, $\mathbf{Q}(K) > 1-\varepsilon$. For $g \in \mathcal{C}_b(E)$, according to the Stone-Weierstrass Theorem 9.23 there is a sequence $(g_n)_{n=1,2,\ldots}$ in $\mathcal{M}$ with

$$\sup_{x \in K} |g_n(x) - g(x)| \xrightarrow{n\to\infty} 0. \tag{9.6}$$

Now,

$$\begin{aligned}
\big|\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{Q}[ge^{-\varepsilon g^2}]\big| \leq\ & \big|\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{P}[ge^{-\varepsilon g^2}; K]\big| \\
& + \big|\mathbf{P}[ge^{-\varepsilon g^2}; K] - \mathbf{P}[g_n e^{-\varepsilon g_n^2}; K]\big| \\
& + \big|\mathbf{P}[g_n e^{-\varepsilon g_n^2}; K] - \mathbf{P}[g_n e^{-\varepsilon g_n^2}]\big| \\
& + \big|\mathbf{P}[g_n e^{-\varepsilon g_n^2}] - \mathbf{Q}[g_n e^{-\varepsilon g_n^2}]\big| \\
& + \big|\mathbf{Q}[g_n e^{-\varepsilon g_n^2}] - \mathbf{Q}[g_n e^{-\varepsilon g_n^2}; K]\big| \\
& + \big|\mathbf{Q}[g_n e^{-\varepsilon g_n^2}] - \mathbf{Q}[ge^{-\varepsilon g^2}; K]\big| \\
& + \big|\mathbf{Q}[ge^{-\varepsilon g^2}; K] - \mathbf{Q}[ge^{-\varepsilon g^2}]\big|
\end{aligned}$$

We restrict the first term by

$$\big|\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{P}[ge^{-\varepsilon g^2}; K]\big| \leq \frac{C}{\sqrt{\varepsilon}} \mathbf{P}(K^c) \leq C\sqrt{\varepsilon}$$

with $C = \sup_{x\geq 0} xe^{-x^2}$; analogous to the third, fifth and last terms. The second and penultimate terms converge to 0 for $n \to \infty$ due to (9.6). Since $\mathcal{M}$ is an algebra, $g_n e^{-\varepsilon g_n^2}$ can be approximated by functions in $\mathcal{M}$, which means that the fourth term for $n \to \infty$ converges to 0. This means that

$$\big|\mathbf{P}[g] - \mathbf{Q}[g]\big| = \lim_{\varepsilon\to 0} \big|\mathbf{P}[ge^{-\varepsilon g^2}] - \mathbf{Q}[ge^{-\varepsilon g^2}]\big| \leq 4C \lim_{\varepsilon \to 0} \sqrt{\varepsilon} = 0.$$

Since g was arbitrary and $\mathcal{C}_b(E)$ is separating, $\mathbf{P} = \mathbf{Q}$ follows. □

We now come back to the characteristic function and the Laplace transform. As already mentioned, the usefulness of the characteristic function and the Laplace transforms is due to the fact that they are distribution-determining.

Proposition 9.25 (Characteristic function distribution-determining). *A probability measure $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ ($\mathbf{P} \in \mathcal{P}(\mathbb{R}_+^d)$) is uniquely characterized by the characteristic function $\psi_{\mathbf{P}}$ (the Laplace transform $\mathscr{L}_{\mathbf{P}}$).*

Proof. We show the statement only for characteristic functions that are proven for Laplace transforms is proven analogously. We establish that the set $\mathcal{M} := \{x \mapsto e^{itx}; t \in \mathbb{R}^d\}$ in $\mathbb{R}^d$ separates points. Since $\mathcal{M} \subseteq \mathcal{C}_b(\mathbb{R}^d)$ and is closed under product formation, it is also separating according to theorem 9.24. This finishes the proof. □

Corollary 9.26 (Independence and characteristic function).

1. *A family $(X_j)_{j\in I}$ of real-valued random variables is is independent if and only if for all $J \subseteq_f I$*

$$\mathbf{E}\Big[\prod_{j\in J} e^{it_jX_j}\Big] = \prod_{j\in J} \mathbf{E}[e^{it_jX_j}] \tag{9.7}$$

   for all $(t_j)_{j\in J} \in \mathbb{R}^J$ is valid.

2. *A family $(X_j)_{j\in I}$ of random variables with values in $\mathbb{R}_+$ is independent if and only if for all $J \subseteq_f I$*

$$\mathbf{E}\Big[\prod_{j\in I} e^{-t_jX_j}\Big] = \prod_{j\in J} \mathbf{E}[e^{-t_jX_j}]$$

   for all $(t_j)_{j\in J} \in \mathbb{R}^J$ applies.

Proof. We only show the first statement, the second follows analogously. If $(X_j)_{j\in I}$ is independent, then according to Lemma 8.4, the random variables $(e^{it_jX_j})_{j\in I}$ for all $(t_j)_{j\in J} \in \mathbb{R}^J$ are independent. Thus, (9.7) follows from Proposition 8.5. Conversely, the following applies. On the one hand, the left-hand side of (9.7) represents the characteristic function of the distribution $((X_j)_{j\in J})_*\mathbf{P}$. On the other hand, the right side of (9.7) is the characteristic function of $\bigotimes_{j\in J}(X_j)_*\mathbf{P}$. Since the characteristic function according to Proposition 9.25 is the joint distribution of $(X_j)_{j\in J}$ is uniquely determined, $((X_j)_{j\in J})_*\mathbf{P} = \bigotimes_{j\in J}(X_j)_*\mathbf{P}$. The independence of $(X_j)_{j\in I}$ thus follows from Proposition 8.2. □

9.4 Lévy's theorem

We now want to analyze the relationship between weak convergence and the convergence of the characteristic functions of the underlying distributions. Let $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R}^d)$. How to get from Proposition 9.27, the weak convergence follows $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$ follows from the pointwise convergence of the characteristic functions, $\psi_{\mathbf{P}_n}(t) \xrightarrow{n\to\infty} \psi_{\mathbf{P}}(t)$, $t \in \mathbb{R}^d$, given $(\mathbf{P}_n)_{n\in\mathbb{N}}$ is tight. The decisive factor is that the tightness of the family $(\mathbf{P}_n)_{n\in\mathbb{N}}$ can also be read from the characteristic functions as we will show in Proposition 9.32. This leads to the statement of Lévy's continuity theorem (Theorem 9.33), which states when the pointwise limit of characteristic functions is again a characteristic function of a probability measure.

Proposition 9.27 (Separating class of functions and weak convergence). *Let (E, r) be complete and separable and $\mathbf{P}, \mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(E)$. Then the following are equivalent:*

1. $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$.

2. $(\mathbf{P}_n)_{n=1,2,\dots}$ *is tight and there is a separating family* $\mathcal{M} \subseteq \mathcal{C}_b(E)$ *with*

$$\mathbf{P}_n[f] \xrightarrow{n\to\infty} \mathbf{P}[f] \text{ for all } f \in \mathcal{M}.$$

Proof. 1. $\Rightarrow$ 2. According to Corollary 9.18, we have that $(\mathbf{P}_n)_{n=1,2,\dots}$ is tight. The second part of 2. holds because of the definition of weak convergence.
2. $\Rightarrow$ 1. Suppose $(\mathbf{P}_n)_{n=1,2,\dots}$ is tight and $\mathbf{P}_1, \mathbf{P}_2, \dots$ does not converge weakly to $\mathbf{P}$. Then there is $\varepsilon > 0$, some $f \in \mathcal{C}_b(E)$ and a subsequence $(n_k)_{k=1,2,\dots}$ such that

$$\mathbf{P}_{n_k}[f] - \mathbf{P}[f]| > \varepsilon \text{ for all } k. \tag{9.8}$$

According to theorem 9.19 there is a subsequence $(n_{k_\ell})_{\ell=1,2,\dots}$ and a $\mathbf{Q} \in \mathcal{P}(E)$, such that $\mathbf{P}_{n_{k_\ell}} \xRightarrow{\ell\to\infty} \mathbf{Q}$. Because of (9.8),

$$|\mathbf{P}[f] - \mathbf{Q}[f]| \geq |\liminf_{\ell\to\infty} \big(\mathbf{P}[f] - \mathbf{P}_{n_{k_\ell}}[f]|\big) + \liminf_{\ell\to\infty}(\mathbf{P}_{n_{k_\ell}}[f] - \mathbf{Q}[f])| > \varepsilon,$$

in particular $\mathbf{P} \neq \mathbf{Q}$. On the other hand, for all $g \in \mathcal{M}$ we have

$$\mathbf{P}[g] = \lim_{\ell\to\infty} \mathbf{P}_{n_{k_\ell}}[g] = \mathbf{Q}[g].$$

Since $\mathcal{M}$ is separating, this is a contradiction and 1. is shown. □

Let $\mathbf{P} \in \mathcal{P}(\mathbb{R})$ and $\psi_{\mathbf{P}}$ be its characteristic function. We first show an estimate, which is important to relate tightness and $\psi_{\mathbf{P}}$.

Lemma 9.28 (Tightness and the characteristic function function)**.** *Let* $\mathbf{P} \in \mathcal{P}(\mathbb{R})$*. Then for all* $r > 0$

$$\mathbf{P}((-\infty; -r] \cup [r; \infty)) \leq \frac{r}{2} \int_{-2/r}^{2/r} (1 - \psi_{\mathbf{P}}(t))dt, \tag{9.9}$$

Proof. It is $\sin(x)/x \leq 1$ for $x \leq 2$ and $\sin x \leq x/2$ for $x \geq 2$. Let X be a random variable with distribution $\mathbf{P}$. Therefore, for every $c > 0$ according to Fubini,

$$\begin{aligned}
\int_{-c}^{c} (1 - \psi_{\mathbf{P}}(t))dt = \mathbf{P}\Big[\int_{-c}^{c} (1 - e^{itX})dt\Big] &= \mathbf{P}\Big[2c - \frac{1}{iX} e^{itX}\Big|_{t=-c}^{c}\Big] \\
&= 2c\mathbf{P}\Big[1 - \frac{\sin(cX)}{cX}\Big] \\
&\geq 2c\mathbf{P}\Big[1 - \frac{\sin(cX)}{cX}; |cX| \geq 2\Big] \\
&\geq c \cdot \mathbf{P}(|cX| \geq 2) = c\mathbf{P}((-\infty; -\tfrac{2}{c}] \cup [\tfrac{2}{c}; \infty)),
\end{aligned}$$

and the assertion follows with $c = 2/r$. □

Definition 9.29 (Uniform continuity)**.** *We repeat a definition from calculus. A set* $\mathcal{M} \subseteq \mathcal{C}(\mathbb{R}^d)$ *is called uniformly continuous in* $x \in \mathbb{R}^d$ *if*

$$\sup_{f\in\mathcal{M}} |f(y) - f(x)| \xrightarrow{y\to x} 0.$$

Remark 9.30 (Equivalent condition for sequences). *If $\mathcal{M} = \{f_1, f_2, \dots\}$, then the condition*

$$\limsup_{n\to\infty} |f_n(y) - f_n(x)| \xrightarrow{y\to x} 0$$

equivalent.

Lemma 9.31 (Uniform integrability and convergence). *Let $f_1, f_2, \dots \in \mathcal{C}(\mathbb{R}^d)$, so that $f_n \xrightarrow{n\to\infty} f$ pointwise for a function $f : \mathbb{R}^d \to \mathbb{R}$. Then f is continuous in 0 iff $(f_n)_{n=1,2,\dots}$ is uniformly continuous in 0.*

Proof. If $(f_n)_{n=1,2,\dots}$ is uniformly continuous in 0, then

$$|f(t) - f(0)| = |\lim_{n\to\infty}(f_n(t) - f_n(0))| \le \limsup_{n\to\infty} |f_n(t) - f_n(0)| \xrightarrow{t\to 0} 0.$$

Conversely, if f is continuous in 0, then

$$\limsup_{n\to\infty} |f_n(t)-f_n(0)| \le \limsup_{n\to\infty} |f_n(t)-f(t)|+|f(t)-f(0)|+|f(0)-f_n(0) = |f(t)-f(0)| \xrightarrow{t\to 0} 0.$$

□

Proposition 9.32 (Tightness and uniformity continuity). *Let $(\mathbf{P}_i)_{i\in I}$ be a family in $\mathcal{P}(\mathbb{R}^d)$. If $(\psi_{\mathbf{P}_i})_{i\in I}$ is uniformly continuous in 0, then $(\mathbf{P}_i)_{i\in I}$ is tight.*

Proof. It suffices to show that $((\pi_k)_*\mathbf{P}_i)_{i\in I}$ is tight for all projections $\pi_1, \dots, \pi_d$. Apparently, $\psi_{(\pi_k)_*\mathbf{P}_i}(t) = \psi_{\mathbf{P}_i}(te_k)$, if e_k is the k-th unit vector. It is therefore sufficient to prove the assertion in the case $d = 1$. Since $\psi_{\mathbf{P}_i}(0) = 1$ for all $i \in I$, we conclude from uniform continuity that

$$\sup_{i\in I} |1 - \psi_{\mathbf{P}_i}(t)| \xrightarrow{t\to 0} 0,$$

thus, see Remark 9.15,

$$\begin{aligned}
\sup_{r>0} \inf_{i\in I} \mathbf{P}_i([-r;r]) &\ge 1 - \inf_{r>0} \sup_{i\in I} \frac{r}{2} \int_{-2/r}^{2/r} (1 - \psi_{\mathbf{P}_i}(t))dt \\
&\ge 1 - \inf_{r>0} \frac{r}{2} \int_{-2/r}^{2/r} \sup_{i\in I} |1 - \psi_{\mathbf{P}_i}(t)|dt \\
&\ge 1 - 2 \inf_{r>0} \sup_{t\in[0;2/r]} \sup_{i\in I} |1 - \psi_{\mathbf{P}_i}(t)| = 1.
\end{aligned}$$

This shows the assertion. □

Theorem 9.33 (Lévy's continuity theorem). *Let $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R}^d)$ and $\psi : \mathbb{R}^d \to \mathbb{C}$, so that $\psi_{\mathbf{P}_n}(t) \xrightarrow{n\to\infty} \psi(t)$ for all $t \in \mathbb{R}^d$. If ψ is continuous in 0, then $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$ for a $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ with $\psi_{\mathbf{P}} = \psi$.*

Proof. Since $\psi_{\mathbf{P}_n}$ converges pointwise to a function ψ which is continuous in 0, it follows from Lemma 9.31 that $(\psi_{\mathbf{P}_n})_{n=1,2,\dots}$ is uniformly continuous in 0. With Proposition 9.32 it follows that $(\mathbf{P}_n)_{n=1,2,\dots}$ is tight. Let $(n_k)_{k=1,2,\dots}$ be a subsequence and $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ such that $\mathbf{P}_{n_k} \xRightarrow{k\to\infty} \mathbf{P}$. Since $x \mapsto e^{itx}$ is a continuous, bounded function, it follows that $\psi_{\mathbf{P}_{n_k}}(t) \xrightarrow{k\to\infty} \psi_{\mathbf{P}}(t)$ for all $t \in \mathbb{R}^d$. On the other hand, since $\psi_{\mathbf{P}_n}(t) \xrightarrow{n\to\infty} \psi(t)$, and $\psi_{\mathbf{P}} = \psi$ follows. This identifies ψ as a characteristic function of $\mathbf{P}$ and since this uniquely determines $\mathbf{P}$, we find $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$. □

Example 9.34 (Theorem of deMoivre-Laplace)**.** *Let $S_n \sim B(n,p)$. The Theorem of deMoivre-Laplace states that*

$$S_n^* := \frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{n\to\infty} N(0,1). \tag{9.10}$$

We now want to show this again with the help of characteristic functions, i.e. $\psi_{S_n^*} \xrightarrow{n\to\infty} \psi_{N(0,1)}$ pointwise. To do this, we use Proposition 6.12.3 and write with $q := 1-p$ and $C_1, C_2, \dots \in \mathbb{C}$ with $\limsup_{n\to\infty} |C_n| < \infty$

$$\begin{aligned}
\psi_{S_n^*}(t) &= \exp\Big(-it\sqrt{\frac{np}{q}}\Big) \cdot \psi_{B(n,p)}\Big(\frac{t}{\sqrt{npq}}\Big) \\
&= \exp\Big(-it\sqrt{\frac{np}{q}}\Big)\Big(q + p\exp\Big(\frac{it}{\sqrt{npq}}\Big)\Big)^n \\
&= \Big(q\exp\Big(-it\sqrt{\frac{p}{nq}}\Big) + p\exp\Big(it\sqrt{\frac{q}{np}}\Big)\Big)^n \\
&= \Big(1 - qit\sqrt{\frac{p}{nq}} - q\frac{t^2}{2}\frac{p}{nq} + pit\sqrt{\frac{q}{np}} - p\frac{t^2}{2}\frac{q}{np} + \frac{C_n}{n^{3/2}}\Big)^n \\
&= \Big(1 - \frac{t^2}{2}\frac{1}{n} + \frac{C_n}{n^{3/2}}\Big)^n \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}} = \psi_{N(0,1)}(t).
\end{aligned}$$

The result now follows from Theorem 9.33.

Lévy's continuity theorem can also be formulated with Laplace transforms. We state the theorem without proof:

Theorem 9.35 (Lévy's continuity theorem for Laplace transforms)**.** *Let $\mathbf{P}_1, \mathbf{P}_2, \dots \in \mathcal{P}(\mathbb{R}_+^d)$ and $\mathscr{L} : \mathbb{R}^d \to [0,1]$, so that $\mathscr{L}_{\mathbf{P}_n}(t) \xrightarrow{n\to\infty} \mathscr{L}(t)$ for all $t \in \mathbb{R}^d$. If $\mathscr{L}$ is continuous in 0, then $\mathbf{P}_n \xRightarrow{n\to\infty} \mathbf{P}$ for a $\mathbf{P} \in \mathcal{P}(\mathbb{R}^d)$ with $\mathscr{L}_{\mathbf{P}} = \mathscr{L}$.*

Example 9.36 (Convergence of the geometric to the exponential distribution)**.** *Let $X_n \sim \mu_{geo(p_n)}$ be distributed and $n \cdot p_n \xrightarrow{n\to\infty} \lambda$. Then*

$$\begin{aligned}
\mathscr{L}_{X_n/n}(t) = \mathbf{P}[e^{-tX_n/n}] &= \sum_{k=1}^{\infty} (1-p_n)^{k-1} p_n e^{-tk/n} \\
&= p_n e^{-t/n} \frac{1}{1-(1-p_n)e^{-t/n}} \\
&= \frac{\lambda}{n(1-(1-\lambda/n)(1-t/n))} + o(1/n) \\
&\xrightarrow{n\to\infty} \frac{\lambda}{\lambda + t}.
\end{aligned}$$

Therefore, $\frac{X_n}{n} \xRightarrow{n\to\infty} Y$, where $Y \sim \mu_{exp(\lambda)}$, since

$$\mathscr{L}_{exp(\lambda)}(t) = \int_0^\infty \lambda e^{-\lambda a} e^{-ta} da = \frac{\lambda}{\lambda + t}.$$

10 Weak limit laws

We will now apply our knowledge of weak convergence and characteristic functions in special situations. In Section 10.1 we are concerned with statements about when the sum of random variables converges against a Poisson distributed random variable. In section 10.2 we will apply the central Lindeberg-Feller's central limit theorem, which provides a characterization for the weak convergence against a normal distribution. Section 10.3 finally deals with extensions for the case of multidimensional random variables.

10.1 Poisson convergence

We already know the statement that $B(n, p_n)$ for $n \cdot p_n \xrightarrow{n\to\infty} \lambda$ converges weakly against $\text{Poi}(\lambda)$ for large n; see Example 10.1. In this section we generalize this statement; see Theorem 10.5.

Example 10.1 (Poisson approximation of the binomial distribution)**.** *Let* $p_1, p_2, \dots \in [0,1]$ *be such that* $n \cdot p_n \xrightarrow{n\to\infty} \lambda$. *Then we already know from* Basic probability *that*

$$B(n, p_n)(\{k\}) \xrightarrow{n\to\infty} Poi(\lambda)(\{k\}).$$

In other words, this is a statement about weak convergence:

$$B(n, p_n) \xRightarrow{n\to\infty} Poi(\lambda). \tag{10.1}$$

Lévy's theorem provides another way to prove this result. We recall the characteristic functions of the binomial and Poisson distribution from Example 6.13. We write directly

$$\begin{aligned}
\psi_{B(n,p_n)}(t) &= \Big(1 - p_n\big(1 - e^{it}\big)\Big)^n \\
&= \Big(1 - \frac{n \cdot p_n}{n}\big(1 - e^{it}\big)\Big)^n \\
&\xrightarrow{n\to\infty} \exp\big(-\lambda(1 - e^{it})\big) = \psi_{Poi(\lambda)}(t).
\end{aligned}$$

In particular, the characteristic functions of the binomial distributions converge pointwise to a function that is continuous in 0, namely the characteristic function of the Poisson distribution. With Theorem 9.33 this imples (10.1)*.*

In the following, we will see that the weak convergence to a Poisson distribution is even more general. For this we will use use generating functions.

Remark 10.2 (Generating function)**.** *Consider a random variable* X *with values in* $\mathbb{Z}_+$ *and define the* generating function

$$z \mapsto \varphi_X(z) := \mathbf{P}[z^X] = \sum_{k=0}^{\infty} z^k \mathbf{P}[X = k].$$

We note that for $z \in [0,1]$ *this is related to the Laplace transform of* X *because (with* $z = e^{-t}$*)*

$$\mathscr{L}_X(t) = \mathbf{P}[e^{-tX}] = \mathbf{P}[z^X] = \varphi_X(z).$$

In particular, the following two properties of Laplace transforms carry ofer to generating functions.

1. Generating functions determine the distribution, see Proposition 9.25: *The distribution of X is uniquely determined by $z \mapsto \varphi_X(z)$ for $z \in [0,1]$.*

2. Weak convergence equivalent to the convergence of the generating functions, see Theorem 9.33: *Let $X_1, X_2, \dots$ be a sequence of random variables with values in $\mathbb{Z}_+$ such that $\varphi_{X_n}(z) \xrightarrow{n\to\infty} \varphi(z)$ for $z \in [0,1]$ for a function φ that is continuous from below in 1. Then $X_n \xRightarrow{n\to\infty} X$ for a random variable X with generating function φ.*

Sometimes generating functions are practical tools. By their definition, they are power series with radius of convergence $r \geq 1$. It is known that inside the radius of convergence, the derivative and sum are interchanged. So if $r > 1$, for example, we write

$$\varphi_X'(1) = \sum_{k=0}^{\infty} kz^{k-1}\mathbf{P}(X=k)\Big|_{z=1} = \sum_{k=0}^{\infty} k\mathbf{P}(X=k) = \mathbf{P}[X].$$

Analogous calculations for higher derivatives are also possible.

Definition 10.3 (Asymptotic negligibility)**.** *A triagonal family of random variables $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ with $m_1, m_2, \dots \in \mathbb{N}$ is* asymptotically negligible*, if the random variables $X_{n1}, \dots, X_{n,m_n}$ are independent for each $n = 1, 2, \dots,$ and*

$$\sup_{j=1,\dots,m_n} \mathbf{P}(|X_{nj}| > \varepsilon) \xrightarrow{n\to\infty} 0 \tag{10.2}$$

for all $\varepsilon > 0$. If $X_{ij} \geq 0$ for all i, j, then $m_n = \infty$ is also permitted.

Remark 10.4 (Equivalent formulation)**.** 1. *For a triagonal family of random variables $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$,* (10.2) *holds iff*

$$\sup_{j=1,\dots,m_n} \mathbf{E}[|X_{nj}| \wedge 1] \xrightarrow{n\to\infty} 0.$$

2. *Let $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ be a triangular of $\mathbb{Z}_+$-valued random variables. Then* (10.2) *holds iff*

$$\inf_{z\in[0,1]} \inf_{j=1,\dots,m_n} \varphi_{X_{nj}}(z) = \inf_{j=1,\dots,m_n} \varphi_{X_{nj}}(0) = \inf_{j=1,\dots,m_n} \mathbf{P}(|X_{nj}| = 0) \xrightarrow{n\to\infty} 1. \tag{10.3}$$

Theorem 10.5 (Poisson convergence)**.** *Let $(X_{nj})_{n=1,2,\dots,n,j=1,\dots,m_n}$ be a family of asymptotically negligible random variables with values in $\mathbb{Z}_+$ and $X \sim Poi(\lambda)$. Then,*

$$\sum_{j=1}^{m_n} X_{nj} \xRightarrow{n\to\infty} X$$

iff

1. $\displaystyle\sum_{j=1}^{m_n} \mathbf{P}(X_{nj} > 1) \xrightarrow{n\to\infty} 0$

2. $\displaystyle\sum_{j=1}^{m_n} \mathbf{P}(X_{nj} = 1) \xrightarrow{n\to\infty} \lambda.$

We prepare the proof with a lemma.

Lemma 10.6. *Let $(\lambda_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ be a triangular family of asymptotically negligible, non-negative constants and $\lambda \in [0;\infty]$. Then,*

$$\prod_{j=1}^{m_n}(1-\lambda_{nj}) \xrightarrow{n\to\infty} e^{-\lambda} \qquad \Longleftrightarrow \qquad \sum_{j=1}^{m_n}\lambda_{nj} \xrightarrow{n\to\infty} \lambda$$

Proof. First note that $\log(1-x) = -x + \varepsilon(x)$ for $x > 0$ with $\varepsilon(x)/x \xrightarrow{x\to 0} 0$. Since $\sup_{j=1,\dots,m_n}\lambda_{nj} < 1$ for large n, the left hand side is equivalent to

$$-\lambda = \lim_{n\to\infty}\sum_{j=1}^{m_n}\log(1-\lambda_{nj}) = -\lim_{n\to\infty}\sum_{j=1}^{m_n}\lambda_{nj}\big(1-\tfrac{\varepsilon(\lambda_{nj})}{\lambda_{nj}}\big) = -\lim_{n\to\infty}\sum_{j=1}^{m_n}\lambda_{nj},$$

as

$$\sup_{j=1,\dots,m_n}\tfrac{\varepsilon(\lambda_{nj})}{\lambda_{nj}} \xrightarrow{n\to\infty} 0.$$

From this, the right hand side is immediate. □

Proof of Theorem 10.5. We denote by $\varphi_{n,j}$ the generating function of $X_{n,j}$. According to Remark 10.2.2, the weak convergence in the theorem is equivalent to pointwise convergence of $\prod_{j=1}^{m_n}\varphi_{nj}(z) \xrightarrow{n\to\infty} e^{-\lambda(1-z)}$, since

$$\varphi_X(z) = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}z^k = e^{-\lambda(1-z)}.$$

By Lemma 10.6 this is true iff

$$A_n(z) := \sum_{j=1}^{m_n}(1-\varphi_{nj}(z)) \xrightarrow{n\to\infty} \lambda(1-z), \tag{10.4}$$

since the family $(1-\varphi_{nj}(z))_{n=1,2,\dots,j=1,\dots,m_n}$ for each $z \in [0,1]$ after (10.3) is asymptotically negligible. We decompose $A_n(z) = A_n^1(z) + A_n^2(z)$ with

$$A_n^1(z) = \sum_{k=1}^{\infty}(1-z)\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}=k) = (1-z)\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}>0),$$
$$A_n^2(z) = \sum_{k=2}^{\infty}(z-z^k)\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}=k).$$

First, $z(1-z) \le z - z^k \le z$ for all $k = 2, 3, \dots$ This means that

$$z(1-z)\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}>1) \le A_n^2(z) \le z\sum_{j=1}^{m_n}\mathbf{P}(X_{nj}>1). \tag{10.5}$$

Let us now turn to the proof of the assertion.

'⇒': Let (10.4) hold. For $z = 0$ this means, since $\varphi_{nj}(0) = \mathbf{P}(X_{nj} = 0)$, that

$$\sum_{j=1}^{m_n} \mathbf{P}(X_{nj} > 0) = \sum_{j=1}^{m_n} (1 - \varphi_{nj}(0)) \xrightarrow{n\to\infty} \lambda.$$

Therefore, $A_n^1(z) \xrightarrow{n\to\infty} \lambda(1-z)$ for $z \in [0,1]$. But then $A_n^2(z) \xrightarrow{n\to\infty} 0$ must apply to $z \in [0,1]$. Because of (10.5) this means that 1. is valid. The statement 2. follows from this by subtraction.
'⇐': So 1. and 2. apply. It is clear that $A_n^2(z) \xrightarrow{n\to\infty} 0$ by (10.5). Then $A_n^1(z) \xrightarrow{n\to\infty} (1-z)\lambda$ by 2., i.e. (10.4) is shown. □

Example 10.7 (Convergence of geometric distributions against Poisson). *Let X_{nj}, $j = 1, \dots, n, n = 1, 2, \dots$ be geometrically distributed with parameter p_n (i.e. $\mathbf{P}(X_{nj} = k) = (1-p_n)^{k-1} p_n$, see Example 2.2.4) and $Y_{nj} = X_{nj} - 1$. (Thus, Y_{nj} is the number of* failures *before the first* success*). We set $Y_n := \sum_{j=1}^n Y_{nj}$, which is as distributed as the number of failures before the nth success. If $Y \sim Poi(\lambda)$ and $(1-p_n) \cdot n \xrightarrow{n\to\infty} \lambda$, then $Y_n \xrightarrow{n\to\infty} Y$. Since*

$$\sum_{j=1}^{n} \mathbf{P}(Y_{nj} = 1) = n(1-p_n)p_n \xrightarrow{n\to\infty} \lambda,$$
$$\sum_{j=1}^{n} \mathbf{P}(Y_{nj} > 1) = n(1-p_n)^2 \xrightarrow{n\to\infty} 0,$$

Theorem 10.5 gives the result.

10.2 The Central Limit Theorem

The central limit theorem, Theorem 10.8, generalizes the Theorem of deMoivre Laplace. The generalization consists of the fact that any sums of independent (not necessarily identically distributed) random variables converge weakly to a normally distributed random variable if they satisfy the *Lindeberg condition* (see 2. in Theorem 10.8).

Theorem 10.8 (Central limit theorem of Lindeberg-Feller). *Let $(X_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ be a family of random variables such that for $n = 1, 2, \dots$ the random variables $X_{n1}, \dots, X_{nm_n}$ are independent. Assume that*

$$\sum_{j=1}^{m_n} \mathbf{E}[X_{nj}] \xrightarrow{n\to\infty} \mu, \qquad \sum_{j=1}^{m_n} \mathbf{V}[X_{nj}] \xrightarrow{n\to\infty} \sigma^2$$

and $X \sim N(\mu, \sigma^2)$. Then the following statements are equivalent:

1. $\displaystyle\sum_{j=1}^{m_n} X_{nj} \xRightarrow{n\to\infty} X$ *and* $\displaystyle\sup_{j=1,\dots,m_n} \mathbf{V}[X_{nj}] \xrightarrow{n\to\infty} 0,$

2. $\displaystyle\sum_{j=1}^{m_n} \mathbf{E}[(X_{nj} - \mathbf{E}[X_{nj}])^2; |X_{nj} - \mathbf{E}[X_{nj}]| > \varepsilon] \xrightarrow{n\to\infty} 0$ *for all* $\varepsilon > 0$.

Before we prove the central limit theorem, we refer to the special case of identically distributed random variables, which was already discussed in the lecture *Basic Probability*.

Corollary 10.9 (Central limit theorem for identically distributed random variables)**.** *Let $X_1, X_2, \ldots$ be independent and identically distributed with $\mathbf{E}[X_1] = \mu, \mathbf{V}[X_1] = \sigma^2 > 0$. Let $S_n := \sum_{k=1}^n X_k$ and $X \sim N(0,1)$. Then,*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xRightarrow{n\to\infty} X.$$

Proof. Let $m_n = n$ and $X_{nj} = \frac{X_j - \mu}{\sqrt{n\sigma^2}}$. Then the family $(X_{nj})_{n=1,2,\ldots,j=1,\ldots,n}$ fulfills the conditions of Theorem 10.8 with $\mu = 0, \sigma^2 = 1$. Furthermore

$$\sum_{j=1}^n \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] = \frac{1}{\sigma^2}\mathbf{E}[(X_1 - \mu)^2; |X_1 - \mu| > \varepsilon\sqrt{n\sigma^2}] \xrightarrow{n\to\infty} 0$$

due to dominated convergence. □

The Lindeberg condition is often not easy to verify. The stronger Lyapunoff condition is often simpler.

Remark 10.10 (Lyapunoff condition)**.** *The family $(X_{nj})_{n=1,2,\ldots,j=1,\ldots,m_n}$ from Theorem 10.8 satisfies the Lyapunoff condition if for some $\delta > 0$*

$$\sum_{j=1}^{m_n} \mathbf{E}\big[|X_{nj} - \mathbf{E}[X_{nj}]|^{2+\delta}\big] \xrightarrow{n\to\infty} 0.$$

Under the conditions of Theorem 10.8, the Lyapunoff condition implies the Lindeberg condition. To see this, let wlog $\mathbf{E}[X_{nj}] = 0$. For all $\varepsilon > 0$,

$$x^2 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta} 1_{|x|>\varepsilon} \leq \frac{|x|^{2+\delta}}{\varepsilon^\delta}.$$

If the Lyapunoff condition applies, the Lindeberg condition follows from

$$\sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \leq \frac{1}{\varepsilon^\delta} \sum_{j=1}^{m_n} \mathbf{E}[|X_{nj}|^{2+\delta}] \xrightarrow{n\to\infty} 0.$$

The proof of Theorem 10.8 is based on the clever use of the characteristic functions of the random variable random variable X_{nj} and Taylor approximations. We prepare the proof of the theorem with two lemmas.

Lemma 10.11 (An estimate)**.** *For complex numbers $z_1, \ldots, z_n, z_1', \ldots, z_n'$ with $|z_i| \leq 1, |z_i'| \leq 1$ for $i = 1, \ldots, n$,*

$$\Big|\prod_{k=1}^n z_k - \prod_{k=1}^n z_k'\Big| \leq \sum_{k=1}^n |z_k - z_k'|. \tag{10.6}$$

Proof. For $n = 1$ the equation is obviously correct. Moreover, if (10.6) is valid for an n, then

$$\Big|\prod_{k=1}^{n+1} z_k - \prod_{k=1}^{n+1} z'_k\Big| \leq \Big|z_{n+1}\Big(\prod_{k=1}^{n} z_k - \prod_{k=1}^{n} z'_k\Big)\Big| + \Big|(z_{n+1} - z'_{n+1})\prod_{k=1}^{n} z'_k\Big|$$
$$\leq \sum_{k=1}^{n} |z_k - z'_k| + |z_{n+1} - z'_{n+1}|.$$

From this the assertion follows. □

Lemma 10.12 (Taylor approximation of the exponential function)**.** *Let $t \in \mathbb{C}$ and $n \in \mathbb{Z}_+$. Then,*

$$\Big|e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!}\Big| \leq \frac{2|t|^n}{n!} \wedge \frac{|t|^{n+1}}{(n+1)!}. \tag{10.7}$$

Proof. Denote by $h_n(t)$ the difference on the left-hand side. For $n = 0$, (10.7) follows from

$$|h_0(t)| = \Big|\int_0^t e^{is}ds\Big| \leq \int_0^t |e^{is}|ds = |t|$$

and

$$|h_0(t)| \leq |e^{it}| + 1 = 2.$$

In general, the following applies to $t \in \mathbb{R}, n \in \mathbb{N}$

$$\Big|\int_0^t h_n(s)ds\Big| = \Big| - i(e^{it} - 1) + i\sum_{k=0}^{n} \frac{(it)^{k+1}}{(k+1)!}\Big| = \Big|ie^{it} - i\sum_{k=0}^{n+1} \frac{(it)^k}{k!}\Big| = |h_{n+1}(t)|,$$

and (10.7) follows by induction. □

Remark 10.13 (notation)**.** *In the following proof, we will use for functions a and b the notation $a \lesssim b$ iff there is a constant C with $a \leq Cb$.*

Proof of theorem 10.8. Wlog let $\mathbf{E}[X_{nj}] = \mu = 0$ and $\sigma^2 = 1$; otherwise we replace X_{nj} by $\frac{X_{nj}-\mathbf{E}[X_{nj}]}{\sqrt{\sigma^2}}$. Let $\sigma_{nj}^2 := \mathbf{V}[X_{nj}]$ and $\sigma_n^2 := \sum_{j=1}^{m_n} \sigma_{nj}^2 \xrightarrow{n\to\infty} 1$. Denote by ψ_{nj} the characteristic function of X_{nj}.
2. ⇒ 1. Since for every $\varepsilon > 0$

$$\sup_{j=1,\dots,m_n} \sigma_{nj}^2 \leq \varepsilon^2 + \sup_{j=1,\dots,m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \leq \varepsilon^2 + \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] \xrightarrow{n\to\infty} \varepsilon^2, \tag{10.8}$$

the second part of 1. is already shown.

Let $(Z_{nj})_{n=1,2,\dots,j=1,\dots,m_n}$ be independent random variables with $Z_{nj} \sim N(0, \sigma_{nj}^2)$. This means that $Z_n = \sum_{j=1}^{m_n} Z_{nj} \sim N(0, \sigma_n^2)$. In particular, the following applies thus $Z_n \xRightarrow{n\to\infty} X$, which can be derived directly from the form of the characteristic functions of the normal distribution, Example 6.13.3 can be read off. Let $\widetilde{\psi}_{nj}$ be the characteristic function of Z_{nj}. Then it suffices to show, see Theorem 9.33, that

$$\prod_{j=1}^{n_j} \psi_{nj}(t) - \prod_{j=1}^{m_n} \widetilde{\psi}_{nj}(t) \xrightarrow{n\to\infty} 0 \tag{10.9}$$

for all t. Using Lemma 10.11 and Lemma 10.12 we write

$$\begin{aligned}\Big|\prod_{j=1}^{m_n}\psi_{nj}(t)-\prod_{j=1}^{m_n}\widetilde{\psi}_{nj}(t)\Big| &\le \sum_{j=1}^{m_n}|\psi_{nj}(t)-\widetilde{\psi}_{nj}(t)|\\ &\le \sum_{j=1}^{m_n}|\psi_{nj}(t)-1+\tfrac{1}{2}t^2\sigma_{nj}^2|+\sum_{j=1}^{m_n}|\widetilde{\psi}_{nj}(t)-1+\tfrac{1}{2}t^2\sigma_{nj}^2|\\ &\lesssim 2\sum_{j=1}^{m_n}\mathbf{E}[X_{nj}^2(1\wedge|X_{nj}|)+\sum_{j=1}^{m_n}|e^{-\frac{1}{2}\sigma_{nj}^2t^2}-1+\tfrac{1}{2}t^2\sigma_{nj}^2|.\end{aligned}$$

Furthermore,

$$\sum_{j=1}^{m_n}\mathbf{E}[X_{nj}^2(1\wedge|X_{nj}|)]\le\varepsilon\sum_{j=1}^{m_n}\sigma_{nj}^2+\sum_{j=1}^{m_n}\mathbf{E}[X_{nj}^2;|X_{nj}|>\varepsilon]\xrightarrow{n\to\infty}\varepsilon$$

and

$$\sum_{j=1}^{m_n}|e^{-\frac{1}{2}\sigma_{nj}^2t^2}-1+\tfrac{1}{2}t^2\sigma_{nj}^2|\lesssim\sum_{j=1}^{m_n}\sigma_{nj}^4\le\sigma_n^2\sup_{j=1,\dots,m_n}\sigma_{nj}^2\xrightarrow{n\to\infty}0$$

because of (10.8). This means (10.9) is already proven.
1. $\Rightarrow$ 2. According to the second part of 1. for each $\varepsilon>0$ with the Chebyshev inequality

$$\sup_{j=1,\dots,m_n}\mathbf{P}[|X_{nj}|>\varepsilon]\le\sup_{j=1,\dots,m_n}\frac{\sigma_{nj}^2}{\varepsilon^2}\xrightarrow{n\to\infty}0. \tag{10.10}$$

With Lemma 10.12,

$$\sup_{j=1,\dots,m_n}|\psi_{nj}(t)-1|\le\sup_{j=1,\dots,m_n}\mathbf{E}[2\wedge|t\cdot X_{nj}|]\le 2\sup_{j=1,\dots,m_n}\mathbf{P}[|X_{nj}|>\varepsilon]+\varepsilon|t|\xrightarrow{n\to\infty}\varepsilon|t|.$$

In particular, $\sum_{j=1}^{m_n}\log\psi_{nj}(t)$ is defined for every t if n is large enough. From 1.

$$\sum_{j=1}^{m_n}\log\psi_{nj}(t)\xrightarrow{n\to\infty}-\frac{t^2}{2}. \tag{10.11}$$

Furthermore, because $\psi_{nj}'(0)=i\mathbf{E}[X_{nj}]=0,\psi_{nj}''(0)=-\mathbf{V}[X_{nj}]=-\sigma_{nj}^2$ with the help of a Taylor expansion of ψ_{nj} around 0

$$|\psi_{nj}(t)-1|\lesssim\sigma_{nj}^2|t|^2$$

and

$$\begin{aligned}\Big|\sum_{j=1}^{m_n}\log\psi_{nj}(t)-\sum_{j=1}^{m_n}(\psi_{nj}(t)-1)\Big| &\lesssim\sum_{j=1}^{m_n}|\psi_{nj}(t)-1|^2\\ &\lesssim\sum_{j=1}^{m_n}(\sigma_{nj}^2)^2|t|^4\lesssim|t|^4\sup_{j=1,\dots,m_n}\sigma_{nj}^2\xrightarrow{n\to\infty}0.\end{aligned} \tag{10.12}$$

Since the convergence of an imaginary series follows from the convergence of its real and imaginary parts, we deduce from (10.11) and (10.12) because $\text{Re}(\psi_{nj}(t)) = \mathbf{E}[\cos(tX_{nj})]$

$$\sum_{j=1}^{m_n} \mathbf{E}[\cos(tX_{nj}) - 1] \xrightarrow{n\to\infty} -\frac{t^2}{2}$$

For $\varepsilon > 0$ is now because of $0 \leq 1 - \cos(\theta) \leq \frac{\theta^2}{2}$

$$\begin{aligned}
0 \leq \limsup_{n\to\infty} \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| > \varepsilon] &= \limsup_{n\to\infty}\ 1 - \sum_{j=1}^{m_n} \mathbf{E}[X_{nj}^2; |X_{nj}| \leq \varepsilon] \\
&\leq \limsup_{n\to\infty}\ 1 - \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{E}[1 - \cos(tX_{nj}); |X_{nj}| \leq \varepsilon] \\
&= \limsup_{n\to\infty} \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{E}[1 - \cos(tX_{nj}); |X_{nj}| > \varepsilon] \qquad (10.13) \\
&\leq \limsup_{n\to\infty} \frac{2}{t^2} \sum_{j=1}^{m_n} \mathbf{P}[|X_{nj}| > \varepsilon] \\
&\leq \frac{2}{\varepsilon^2 t^2} \limsup_{n\to\infty} \sum_{j=1}^{m_n} \sigma_{nj} = \frac{2}{\varepsilon^2 t^2}.
\end{aligned}$$

Since $t, \varepsilon > 0$ were arbitrary, 2. is shown, if in the the last inequality chain $t \to \infty$ is considered. □

10.3 Multidimensional limit laws

So far, we have only considered weak limit theorems (Theorems 10.5 and 10.8) for the case of $\mathbb{R}$–valued random variables. We now generalize this to $\mathbb{R}^d$–valued random variables. In particular, we give a variant of the multidimensional central limit theorem.

Definition 10.14 (Multidimensional normal distribution)**.** *Let $\mu \in \mathbb{R}^d$ and $C \in \mathbb{R}^{d\times d}$ be a strictly positive definite symmetric matrix.*[15,16]*. The d-*dimensional normal distribution *with expected value μ and covariance matrix C is the probability measure $N_{\mu,C}$ on $\mathbb{R}^d$ with density*

$$f_{\mu,C}(x) = \frac{1}{\sqrt{(2\pi)^d \det(C)}} \exp\Big(-\tfrac{1}{2}(x-\mu)C^{-1}(x-\mu)^\top \Big).$$

Proposition 10.15 (Properties of the multidimensional normal distribution)**.** *Let $\mu \in \mathbb{R}^d$, $C = AA^\top \in \mathbb{R}^{d\times d}$ a strictly positive definite symmetric matrix and I the d-dimensional unit matrix. The following are equivalent:*

1. *$X \sim N_{\mu,C}$;*
2. *$tX^\top \sim N_{t\mu^\top, tCt^\top}$ for each $t \in \mathbb{R}^d$;*

[15] We denote row vectors by x and column vectors by $x^\top$.

[16] Strictly positive definite means $xCx^\top > 0$ for all $x \in \mathbb{R}^d$. From linear algebra it is known that for a strictly positive definite matrix C there is always an invertible matrix A with $C = AA^\top$

3. $\psi_X(t) = e^{it\mu^\top} e^{-\frac{1}{2}tCt^\top}$ *for each* $t \in \mathbb{R}^d$.

In each of these cases

4. $X \stackrel{d}{=} AY + \mu$ *for* $Y \sim N_{0,I}$,

5. $\mathbf{E}[X_i] = \mu_i$ *for* $i = 1, \dots, d$,

6. $\mathbf{COV}[X_i, X_j] = C_{ij}$ *for* $i, j = 1, \dots, d$.

Proof. First, let $X \sim N_{\mu,C}$. We first show 4.-6. The property 4. is an application of the transformation theorem. For $B \in \mathcal{B}(\mathbb{R}^d)$ and $T : y \mapsto Ay^\top + \mu^\top$,

$$\begin{aligned} N_{0,I}(T^{-1}(B)) &= \frac{1}{\sqrt{(2\pi)^d}} \int_{T^{-1}(B)} e^{-\frac{1}{2}yy^\top} dy \\ &\stackrel{y=A^{-1}(x-\mu)}{=} \frac{1}{\sqrt{(2\pi)^d}} \frac{1}{\det A} \int_B \exp\big(-\tfrac{1}{2}(x-\mu)(A^\top)^{-1}A^{-1}(x-\mu)^\top \big) dx \\ &= \frac{1}{\sqrt{(2\pi)^d \det C}} \int_B \exp\big(-\tfrac{1}{2}(x-\mu)C^{-1}(x-\mu)^\top \big) dx \\ &= N_{\mu,C}(B). \end{aligned}$$

5. follows from 4. with

$$\mathbf{E}[X_i] = \mathbf{E}[\pi_i(AY + \mu)] = \pi_i \mu = \mu_i,$$

where π_i is the projection onto the i-th coordinate.
6. also follows from 4. with

$$\begin{aligned} \mathbf{COV}[X_i, X_j] &= \mathbf{E}[(\pi_i AY^\top)(\pi_j AY^\top)] = \mathbf{E}[(A_{i\cdot}Y^\top)(A_{j\cdot}Y^\top)] = \mathbf{E}[A_{i\cdot}Y^\top Y A_{j\cdot}^\top] \\ &= A_{i\cdot}A_{j\cdot}^\top = (AA^\top)_{ij} = C_{ij}. \end{aligned}$$

We now come to the equivalence of 1.-3.: '1. $\Rightarrow$ 2.': Since $X \stackrel{d}{=} AY^\top + \mu^\top$ as in 4. $tX^\top = tAY^\top + t\mu^\top$ as a linear combination of (one-dimensional) normal distributions is normally distributed again. The expected value is obviously $t\mu^\top$ and the variance

$$\mathbf{V}[tX^\top] = \mathbf{E}[(tAY^\top)^2] = \mathbf{E}[tAY^\top Y A^\top t^\top] = tAA^\top t^\top = tCt^\top.$$

'2. $\Rightarrow$ 3.': Since $tX^\top \sim N_{t\mu^\top, tCt^\top}$, the statement follows from example 6.13.3.
'3. $\Rightarrow$ 1.': This follows from Proposition 9.25. □

Remark 10.16 (Special cases). *1. If C in Definition 10.14 is positive, but not strictly positive definite (i.e. there is $x \in \mathbb{R}^d$ with $x \neq 0$ and $xCx = 0$), one cannot determine $N_{\mu,C}$ by specifying the density as in the definition above. In this case $N_{\mu,C}$ is defined by specifying the characteristic function, i.e. function, i.e. $N_{\mu,C}$ is the uniquely determined distribution on $\mathbb{R}^d$ with $\psi_{N_{\mu,C}}(t) = e^{it\mu} e^{-\frac{1}{2}tCt^\top}$.*

2. If $Y \sim N_{0,I}$ and A is an orthogonal matrix, then also $X := AY \sim N_{0,I}$. This follows from Proposition 10.15, if you write $I = AA^\top$ and use 4. is used.

Proposition 10.17 (Cramér-Wold Device)**.** *If $X, X_1, X_2, \dots$ are random variables with values in $\mathbb{R}^d$. Then $X_n \xRightarrow{n\to\infty} X$ applies if and only if $tX_n \xRightarrow{n\to\infty} tX$ for all $t \in \mathbb{R}^d$ (where $(t, x) \mapsto tx$ is the scalar product in $\mathbb{R}^d$).*

Proof. '$\Rightarrow$': Let $t \in \mathbb{R}^d$ and $f \in \mathcal{C}_b(\mathbb{R})$. Then $f(t\cdot) \in \mathcal{C}_b(\mathbb{R}^d)$. This means that $\mathbf{E}[f(tX_n)] \xrightarrow{n\to\infty} \mathbf{E}[f(tX)]$, i.e. $tX_n \xrightarrow{n\to\infty} tX$.
'$\Leftarrow$': Let π_i be the projection onto the ith coordinate. Since $(\pi_i X_n)_{n=1,2,\dots}$ according to Corollary 9.18 is tight for all i, you can see that $(X_n)_{n=1,2,\dots}$ is tight. Since $\{x \mapsto e^{itx} : t \in \mathbb{R}^d\}$ is a separating class of functions, the assertion follows from $\mathbf{E}[e^{itX_n}] \xrightarrow{n\to\infty} \mathbf{E}[e^{itX}]$ for all $t \in \mathbb{R}^d$ and Proposition 9.27. □

Theorem 10.18 (Multidimensional central limit theorem)**.** *Let $X_1, X_2, \dots$ be independent, identical distributed random variables with values in $\mathbb{R}^d$ with $\mathbf{E}[X_n] = \mu \in \mathbb{R}^d$ and $\mathbf{COV}[X_{n,i}, X_{n,j}] = C_{ij}$ for $i, j = 1, \dots, d$ and $S_n = \sum_{i=1}^n X_i$. If $X \sim N_{0,C}$, then*

$$\frac{S_n - n\mu}{\sqrt{n}} \xRightarrow{n\to\infty} X.$$

Proof. We apply the one-dimensional central limit theorem, Corollary 10.9, to the independent, identically distributed random variables $tX_1, tX_2, \dots$. This provides

$$t\frac{S_n - n\mu}{\sqrt{n}} \xRightarrow{n\to\infty} tX.$$

Since t was arbitrary, the statement follows from Proposition 10.17. □

11 The conditional expectation

Let $(\Omega, \mathcal{A}, \mathbf{P})$ be a probability space. We write $\mathcal{L}^1 := \mathcal{L}^1(\mathbf{P})$ for the set of all real random variables whose expected value exists. In this chapter we again use the notation $\mathbb{E}[\cdot]$ for the integral with respect to the probability measure $\mathbf{P}$, as well as $\mathcal{L}^p := \mathcal{L}^p(\mathbf{P})$.

11.1 Motivation

Define as in *Elementary Probability* for $A, G \in \mathcal{A}$ and $\mathbf{P}(G) > 0$

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

and analogously the *conditional expectation*

$$\mathbf{E}[X|G] := \frac{\mathbf{E}[X; G]}{\mathbf{P}(G)}.$$

Then $\mathbf{P}(A|G) = \mathbf{E}[1_A|G]$. This relationship means that conditional expectations can be used to calculate conditional probabilities. In particular, the notion of conditional expectation is more general than the notion of conditional probability.

In this chapter, we will use the conditional expectation $\mathbf{E}[X|\mathcal{G}]$ for a random variable X and a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. Here, $\mathbf{E}[X|\mathcal{G}]$ is a $\mathcal{G}$-measurable random variable. As a simple example, $\{G_1, G_2, \dots\} \subseteq \mathcal{F}$ is a partition of Ω with $\mathbf{P}(G_i) > 0$ for $i = 1, 2, \dots$ and $\mathcal{G}$ the generated σ algebra. Then we set for $X \in \mathcal{L}^1$

$$\mathbf{E}[X|\mathcal{G}](\omega) := \sum_{i=1}^{\infty} \mathbf{E}[X|G_i] \cdot 1_{G_i}(\omega). \tag{11.1}$$

The following therefore applies: for $\omega \in G_i$, the random variable $\mathbf{E}[X|\mathcal{G}]$ is given by $\mathbf{E}[X|\mathcal{G}](\omega) = \mathbf{E}[X|G_i] = \mathbf{E}[X; G_i]/\mathbf{P}(G_i)$. In particular it is constant on G_i, $i = 1, 2, \dots$ In other words, $\mathbf{E}[X|\mathcal{G}]$ is measurable with respect to $\mathcal{G}$. The following also applies to $J \subseteq \mathbb{N}$ and $A = \bigcup_{j \in J} G_j \in \mathcal{G}$

$$\begin{aligned} \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] &= \mathbf{E}\Big[\sum_{i=1}^{\infty} \mathbf{E}[X|G_i] 1_{G_i} 1_A\Big] \\ &= \sum_{j \in J} \mathbf{E}\big[\mathbf{E}[X|G_j] 1_{G_j}\big] \\ &= \sum_{j \in J} \mathbf{E}[X|G_j] \cdot \mathbf{P}(G_j) \\ &= \mathbf{E}[X; A]. \end{aligned} \tag{11.2}$$

In particular, with $J = \mathbb{N}$ therefore $\mathbf{E}\big[\mathbf{E}[X|\mathcal{F}]\big] = \mathbf{E}[X]$. The definition of the conditional expectation (11.1) can be generalized with the help of the property (11.2) to any σ-algebras $\mathcal{G} \subseteq \mathcal{F}$.

Example 11.1 (Binomial distribution with random success probability)**.** *Let X be uniformly distributed on $[0, 1]$, i.e. the distribution of X has density $1_{[0;1]}$. Given $X = x$ let $Y_1, \dots, Y_n$ be a sequence of Bernoulli distributed random variables with probability of success x. Therefore, $Y = Y_1 + \cdots + Y_n$ is binomially distributed with n and x, i.e. Y counts the number of successes in n independent experiments with probability of success x. Intuitively, it is clear what*

$$\mathbf{P}(Y = k|X) = \binom{n}{k} X^k (1 - X)^{n-k}$$

should mean. However, this has not yet been defined, since $\mathbf{P}(X = x) = 0$. However, it is worth noting that the right side is a $\sigma(X)$-measurable random variable (since it is a function of X; see Lemma 6.2).

11.2 Definition and properties

We now formally define the conditional expectation $\mathbf{E}[X|\mathcal{G}]$ for $\mathcal{G} \subseteq \mathcal{F}$. As mentioned above, this is a $\mathcal{G}$-measurable random variable whose expectations are as in (11.2) match those of X.

Theorem 11.2 (Existence and properties of the conditional expectation)**.** *Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. Then there is an almost surely unique linear operator $\mathbf{E}[.|\mathcal{G}] : \mathcal{L}^1 \to \mathcal{L}^1$ such that $\mathbf{E}[X|\mathcal{G}]$ for all $X \in \mathcal{L}^1$ a $\mathcal{G}$-measurable random variable with*

1. $\mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]; A\big] = \mathbf{E}[X; A]$ *for all* $A \in \mathcal{G}$.

Further,

2. $\mathbf{E}[X|\mathcal{G}] \geq 0$ *if* $X \geq 0$.

3. $\mathbf{E}\big[|\mathbf{E}[X|\mathcal{G}]|\big] \leq \mathbf{E}[|X|]$.

4. *If* $0 \leq X_n \uparrow X$ *for* $n \to \infty$*, then also* $\mathbf{E}[X_n|\mathcal{G}] \uparrow \mathbf{E}[X|\mathcal{G}]$ *in* $\mathcal{L}^1$ *if all expectations exist.*

5. *If* X *is a* $\mathcal{G}$*-measurable function, then* $\mathbf{E}[XY|\mathcal{G}] = X\mathbf{E}[Y|\mathcal{G}]$ *if all expectations exist.*

6. $\mathbf{E}\big[X\mathbf{E}[Y|\mathcal{G}]\big] = \mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]Y\big] = \mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]\mathbf{E}[Y|\mathcal{G}]\big]$ *if all expectations exist.*

7. *If* $\mathcal{H} \subseteq \mathcal{G}$*, then* $\mathbf{E}\big[\mathbf{E}[X|\mathcal{G}]|\mathcal{H}\big] = \mathbf{E}[X|\mathcal{H}]$.

8. *If* X *is independent of* $\mathcal{G}$*, then* $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$.

Proof. 1. in the case $X \in \mathcal{L}^2$: Let M be the closed linear subspace of $\mathcal{L}^2$, which consists of all functions which, except for a zero set, correspond to a $\mathcal{G}$-measurable function. According to Proposition 4.10 there are almost surely unique functions $Y \in M, Z \perp M$ with $X = Y + Z$. We define $\mathbf{E}[X|\mathcal{G}] := Y$. This means that $X - \mathbf{E}[X|\mathcal{G}] \perp M$, i.e. $\mathbf{E}[X - \mathbf{E}[X|\mathcal{G}]; A] = 0$ for $A \in \mathcal{G}$, from which 1. for $X \in \mathcal{L}^2$ follows.
3. in the case $X \in \mathcal{L}^2$: Choose $A := \{\mathbf{E}[X|\mathcal{G}] \geq 0\}$. According to 1.,

$$\mathbf{E}[|\mathbf{E}[X|\mathcal{G}]|] = \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] - \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A^c] = \mathbf{E}[X; A] - \mathbf{E}[X; A^c] \leq \mathbf{E}[|X|].$$

1. in the case $X \in \mathcal{L}^1$: If $X \in \mathcal{L}^1 \supset \mathcal{L}^2$, then choose $X_1, X_2, \dots \in \mathcal{L}^2$ with $||X_n - X||_1 \xrightarrow{n\to\infty} 0$ (such that $|X_n| := |X| \wedge n$), and define $\mathbf{E}[X|\mathcal{G}] := \lim_{n\to\infty} \mathbf{E}[X_n|\mathcal{G}]$. This limit value exists in $\mathcal{L}^1$, since because of 3.

$$\mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X_m|\mathcal{G}]|] = \mathbf{E}[|\mathbf{E}[X_n - X_m|\mathcal{G}]|] \leq \mathbf{E}[|X_n - X_m|] \xrightarrow{n,m\to\infty} 0$$

the sequence $(\mathbf{E}[X_n|\mathcal{G}])_{n=1,2,\dots}$ is a Cauchy sequence and $\mathcal{L}^1$ is complete. Furthermore, this means that $||\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]||_1 \xrightarrow{n\to\infty} 0$. Furthermore, for $A \in \mathcal{G}$

$$\begin{aligned}|\mathbf{E}[X - \mathbf{E}[X|\mathcal{G}]; A]| \leq \mathbf{E}[|X1_A - X_n1_A|]& \\ + |\mathbf{E}[X_n - \mathbf{E}[X_n|\mathcal{G}]; A]]|& \\ + \mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}]1_A - \mathbf{E}[X|\mathcal{G}]1_A|]& \\ \xrightarrow{n\to\infty} 0&\end{aligned}$$

due to dominated convergence and 1. follows in the case $X \in \mathcal{L}^1$.
3. in the case $X \in \mathcal{L}^1$. Here, too, you can see through an approximation argument if $X_1, X_2, \dots \in \mathcal{L}^2$ with $X_n \xrightarrow{n\to\infty}_{\mathcal{L}^1} X$,

$$\mathbf{E}[|\mathbf{E}[X|\mathcal{G}]|] = \lim_{n\to\infty} \mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}]|] \leq \lim_{n\to\infty} \mathbf{E}[|X_n|] = \mathbf{E}[|X|],$$

since, due to the inverse triangle inequality, approximately,

$$\mathbf{E}[||\mathbf{E}[X_n|\mathcal{G}]| - |\mathbf{E}[X|\mathcal{G}]||] \leq \mathbf{E}[|\mathbf{E}[X|\mathcal{G}] - \mathbf{E}[X_n|\mathcal{G}]|] \xrightarrow{n\to\infty} 0.$$

.
2. set $A = \{\mathbf{E}[X|\mathcal{G}] \leq 0\}$ and thus

$$0 \geq \mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A] \geq 0,$$

thus because of $\mathbf{E}[X|\mathcal{G}]1_A \leq 0$ also $\mathbf{E}[X|\mathcal{G}]1_A = 0$ is almost sure.

4. Due to monotone convergence, $||X_n - X||_1 \xrightarrow{n\to\infty} 0$, i.e. with 3.,

$$\mathbf{E}[|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]|] = \mathbf{E}[|\mathbf{E}[X_n - X|\mathcal{G}]|] \le \mathbf{E}[|X_n - X|] \xrightarrow{n\to\infty} 0.$$

6. in the case $X, Y \in \mathcal{L}^2$. According to the definition of the conditional expectation, $\mathbf{E}[X|\mathcal{G}], \mathbf{E}[Y|\mathcal{G}] \in M$ if M is the linear subspace of $\mathcal{L}^2$ which contains functions which, apart from a zero set with a $\mathcal{G}$-measurable function. Furthermore $X - \mathbf{E}[X|\mathcal{G}] \perp M$. Thus

$$\mathbf{E}[(X - \mathbf{E}[X|\mathcal{G}])\mathbf{E}[Y|\mathcal{G}]] = 0.$$

6. in the case $X, Y \in \mathcal{L}^1$. Choose $X_1, Y_1, X_2, Y_2, \dots \in \mathcal{L}^2$ with $X_n \uparrow X, Y_n \uparrow Y$. Because of 4. and dominated convergence, if all expectations exist,

$$\mathbf{E}[(X - \mathbf{E}[X|\mathcal{G}])\mathbf{E}[Y|\mathcal{G}]] = \lim_{n\to\infty} \mathbf{E}[(X_n - \mathbf{E}[X_n|\mathcal{G}])\mathbf{E}[Y_n|\mathcal{G}]] = 0.$$

5. because of 1. is $\mathbf{E}[X|\mathcal{G}]1_A = X1_A$ for $A \in \mathcal{G}$, almost surely. This means that

$$\mathbf{E}[XY; A] = \mathbf{E}[X\mathbf{E}[Y|\mathcal{G}]; A]$$

after 6. from this follows after 1. already $\mathbf{E}[XY|\mathcal{G}] = X\mathbf{E}[Y|\mathcal{G}]$.

Since $\mathcal{H} \subseteq \mathcal{G}$, for $A \in \mathcal{H}$,

$$\mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A] = \mathbf{E}[\mathbf{E}[X|\mathcal{H}]; A]$$

after 1. From here follows but $\mathbf{E}[\mathbf{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbf{E}[X|\mathcal{H}]$.

8. Certainly, $\mathbf{E}[X]$ is measurable with respect to $\mathcal{G}$. For $A \in \mathcal{G}$,

$$\mathbf{E}[\mathbf{E}[X|\mathcal{G}]; A] = \mathbf{E}[X; A] = \mathbf{E}[X]\mathbf{E}[1_A] = \mathbf{E}\big[\mathbf{E}[X]; A\big]$$

and thus $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$. □

Remark 11.3 (Interpretation and alternative proof). *1. Let $X \in \mathcal{L}^2$. As the proof of 1. in Theorem 11.2 shows, $X - \mathbf{E}[X|\mathcal{G}]$ is perpendicular to the linear subspace of all $\mathcal{G}$-measurable functions. In particular $\mathbf{E}[X|\mathcal{G}]$ is the $\mathcal{G}$-measurable random variable that (in terms of the $\mathcal{L}^2$ norm) is closest to the random variable X comes closest. Therefore, we can say that $\mathbf{E}[X|\mathcal{G}]$ is the best estimate of X if information from the σ algebra $\mathcal{G}$ is available.*

2. The almost surely unambiguous existence of the conditional expectation with the property 1. in Theorem 11.2 can be proved differently than above with the help of the theorem of Radon-Nikodým (Corollary 4.17):

Let $X \ge 0$ first. Set $\widetilde{\mathbf{P}} := \mathbf{P}|_{\mathcal{G}}$, the restriction of $\mathbf{P}$ to $\mathcal{G}$, and $\mu(.) := \widetilde{\mathbf{E}}[X; .]$ a finite measure. Then obviously $\mu \ll \widetilde{\mathbf{P}}$ applies. The theorem of Radon-Nikodým ensures that μ is a density with respect to $\widetilde{\mathbf{P}}$, i.e. there is a $\mathcal{G}$-measurable random variable Z with

$$\mathbf{E}[X; A] = \widetilde{\mathbf{E}}[X; A] = \mu(A) = \widetilde{\mathbf{E}}[Z; A] = \mathbf{E}[Z; A]$$

for all $A \in \mathcal{G}$. Thus Z fulfills the properties of 1. from theorem 11.2. The general case (i.e. X can also take can also assume negative values) then follows with the decomposition $X = X^+ - X^-$.

To prove the (almost sure) uniqueness of the conditional expectation, let Z' be another $\mathcal{G}$-measurable random variable with random variable with $\mathbf{E}[Z'; A] = \mathbf{E}[X; A]$ for all $A \in \mathcal{G}$. Then $B := \{Z' - \mathbf{E}[X|\mathcal{G}] > 0\} \in \mathcal{G}$ and $\mathbf{E}[\mathbf{E}[X|\mathcal{G}] - Z'; B] = \mathbf{E}[X - X; B] = 0$ and likewise $\mathbf{E}[\mathbf{E}[X|\mathcal{G}] - Z'; B^c] = 0$. This therefore means $Z' = \mathbf{E}[X|\mathcal{G}]$, almost surely.

Proposition 11.4 (Jensen's inequality for conditional expectations). *Let I be an open interval, $\mathcal{G} \subseteq \mathcal{A}$ and $X \in \mathcal{L}^1$ with values in I and $\varphi : I \to \mathbb{R}$ is convex. Then,*

$$\mathbf{E}[\varphi(X)|\mathcal{G}] \geq \varphi(\mathbf{E}[X|\mathcal{G}]).$$

Proof. The proof is analogous to that of Jensen's inequality in the unconditional case, Proposition 6.6: Since I is open, $\mathbf{E}[X|\mathcal{G}] \in I$, almost surely. We recall the definition of λ in (6.4). Further, as in (6.5) for $x \in I$,

$$\varphi(x) \geq \varphi(\mathbf{E}[X|\mathcal{G}]) + \lambda(\mathbf{E}[X|\mathcal{G}])(x - \mathbf{E}[X|\mathcal{G}])$$

and thus

$$\begin{aligned}\mathbf{E}[\varphi(X)|\mathcal{G}] &\geq \mathbf{E}[\varphi(\mathbf{E}[X|\mathcal{G}])|\mathcal{G}] + \mathbf{E}[\lambda(\mathbf{E}[X|\mathcal{G}]) \cdot (X - \mathbf{E}[X|\mathcal{G}])|\mathcal{G}] \\ &= \varphi\big(\mathbf{E}[X|\mathcal{G}]\big).\end{aligned}$$

□

Lemma 11.5 (Uniform integrability and conditional expectation). *Let $X \in \mathcal{L}^1$. Then the family $(\mathbf{E}[X|\mathcal{G}])_{\mathcal{G} \subseteq \mathcal{A}}$ is uniformly integrable.*

Proof. Since $\{X\}$ is uniformly integrable, according to Lemma 7.9 there is a monotonically increasing convex function $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ with $\frac{\varphi(x)}{x} \xrightarrow{x\to\infty} \infty$ and $\mathbf{E}[\varphi(|X|)] < \infty$. With Theorem 11.2.3, we obtain

$$\sup_{\mathcal{F} \subseteq \mathcal{A}} \mathbf{E}[\varphi(|\mathbf{E}[X|\mathcal{F}]|)] \leq \mathbf{E}[\varphi(|X|)] < \infty.$$

This means that $\{\mathbf{E}[X|\mathcal{F}] : \mathcal{F} \subseteq \mathcal{A}\ \sigma\text{-algebra}\}$ is uniformly integrable, again according to Lemma 7.9. □

Theorem 11.6 (Dominated and monotone convergence for conditional expectations). *Let $\mathcal{G} \subseteq \mathcal{F}$ and $X_1, X_2, \dots \in \mathcal{L}^1$. Assume one of the following:*

1. *Let $X \in \mathcal{L}^1$ such that $X_n \uparrow X$, almost surely.*
2. *If $Y \in \mathcal{L}^1$ such that $|X_n| \leq |Y|$ for all n, and $X_n \xrightarrow{n\to\infty} X$ almost surely.*

Then

$$\mathbf{E}[X_n|\mathcal{G}] \xrightarrow{n\to\infty} \mathbf{E}[X|\mathcal{G}]$$

almost surely and in $\mathcal{L}^1$.

Proof. For the $\mathcal{L}^1$-convergence one has in both cases with Theorem 11.2.3

$$\begin{aligned}\mathbf{E}\big[\big|\mathbf{E}[X_n|\mathcal{G}] - \mathbf{E}[X|\mathcal{G}]\big|\big] &= \mathbf{E}\big[\big|\mathbf{E}[X_n - X|\mathcal{G}]\big|\big] \\ &\leq \mathbf{E}[|X_n - X|] \xrightarrow{n\to\infty} 0.\end{aligned}$$

We divide the almost sure convergence into the two cases: in case 1. it is clear from Theorem 11.2.2 that $\mathbf{E}[X_n|\mathcal{G}]$ grows monotonically. Furthermore, for $A \in \mathcal{F}$ with the theorem of monotone convergence

$$\mathbf{E}\big[\sup_n \mathbf{E}[X_n|\mathcal{G}]; A\big] = \sup_n \mathbf{E}\big[\mathbf{E}[X_n|\mathcal{G}]; A\big] = \sup_n \mathbf{E}[X_n; A] = \mathbf{E}[\sup_n X_n; A] = \mathbf{E}[X; A].$$

However, this shows that $\sup_n \mathbf{E}[X_n|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}]$, almost surely. In case 2. we set

$$\begin{aligned}
Y_n &:= \sup_{k\geq n} X_k \downarrow \limsup_n X_n = X \text{ almost surely,}\\
Z_n &:= \inf_{k\geq n} X_k \uparrow \liminf_n X_n = X \text{ almost surely.}
\end{aligned}$$

Thus $-Y \leq Z_n \leq X_n \leq Y_n \leq Y$, i.e. in particular $Y_1, Z_1, Y_2, Z_2, \dots \in \mathcal{L}^1$, so according to 1.,

$$\mathbf{E}[X|\mathcal{G}] = \lim_{n\to\infty} \mathbf{E}[Z_n|\mathcal{G}] \leq \lim_{n\to\infty} \mathbf{E}[X_n|\mathcal{G}] \leq \lim_{n\to\infty} \mathbf{E}[Y_n|\mathcal{G}] = \mathbf{E}[X|\mathcal{G}],$$

almost surely. In particular, $\mathbf{E}[X_n|\mathcal{G}] \xrightarrow{n\to\infty} \mathbf{E}[X|\mathcal{G}]$, almost surely. □

11.3 The case $\mathcal{G} = \sigma(X)$

In the case $\mathcal{G} = \sigma(X)$, $\mathbf{E}[Y|X] := \mathbf{E}[Y|\sigma(X)]$ is the expectation of Y, given that the random variable X is fixed. This is a function of X, as Proposition 11.7 shows.

Proposition 11.7 (Conditioning on a random variable)**.** *Let $(\Omega', \mathcal{F}')$ be a measurable space, X a random variable with values in Ω' and $Y \in \mathcal{L}^1$. Then there exists a $\mathcal{F}'/\mathcal{B}(\mathbb{R})$-measurable mapping $\varphi : \Omega' \to \mathbb{R}$ with $\mathbf{E}[Y|X] = \varphi(X)$.*

Proof. Clear according to Lemma 6.2. □

Example 11.8 (Random success probability)**.** *Let us consider the question posed in Example 11.1 regarding the existence of the conditional probability $\mathbf{P}(Y = k|X)$, where X is uniform on $[0, 1]$ and X is independently binomially distributed with n and X. We now show (the intuitive equation)*

$$\mathbf{P}(Y = k|X) = \binom{n}{k} X^k (1 - X)^{n-k}. \tag{11.3}$$

Let $A = \{X \in I\}$ for $I \in \mathcal{B}([0, 1])$, i.e. A is a $\sigma(X)$-measurable quantity. Then,

$$\mathbf{E}[1_{Y=k}; A] = \mathbf{P}(Y = k, X \in I) = \int_I \binom{n}{k} x^k (1 - x)^{n-k} dx = \mathbf{E}\Big[\binom{n}{k} X^k (1 - X)^{n-k}; A\Big]$$

However, this means that (11.3) *is true.*

Example 11.9 (Sums of independent identically distributed random variables)**.** *Let $X_1, X_2, \dots$ be a sequence of independent, identically distributed random variables, $\mu = \mathbf{E}[X_1]$ and $S_n := X_1 + \dots + X_n$. Then*

$$\begin{aligned}
\mathbf{E}[S_n|X_1] &= \mathbf{E}[X_1|X_1] + \mathbf{E}[X_2 + \dots + X_n|X_1] = X_1 + (n - 1)\mu,\\
\mathbf{E}[X_1|S_n] &= \tfrac{1}{n} \sum_{i=1}^n \mathbf{E}[X_i|S_n] = \tfrac{1}{n}\mathbf{E}[S_n|S_n] = \tfrac{1}{n} S_n.
\end{aligned}$$

In the second calculation, for example, for $X = S_n$ and $Y = X_1$ the function φ from Proposition 11.7 is given by $\varphi(x) = \frac{1}{n}x$.

Example 11.10 (Buffon's needle problem)**.** *On a plane, vertical lines are at a horizontal distance of 1. Needles, also of length 1, are thrown onto the plane; see Figure 2. Let us consider a needle. We set*

$$Z := \begin{cases} 1, & \textit{if the needle intersects a straight line} \\ 0, & \textit{otherwise} \end{cases}.$$

The center of the needle X is away from the left straight line and the (extension of the)

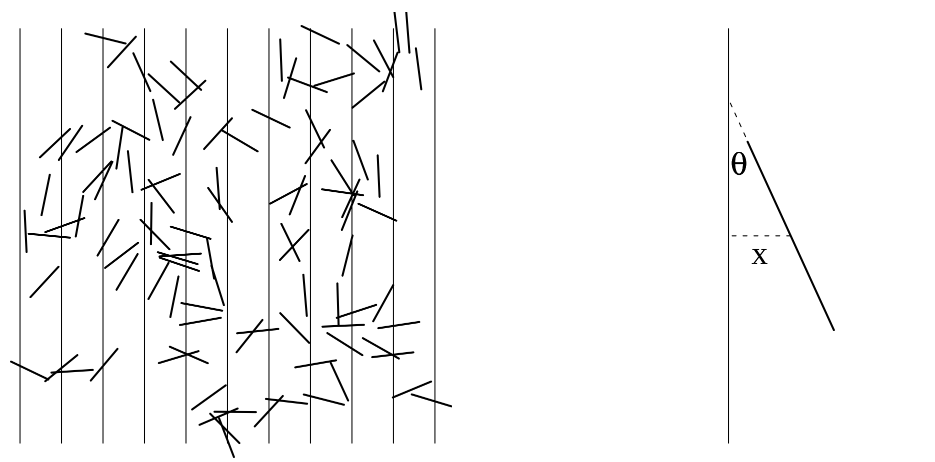


Figure 2: Sketch of Buffon's needle problem

needle makes an angle Θ with the straight line. This means that X is uniform on $[0;1]$, Θ is uniformly independent on $[0;\frac{\pi}{2}]$ and

$$\mathbf{P}(Z = 1|\Theta) = \mathbf{P}(X \leq \tfrac{1}{2}\sin(\Theta) \textit{ or } X \geq 1 - \tfrac{1}{2}\sin(\Theta) Theta) = \sin(\Theta).$$

This means that

$$\mathbf{P}(Z = 1) = \mathbf{E}[\mathbf{P}(Z = 1|\Theta)] = \mathbf{E}[\sin(\Theta)] = \frac{2}{\pi}\int_0^{\pi/2} (\sin(\theta) d\theta = \frac{2}{\pi}.$$

This can be interpreted as follows: if you want to determine by simulation (i.e. by a Monte Carlo method) to find the numerical value of π you can simulate Buffon's needles. Since each individual needle has the probability $\frac{2}{\pi}$ of hitting a vertical line, is approximately

$$\pi \approx \frac{2}{\textit{proportion of needles that hit a vertical line}}$$

according to the law of large numbers.

Example 11.11 (Search in lists)**.** *Consider n names of people who come from r different cities. Each person comes (independently of any other) with probability p_j from city j, $j = 1, \dots, r$. The names (together with other personal data) are entered in r different (unordered) lists. If you now want a (random, according to the probabilities $p_1, \dots, p_r$) person in the list,* *first determine the list, you first determine the city j from which the person comes from and then search the list j for the person's name. Until you realize that the name does not appear in the list you have to compare the person to be found with names on the list. The question now is: How many times on average do you have to compare the name of the person to be found with names on the list without success until you finally know that the person is not on the list?*

We first define a few random variables:
J *: number of the city from which the person to be searched comes*
L *: number of unsuccessful comparisons until the name of the person to be found is found*
Z_j *: number of people from city j*
and $Z = (Z_1, \dots, Z_r)$. In order to determine $\mathbf{E}[L]$, we first determine

$$\mathbf{P}(L = a|J, Z) = 1_{Z_J = a}$$

and thus

$$\mathbf{P}(L = a|Z) = \sum_{j=1}^{r} p_j 1_{Z_j = a}.$$

From this we conclude

$$\mathbf{E}[L|Z] = \sum_{a=1}^{\infty} \sum_{j=1}^{r} a \cdot p_j \cdot 1_{Z_j = a} = \sum_{j=1}^{r} p_j Z_J$$

and therefore

$$\mathbf{E}[L] = \mathbf{E}[\mathbf{E}[L|Z]] = \sum_{j=1}^{r} p_j \mathbf{E}[Z_j] = n \cdot \sum_{j=1}^{r} p_j^2.$$

Example 11.12 (Mixture of Poisson distributions)**.** *Let $\lambda > 0$ and $\lambda \sim \exp(\lambda)$ and for a given λ let $X \sim Poi(\lambda)$. We now show that $X + 1 \sim geo(1/(1+\lambda))$.*
Because: According of Proposition 9.25, the distribution is determined by the characteristic function. First of all, the characteristic function of $Y \sim geo(p)$

$$t \mapsto \mathbf{E}[e^{itY}] = \sum_{k=0}^{\infty} (1-p)^{k-1} p e^{itk} = p e^{it} \sum_{k=0}^{\infty} \big((1-p)e^{it}\big)^k = \frac{p e^{it}}{1 - (1-p)e^{it}} = \frac{\frac{p}{1-p} e^{it}}{\frac{1}{1-p} - e^{it}}.$$

We calculate with Example 6.13.2 for $t \in \mathbb{R}$

$$\mathbf{E}[e^{it(X+1)}] = e^{it} \mathbf{E}\big[\mathbf{E}[e^{itX}|\Lambda]\big] = e^{it} \mathbf{E}\big[e^{-(1-e^{it})\Lambda}\big] = \frac{\lambda e^{it}}{1 + \lambda - e^{it}},$$

so that the assertion with $\lambda = p/(1-p)$ or $p = 1/(1+\lambda)$ follows.

11.4 Conditional independence

In Section 8, we have already learned about the independence of σ algebras (or of random variables). Conditional expectations and independence are closely related, as the next lemma shows.

Lemma 11.13 (Conditional probability and independence)**.** *The σ-algebras $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ are independent if and only if $\mathbf{P}(G|\mathcal{H}) = \mathbf{P}(G)$ for all $G \in \mathcal{G}$.*

Proof. '⇒': Let $\mathcal{G}$ and $\mathcal{H}$ be independent. Then, for $G \in \mathcal{G}, H \in \mathcal{H}$,

$$\mathbf{E}[\mathbf{P}(G), H] = \mathbf{P}(G \cap H) = \mathbf{E}[\mathbf{P}(G|\mathcal{H}), H].$$

This means that $\mathbf{P}(G|\mathcal{H}) = \mathbf{P}(G)$ according to the definition of the conditional expectation.

'⇐': So if $\mathbf{P}(G|\mathcal{H}) = \mathbf{P}(G)$, it follows for $H \in \mathcal{H}$

$$\mathbf{P}(G \cap H) = \mathbf{E}[1_G, H] = \mathbf{E}[\mathbf{P}(G|\mathcal{H}), H] = \mathbf{E}[\mathbf{P}(G), H] = \mathbf{P}(G) \cdot \mathbf{P}(H).$$

□

The concept of independence is often also required in a conditional form. Let's start with an important example.

Example 11.14 (Markov chains)**.** *Let E be a countable set. A Markov chain $\mathcal{X} = (X_t)_{t=0,1,2,\dots}$ is a family of E-valued random variables such that for all $A \subseteq E$*

$$\mathbf{P}(X_{t+1} \in A|X_0, \dots, X_t) = \mathbf{P}(X_{t+1} \in A|X_t). \tag{11.4}$$

This means: if you want to know the distribution of X_{t+1}, and the information of the random variable X_t is already available, the information about the random variables $X_0, \dots, X_{t-1}$ does not provide any additional information. One also says:

> *Given X_t, X_{t+1} is independent of $X_0, \dots, X_{t-1}$.*

Or in terms of σ-algebras:

> *Given $\sigma(X_t)$, $\sigma(X_{t+1})$ is independent of $\sigma(X_0, \dots, X_{t-1})$.*

One can also say in this case: given the present (that is the state at time t, X_t) the future (i.e. X_{t+1}) is independent of the past (these are the states $X_0, \dots, X_{t-1}$).

A simple example of a Markov chain is the one-dimensional random walk: let $Y_1, Y_2, \dots$ be independent and identically distributed such that $\mathbf{P}(Y_1 = 1) = p$ and $\mathbf{P}(Y_1 = -1) = q$ for a $p \in [0, 1]$. Further, let $X_0 = 0$ and $X_t = Y_1 + \dots + Y_t$. Then $(X_t)_{t \geq 0}$ is a Markov chain, because

$$\mathbf{P}(X_{t+1} = k|X_0, \dots, X_t) = \begin{cases} p, & k = X_t + 1, \\ q, & k = X_t - 1. \end{cases}$$

In particular, the right-hand side defines an X_t-measurable random variable and is therefore equal to $\mathbf{P}(X_{t+1} = k|X_t)$.

Definition 11.15 (Conditional independence)**.** *Let $\mathcal{G} \subseteq \mathcal{F}$. A family $(\mathcal{C}_i)_{i \in I}$ of set systems with $\mathcal{C}_i \subseteq \mathcal{F}$ is called* independently given $\mathcal{G}$ *if*

$$\mathbf{P}\Big(\bigcap_{j \in J} A_j|\mathcal{G}\Big) = \prod_{j \in J} \mathbf{P}(A_j|\mathcal{G}) \tag{11.5}$$

applies to all $J \subseteq_f I$ and $A_j \in \mathcal{C}_j, j \in J$.

Similarly, conditional independence is de fined for random variables. Let Y be a random variable. A family $(X_i)_{i \in I}$ of random variables is independent given $\mathcal{G}$ (or Y) if $(\sigma(X_i))_{i \in I}$ is independent given $\mathcal{G}$ (resp. $\sigma(Y)$).

Example 11.16 (Simple cases). *Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra and $(\mathcal{C}_i)_{i\in I}$ a family of set systems.*

1. *If $\mathcal{G} = \mathcal{F}$, then $(\mathcal{C}_i)_{i\in I}$ is always independent given $\mathcal{G}$.*
2. *If $\mathcal{G} = \{\emptyset, \Omega\}$, then $(\mathcal{C}_i)_{i\in I}$ is independent given $\mathcal{G}$ if and only if $(\mathcal{C}_i)_{i\in I}$ are independent.*

Example 11.17 (Binomial distribution with random success probability). *We look again at the coin toss with random success probability from Example 11.1 and 11.8. Here X was uniformly distributed on $[0,1]$ and, given X, $Y_1, \dots, Y_n$ are Bernoulli distributed. Now it should hold that $(Y_1, \dots, Y_n)$ are independent given X. Just like in Example 11.8, we calculate for $A = \{X \in I\}$ and for some $I \in \mathcal{B}([0,1])$ and $y_1, \dots, y_n \in \{0,1\}$ and $k := y_1 + \cdots + y_n$*

$$\begin{aligned}\mathbf{E}[1_{Y_1=y_1,\dots,Y_n=y_n}, A] &= \mathbf{P}(Y_1 = y_1, \dots, Y_n = y_n, X \in I) \\ &= \int_I x^{y_1+\cdots+y_n}(1-x)^{n-y_1-\cdots-y_n}dx = \mathbf{E}\big[X^k(1-X)^{n-k}, A\big],\end{aligned}$$

so

$$\mathbf{P}(Y_1 = y_1, \dots, Y_n = y_n|X) = X^k(1-X)^{n-k}.$$

Analogously, one shows for $i = 1, \dots, n$

$$\mathbf{P}(Y_i = y_1|X) = X^{y_i}(1-X)^{1-y_i}.$$

From this follows

$$\mathbf{P}(Y_1 = y_1, \dots, Y_n = y_n|X) = \prod_{i=1}^n \mathbf{P}(Y_i = y_i|X),$$

so $(Y_1, \dots, Y_n)$ are independent given X.

Lemma 11.13 also exists in the following version, in which the independence is replaced by conditional independence.

Proposition 11.18 (Conditional probability and conditional independence). *Let $\mathcal{K} \subseteq \mathcal{F}$ be a σ-algebra. The σ-algebras $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$ are independent given $\mathcal{K}$ if and only if $\mathbf{P}(G|\sigma(\mathcal{H},\mathcal{K})) = \mathbf{P}(G|\mathcal{K})$ for all $G \in \mathcal{G}$.*

Proof. '⇒': If $\mathcal{G}$ and $\mathcal{H}$ are independent given $\mathcal{K}$, then for $G \in \mathcal{G}, H \in \mathcal{H}, K \in \mathcal{K}$

$$\mathbf{E}[\mathbf{P}(G|\mathcal{K}), H \cap K] = \mathbf{E}[\mathbf{P}(G|\mathcal{K})\mathbf{P}(H|\mathcal{K}), K] = \mathbf{E}[\mathbf{P}(G \cap H|\mathcal{K}), K] = \mathbf{P}(G \cap H \cap K).$$

Now we can show that the set system

$$\mathcal{D} := \{A \in \sigma(\mathcal{H},\mathcal{K}) : \mathbf{E}[\mathbf{P}(G|\mathcal{K}), A] = \mathbf{P}(G \cap A)\}$$

is a $\cap$-stable Dynkin system with $\mathcal{D} \supseteq \mathcal{H}, \mathcal{K}$. Now it follows from Theorem 1.13 that $\mathcal{D} = \sigma(\mathcal{H},\mathcal{K})$, from which $\mathbf{P}(G|\sigma(\mathcal{H},\mathcal{K})) = \mathcal{P}(G|\mathcal{K})$ follows.
'⇐': So if $\mathbf{P}(G|\sigma(\mathcal{H},\mathcal{K})) = \mathbf{P}(G|\mathcal{K})$, it follows for $H \in \mathcal{H}$

$$\mathbf{P}(G \cap H|\mathcal{K}) = \mathbf{E}[\mathbf{P}(G|\sigma(\mathcal{H},\mathcal{K})), H|\mathcal{K}] = \mathbf{E}[\mathbf{P}(G|\mathcal{K}), H|\mathcal{K}] = \mathbf{P}(G|\mathcal{K}) \cdot \mathbf{P}(H|\mathcal{K}).$$

□

Example 11.19 (Markov chains). *Let's look again at the Markov chain $(X_t)_{t=0,1,2,\dots}$ from Example 11.14. For fixed t we set $\mathcal{G} = \sigma(X_{t+1}), \mathcal{H} = \sigma(X_0, \dots, X_{t-1}), \mathcal{K} = \sigma(X_t)$. The Markov property* (11.4) *now says for $G \in \mathcal{G}, H \in \mathcal{H}, K \in \mathcal{K}$ that $\mathbf{P}(G|\sigma(\mathcal{H},\mathcal{K})) = \mathbf{P}(G|\mathcal{K})$. According to Proposition 11.18 this means that X_{t+1} and $(X_0, \dots, X_{t-1})$ are independent given X_t.*

11.5 Regular version of the conditional distribution

We have seen in Section 11.1 how the conditional probability $\mathbf{P}(A|\mathcal{G}) := \mathbf{E}[1_A|\mathcal{G}]$ for a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ is defined. However, this does *not* mean that we have a probability measure $A \mapsto \mathbf{P}(A|\mathcal{G})$; see the next remark. In most cases, however, one can define such a (random, $\mathcal{G}$-measurable) measure, the (or better: a) regular version of the conditional distribution.

Remark 11.20 (Conditional probabilities and conditional distributions)**.** *Let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra and $A_1, A_2, \dots \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$. Then, for $B \in \mathcal{G}$*

$$\begin{aligned}\mathbf{E}\Big[\mathbf{P}\Big(\bigcup_{n=1}^\infty A_n|\mathcal{G}\Big); B\Big] &= \mathbf{E}\big[\mathbf{E}[1_{\bigcup_{n=1}^\infty A_n}|\mathcal{G}]; B\big] = \mathbf{E}[1_{\bigcup_{n=1}^\infty A_n}; B]\\ &= \mathbf{E}\Big[\sum_{n=1}^\infty 1_{A_n}; B\Big] = \sum_{n=1}^\infty \mathbf{E}[1_{A_n}; B]\\ &= \sum_{n=1}^\infty \mathbf{E}\big[\mathbf{P}(A_n|\mathcal{G}); B\big] = \mathbf{E}\Big[\sum_{n=1}^\infty \mathbf{P}(A_n|\mathcal{G}); B\Big]\end{aligned}$$

and therefore

$$\mathbf{P}\Big(\bigcup_{n=1}^\infty A_n|\mathcal{G}\Big) = \sum_{n=1}^\infty \mathbf{P}(A_n|\mathcal{G}) \tag{11.6}$$

$\mathbf{P}$*-almost surely. This means that there is a zero set (depending on $A_1, A_2, \dots$) so that* (11.6) *applies to all ω outside this zero set. However, since there are uncountably many sequences $A_1, A_2, \dots \in \mathcal{F}$, there does not have to be a zero set N, so that* (11.6) *holds for every choice of $A_1, A_2, \dots \in \mathcal{F}$ outside of N. However, if there is such an N, we will say that a* regular version of the conditional distribution *of $\mathbf{P}$ given $\mathcal{G}$ exists. We will give conditions for this in Theorem 11.23.*

We recall the concept of the stochastic kernel; see Definition 5.9.

Definition 11.21 (Regular version of the conditional distribution)**.** *Let $(\Omega', \mathcal{F}')$ be a measurable space, Y an Ω'-valued measurable random variable and $\mathcal{G} \subseteq \mathcal{F}$. A stochastic kernel $\kappa_{Y,\mathcal{G}}$ from $(\Omega, \mathcal{G})$ to $(\Omega', \mathcal{F}')$ is called* regular version of the conditional distribution *of Y, given $\mathcal{G}$, if*

$$\kappa_{Y,\mathcal{G}}(\omega, B) = \mathbf{P}(Y \in B|\mathcal{G})(\omega)$$

for $\mathbf{P}$-almost all ω and every $B \in \mathcal{F}'$.

Remark 11.22 (Distribution conditional on a random variable)**.** *42*

1. *For the stochastic kernel from Definition 11.21 it is sufficient to use property (ii) from Definition 5.9 only for a $\cap$-stable generator $\mathcal{C}$ of $\mathcal{F}$. This is because*

$$\mathcal{D} := \{A' \in \mathcal{F}' : \omega \mapsto \kappa(\omega, A') \text{ is } \mathcal{A}\text{-measurable}\}$$

is alwaysa Dynkin system. Thus, according to Theorem 1.13, $\mathcal{D} = \sigma(\mathcal{C})$.

2. *Let $\mathcal{G} = \sigma(X)$ for a random variable X in Definition 11.21.2. Then, if $\kappa_{Y,\sigma(X)}$ is a regular version of the conditional expectation of Y given $\sigma(X)$, then $\omega \mapsto \kappa_{Y,\sigma(X)}(\omega, A')$ $\sigma(X)$-is measurable for all $A' \in \mathcal{A}'$. This means that, according to Proposition 11.7, there is a $\sigma(X)/\mathcal{B}([0;1])$-measurable map $\varphi_{A'} : \Omega \to [0;1]$ with $\varphi_{A'} \circ X = \kappa_{Y,\sigma(X)}(., A')$. We then set*

$$\kappa_{Y,X}(x, A') := \varphi_{A'}(x)$$

and say $\kappa_{Y,X}$ is the regular version of the conditional distribution of Y given X.

Theorem 11.23 (Existence of the regular version of the conditional distribution)**.** *Let (E, r) be a complete and separable metric space space equipped with Borel's σ-algebra, $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra and Y a (according to $\mathcal{F}$ measurable) random variable with values in E. Then there exists a regular version of the conditional distribution of Y given $\mathcal{G}$.*

Before we can prove the theorem, we need a property (Proposition 11.25) over complete, separable metric spaces.

Definition 11.24 (Borel space)**.** 1. *Two metric spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ are called* isomorphic *if there is a bijective, according to $\mathcal{F}/\mathcal{F}'$-measurable mapping $\varphi : \Omega \to \Omega'$ exists such that φ^{-1} is $\mathcal{F}'/\mathcal{F}$-measurable.*

2. *A measurable space $(\Omega, \mathcal{F})$ is called* Borel space *if there is a Borel set $A \in \mathcal{B}(\mathbb{R})$ exists such that $(\Omega, \mathcal{F})$ and $(A, \mathcal{B}(A))$ are isomorphic.*

Proposition 11.25 (Polish and Borel spaces)**.** *Every complete and separable metric space (E, r), equipped with the Borel's σ-algebra, is a Borel space.*

Proof. See, for example, Dudley, Real analysis and probability, Theorem 13.1.1. □

Proof of theorem 11.23. We prove the theorem under the weaker condition that E, equipped with the Borel σ-algebra, is a Borel space. Wlog, we can therefore assume that $E \in \mathcal{B}(\mathbb{R})$ is. The strategy of our proof consists of finding a distribution function of the conditional distribution by first fixing it for rational values before extending it to all real numbers.

For $r \in \mathbb{Q}$, let F_r be a version of $\mathbf{P}(Y \leq r|\mathcal{G})$ (i.e. $F_r = \mathbf{P}(Y \leq r|\mathcal{G})$ almost surely. Let $A \in \mathcal{F}$ be such that for $\omega \in A$ the mapping $r \mapsto F_r(\omega)$ is non-increasing with limits 1 and 0 at $\pm\infty$. Since A is given by countably many conditions, all of which are almost certainly fulfilled, $\mathbf{P}(A) = 1$. Now define for $x \in \mathbb{R}$

$$F_x(\omega) := 1_A(\omega) \cdot \inf_{r>x} F_r(\omega) + 1_{A^c}(\omega) \cdot 1_{x\geq 0}.$$

Thus, $x \mapsto F_x(\omega)$ is a distribution function for all ω. Define

$$\kappa(\omega, .) := \text{measure defined by } x \mapsto F_x(\omega).$$

For $r \in \mathbb{Q}$ and $B = (-\infty; r]$,

$$\omega \mapsto \kappa(\omega, B) = 1_A(\omega) \cdot \mathbf{P}(Y \leq r|\mathcal{G})(\omega) + 1_{A^c}(\omega) \cdot 1_{r\geq 0} \tag{11.7}$$

is $\mathcal{F}$-measurable. Since $\{(-\infty; r] : r \in \mathbb{Q}\}$ is a $\cap$-stable generator of $\mathcal{B}(\mathbb{R})$, according to Remark 11.22 the mapping $\omega \mapsto \kappa(\omega, B)$ is measurable for all $B \in \mathcal{F}$. Therefore, κ is a stochastic kernel.

It remains to show that κ is a regular version of the conditional distribution. Since (11.7) is based on a $\cap$-stable generator of $\mathcal{E}$, for $\omega \in A$

$$\kappa(\omega, B) = \mathbf{P}(Y \in B|\mathcal{G})(\omega).$$

In other words, κ is a regular version of the conditional distribution. $\square$

Part III
Stochastic Processes

Stochastic processes play a central role in modern stochastics. They are used in various application fields, including financial mathematics, as well as in biology and physics. Stochastic processes are always used when a variable - for example a stock price, the frequency of an allele in a population or the position of a small particle – changes randomly over time.

The aim here is to provide important tools for dealing with stochastic processes. We will deal with important examples, such as the Poisson process or Brownian motion. The latter also plays a decisive role in the construction of stochastic integrals.

The following books have guided me as references for the purpose of this manuscript.

- Durrett, Rick. Probability: Theory and Examples, Cambridge Series in Statistical and Probabilistic Mathematics, 2019
- Kallenberg, Olaf. Foundations of Modern Probability Theory. Springer, third edition, 2021
- Klenke, Achim. Probability theory. A comprehensive course. Springer, 2014

This manuscript is based on the courses in Measure Theory and Probability Theory, which cover Sections 1–3, and 4–12, respectively.

The present english version of this manuscript was written based on the German version with the help of DeepL.

12 Introduction

Stochastic processes are nothing more than families of random variables. It is important to realize that this family is indexed by with time. In the course of time, more and more random variables are realized.

In the following, let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, (E, r) a complete and separable metric space with with Borel's σ-algebra $\mathcal{B}(E)$ and I an ordered subset of $\overline{\mathbb{R}}$, which we also call index set. We will always consider the two cases $I \subseteq \overline{\mathbb{Z}}$ and $I \subseteq \overline{\mathbb{R}}$. Note already here that an uncountable index set, such as $I = \mathbb{R}$, raises new questions, as probability measures are only known to be able to deal with countably number of events.

12.1 Definition and existence

First of all, we take care of the elementary question of what a stochastic process is and how it can be defined in an unambiguous way.

Definition 12.1 (Stochastic process)**.** *1. Let $\mathcal{X} = (X_t)_{t \in I}$ such that $X_t : \Omega \to E$ is $\mathcal{F}/\mathcal{B}(E)$-measurable. Then, $\mathcal{X}$ is called an E-valued* (stochastic) process. *For $\omega \in \Omega$, the mapping given by $X(\omega) : t \mapsto X_t(\omega)$ is called a* path *of X.*

2. *If in 1., the probability space $\Omega = E^I$ and $X_t = \pi_t$ is the projection, then $\mathcal{X}$ is called* canonical process.

3. *Let $0 < p < \infty$. A real-valued process $\mathcal{X} = (X_t)_{t \in I}$ is called p-fold integrable if $\mathbf{E}[|X_t|^p] < \infty$ for all $t \in I$. It is called L^p-bounded, if $\sup_{t \in I} \mathbf{E}[|X_t|^p] < \infty$.*

In the Sections 12.2 and 12.3, we will become familiar with two examples of stochastic processes. In particular, the Poisson process (see Section 12.2) is the first process with an uncountable index set $I = [0, \infty)$.

Example 12.2 (Sums of independent random variables and Markov chains)**.** *From the lecture* Elementary Probability Theory*, some stochastic processes are already known, even if they were not called stochastic processes.*

1. *Let $(X_t)_{t \in I}$ be independent. Then $\mathcal{X} = (X_t)_{t \in I}$ is a (very simple) stochastic process.*

2. *Let $X_1, X_2, ...$ be real-valued, independent, identically distributed random variables. Then, $\mathcal{S} = (S_t)_{t=0,1,2,...}$ with $S_0 = 0$ and*

$$S_t = \sum_{i=1}^{t} X_i$$

   for $t = 1, 2, ...$ is a real-valued, stochastic process with index set $I = \{0, 1, 2, ...\}$. In particular, if $\mathbf{P}(X_i = \pm 1) = 1/2$, then $\mathcal{S}$ is called a one-dimensional, simple random walk; see Figure 3.

3. *Let $\kappa(.,.)$ be a stochastic kernel (see Definition 5.9) from $(E, \mathcal{B}(E))$ to $(E, \mathcal{B}(E))$. Further, let X_0 be an E-valued random variable and given X_t, X_{t+1} has the distribution $\kappa(X_t, .)$, $t = 0, 1, 2, ...$ Then $(X_t)_{t=0,1,...}$ is called an E-valued* Markov chain.

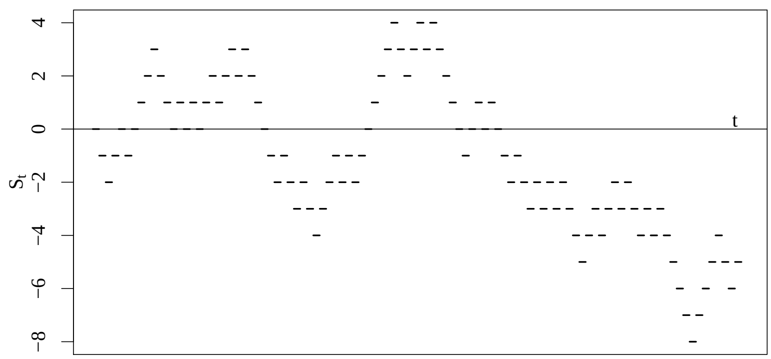


Figure 3: A path of a one-dimensional random walk.

Remark 12.3 (Repetition: Existence of stochastic Processes)**.**

1. *Recall from Section 5: the product σ-algebra on the space E^I is defined as the smallest σ-algebra with respect to which all projections $\pi_t, t \in I$ are measurable. In particular, for an E-valued stochastic process $\mathcal{X} = (X_t)_{t\in I}$, the mapping $\omega \mapsto X(\omega)$ is $\mathcal{F}/\mathcal{B}(E)^I$-measurable. Furthermore, a projective family on $\mathcal{F}$ is a family of distributions $(\mathbf{P}_J)_{J\subseteq_f I}$ with $\mathbf{P}_H = (\pi_H^J)_* \mathbf{P}_J$ for $H \subseteq J$, where π_H^J is the projection of E^J onto E^H.*

2. *Often, the finite-dimensional distributions of a stochastic process $\mathcal{X} = (X_t)_{t\in I}$, i.e. the joint distribution of $(X_{t_1}, ..., X_{t_n})$ for any $t_1, ..., t_n \in I$, are given. For example, in Sections 12.2 and 12.3, the Poisson process and Brownian motion are given by specifying the joint distribution of $(X_{t_1}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}})$. This also uniquely defines the finite-dimensional distributions. In order to ensure that there is a stochastic process for these finite-dimensional distributions, we need Kolmogorov's extension theorem; see Theorem 5.24. It should be noted that finite dimensional distributions of stochastic processes are always projective; see also Example 5.22.2.*

Definition 12.4 (Equality of stochastic processes)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ and $\mathcal{Y} = (Y_t)_{t\in I}$ be two E-valued stochastic processes.*

1. *If $\mathcal{X} \stackrel{d}{=} \mathcal{Y}$, then $\mathcal{Y}$ is a* version *of $\mathcal{X}$ (and vice versa).*

2. *If $\mathcal{X}$ and $\mathcal{Y}$ are defined on the same probability space and $\mathbf{P}(X_t = Y_t) = 1$ for all $t \in I$, then $\mathcal{X}$ is called a* modification *of $\mathcal{Y}$ (and vice versa).*

3. *If $\mathcal{X}$ and $\mathcal{Y}$ are defined on the same probability space and $\mathbf{P}(X_t = Y_t \text{ for all } t \in I) = 1$, then $\mathcal{X}$ and $\mathcal{Y}$ are called* indistinguishable.

The paths $t \mapsto X_t(\omega)$ of a stochastic process can have have certain properties. For example, they can be continuous functions $I \to E$. In addition to processes with continuous paths, we will need processes with right-continuous paths and left-limits.

Definition 12.5 (Right-continuous functions, left limits)**.** *A function $f : I \to E$ is called* *right-continuous in $t \in I$ with left-sided limit value*[17] *if*

$$f(t) = \lim_{s \downarrow t} f(s) \text{ and } \lim_{s \uparrow t} f(s) \text{ exists.}$$

It is called right-continuous with left limit values if this property holds for all $t \in I$. The set of right-continuous functions with left limits is denoted by $\mathcal{D}_E(I)$.

Proposition 12.6 (Versions, modifications, indistinguishable processes)**.** *Let $\mathcal{X} = (X_t)_{t \in I}$ and $\mathcal{Y} = (Y_t)_{t \in I}$ be stochastic processes with values in E.*

1. *The process $\mathcal{Y}$ is a version of $\mathcal{X}$ (and vice versa) if both processes have the same finite-dimensional distributions, i.e. $(X_{t_1}, ..., X_{t_n}) \stackrel{d}{=} (Y_{t_1}, ..., Y_{t_n})$ for any choice of $n \in \mathbb{N}$ and $t_1, ..., t_n \in I$.*

2. *If $\mathcal{X}$ and $\mathcal{Y}$ are indistinguishable, then $\mathcal{X}$ is a modification of $\mathcal{Y}$ (or vice versa). If $\mathcal{X}$ is a modification of $\mathcal{Y}$, then $\mathcal{X}$ is a version of $\mathcal{Y}$.*

3. *If I is at most countable and $\mathcal{X}$ is a modification of $\mathcal{Y}$ (or vice versa), then $\mathcal{X}$ and $\mathcal{Y}$ are indistinguishable.*

4. *If $I = [0, \infty)$ and $\mathcal{X}$ and $\mathcal{Y}$ have almost surely right-continuous paths and $\mathcal{X}$ is a modification of $\mathcal{Y}$, then $\mathcal{X}$ and $\mathcal{Y}$ are indistinguishable.*

Proof. 1. '$\Rightarrow$': clear. '$\Leftarrow$': Let $(\Omega, \mathcal{F}, \mathbf{P})$ and $(\Omega', \mathcal{F}', \mathbf{P}')$ the probability spaces on which probability spaces on which $\mathcal{X}$ and $\mathcal{Y}$ are defined. We consider the $\cap$-stable generator

$$\mathcal{C} := \{\pi_J^{-1}(A) : A \in \mathcal{B}(E)^{|J|}; J \subseteq_f I\} \subseteq \mathcal{B}(E)^I$$

of $\mathcal{B}(E)^I$. Further, for $J \subseteq_f I$, $A \in \mathcal{B}(E)^{|J|}$,

$$\mathbf{P}((X_t)_{t \in J} \in A) = \mathbf{P}'_I((Y_t)_{t \in J} \in A),$$

i.e. $\mathcal{X}_*\mathbf{P}$ and $\mathcal{Y}_*\mathbf{P}'$ coincide on $\mathcal{C}$. According to Theorem 2.11, this means that $\mathcal{X}_*\mathbf{P} = \mathcal{Y}_*\mathbf{P}'$. So $\mathcal{Y}$ is a version of $\mathcal{X}$.
2. Let $t \in I$. If $\mathcal{X}$ and $\mathcal{Y}$ are indistinguishable, then $\mathbf{P}(X_t \neq Y_t) \leq \mathbf{P}(X_s \neq Y_s \text{ for a } s \in I) = 0$. If $\mathcal{X}$ and $\mathcal{Y}$ are modifications and $t_1, ..., t_n \in I$, then

$$\mathbf{P}(X_{t_1} = Y_{t_1}, ..., X_{t_n} = Y_{t_n}) = 1$$

since finite unions of null-sets are null-sets. In particular, $\mathcal{X}$ and $\mathcal{Y}$ have the same finite-dimensional distributions. According to 1. $\mathcal{Y}$ is therefore is a version of $\mathcal{X}$.
3. The statement is clear because of the σ-subadditivity of probability measures,

$$\mathbf{P}(X_t \neq Y_t \text{ for a } t \in I) \leq \sum_{t \in I} \mathbf{P}(X_t \neq Y_t) = 0.$$

4. Let R be a set with $\mathbf{P}(R) = 1$ such that $\mathcal{X}$ and $\mathcal{Y}$ have right-continuous paths on R and $N_t := \{X_t \neq Y_t\}$. Further, let $I' = I \cap \mathbb{Q}$. Then, $\mathbf{P}(\bigcup_{t \in I'} N_t) = 0$ and

$$\mathbf{P}\Big(\bigcup_{t \in I} N_t\Big) \leq \mathbf{P}\Big(R \cap \bigcup_{t \in I} \bigcup_{r \geq t, r \in I'} N_r\Big) = \mathbf{P}\Big(R \cap \bigcup_{r \in I'} N_r\Big) = 0.$$

□

[17] Such functions are also called rcll (right-continuous with left limits) or càdlàg (continue à droite, limite à gauche)

Remark 12.7 (Versions with different path properties)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be an E-valued stochastic process and $I = [0, \infty)$. Each path $t \mapsto X_t(\omega)$ is therefore a mapping $I \to E$. A distinction is made between stochastic processes according to their path properties. For example, if $t \mapsto X_t(\omega)$ is a continuous function for almost all ω, we say that $\mathcal{X}$ has (almost certainly) continuous paths. It is important to realize that the property of the process to have continuous paths cannot be read from its distribution:*

Let $\mathcal{Y} = (Y_t)_{t\in I}$ with $Y_t = 0$, and $T \sim \exp(1)$ and $\mathcal{X} = (X_t)_{t\in I}$ given by

$$X_t = \begin{cases} 1, & t = T, \\ 0, & \textit{otherwise.} \end{cases}$$

Then $\mathbf{P}(X_t = Y_t) = \mathbf{P}(T \neq t) = 1$ for each $t \in I$. So $\mathcal{X}$ is a modification of $\mathcal{Y}$. In particular, according to the last proposition, the distributions of $\mathcal{X}$ and $\mathcal{Y}$ coincide. However, only $\mathcal{Y}$ has continuous paths, but every path of $\mathcal{X}$ is discontinuous (at T). In particular, $\mathcal{X}$ and $\mathcal{Y}$ are not indistinguishable.

Theorem 12.8 (Continuous modifications; Kolmogorov, Chentsov)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be an E-valued stochastic process with $I = \mathbb{R}$ or $I = [0, \infty)$. For every $\tau > 0$ there are numbers $\alpha, \beta, C > 0$ with*

$$\mathbf{E}[r(X_s, X_t)^\alpha] \leq C|t - s|^{1+\beta}$$

for all $0 \leq s, t \leq \tau$. Then there is a modification $\widetilde{\mathcal{X}} = (\widetilde{X}_t)_{t\in I}$ of $\mathcal{X}$ with continuous paths. The paths are even almost surely local Hölder-continuous of any order $\gamma \in (0, \beta/\alpha)$.[18]

Proof. It is sufficient to show the statement for $I = [0, 1]$. The general case follows by dividing I into countably many intervals of length 1. We consider the set of time points

$$D_n := \{0, 1, ..., 2^n\} \cdot 2^{-n}$$

for $n = 0, 1, ...$, $D = \bigcup_{n=0}^\infty D_n$ and the random variable

$$\xi_n := \max\{r(X_s, X_t) : s, t \in D_n, |t - s| = 2^{-n}\}.$$

Let $0 < \gamma < \beta/\alpha$. Then for some $C > 0$,

$$\begin{aligned} \mathbf{E}\Big[\sum_{n=0}^\infty (2^{\gamma n}\xi_n)^\alpha\Big] &= \sum_{n=0}^\infty 2^{\alpha\gamma n}\mathbf{E}[\xi_n^\alpha] \leq \sum_{n=0}^\infty 2^{\alpha\gamma n} \sum_{s,t\in D_n, |t-s|=2^{-n}} \mathbf{E}[r(X_s, X_t)^\alpha] \\ &\leq C\sum_{n=0}^\infty 2^{\alpha\gamma n}2^n 2^{-n(1+\beta)} = C\sum_{n=0}^\infty 2^{(\alpha\gamma-\beta)n} < \infty. \end{aligned} \tag{12.1}$$

Therefore, there is a random variable C' with $\xi_n \leq C' 2^{-\gamma n}$ for all $n = 0, 1, ...$ Now let $m \in \{0, 1, ...\}$ and $r \in [2^{-m-1}, 2^{-m}] \cap D$. Then,

$$\begin{aligned} \sup\big\{r(X_s, X_t) : s, t \in D, |s - t| \leq r\big\} &= \sup_{n\geq m}\big\{r(X_s, X_t) : s, t \in D_n, |s - t| \leq r\big\} \\ &\leq 2\sum_{n\geq m} \xi_n \leq 2C' \sum_{n\geq m} 2^{-\gamma n} \leq C'' 2^{-\gamma(m-1)} \leq C'' r^\gamma. \end{aligned} \tag{12.2}$$

[18] As a reminder: a function $f : I \to E$ is locally Hölder-continuous of order γ, if for every $\tau > 0$ there is a C with $r(f(s), f(t)) \leq C|t - s|^\gamma$ for all $0 \leq s, t \leq \tau$.

for a random variable C''. It follows that almost every path on D is Hölder-continuous to the parameter γ. This means that $\mathcal{X}$ can be extended Hölder-continuously to I. We call this continuous extension $\mathcal{Y} = (Y_t)_{t\in I}$. To show that $\mathcal{Y}$ is a modification of $\mathcal{X}$, we consider a $t \in I$ and a sequence $t_1, t_2, ... \in D$ with $t_n \to t$ with $n \to \infty$. Because of the condition, $\mathbf{P}(r(X_{t_n}, X_t) > \varepsilon) \le \mathbf{E}[r(X_{t_n}, X_t)^\alpha]/\varepsilon^\alpha \xrightarrow{n\to\infty} 0$ for each $\varepsilon > 0$, i.e. $X_{t_n} \xrightarrow{n\to\infty}_p X_t$. Furthermore, due to the continuity of $\mathcal{Y}$, we find $Y_{t_n} \xrightarrow{n\to\infty}_{fs} Y_t$. In particular, $\mathbf{P}(X_t = Y_t) = 1$. This completes the proof. □

12.2 Example 1: The Poisson process

For the first time, we consider a concrete stochastic process with process with index set $I = [0, \infty)$. A path of the Poisson process is shown in Figure 4.

Remark 12.9 (Modeling by a Poisson process). *We want to model clicks of a Geiger counter, calls to a call-center, mutation events along ancestral lines, or something else which has events randomly occurring in time. We want to analyze such counting processes with the help of a stochastic process $\mathcal{X} = (X_t)_{t\in I}$ with $I = [0, \infty)$. Let X_t be the number of clicks/calls/mutations up to time t. For such a process it makes sense to make a few assumptions:*

1. Independent increments: *If $0 = t_0 < t_1 < ... < t_n$, then $(X_{t_i} - X_{t_{i-1}} : i = 1, ..., n)$ is an independent family.*
2. Identically distributed increments: *If $0 < t_1 < t_2$, then $X_{t_2} - X_{t_1} \stackrel{d}{=} X_{t_2 - t_1} - X_0$.*
3. No double-points: *It is $\limsup_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathbf{P}(X_\varepsilon - X_0 > 1) = 0$.*

Definition 12.10 (Poisson process). *A real-valued stochastic process $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ with $X_0 = 0$ is called a Poisson process with intensity λ if the following applies:*

1. *For $0 = t_0 < ... < t_n$, the family $(X_{t_i} - X_{t_{i-1}} : i = 1, ..., n)$ is independent.*
2. *For $0 \le t_1 < t_2$ is $X_{t_2} - X_{t_1} \sim Poi(\lambda(t_2 - t_1))$.*

Proposition 12.11 (Existence of Poisson processes). *Let $\lambda \ge 0$. Then there is exactly one distribution $\mathbf{P}_I$ on $(\mathcal{B}(\mathbb{R}))^I$ such that the canonical process with respect to $\mathbf{P}_I$ is a Poisson process with intensity λ.*

Proof. As for uniqueness: The finite-dimensional distributions of $\mathbf{P}_I$ as given by 1. and 2. from Definition 12.10 are uniquely defined. Therefore the uniqueness follows from Proposition 12.6.1.

For existence, we define the Poisson process as a projective limit. For $J = \{t_1, ..., t_n\} \subseteq_f I$ with $0 = t_0 < t_1 < ... < t_n$ we set for $x_0 = 0$

$$S^n : (x_1 - x_0, ..., x_n - x_{n-1}) \mapsto (x_1, ..., x_n).$$

Further,

$$\mathbf{P}_J := S^n_* \bigotimes_{i=1}^n \mathrm{Poi}(\lambda(t_i - t_{i-1})). \tag{12.3}$$

In other words: If $Y_{t_i - t_{i-1}}$ for $i = 1, ..., n$ are independently Poisson distributed with parameter $\lambda(t_i - t_{i-1})$, then $S^n(Y_{(t_1-t_0)}, ..., Y_{t_n-t_{n-1})}) \sim \mathbf{P}_J$.

We now show that the family $(\mathbf{P}_J : J \subseteq_f I)$ is projective: let $J = \{t_1, ..., t_n\}$ as above and $H = J \setminus \{t_i\}$ for one i. Then,

$$\text{Poi}(\lambda(t_{i+1} - t_i)) * \text{Poi}(\lambda(t_i - t_{i-1})) = \text{Poi}(\lambda(t_{i+1} - t_{i-1}))$$

and therefore

$$(\pi_H^J)_* \mathbf{P}_J = (\pi_H^J \circ S^n)_* \bigotimes_{j=1}^n \text{Poi}(\lambda(t_j - t_{j-1})) = \mathbf{P}_H.$$

According to Theorem 5.24, there is the projective limit $\mathbf{P}_I$. Let us consider the canonical process $\mathcal{X} = (X_t)_{t \in I}$ with respect to $\mathbf{P}_I$. It has the finite-dimensional distributions $(\mathbf{P}_J : J \subseteq_f I)$. In particular, because of (12.3) increments are independent and Poisson distributed. Thus $\mathcal{X}$ fulfills the conditions 1. and 2. from Definition 12.10. □

Proposition 12.12 (Characterization of Poisson processes)**.** *A non-decreasing stochastic process $\mathcal{X} = (X_t)_{t \in I}$ with $X_0 = 0$ and values in $\mathbb{Z}_+$ is a Poisson process with intensity λ iff $\lambda = \mathbf{E}[X_1 - X_0] < \infty$ and 1.-3. from Remark 12.9 are fulfilled.*

Proof. '⇒': 1. and 2. from remark 12.9 are clearly fulfilled. For 3. we calculate directly

$$\tfrac{1}{\varepsilon}\mathbf{P}(X_\varepsilon > 1) = \frac{1 - e^{-\lambda\varepsilon}(1 + \lambda\varepsilon)}{\varepsilon} \leq \frac{1 - (1 - \lambda\varepsilon)(1 + \lambda\varepsilon)}{\varepsilon} \xrightarrow{\varepsilon \to 0} 0.$$

'⇐': 1. from Definition 12.10 is fulfilled. It remains to show that $X_t \sim \text{Poi}(\lambda t)$. Let for $n \in \mathbb{N}, k = 1, ..., n$,

$$Z_k^n := (X_{tk/n} - X_{t(k-1)/n}) \wedge 1, \qquad X_t^n = \sum_{k=1}^n Z_k^n.$$

This means that Z_k^n indicates whether in the interval $(t(k-1)/n; tk/n]$ at least one jump has taken place. Then, since X_t^n is monotonic in n,

$$\begin{aligned}
\mathbf{P}(\lim_{n\to\infty} X_t^n \neq X_t) &= \lim_{n\to\infty} \mathbf{P}(X_t^n \neq X_t) \\
&= \lim_{n\to\infty} \mathbf{P}(X_{tk/n} - X_{t(k-1)/n} > 1 \text{ for a } k) \\
&\leq \lim_{n\to\infty} \sum_{k=1}^n \mathbf{P}(X_{tk/n} - X_{t(k-1)/n} > 1) \\
&= \lim_{n\to\infty} n\mathbf{P}(X_{t/n} > 1) \xrightarrow{n\to\infty} 0
\end{aligned}$$

from 3. Further, X_t^n is binomially distributed with n and probability of success $p_n := \mathbf{P}(X_{t/n} > 0)$. Because of the linearity of the mapping $t \mapsto \mathbf{E}[X_t]$ and, since $X_t^n \uparrow X_t$, it follows from the theorem on monotone convergence,

$$\lambda t = \mathbf{E}[X_t] = \lim_{n\to\infty} \mathbf{E}[X_t^n] = \lim_{n\to\infty} np_n.$$

By a Poisson approximation (see Example 10.1),

$$\begin{aligned}\mathbf{P}(X_t = k) &= \lim_{n\to\infty} \mathbf{P}(X_t^n = k) - \mathbf{P}(X_t^n = k; X_t \neq X_t^n) + \mathbf{P}(X_t = k; X_t \neq X_t^n)\\ &= \lim_{n\to\infty} \mathbf{P}(X_t^n = k) = \mathrm{Poi}(\lambda t)(k),\end{aligned}$$

i.e. $X_t \sim \mathrm{Poi}(\lambda t)$ and the assertion follows. □

Proposition 12.13 (Construction by exponential distributions)**.** *Let $S_1, S_2, ...$ be independent, exponentially distributed with parameter λ. Further, let $\mathcal{X} = (X_t)_{t\in I}$ be given by*

$$X_t := \max\{i : S_1 + ... + S_i < t\}$$

with $\max\emptyset = 0$. *Then $\mathcal{X}$ is a Poisson process with intensity λ.*

Proof. We must show that for $0 = t_0 < ... < t_n$, $k_1, ..., k_n \in \mathbb{N}$

$$\mathbf{P}(X_{t_1} - X_{t_0} = k_1, ..., X_{t_n} - X_{t_{n-1}} = k_n) = \prod_{j=1}^n \mathrm{Poi}(\lambda(t_j - t_{j-1}))(k_j).$$

This will only be calculated for the case $n = 2$, the general case follows analogously. In the following calculation, let $0 \leq s < t$ and $U_1, U_2, ...$ uniformly distributed random variables on $[0, t]$. We calculate

$$\begin{aligned}&\mathbf{P}(X_s - X_0 = k, X_t - X_s = \ell)\\ &\quad= \int_0^s \int_{s_1}^s \cdots \int_{s_{k-1}}^s \int_s^t \int_{s_{k+1}}^t \cdots \int_{s_{k+\ell-1}}^t \int_t^\infty \lambda^{k+\ell+1} e^{-\lambda s_1} e^{-\lambda(s_2-s_1)} \cdots\\ &\qquad\qquad \cdots e^{-\lambda(s_{k+\ell+1}-s_{k+\ell})} ds_{k+\ell+1} ... ds_1\\ &\quad= \lambda^{k+\ell} \int_0^s \int_{s_1}^s \cdots \int_{s_{k-1}}^s \int_s^t \int_{s_{k+1}}^t \cdots \int_{s_{k+\ell-1}}^t \Big(\int_t^\infty \lambda e^{-\lambda s_{k+\ell+1}} ds_{k+\ell+1}\Big) ds_{k+\ell} \cdots ds_1\\ &\quad= e^{-\lambda t}\lambda^{k+\ell} t^{k+\ell} \mathbf{P}[U_1 < ... < U_k < s < U_{k+1} < ... < U_{k+\ell}]\\ &\quad= e^{-\lambda t}\lambda^\ell t^\ell \binom{k+\ell}{k} \Big(\frac{s}{t}\Big)^k \Big(\frac{t-s}{t}\Big)^\ell \frac{1}{(k+\ell)!}\\ &\quad= e^{-\lambda s}\frac{(\lambda s)^k}{k!} \cdot e^{-\lambda(t-s)}\frac{(\lambda(t-s))^\ell}{\ell!},\end{aligned}$$

and the assertion follows. □

Example 12.14 (Left- and right-continuous Poisson process)**.** *Let, similar to Proposition 12.13, the stochastic process process $\mathcal{Y} = (Y_t)_{t\in I}$ given by*

$$Y_t := \max\{i : S_1 + ... + S_i \leq t\}.$$

Paths of the processes $\mathcal{X}$ from proposition 12.13 and $\mathcal{Y}$ can be seen in Figure 4. The two processes differ in that $\mathcal{X}$ is right-continuous and $\mathcal{Y}$ is left-continuous. However, both processes are Poisson processes with intensity λ, as you can easily see. This is because $\mathbf{P}(X_t = Y_t) = 1$ applies for all $t \in [0, \infty)$ and thus $\mathcal{Y}$ is a version of $\mathcal{X}$ according to Proposition 12.6. As you can see from this example, two processes with the same distribution can have completely different paths.

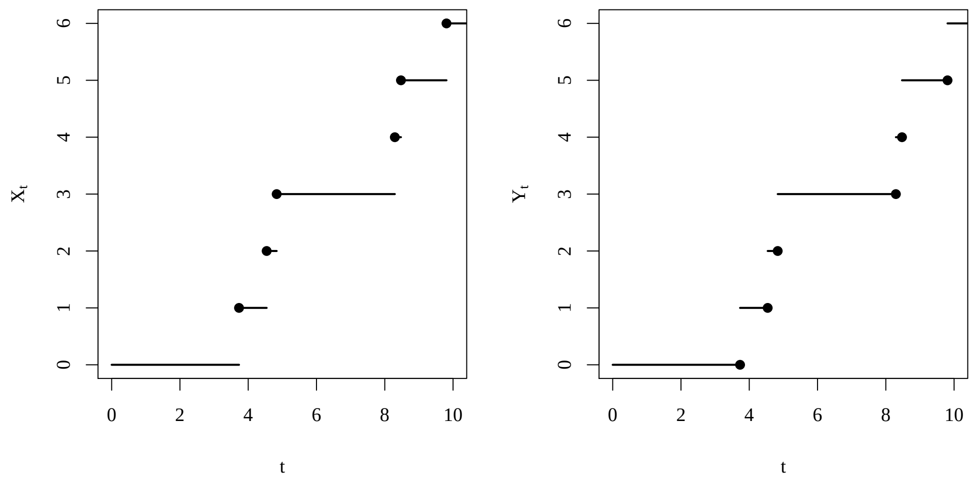


Figure 4: The right-continuous Poisson process $\mathcal{X}$ and the left-continuous Poisson process $\mathcal{Y}$.

12.3 Example 2: Brownian motion

Brownian motion is named after the botanist Robert Brown who observed in a microscope how pollen appears to move under thermal fluctuations seem to move erratically. We will give a mathematical definition for this process, that will be particularly important in stochastic analysis. Moreover, the normal distribution will play an important role in this process. A path of a one-dimensional Brownian motion can be found in Figure 5.

This section only serves to introduce Brownian motion. We will learn more about properties of Brownian motion later.

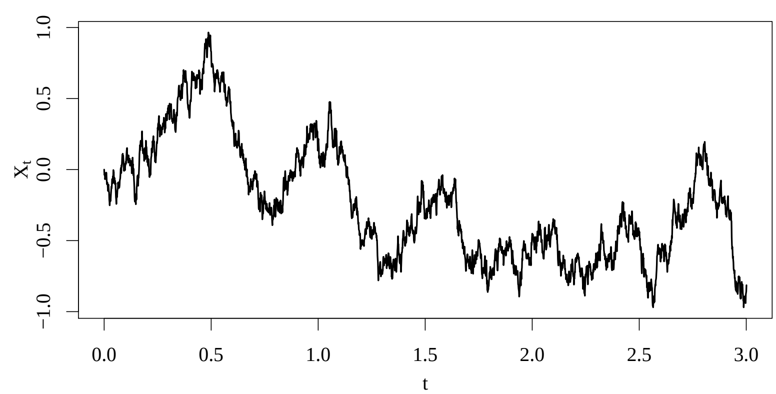


Figure 5: A path of a Brownian motion.

Definition 12.15 (Brownian motion and Gaussian processes)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process with values in $\mathbb{R}$.*

1. *The process $\mathcal{X}$ is called* Gaussian *if $c_1X_{t_1} + \cdots + c_nX_{t_n}$ for each choice of $c_1, ..., c_n \in \mathbb{R}$ and $t_1, ..., t_n \in I$ is normally distributed. For a Gaussian process, $t \mapsto \mathbb{E}[X_t]$ denotes its*

expectation and $(s,t) \mapsto \mathbf{COV}(X_s, X_t)$ its covariance structure.

2. *If $I = [0,\infty)$, then $\mathcal{X}$ is called a* Brownian motion *with start in x, if the process has continuous paths and if for each choice of $0 = t_0 < t_1 < \cdots < t_n$ it holds that $X_{t_0} = x$ and $X_{t_i} - X_{t_{i-1}}$ are independently distributed according to $N(0, t_i - t_{i-1})$, $i = 1, ..., n$. If $x = 0$, then $\mathcal{X}$ is also called* standardized *or also* Wiener process.

3. *Let $\mathcal{X}^1 = (X_t^1)_{t\in[0,\infty)}, ..., \mathcal{X}^d = (X_t^d)_{t\in[0,\infty)}$ Brownian motion. Then the $\mathbb{R}^d$-valued process $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ with $X_t = (X_t^1, ..., X_t^d)$ is called a*d-dimensional Brownian motion.

Remark 12.16 (Continuity of Brownian motion). *According to Theorem 5.24 it is clear that there is a process whose increments are normally distributed as specified in Definition 12.15.2. The specified distributions $(X_{t_1}, ..., X_{t_n})_{n\in\mathbb{N}, t_1,...,t_n \in I}$ are in fact a projective family. For example, if $X_{t_i} - X_{t_{i-1}} \sim N(0, t_i - t_{i-1})$ and $X_{t_{i-1}} - X_{t_{i-2}} \sim N(0, t_{i-1} - t_{i-2})$, then $X_{t_i} - X_{t_{i-2}} = X_{t_i} - X_{t_{i-1}} + X_{t_{i-1}} - X_{t_{i-2}} \sim N(0, t_i - t_{i-2})$ because of example 5.20. However, it is less clear whether there is there is also a process with such increments that has continuous paths has. To check this, we use the criterion from Theorem 12.8.*

Proposition 12.17 (Existence of Brownian motion). *Let $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ be a real-valued stochastic process such that for any choice of $0 = t_0 < t_1 < ... < t_n$ it holds that $X_{t_0} = x$ and $X_{t_i} - X_{t_{i-1}}$ are independent and are distributed according to $N(0, t_i - t_{i-1})$, $i = 1, ..., n$. Then there exists a modification $\mathcal{Y}$ of $\mathcal{X}$ with continuous paths. In other words, $\mathcal{Y}$ is a Brownian motion. The process $\mathcal{Y}$ is even locally Hölder continuous for every parameter $\gamma < 1/2$. Furthermore, the covariance structure of Brownian motion $\mathcal{Y}$ is given by $\mathbf{COV}(X_s, X_t) = s \wedge t$.*

Proof. Wlog let $x = 0$. The existence and uniqueness of a process with independent normally distributed increments follows as in the proof of Proposition 12.11. Since $X_s \sim N(0, s)$, $X_s \stackrel{d}{=} s^{1/2} X_1$, as can be seen, for example, from Example 6.13.3. For $a > 2$,

$$\mathbf{E}[|X_t - X_s|^a] = \mathbf{E}[|X_{t-s}|^a] = \mathbf{E}[((t-s)^{1/2}|X_1|)^a] = (t-s)^{a/2}\mathbf{E}[|X_1|^a].$$

According to Theorem 12.8, there is therefore a modification of $\mathcal{X}$ with continuous paths. According to the above calculation, these are Hölder-continuous for each parameter $\gamma \in (0, ((a/2) - 1)/a) = (0, (a-2)/(2a))$. Since a was arbitrary, the Hölder continuity follows for each $\gamma \in (0, 1/2)$.

To determine the covariance structure of $\mathcal{X}$, we calculate for $0 \le s \le t$

$$\mathbf{COV}(X_s, X_t) = \mathbf{COV}(X_s, X_s) + \mathbf{COV}(X_s, X_t - X_s) = \mathbf{V}[X_s] = s.$$

An analogous calculation for $t < s$ provides the result $\mathbf{COV}(X_s, X_t) = s \wedge t$. □

Lemma 12.18 (Characterization of Gaussian processes).
Let $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ and $\mathcal{Y} = (Y_t)_{t\in[0,\infty)}$ be Gaussian processes with $\mathbf{E}[X_t] = \mathbf{E}[Y_t]$ and $\mathbf{COV}(X_s, X_t) = \mathbf{COV}(Y_s, Y_t)$. Then $\mathcal{Y}$ is a version of $\mathcal{X}$ (and vice versa).

Proof. Let $n \in \mathbb{N}$ and $c_1, ..., c_n \in \mathbb{R}$ be arbitrary. Then, for each choice of $t_1, ..., t_n \in I$ both $Z_X := c_1 X_{t_1} + \cdots + c_n X_{t_n}$ as well as $Z_Y := c_1 Y_{t_1} + \cdots c_n Y_{t_n}$ are normally distributed. Furthermore, according to the assumption,

$$\mathbf{E}[Z_X] = c_1\mathbf{E}[X_{t_1}] + \cdots + c_n\mathbf{E}[X_{t_n}] = c_1\mathbf{E}[Y_{t_1}] + \cdots + c_n\mathbf{E}[Y_{t_n}] = \mathbf{E}[Z_Y]$$

and

$$\mathbf{V}(Z_X) = \sum_{i,j=1}^n c_i c_j \mathbf{COV}(X_{t_i}, X_{t_j}) = \sum_{i,j=1}^n c_i c_j \mathbf{COV}(Y_{t_i}, Y_{t_j}) = \mathbf{V}(Z_Y).$$

This means that $Z_X \stackrel{d}{=} Z_Y$. Since $c_1, ..., c_n$ were arbitrary, it follows from Proposition 10.17 that $(X_{t_1}, ..., X_{t_n}) \stackrel{d}{=} (Y_{t_1}, ..., Y_{t_n})$. With Proposition 12.6.1 the statement follows. □

Theorem 12.19 (Brownian scaling)**.** *Let $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ be a Brownian motion. Then, the processes $(X_{c^2t}/c)_{t\in[0,\infty)}$ are for each $c > 0$ and $(tX_{1/t})_{t\in[0,\infty)}$ also Brownian motions.*

Proof. It is clear that both $(X_{c^2t}/c)_{t\in[0,\infty)}$ and $(tX_{1/t})_{t\in[0,\infty)}$ are Gaussian processes. Furthermore,

$$\begin{aligned} \mathbf{E}[X_{c^2t}/c] &= 0, \\ \mathbf{E}[tX_{1/t}] &= 0, \end{aligned}$$

and for $s, t \geq 0$

$$\begin{aligned} \mathbf{COV}[X_{c^2s}/c, X_{c^2t}/c] &= \frac{1}{c^2}(c^2 s \wedge c^2 t) = s \wedge t, \\ \mathbf{COV}[sX_{1/s}, tX_{1/t}] &= st\Big(\frac{1}{s} \wedge \frac{1}{t}\Big) = s \wedge t. \end{aligned}$$

Now the assertion follows with Lemma 12.18. □

12.4 Filtrations and stopping times

In a stochastic process, more and more of the underlying random variables of the underlying random variables are realized as time goes by. This means that more and more information about the path of the process becomes process becomes visible. Now information is synonymous with the measurability with respect to a σ-algebra, as can be seen from the lecture *Probability Theory.* Since the information grows over over time, a stochastic process involves an increasing family of σ-algebras, which we will call a filtration in the in the following.

Definition 12.20 (Filtrations, stopping times)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be an E-valued stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$.*

1. *A family $(\mathcal{F}_t)_{t\in I}$ of σ-algebras with $\mathcal{F}_t \subseteq \mathcal{F}, t \in I$, is called* filtration *if $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$.*
2. *The filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in I}$ with $\mathcal{F}_t = \sigma(X_s : s \leq t)$ is called the filtration generated by $\mathcal{X}$.*
3. *The stochastic process $(X_t)_{t\in I}$ is called* adapted *to the filtration $(\mathcal{F}_t)_{t\in I}$ if X_t is a $\mathcal{F}_t/\mathcal{B}(E)$-measurable random variable for all $t \in I$.*

Now let $\mathcal{F} = (\mathcal{F}_t)_{t\in I}$ be a filtration.

5. *A random time is a random variable with values in $\bar{I}$ (the completion of I). A random time T is called $((\mathcal{F}_t)_{t\in I}$-)*stopping time *if $\{T \le t\} \in \mathcal{F}_t$ for all $t \in I$. If $I = [0, \infty)$, then a random time T is called $((\mathcal{F}_t)_{t\in I}$-)*option time *if $\{T < t\} \in \mathcal{F}_t$ for all $t \in I$. (In the case $I = \{0, 1, 2, ...\}$ we do not need this term.)*

6. *Each stopping time T defines the σ-algebra*

$$\mathcal{F}_T := \{A \in \mathcal{A} : A \cap \{T \le t\} \in \mathcal{F}_t, t \in I\}$$

of the T-past.

7. *For a random time T, X_T is defined by $\omega \mapsto X_{T(\omega)}(\omega)$. Further, $\mathcal{X}^T := (X_{T\wedge t})_{t\in I}$ is the process stopped at T.*

Remark 12.21 (Interpretation of the definition of stopping times)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process and $(\mathcal{F}_t)_{t\in I}$ the canonical filtration. One can view $\mathcal{F}_t$ as the information that is available at time t through knowledge of $(X_s)_{0\le s\le t}$. If T is a stopping time, then $\{T \le t\} \in \mathcal{F}_t$. So the occurrence of the event $\{T \le t\}$ can be predicted by knowledge of $(X_s)_{s\le t}$. In other words by knowing the stochastic process up to time t it can be decided whether the stopping time T has occurred by now at the latest. If T is an option time, then by knowing the stochastic process up to time t, it can be decided whether the stopping time T has already occurred in the past of t.*

Example 12.22 (Hitting times in the Poisson process)**.** *Let $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ and $\mathcal{Y} = (Y_t)_{t\in[0,\infty)}$ the right and left continuous Poisson process from example 12.14, respectively, and $(\mathcal{F}_t^{\mathcal{X}})_{t\in[0,\infty)}$ and $(\mathcal{F}_t^{\mathcal{Y}})_{t\in 0,\infty)}$ the corresponding filtrations. Further, let*

$$T_1 := \inf\{t \ge 0 : X_t = 1\} = \inf\{t \ge 0 : Y_t = 1\}$$

be the hitting time *of* 1. *(The last equality holds because the processes $\mathcal{X}$ and $\mathcal{Y}$ jump from 0 to 1 at the same time). Then:*

- *T_1 is both $(\mathcal{F}_t^{\mathcal{X}})_{t\in[0,\infty)}$-stopping time, as well as a $(\mathcal{F}_t^{\mathcal{X}})_{t\in[0,\infty)}$ option time.*

Indeed: If $T_1 = t$ is the jump time from 0 to 1, then $X_t = 1$, i.e. $\{T_1 \le t\} = \{X_t \ge 1\} \in \sigma((X_s)_{s\le t}) = \mathcal{F}_t^{\mathcal{X}}$ and $\{T_1 < t\} = \{X_{t-} \ge 1\} \in \sigma((X_s)_{s<t}) \subseteq \mathcal{F}_t^{\mathcal{X}}$.

- *T_1 is indeed an $(\mathcal{F}_t^{\mathcal{Y}})_{t\in[0,\infty)}$ option time, but not a $(\mathcal{F}_t^{\mathcal{Y}})_{t\in[0,\infty)}$ stopping time.*

Indeed: If $T_1 = t$ is the jump time from 0 to 1, then $Y_t = 0$, but $Y_{t+} = 1$, i.e. $\{T_1 \le t\} = \{X_{t+} \ge 1\} \in \sigma((Y_s)_{s\le t+h})$ for every $h > 0$, but not $\{T_1 \le t\} \in \mathcal{F}_t^{\mathcal{Y}}$. However, still $\{T_1 < t\} = \{Y_t \ge 1\} \in \sigma((Y_s)_{s\le t}) \subseteq \mathcal{F}_t^{\mathcal{Y}}$.

Lemma 12.23 (Simple properties of stopping times)**.** *Let $(\mathcal{F}_t)_{t\in I}$ be a filtration.*

1. *Each time $T = s \in I$ is a stopping time*

2. *For stopping times S, T, the times $S \wedge T$ and $S \vee T$ are also stopping times.*

3. *For stopping times $S, T \ge 0$, $S + T$ is a stopping time.*

4. *Each stopping time T is $\mathcal{F}_T$ measurable.*

5. *For stopping times S, T with $S \le T$ is $\mathcal{F}_S \subseteq \mathcal{F}_T$.*

Proof. 1. for $t \in I$ is $\{s \le t\} \in \{\emptyset, \Omega\} \subseteq \mathcal{F}_t$, i.e. $T = s$ is a stopping time.
2. for $t \in I$ is $\{S \wedge T \le t\} = \{S \le t\} \cup \{T \le t\} \in \mathcal{F}_t$ and $\{S \vee T \le t\} = \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_t$.
3. Let $t \in I$. There are $S \wedge t$ and $T \wedge t$ stopping times, i.e. for $s \le t$ is $\{S \wedge t \le s\} \in \mathcal{F}_s \subseteq \mathcal{F}_t$. For $s > t$, $\{S \wedge t \le s\} = \Omega \in \mathcal{F}_t$, i.e. $S \wedge t$ is $\mathcal{F}_t$-measurable. Analogously, it follows that $T \wedge t$ is $\mathcal{F}_t$-measurable. Furthermore, $1_{\{S>t\}}, 1_{\{T>t\}}$ $\mathcal{F}_t$-measurable. If we set $S' = S \wedge t + 1_{\{S>t\}}, T' = T \wedge t + 1_{\{T>t\}}$, then $S' + T'$ is $\mathcal{F}_t$-measurable and $\{S + T \le t\} = \{S' + T' \le t\} \in \mathcal{F}_t$.
4. Since T is a stopping time, $\{T \le t\} \in \mathcal{F}_t$. According to the definition of $\mathcal{F}_T$ this means $\{T \le t\} \in \mathcal{F}_T$. Since $\mathcal{H} := \{(-\infty; t] : t \in \mathbb{R}\}$ is a generator of $\mathcal{B}(\mathbb{R})$, so the assertion follows.
5. Let $A \in \mathcal{F}_S$ and $t \in I$. Since $B := A \cap \{S \le t\} \in \mathcal{F}_t$,

$$A \cap \{T \le t\} = B \cap \{T \le t\} \in \mathcal{F}_t,$$

i.e. $A \in \mathcal{F}_T$. □

Definition 12.24 (Continuous and complete filtration). 1. *Let $(\mathcal{F}_t)_{t \in [0,\infty)}$ be a filtration. We define $(\mathcal{F}_t^+)_{t \in [0,\infty)}$ by $\mathcal{F}_t^+ := \bigcap_{s>t} \mathcal{F}_s$. Further, $(\mathcal{F}_t)_{t \in [0,\infty)}$ is* continuous *if $\mathcal{F}_t^+ = \mathcal{F}_t$.*

2. *Let $\mathcal{N} = \{A :$ there exists a $N \supseteq A$ with $N \in \mathcal{F}$ and $\mathbf{P}(N) = 0\}$. Then, the filtration $(\mathcal{F}_t)_{t \in I}$ is called* complete *if $\mathcal{N} \subseteq \mathcal{F}_t$ for each $t \in I$.*

Lemma 12.25 (Usual completion of a filtration). *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $(\mathcal{F}_t)_{t \in [0,\infty)}$ a filtration and $\mathcal{N}$ as in Definition 12.24. Then there is a smallest continuous and complete filtration $(\mathcal{G}_t)_{t \in [0,\infty)}$ with $\mathcal{F}_t \subseteq \mathcal{G}_t, t \in [0,\infty)$. This is given by*

$$\mathcal{G}_t = \sigma(\mathcal{F}_t^+, \mathcal{N}).$$

Furthermore, $\sigma(\mathcal{F}_t^+, \mathcal{N}) = \sigma(\mathcal{F}_t, \mathcal{N})^+$.

Proof. First we show the last equation. It is clear that

$$\sigma(\mathcal{F}_t^+, \mathcal{N}) \subseteq \sigma(\sigma(\mathcal{F}_t, \mathcal{N})^+, \mathcal{N}) = \sigma(\mathcal{F}_t, \mathcal{N})^+.$$

Conversely, let $A \in \sigma(\mathcal{F}_t, \mathcal{N})^+$. Then, $A \in \sigma(\mathcal{F}_{t+h}, \mathcal{N})$ for all $h > 0$. So there is an $A_h \in \mathcal{F}_{t+h}$ with $\mathbf{P}((A \setminus A_h) \cup (A_h \setminus A)) = 0$. Now choose $h_1, h_2, ...$ with $h_n \downarrow 0$ and

$$A' = \{A_{h_n} \text{ infinitely often}\}.$$

Then, obviously, $A' \in \mathcal{F}_t^+$ and $\mathbf{P}((A \setminus A') \cup (A' \setminus A)) = 0$, i.e. $A \in \sigma(\mathcal{F}_t^+, \mathcal{N})$. From this follows $\sigma(\mathcal{F}_t, \mathcal{N})^+ \subseteq \sigma(\mathcal{F}_t^+, \mathcal{N})$.

To prove the minimality of $(\mathcal{G}_t)_{t \in [0,\infty)}$ let $(\mathcal{H}_t)_{t \in [0,\infty)}$ be another right-continuous complete extension of $(\mathcal{F}_t)_{t \in [0,\infty)}$. Then,

$$\mathcal{G}_t = \sigma(\mathcal{F}_t^+, \mathcal{N}) \subseteq \sigma(\mathcal{H}_t, \mathcal{N}) = \mathcal{H}_t$$

for all $t \in [0,\infty)$. □

Lemma 12.26 (Option and stopping times). *Let $(\mathcal{F}_t)_{t \in [0,\infty)}$ be a filtration. A random time T is an $(\mathcal{F}_t)_{t \in [0,\infty)}$ option time iff T is a $(\mathcal{F}_t^+)_{t \in [0,\infty)}$-stopping time. In this case,*

$$\mathcal{F}_T^+ = \{A \in \mathcal{F} : A \cap \{T < t\} \in \mathcal{F}_t, t > 0\}.$$

In particular, if $(\mathcal{F}_t)_{t \in [0,\infty)}$ is continuous, then every random time is a $(\mathcal{F}_t)_{t \in [0,\infty)}$-stopping time if it is a $(\mathcal{F}_t)_{t \in [0,\infty)}$ option time.

Proof. First,

$$\{T \leq t\} = \bigcap_{\mathbb{Q} \ni s > t} \{T < s\}, \qquad \{T < t\} = \bigcup_{\mathbb{Q} \ni s < t} \{T \leq s\}.$$

If T is a $(\mathcal{F}_t^+)_{t \in [0,\infty)}$-stopping time and $A \cap \{T \leq t\} \in \mathcal{F}_t^+$. Then,

$$A \cap \{T < t\} = \bigcup_{\mathbb{Q} \ni s < t} (A \cap \{T \leq s\}) \in \mathcal{F}_t.$$

If, on the other hand, $A \cap \{T < t\} \in \mathcal{F}_t$, then

$$A \cap \{T \leq t\} = \bigcap_{h>0} \bigcap_{t<s<t+h} (A \cap \{T < s\}) \in \bigcap_{h>0} \mathcal{F}_{t+h} = \mathcal{F}_t^+.$$

If you set $A = \Omega$ in the last two equations, the first assertion follows. For general A the second also follows. □

Lemma 12.27 (Suprema and infima of stopping times)**.** *Let $T_1, T_2, ...$ be random times and $(\mathcal{F}_t)_{t \in I}$ a filtration. Then the following applies:*

1. *If $T_1, T_2, ...$ are stopping times, then $T := \sup_n T_n$ is also a stopping time.*
2. *If $I = \{0, 1, 2, ...\}$ and $T_1, T_2, ...$ are stopping times, then $T := \inf_n T_n$ is also a stopping time.*
3. *If $I = [0, \infty)$ and $T_1, T_2, ...$ are option times, then $T := \inf_n T_n$ is also an option time. In addition, $\mathcal{F}_T^+ = \bigcap_n \mathcal{F}_{T_n}^+$.*

Proof. 1. We have $\{T \leq t\} = \bigcap_n \{T_n \leq t\} \in \mathcal{F}_t$ and the assertion follows.
2. It holds $\{T \leq t\} = \bigcup_n \{T_n \leq t\} \in \mathcal{F}_t$, from which the assertion follows.
3. Here, $\{T < t\} = \bigcup_n \{T_n < t\} \in \mathcal{F}_t$. Since $T \leq T_n$, $\mathcal{F}_T^+ \subseteq \bigcap_n \mathcal{F}_{T_n}^+$ according to Lemma 12.23.5. If, on the other hand, $A \in \bigcap_n \mathcal{F}_{T_n}^+$, then

$$A \cap \{T < t\} = A \cap \bigcup_n \{T_n < t\} = \bigcup_n (A \cap \{T_n < t\}) \in \mathcal{F}_t.$$

Thus $A \in \mathcal{F}_T^+$. □

Proposition 12.28 (Approximation by countable stopping times)**.** *If $I = [0, \infty)$, each option time T can be replaced by a sequence of stopping times $T_1, T_2, ...$, such that T_n only assumes values in a discrete (in particular countable) quantity and $T_n \downarrow T$.*

Proof. We define $T_n = 2^{-n}[2^n T + 1]$. Then $T_1, T_2, ...$ is a sequence decreasing towards T, where T_n only contains the values $\{1, 2, ...\} \cdot 2^{-n}$, $n = 1, 2, ...$ Further, $\{T_n \leq k2^{-n}\} = \{T < k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$, so T_n is a stopping time, $n = 1, 2, ...$ □

Definition 12.29 (Hitting time)**.** *Let $B \in \mathcal{B}(E)$. Then the* hitting time of B *is given by*

$$T_B := \inf\{t : X_t \in B\}.$$

To find out whether the hitting time T_B is a stopping (or option) time, the following result is important.

Proposition 12.30 (Hitting times as option and stopping times)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be an E-valued process that is adapted with respect to a filtration $(\mathcal{F}_t)_{t\in I}$. Then, for $B \in \mathcal{B}(E)$:*

1. *If $I = \{0, 1, 2, ...\}$, then the time T_B is a stopping time.*
2. *If $I = [0, \infty)$, B is open and $\mathcal{X}$ has right-continuous paths, then T_B is an option time.*
3. *If $I = [0, \infty)$, B is closed and $\mathcal{X}$ has continuous paths, then T_B is a stopping time.*

Proof. 1. Here,

$$\{T_B \le t\} = \bigcup_{s\le t}\{X_s \in B\} \in \mathcal{F}_t.$$

For 2. we write

$$\{T_B < t\} = \bigcup_{\mathbb{Q}\ni s<t} \{X_s \in B\} \in \mathcal{F}_t.$$

For 3. with $B_n := \{x \in E : r(x, B) < 1/n\}$

$$\{T_B \le t\} = \bigcap_n \{T_{B_n} \le t\} = \bigcap_n (\{T_{B_n} < t\} \cup \{X_t \in \overline{B_n}\}) \in \mathcal{F}_t.$$

This shows all assertions. □

12.5 Progressive measurability

By definition, for a stochastic process $\mathcal{X} = (X_t)_{t\in I}$, each of the variables X_t is measurable, $t \in I$. However, it is (still) unclear when exactly for a random time T the quantity $X_T : \omega \mapsto X_{T(\omega)}(\omega)$ is measurable and therefore a random variable. For this we need a stronger measurability concept for the process $\mathcal{X}$.

Definition 12.31 (Progressive measurability)**.** *Let $(\mathcal{F}_t)_{t\in I}$ be a filtration and $\mathcal{X} = (X_t)_{t\in I}$ a stochastic process adapted to it. Then $\mathcal{X}$ is called* progressively measurable with respect to $(\mathcal{F}_t)_{t\in I}$*, if for all $t \in I$ the mapping*

$$\begin{cases} I \cap [0, t] \times \Omega & \to E \\ (s, \omega) & \mapsto X_s(\omega) \end{cases}$$

is measurable with respect to $I \cap \mathcal{B}([0, t]) \otimes \mathcal{F}_t/\mathcal{B}(E)$.

Lemma 12.32 (Right-continuous paths and progressive measurability)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process adapted to the filtration $(\mathcal{F}_t)_{t\in I}$. If either I is countable, or $\mathcal{X}$ has right-continuous paths, then $\mathcal{X}$ is progressively measurable with respect to $(\mathcal{F}_t)_{t\in I}$.*

Proof. Let $t \in I$. We consider the mapping

$$Y : \begin{cases} I \cap [0, t] \times \Omega & \to E \\ (s, \omega) & \mapsto X_s(\omega). \end{cases}$$

First, let I be countable and $B \in \mathcal{B}(E)$. Then,

$$Y^{-1}(B) = \bigcup_{s\in I, s\le t} \{s\} \times X_s^{-1}(B) \in \mathcal{B}(I \cap [0, t]) \otimes \mathcal{F}_t.$$

Next, let I be uncountable and let $\mathcal{X}$ have right-continuous paths. Consider the processes $\mathcal{X}^n = (X_s^n)_{t\in I\cap[0,t]}, n = 1, 2, ...$ with $X_s^n := X_{(2^{-n}\lceil 2^n s\rceil)\wedge t}$ and the corresponding mappings Y_n. Due to the right continuity of the paths, $Y_n \xrightarrow{n\to\infty}_{as} Y$. Furthermore

$$\begin{aligned} Y_n^{-1}(B) &= \bigcup_{k:(k+1)2^{-n}\leq t} [k2^{-n}, (k+1)2^{-n}) \times X^{-1}_{(k+1)2^{-n}}(B) \cup [2^{-n}\lfloor 2^n t\rfloor, t] \times X_t^{-1}(B) \\ &\in \mathcal{B}([0,t]) \otimes \mathcal{F}_t. \end{aligned}$$

□

Proposition 12.33 (Measurability of X_T)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be adapted to the filtration $(\mathcal{F}_t)_{t\in I}$, progressively measurable, and T a $(\mathcal{F}_t)_{t\in I}$ stopping time. Then*

$$X_T : \begin{cases} \{T < \infty\} & \to E \\ \omega & \mapsto X_{T(\omega)}(\omega) \end{cases}$$

is measurable with respect to $\{T < \infty\} \cap \mathcal{F}_T/\mathcal{B}(E)$.

Proof. We have to show that $\{X_T \in B, T \leq t\} \in \mathcal{F}_t$ for $B \in \mathcal{B}(E)$ holds, $t \in I$. By definition of $\mathcal{F}_T$, it then holds that $\{X_T \in B\} \in \mathcal{F}_T$, from which the assertion follows. However, since $\{X_T \in B, T \leq t\} = \{X_{T\wedge t} \in B, T \leq t\}$, it suffices to show that $X_{T\wedge t}$ is measurable with respect to $\mathcal{F}_t$, $t \in I$. We can therefore wlog assume that $T \leq t$ applies. We write $X_T = Y_t \circ \psi$, where $\psi(\omega) := (T(\omega), \omega)$ is measurable with respect to $\mathcal{F}_t/(I\cap\mathcal{B}([0,t])\otimes\mathcal{F}_t)$ and $Y_t(s,\omega) = X_s(\omega)$ according to condition $I \cap \mathcal{B}([0,t]) \otimes \mathcal{F}_t/\mathcal{B}(E)$-measurable. The assertion now follows with Lemma 3.6.2. □

13 Martingales

We now begin to deal with a particular class of stochastic processes, martingales. They are often referred to as *fair games*. Simply put, a martingale is a real-valued stochastic process whose increments vanish on average.

13.1 Introduction

Throughout the section, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, (E, r) a complete and separable metric space and $I \subseteq \mathbb{R}$ an ordered index set (usually $I = \{0, 1, 2, ...\}$ or $I = \mathbb{R}_+$). In addition, let a filtration $(\mathcal{F}_t)_{t\in I}$ be given. Adaptedness of a stochastic process is always with respect to $(\mathcal{F}_t)_{t\in I}$.

Example 13.1 (A simple martingale)**.** *For a $\mathcal{F}$-measurable random variable X one can define a stochastic process, namely $\mathcal{X} = (X_t)_{t\in I}$ with*

$$X_t := \mathbf{E}[X|\mathcal{F}_t]. \tag{$*$}$$

Of course, because of Theorem 11.2.7,

$$\mathbf{E}[X_t|\mathcal{F}_s] = \mathbf{E}[\mathbf{E}[X|\mathcal{F}_t]|\mathcal{F}_s] = \mathbf{E}[X|\mathcal{F}_s] = X_s.$$

Stochastic processes $\mathcal{X}$ with this property will be called martingales. In Section 13.4, we will then (among other things) deal with when a martingale $(X_t)_{t\in I}$ is associated with a random variable X so that $(*)$ applies.

Definition 13.2 ((Sub-, Super-)martingale). *Let $\mathcal{X} = (X_t)_{t\in I}$ be an adapted, real-valued stochastic process with $\mathbf{E}[|X_t|] < \infty$, $t \in I$. Then $\mathcal{X}$ is called*

$$\begin{aligned} \textit{martingale, if } \mathbf{E}[X_t|\mathcal{F}_s] = X_s \textit{ for } s,t \in I,\ s < t,\\ \textit{sub-martingale, if } \mathbf{E}[X_t|\mathcal{F}_s] \geq X_s \textit{ for } s,t \in I,\ s < t,\\ \textit{super-martingale, if } \mathbf{E}[X_t|\mathcal{F}_s] \leq X_s \textit{ for } s,t \in I,\ s < t. \end{aligned}$$

More precisely, we say that $\mathcal{X}$ is a $(\mathcal{F}_t)_{t\in I}$-(sub, super)-martingale.

Remark 13.3 (Martingale property with discrete index set). *If I is discrete, for example $I = \{0, 1, 2, ...\}$, then a real-valued stochastic process $\mathcal{X} = (X_t)_{t\in I}$ is a martingale iff $\mathbf{E}[|X_t|] < \infty$, $t \in I$ and $\mathbf{E}[X_t|\mathcal{F}_{t-1}] = X_{t-1}$ for all $t = 1, 2, ...$ Then, for $s, t \in I, s \leq t$,*

$$\mathbf{E}[X_t|\mathcal{F}_s] = \mathbf{E}[\cdots \mathbf{E}[\mathbf{E}[X_t|\mathcal{F}_{t-1}]|\mathcal{F}_{t-2}]\cdots \mathcal{F}_s] = X_s$$

according to theorem 11.2.7 The same holds to sub- and super martingales.

Example 13.4 (Sums and products of integrable random variables).

1. *Let $X_1, X_2, ...$ be a sequence of independent, integrable random variables with $\mathbf{E}[X_i] = 0, i = 1, 2, ...$ and $\mathcal{F}_t := \sigma(X_1, ..., X_t)$. Further, let $S_0 := 0$ and for $t = 1, 2, ...$*

$$S_t := \sum_{i=1}^{t} X_i.$$

   Then,

$$\mathbf{E}[S_t|\mathcal{F}_{t-1}] = \mathbf{E}[S_{t-1} + X_t|\mathcal{F}_{t-1}] = S_{t-1} + \mathbf{E}[X_t|\mathcal{F}_{t-1}] = S_{t-1} + \mathbf{E}[X_t] = S_{t-1},$$

   i.e. $(S_t)_{t=0,1,2,...}$ is a martingale.

   If $\mathbf{E}[X_i] \geq 0$ for all $i = 1, 2, ...,$ then $(S_t)_{t\geq 0}$ is a sub-martingale.

2. *Let $I = \{-1, -2, ...\}$ and $X_1, X_2, ...$ be integrable, independent, identically distributed random variables. Further, we set for $t \in I$*

$$S_t := \frac{1}{|t|}\sum_{i=1}^{|t|} X_i$$

   and $\mathcal{F}_t := \sigma(..., S_{t-1}, S_t)$. Then for $t \in I$,

$$\begin{aligned} \mathbf{E}[S_t|\mathcal{F}_{t-1}] &= \mathbf{E}\Big[\frac{1}{|t|}\sum_{i=1}^{|t|} X_i\Big|S_{t-1}, S_{t-2}, ...\Big] \\ &= \frac{1}{|t|}\sum_{i=1}^{|t|} \mathbf{E}\Big[X_i\Big|\sum_{i=1}^{|t|+1} X_i\Big] \\ &= \mathbf{E}\Big[X_1\Big|\sum_{i=1}^{|t|+1} X_i\Big] \\ &= \frac{1}{|t-1|}\sum_{i=1}^{|t-1|} X_i \\ &= S_{t-1} \end{aligned}$$

according to Example 11.9. Specifically,

$$\begin{aligned}\mathbf{E}[X_1|\mathcal{F}_t] &= \mathbf{E}\Big[X_1\Big|\sum_{i=1}^{|t|} X_i\Big]\\ &= \frac{1}{|t|}\sum_{i=1}^{|t|} X_i\\ &= S_t.\end{aligned}$$

3. *Let $I = \{0,1,2,...\}$ and $X_1, X_2, ...$ be a sequence of independent, integrable random variables with $\mathbf{E}[X_i] = 1, i = 1,2,...$ and $\mathcal{F}_t := \sigma(X_1,...,X_t)$. Further, $S_0 := 1$ and for $t = 1,2,...$*

$$S_t := \prod_{i=1}^{t} X_i.$$

Then, $S_1, S_2, ...$ are integrable and

$$\mathbf{E}[S_t|\mathcal{F}_{t-1}] = \mathbf{E}[S_{t-1}X_t|\mathcal{F}_{t-1}] = S_{t-1}\cdot\mathbf{E}[X_t|\mathcal{F}_{t-1}] = S_{t-1}\cdot\mathbf{E}[X_t] = S_{t-1},$$

i.e. $(S_t)_{t\in I}$ is a martingale.

If $\mathbf{E}[X_i] \geq 1$ for all $i = 1,2,...$, then $(S_t)_{t\in I}$ is a sub-martingale.

Example 13.5 (Branching processes in discrete time)**.** *We consider a simple model for a randomly evolving population. evolving population. Let $X_i^{(t)}$) be independent, $\{0,1,2,...\}$-valued random variable and $\mu = \mathbf{E}[X_i^{(t)}]$. Here, $X_i^{(t)}$ stands for the number of offspring of the ith individual of generation t with $i,t = 0,1,...,t = 1,2,....$ Starting with $Z_0 = k$ we set*

$$Z_{t+1} = \sum_{i=1}^{Z_t} X_i^{(t)},$$

so $\mathcal{Z} = (Z_t)_{t=0,1,2,...}$ is the stochastic process of the total process of the total number of individuals. The distribution of $X_i^{(t)}$ is also called the offspring distribution.

The process $\mathcal{Z}$ a (non-negative) martingale (with respect to the filtration generated by $\mathcal{Z}$), iff $\mathbf{E}[X_i^{(t)}] = 1$, i.e. each individual has on average has one offspring. Then, for $t = 1,2,...$

$$\mathbf{E}[Z_{t+1} - Z_t|\mathcal{F}_t] = \mathbf{E}\Big[\sum_{i=1}^{Z_t} X_i^{(t)} - Z_t|\mathcal{F}_t\Big] = (\mu-1)Z_t.$$

If $\mu > 1$, $\mathcal{Z}$ is a sub-martingale, and if $\mu < 1$, $\mathcal{Z}$ is a super-martingale. Also, we call[19]*.*

$$\begin{aligned}&\mathcal{Z} \text{ a critical branching process if } \mu = 1,\\ &\mathcal{Z} \text{ a super-critical branching process if } \mu > 1,\\ &\mathcal{Z} \text{ a sub-critical branching process if } \mu < 1.\end{aligned}$$

[19]It may seem irritating that a *supercritical branching process* is a sub-martingale and a *subcritical branching process* is a super martingale

In general, $(Z_t/\mu^t)_{t=0,1,2,...}$ is a (non-negative) martingale, because just like in the last calculation,

$$\mathbf{E}[Z_{t+1} - \mu Z_t | \mathcal{F}_t] = \mu Z_t - \mu Z_t = 0.$$

It is also worth noting that $\mathbf{E}[Z_{t+1}|\mathcal{F}_t] = \mu Z_t$, from which one can recursively conclude that

$$\mathbf{E}[Z_t] = \mu^t.$$

□

Example 13.6 (Martingales derived from Markov chains)**.** *With (discrete-time) Markov chains, we have already described a fairly simple dependency structure between random variables. Let $I = \{0, 1, 2, ...\}$, E at most countable and $P = (p_{xy})_{x,y \in E}$ a stochastic matrix, i.e.*

$$p_{xy} \geq 0, \qquad \sum_{z \in E} p_{xz} = 1$$

for all $x, y \in E$. We find for $f : E \to \mathbb{R}$ bounded and all $s = 0, 1, 2, ...,$

$$\mathbf{E}[f(X_{s+1}) - f(X_s)|\mathcal{F}_s] = \mathbf{E}[f(X_{s+1}) - f(X_s)|X_s] = \sum_{x \in E} p_{X_s,y}(f(y) - f(X_s)).$$

Therefore, setting $\mathcal{M} = (M_t)_{t=0,1,2,...}$ with

$$M_t = f(X_t) - \sum_{s=1}^{t-1} \mathbf{E}[f(X_{s+1}) - f(X_s)|X_s],$$

we have

$$\mathbf{E}[M_t - M_{t-1}|\mathcal{F}_{t-1}] = \mathbf{E}[f(X_t) - f(X_{t-1})|\mathcal{F}_{t-1}] - \mathbf{E}[f(X_t) - f(X_{t-1})|X_{t-1}] = 0.$$

In other words, $\mathcal{M}$ is a martingale.

We conclude this section with a simple statement on how to obtain further sub-martingales from known (sub)-martingales.

Proposition 13.7 (Convex functions of martingales are sub-martingales)**.** *Let $\mathcal{X} = (X_t)_{t \in I}$ be a stochastic process and $\varphi : \mathbb{R} \to \mathbb{R}$ convex. If $\varphi(X) = (\varphi(X_t))_{t \in I}$ is integrable and one of the two conditions*

1. *$\mathcal{X}$ is a martingale*
2. *$\mathcal{X}$ is a sub-martingale and varphi is non-decreasing*

is satisfied, then $\varphi(\mathcal{X}) = (\varphi(X_t))_{t \in I}$ is a sub-martingale.

Proof. If $\mathcal{X}$ is a martingale, then $\varphi(X_s) = \varphi(\mathbf{E}[X_t|\mathcal{F}_s])$. If $\mathcal{X}$ is a sub-martingale and φ is non-decreasing, $\varphi(X_s) \leq \varphi(\mathbf{E}[X_t|\mathcal{F}_s])$. In both cases, for $s \leq t$ because of Jensen's inequality for conditional expectations, Proposition 11.4,

$$\varphi(X_s) \leq \varphi(\mathbf{E}[X_t|\mathcal{F}_s]) \leq \mathbf{E}[\varphi(X_t)|\mathcal{F}_s],$$

i.e. $\varphi(\mathcal{X})$ is a sub-martingale. □

13.2 The stochastic integral as a martingale

In this section, let always $I = \{0, 1, 2, ...\}$ (whereby all results can be transferred to a discrete index set $I = \{t_0, t_1, ...\}$ with $t_0 < t_1 < ...$). All concepts introduced here have an analogue for processes in continuous time. However, the statements are much more complex to formulate and prove in that case. Some of these analogous statements are first formulated in the lecture *Stochastic analysis.*

Definition 13.8 (Previsible process). *A stochastic process $\mathcal{X}$ is called $(\mathcal{F}_t)_{t\in I}$-*previsible *if $X_0 = 0$ and X_t is $\mathcal{F}_{t-1}$-measurable, $t = 1, 2, ...$*

Proposition 13.9 (Doob decomposition). *Let $I = \{0, 1, 2, ...\}$. Each adapted process $\mathcal{X} = (X_t)_{t\in I}$ has an almost surely unique decomposition $\mathcal{X} = \mathcal{M} + \mathcal{A}$, where $\mathcal{M}$ is a martingale and $\mathcal{A}$ is previsible. In particular, $\mathcal{X}$ is a sub-martingale iff $\mathcal{A}$ almost surely non-decreasing.*

Proof. Define the previsible process $\mathcal{A} = (A_t)_{t\in I}$ by

$$A_t = \sum_{s=1}^{t} \mathbf{E}[X_s - X_{s-1}|\mathcal{F}_{s-1}]. \tag{13.1}$$

Then $\mathcal{M} = \mathcal{X} - \mathcal{A}$ is a martingale, because

$$\mathbf{E}[M_t - M_{t-1}|\mathcal{F}_{t-1}] = \mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}] - (A_t - A_{t-1}) = 0.$$

Now we come to the uniqueness of the representation. If $\mathcal{X} = \mathcal{M} + \mathcal{A}$ for a martingale $\mathcal{M}$ and a previsible process $\mathcal{A}$, then $A_t - A_{t-1} = \mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}]$ for all $t = 1, 2, ...$, i.e. (13.1) is almost surely true. □

Definition 13.10 (Quadratic variation, increasing process). *Let $I = \{0, 1, 2, ...\}$ and $\mathcal{X} = (X_t)_{t\in I}$ be a square integrable martingale. The almost surely uniquely determined, previsible process $(\langle\mathcal{X}\rangle_t)_{t\in I}$, for which $(X_t^2 - \langle\mathcal{X}\rangle_t)_{t\in I}$ is a is a martingale, is the* quadratic variation process *(or also the* increasing process*) of $\mathcal{X}$.*

Proposition 13.11 (Increasing process and variance). *Let $I = \{0, 1, 2, ...\}$, $\mathcal{X} = (X_t)_{t\in I}$ be a martingale with quadratic variation process $\langle\mathcal{X}\rangle = (\langle\mathcal{X}\rangle_t)_{t\in I}$. Then*

$$\langle\mathcal{X}\rangle_t = \sum_{s=1}^{t} \mathbf{E}[X_s^2 - X_{s-1}^2|\mathcal{F}_{s-1}] = \sum_{s=1}^{t} \mathbf{E}[(X_s - X_{s-1})^2|\mathcal{F}_{s-1}]$$

and

$$\mathbf{E}[\langle X\rangle_t] = \mathbf{V}[X_t - X_0].$$

Proof. As in the proof of Proposition 13.9, the process $\langle\mathcal{X}\rangle$ using (13.1) can be written. This immediately results in the first equals sign. The second follows, since $\mathbf{E}[X_s X_{s-1}|\mathcal{F}_{s-1}] = X_{s-1}^2$. Further is

$$\mathbf{E}[\langle\mathcal{X}\rangle_t] = \sum_{s=1}^{t} \mathbf{E}[X_s^2 - X_{s-1}^2] = \mathbf{E}[X_t^2 - X_0^2] = \mathbf{E}[(X_t - X_0)^2] = \mathbf{V}[X_t - X_0].$$

□

Example 13.12 (Increasing processes). *1. Let $\mathcal{S} = (S_t)_{\in I}$ with $S_t = \sum_{s=1}^t X_s$ as in example 13.4.1 with square integrable random variables $X_1, X_2, ...$ Then, with Proposition 13.11*

$$\langle \mathcal{S} \rangle_t = \sum_{s=1}^t \mathbf{E}[X_s^2].$$

In particular, the quadratic variation process of $\mathcal{S}$ is deterministic.

2. *Let $\mathcal{S} = (S_t)_{\in I}$ with $S_t = \prod_{s=1}^t X_s$ as in Example 13.4.3 with square integrable random variables $X_1, X_2, ...$ Then*

$$\langle \mathcal{S} \rangle_t = \sum_{s=1}^t \mathbf{E}[(S_s - S_{s-1})^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t S_{s-1}^2 \mathbf{E}[(X_s - 1)^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t S_{s-1}^2 \mathbf{V}[X_s].$$

In particular, in this example the process $\langle \mathcal{S} \rangle$ is truly stochastic.

3. *Let $I = [0, \infty)$ and $(X_t)_{t \in I}$ be a Brownian motion. Even in continuous time, the increasing process $(\langle \mathcal{X} \rangle_t)_{t \in I}$ is defined such that $(X_t^2 - \langle \mathcal{X} \rangle_t)_{t \in I}$ is a martingale. According to Example 13.47, $\langle \mathcal{X} \rangle_t = t$ is a candidate for the increasing process of the Brownian motion. However, in continuous time it is more difficult to say what the equivalent of a* previsible *process should be.*

Definition 13.13 (Discrete stochastic integral). *Let $I = \{0, 1, 2, ...\}$ and $\mathcal{H} = (H_t)_{t \in I}, \mathcal{X} = (X_t)_{t \in I}$ be a stochastic processes with values in $\mathbb{R}$. If $\mathcal{X}$ is adapted and $\mathcal{H}$ is previsible, then we define the stochastic integral $\mathcal{H} \cdot \mathcal{X} = ((\mathcal{H} \cdot \mathcal{X})_t)_{t \in I}$ by*

$$(\mathcal{H} \cdot \mathcal{X})_t = \sum_{s=1}^t H_s (X_s - X_{s-1})$$

for all $t \in I$. If $\mathcal{X}$ is a martingale, then we call $\mathcal{H} \cdot \mathcal{X}$ a martingale transform of $\mathcal{X}$.

Proposition 13.14 (Stability of stochastic integrals). *Let $I = \{0, 1, 2, ...\}$ and $\mathcal{X} = (X_t)_{t \in I}$ be an adapted, real-valued process with $\mathbf{E}[|X_0|] < \infty$.*

1. *$\mathcal{X}$ is a martingale if and only if for each previsible process $\mathcal{H} = (H_t)_{t \in I}$, the stochastic integral $\mathcal{H} \cdot \mathcal{X}$ is a martingale.*

2. *$\mathcal{X}$ is a sub-martingale (super-martingale) if and only if for every previsible, non-negative process $\mathcal{H} = (H_t)_{t \in I}$ the stochastic integral $\mathcal{H} \cdot \mathcal{X}$ is a sub-martingale (super-martingale).*

Proof. 1. '⇒': We immediately write

$$\begin{aligned}\mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_{t+1} - (\mathcal{H} \cdot \mathcal{X})_t | \mathcal{F}_t] &= \mathbf{E}[H_{t+1}(X_{t+1} - X_t) | \mathcal{F}_t] \\ &= H_{t+1} \mathbf{E}[X_{t+1} - X_t | \mathcal{F}_t] \\ &= 0.\end{aligned}$$

'⇐': Let $t \in I$ and $H_s := 1_{\{s=t\}}$. Then $\mathcal{H} = (H_s)_{s \in I}$ is deterministic, in particular previsible. Since $(\mathcal{H} \cdot \mathcal{X})_{t-1} = 0$, it follows that

$$0 = \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t | \mathcal{F}_{t-1}] = \mathbf{E}[X_t - X_{t-1} | \mathcal{F}_{t-1}] = \mathbf{E}[X_t | \mathcal{F}_{t-1}] - X_{t-1}$$

From this, the assertion follows.
2. follows analogously. □

Example 13.15 (Quadratic variation for stochastic integrals). *Let $I = \{0, 1, 2, ...\}$, $\mathcal{X} = (X_t)_{t\in I}$ a martingale and $\mathcal{H} = (H_t)_{t\in I}$ previsible. Then, because of Proposition 13.11,*

$$\langle \mathcal{H}\cdot\mathcal{X}\rangle_t = \sum_{s=1}^t \mathbf{E}[((\mathcal{H}\cdot\mathcal{X})_s - (\mathcal{H}\cdot\mathcal{X})_{s-1})^2|\mathcal{F}_{s-1}] = \sum_{s=1}^t \mathbf{E}[H_s^2(X_s - X_{s-1})^2|\mathcal{F}_{s-1}]$$
$$= \sum_{s=1}^t H_s^2 \cdot \mathbf{E}[(X_s - X_{s-1})^2|\mathcal{F}_{s-1}].$$

In particular,

$$\mathbf{V}[(\mathcal{H}\cdot\mathcal{X})_t] = \sum_{s=1}^t \mathbf{E}[H_s^2 \cdot (X_s - X_{s-1})^2].$$

Example 13.16 (Payout for games). *Martingale transforms can also be interpreted as payoffs of games. Given a random variable evolves according to the adapted process $\mathcal{X} = (X_t)_{t=0,1,2,...}$. If you bet before time t with a stake H_t (based on the experience gained from $X_0, ..., X_{t-1}$) on the change in the random variable $X_t - X_{t-1}$, then $(\mathcal{H}\cdot\mathcal{X})_t$ is the profit realized up to time time t. Given the underlying process $\mathcal{X}$ is a martingale, Proposition 13.14 shows that the profit realized $\mathcal{H}\cdot\mathcal{X}$ for each strategy $\mathcal{H}$ is a martingale. In particular, the expected profit is 0.*

As an example, consider the Petersburg paradox: a fair coin is tossed infinitely often. In each round, a player places a stake of any amount. If `heads` *comes up, he loses it, if* `tails` *comes up, the stake is doubled and paid out again. The paradox consists of the following strategy: starting with a stake of 1 on the first coin toss, the player doubles his stake with every failure. If the first success comes on the t-th toss, his previous stake is $\sum_{i=1}^t 2^{i-1} = 2^t - 1$. Since the last bet was 2^{t-1}, the player gets 2^t back, so he has certainly made a profit of 1 even though the game was fair.*

To analyze this game using martingales, let $X_1, X_2, ...$ be an independent, identically distributed sequence with $\mathbf{P}(X_1 = -1) = \mathbf{P}(X_1 = 1) = \frac{1}{2}$, and $S_0 = 0, S_t = \sum_{i=1}^t X_i$. Then $\mathcal{S} = (S_t)_{t=0,1,2,...}$ is a martingale. Further, let H_t be the stake in the tth game. Therefore,

$$(\mathcal{H}\cdot\mathcal{S})_t = \sum_{i=1}^t H_i(S_i - S_{i-1}) = \sum_{i=1}^t H_i X_i$$

is the profit after the tth game. Since with $\mathcal{S}$, $\mathcal{H}\cdot\mathcal{S}$ is also a martingale, we find

$$\lim_{t\to\infty} \mathbf{E}[(\mathcal{H}\cdot\mathcal{S})_t] = \mathbf{E}[(\mathcal{H}\cdot\mathcal{S})_1] = \mathbf{E}[X_1] = 0,$$

i.e. the mean profit after a long time is 0, independent of the strategy $\mathcal{H}$. Above we have the bet

$$H_t := 2^{t-1} 1_{\{S_{t-1} = -(t-1)\}} \tag{13.2}$$

and show that for the gain $(\mathcal{H}\cdot\mathcal{S})_t \xrightarrow{t\to\infty}_{fs} 1$ holds.

How do we now evaluate the strategy (13.2)? Let T be the random time of the win, i.e. T is geometrically distributed with parameter $\frac{1}{2}$. In particular, T is almost surely finite. Then

$$\mathbf{E}\Big[\sum_{t=1}^\infty H_t\Big] = \sum_{k=1}^\infty \frac{1}{2^k}(2^k - 1) = \infty,$$

i.e. for the above strategy you may need a lot of capital.

13.3 Stopped martingales

Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process. A stopped stochastic process is given by $\mathcal{X}^T := (X_{T\wedge t})_{t\in I}$, where T is an I-valued random variable. The process $\mathcal{X}^T$ therefore stops when T is reached. Special random variables T are called stopping times, whose occurrence at time t can be decided by means of the σ-algebra $\mathcal{F}_t$. (Consider, for example, a player who plays a fair game and stops at a random time, e.g. when he has won or lost enough. In this section we will learn about the Optional Stopping Theorem, which states that stopped (at a stopping time) martingales are martingales again; see Theorem 13.19. The Optional Sampling Theorem specifies conditions for which the martingale property applies not only to fixed, but also at random stopping times; see Theorem 13.22.

We start by recalling some facts on random times; see also Definition 12.20.

Remark 13.17 (Stopping time). *1. A random time is a random variable with values in $\bar{I}$ (the end of I). A random time T is called $((\mathcal{F}_t)_{t\in I}$-)*stopping time *if $\{T \le t\} \in \mathcal{F}_t$ for all $t \in I$.*

2. Each stopping time T defines the σ-algebra

$$\mathcal{F}_T := \{A \in \mathcal{A} : A \cap \{T \le t\} \in \mathcal{F}_t, t \in I\}$$

of the T-past.

3. Let $B \in \mathcal{B}(E)$. Then the hitting time of B *is defined as*

$$T_B := \inf\{t : X_t \in B\}.$$

4. For a random time T, X_T is defined by $\omega \mapsto X_{T(\omega)}(\omega)$. Further, $\mathcal{X}^T := (X_{T\wedge t})_{t\in I}$ is the process stopped at T.

Remark 13.18 (Interpretation and hitting times). *1. Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process and $(\mathcal{F}_t)_{t\in I}$ the canonical filtration. $\mathcal{F}_t$ can be understood as the information that is available at time t through knowledge of $(X_s)_{0\le s\le t}$. If T is a stopping time, then $\{T \le t\} \in \mathcal{F}_t$. Therefore, the occurrence of the event $\{T \le t\}$ can be decided by knowing $(X_s)_{s\le t}$. In other words, by knowing the stochastic process up to time t, it is possible to decide whether the stopping time T has already occurred.*

2. If I is at most countable and $B \in \mathcal{B}(E)$, then T_B is a stopping time. Indeed, we write

$$\{T_B \le t\} = \bigcup_{s\le t} \underbrace{\{X_s \in B\}}_{\in\mathcal{F}_s\subseteq\mathcal{F}_t} \in \mathcal{F}_t.$$

Proposition 13.19 (Optional Stopping). *Let $I = \{0, 1, 2, ...\}$ and $\mathcal{X} = (X_t)_{t\in I}$ be a (sub-, super-) martingale and T a stopping time. Then $\mathcal{X}^T = (X_{T\wedge t})_{t\in I}$ is a (sub-, super-) martingale.*

Proof. We show the assertion only for the case that $\mathcal{X}$ is a sub-martingale. The other statements follows analogously. For a sub-martingale $\mathcal{X}$ and $\{T > t - 1\} \in \mathcal{F}_t$,

$$\begin{aligned}\mathbf{E}[X_{T\wedge t} - X_{T\wedge(t-1)}|\mathcal{F}_{t-1}] &= \mathbf{E}[(X_t - X_{t-1})1_{\{T>t-1\}}|\mathcal{F}_{t-1}]\\ &= 1_{\{T>t-1\}}\mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}] \ge 0,\end{aligned}$$

i.e. $\mathcal{X}^T$ is a sub-martingale. □

Lemma 13.20 (Conditions on $\mathcal{F}_T$)**.** *Let $I = \{0, 1, 2, ...\}$, $\mathcal{X} = (X_t)_{t\in I}$ be a martingale and T a stopping time bounded by t. Then $X_T = \mathbf{E}[X_t|\mathcal{F}_T]$.*

Proof. According to the definition of the conditional expectation and since X_T is $\mathcal{F}_T$ measurable (see Proposition 12.33), we must show that $\mathbf{E}[X_t; A] = \mathbf{E}[X_T; A]$ for $A \in \mathcal{F}_T$. It is $\{T = s\} \cap A \in \mathcal{F}_s$ for $s \in I$, i.e.

$$\begin{aligned}
\mathbf{E}[X_T; A] &= \sum_{s=1}^{t} \mathbf{E}[X_s; \{T = s\} \cap A] \\
&= \sum_{s=1}^{t} \mathbf{E}[\mathbf{E}[X_t|\mathcal{F}_s]; \{T = s\} \cap A] \\
&= \sum_{s=1}^{t} \mathbf{E}[X_t; \{T = s\} \cap A] \\
&= \mathbf{E}[X_t; A].
\end{aligned}$$

□

Lemma 13.21 (Uniform integrability and stopping times)**.**
Let $I = \{0, 1, 2, ...\}$. A martingale $\mathcal{X} = (X_t)_{t\in I}$ is uniformly integrable if the family $\{X_T :$ T almost surely finite stopping time$\}$ is uniformly integrable.

Proof. '⇐': clear.
'⇒': According to Lemma 7.9 there is a convex function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $\frac{f(x)}{x} \xrightarrow{x\to\infty} \infty$ and $\sup_{t\in I} \mathbf{E}[f(|X_t|)] =: L < \infty$. If T is almost surely a finite stopping time, then according to Lemma 13.20 (applied to the almost surely finite stopping time $T \wedge t$) $\mathbf{E}[X_t|\mathcal{F}_{T\wedge t}] = X_{T\wedge t}$. Since $\{T \le t\} \in \mathcal{F}_{T\wedge t}$, we find with Jensen's inequality

$$\begin{aligned}
\mathbf{E}[f(|X_T|), \{T \le t\}] &= \mathbf{E}[f(|X_{T\wedge t}|), \{T \le t\}] \\
&= \mathbf{E}[f(|\mathbf{E}[X_t|\mathcal{F}_{T\wedge t}]|), \{T \le t\}] \\
&\le \mathbf{E}[\mathbf{E}[f(|X_t|)|\mathcal{F}_{T\wedge t}], \{T \le t\}] \\
&= \mathbf{E}[f(|X_t|), \{T \le t\}] \le L.
\end{aligned}$$

Thus $\mathbf{E}[f(|X_T|)] \le L$, i.e. the assertion follows with lemma 7.9. □

In example 13.16, $\mathcal{H} \cdot \mathcal{S}$ was a martingale, T a stopping time and $\mathbf{E}[(\mathcal{H} \cdot \mathcal{S})_t] = 0 \neq 1 = (\mathcal{H} \cdot \mathcal{S})_T$. If T had been bounded, this inequality would not have been possible, as we now show.

Theorem 13.22 (Optional Sampling Theorem)**.** *Let $I = \{0, 1, 2, ...\}$, $S \le T$ almost certainly finite stopping times and $\mathcal{X} = (X_t)_{t\in I}$ a sub-martingale. If either T is bounded or $\mathcal{X}$ is uniformly integrable, then X_T is is integrable and $X_S \le \mathbf{E}[X_T|\mathcal{F}_S]$.*

Proof. We first carry out the proof in the case of a bounded stopping time T. Let $T \le t$ be for a $t \in I$. We use the Doob decomposition $\mathcal{X} = \mathcal{M} + \mathcal{A}$ of $\mathcal{X}$ into the martingale $\mathcal{M}$ and the monotonically non-decreasing process $\mathcal{A}$. Then with Lemma 13.20 and $\mathcal{F}_S \subseteq \mathcal{F}_T$ according to theorem 11.2.7

$$\begin{aligned}X_S = M_S + A_S &= \mathbf{E}[M_t + A_S|\mathcal{F}_S]\\ &\le \mathbf{E}[M_t + A_T|\mathcal{F}_S]\\ &= \mathbf{E}[\mathbf{E}[M_t|\mathcal{F}_T] + A_T|\mathcal{F}_S]\\ &= \mathbf{E}[M_T + A_T|\mathcal{F}_S]\\ &= \mathbf{E}[X_T|\mathcal{F}_S].\end{aligned}$$

Now let T be unbounded and $\mathcal{X}$ be uniformly integrable. Let $\mathcal{X} = \mathcal{M} + \mathcal{A}$ be the Doob decomposition of $\mathcal{X}$ into the martingale $\mathcal{M}$ and the non-falling previsible process $\mathcal{A} \ge 0$ with $A_0 = 0$. Since

$$\mathbf{E}[|A_t|] = \mathbf{E}[A_t] = \mathbf{E}[X_t - X_0] \le \mathbf{E}[|X_0|] + \sup_{s\in I} \mathbf{E}[|X_s|]$$

we find $A_t \uparrow A_\infty$ for an $A_\infty \ge 0$ with $\mathbf{E}[A_\infty] < \infty$. With Lemma 7.9 one can conclude that $\mathcal{M}$ is also uniformly integrable. We now apply the Optional Sampling Theorem to the bounded stopping times $S \wedge t$, $T \wedge t$ and $\mathcal{M}$. For $A \in \mathcal{F}_S$ is $\{S \le t\} \cap A \in \mathcal{F}_{S\wedge t}$, therefore

$$\mathbf{E}[M_{T\wedge t}, \{S \le t\} \cap A] = \mathbf{E}[\mathbf{E}[M_{T\wedge t}|\mathcal{F}_{S\wedge t}], \{S \le t\} \cap A] = \mathbf{E}[M_{S\wedge t}, \{S \le t\} \cap A].$$

Since according to Lemma 13.21 the set $\{M_{S\wedge t}, M_{T\wedge t} : t \in I\}$ is uniformly integrable, then by Theorem 7.11

$$\mathbf{E}[M_T, A] = \lim_{t\to\infty} \mathbf{E}[M_{T\wedge t}, \{S \le t\} \cap A] = \lim_{t\to\infty} \mathbf{E}[M_{S\wedge t}, \{S \le t\} \cap A] = \mathbf{E}[M_S, A],$$

i.e. $\mathbf{E}[M_T|\mathcal{F}_S] = M_S$. Furthermore,

$$\mathbf{E}[X_T|\mathcal{F}_S] = \mathbf{E}[M_T|\mathcal{F}_S] + A_S + \mathbf{E}[A_T - A_S|\mathcal{F}_S] \ge M_S + A_S = X_S.$$

□

The Optional Sampling Theorem offers a simple way of characterizing martingales.

Lemma 13.23 (Characterization of martingales)**.** *Let $I = \{0, 1, 2, ...\}$, and $\mathcal{X} = (X_t)_{t\in I}$ be an adapted stochastic process. Then $\mathcal{X}$ is a martingale iff $\mathbf{E}[X_S] = \mathbf{E}[X_T]$ for stopping times S, T that only take two values.*

Proof. '⇒': Clear according to the Optional Sampling Theorem.
'⇐': Let $s \le t$, $A \in \mathcal{F}_s$ and $T = s1_A + t1_{A^c}$. Then T is a stopping time and

$$0 = \mathbf{E}[X_t - X_T] = \mathbf{E}[X_t] - \mathbf{E}[X_s, A] - \mathbf{E}[X_t, A^c] = \mathbf{E}[X_t - X_s, A].$$

Since A was arbitrary, it follows that $\mathbf{E}[X_t|\mathcal{F}_s] = X_s$, so $\mathcal{X}$ is a martingale. □

Example 13.24 (Wald's identities, ruin problem)**.** *1. Let $X_1, X_2, ... \in \mathcal{L}^1$ be independent with $\mu := \mathbf{E}[X_1] = \mathbf{E}[X_2] = ...$, and $S_t := \sum_{s=1}^t X_s$. Furthermore, let T be an almost certainly limited stopping time. Then the first Wald identity is*

$$\mathbf{E}[S_T] = \mathbf{E}[T]\mu.$$

Indeed: the process $\mathcal{M} = (M_t)_{t=0,1,2,...}$ *with* $M_0 = 0$, $M_t = S_t - t\mu$ *for* $t = 1, 2, ...$ *is a martingale, and according to the Optional Sampling Theorem*

$$0 = \mathbf{E}[M_T] = \mathbf{E}[S_T] - \mathbf{E}[T]\mu.$$

Furthermore, if $X_1, X_2, ... \in L^2$ *with* $\sigma^2 = \mathbf{V}[X_1] = \mathbf{V}[X_2] = ...$ *and* T *is independent of* $X_1, X_2, ...$, *then the second Wald identity is*

$$\mathbf{V}[S_T] = \mathbf{E}[T]\sigma^2 + \mathbf{V}[T]\mu^2.$$

Indeed: $(M_t^2 - \langle M \rangle_t)_{t=0,1,2,...}$ *is a martingale, and* $\langle M \rangle_t = t\sigma^2$ *according to Example 13.12, thus*

$$0 = \mathbf{E}[M_T^2 - \langle M \rangle_T] = \mathbf{E}[M_T^2] - \mathbf{E}[T]\sigma^2.$$

Furthermore, due to the independence of T *and* $X_1, X_2, ...$,

$$\mathbf{COV}[S_T, T] = \mathbf{E}[\mathbf{E}[X_1 + \cdots + X_T|T]T] - \mu\mathbf{E}[T]^2 = \mu\mathbf{V}[T],$$

as well as

$$\mathbf{E}[M_T^2] = \mathbf{V}[S_T - T\mu] = \mathbf{V}[S_T] + \mu^2\mathbf{V}[T] - 2\mu\mathbf{COV}[S_T, T] = \mathbf{V}[S_T] - \mu^2\mathbf{V}[T].$$

In both Wald identities, the condition that T *is bounded can be weakened.*

2. *Let* $k \in \mathbb{N}$ *and* $X_1, X_2, ...$ *be independent and identically distributed random variables with* $\mathbf{P}(X_1 = 1) = 1 - \mathbf{P}(X_1 = -1) = p := 1 - q$. *For* $N \in \mathbb{N}$ *with* $0 < k < N$ *let* $S_0 = k$ *and* $S_t = S_0 + \sum_{i=1}^t X_i$. *Further, let* $T := \inf\{t : S_t \in \{0, N\}\}$ *and* $p_k := \mathbf{P}(S_T = 0)$. *This means that you play a game, starting with* k *(money) units, until you are either ruined or have* N *units. In each step you win with probability* p *one unit and loses with probability* $q = 1 - p$ *one unit. Then the probability of being ruined (having 0 units) is given by* p_k.

   In the case $p = \frac{1}{2}$, $(S_t)_{t=0,1,2,...}$ *is a martingale, and thus according to the Optional Sampling Theorem*

$$k = \mathbf{E}[S_T] = N(1 - \mathbf{P}(S_T = 0)),$$

   thus

$$\mathbf{P}(S_T = 0) = \frac{N - k}{N}.$$

   A similar calculation allows the determination of p_k *for the case* $p \neq \frac{1}{2}$.

   We now calculate further using the optional sampling theorem for $p \neq \frac{1}{2}$

$$p_k := \mathbf{P}(S_T = 0) = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}. \tag{13.3}$$

   Indeed: the following applies

$$\mathbf{E}\Big[\Big(\frac{q}{p}\Big)^{X_1}\Big] = \frac{q}{p}p + \frac{p}{q}q = 1$$

and thus $\mathcal{Y} = (Y_t)_{t=0,1,2,...}$, *is defined by* $Y_t := \left(\frac{q}{p}\right)^{S_t}$ *according to Example 13.4.3 a martingale. Since* T *is almost surely finite,* $Y_{T\wedge t}$ *is a martingale due to Proposition 13.19, which is bounded by* 1 *and* $\left(\frac{q}{p}\right)^N$. *Because of Theorem 13.22,*

$$\left(\frac{q}{p}\right)^k = \mathbf{E}[Y_0] = \mathbf{E}[Y_T] = p_k + (1-p_k)\left(\frac{q}{p}\right)^N,$$

from which (13.3) *follows.*

3. *Let's consider a fair coin toss. How long does it take until the pattern* `ZKZK` *occurs for the first time? (*`K` *and* `Z` *stand for heads and tails).*

   To calculate this, let's consider the following game: before the first coin toss, a player bets one euro on `Z`*. If she loses, she stops, if she wins, she bets two euros on* `K` *before the next toss. If she loses in the second throw, she stops; if she wins, she bets four euros on euros on* `Z`*. If she loses on the third throw, she stops, if she wins, she bets eight euros on* `K`*. So if she wins on the fourth throw, she has won a total of 15 euros. In all other cases, she loses one euro.*

   Let us now assume that before each coin toss a new player plays according to the above strategy. The game ends when the first player first time a player wins 15 euros.

   Let X_t *be the total winnings of all players up to time* t *and* T *is the time at which the game is stopped because for the first time the pattern* `ZKZK` *has occurred. Certainly,*

$$|X_t| \leq 15 \cdot t, \qquad \mathbf{P}[T > 4t] \leq \frac{15}{16}^t.$$

   This means that $(X_{t\wedge T} : t = 1, 2, ...)$ *has a cominating integrable random variable, so according to Example 7.8.2 it is uniformly integrable. This allows us to apply the optional stopping theorem, i.e.* $(X_{T\wedge t})_{t=1,2,...}$ *is a martingale.*

   It is certain that

$$X_T = 15 - 1 + 3 - 1 - (T - 4)$$

   since the first $T-4$ *players, as well as players* $T-3$ *and* $T-1$ *had to accept a loss of one euro. Player* $T-2$ *currently has at time* T *a profit of three euros and player* $T-4$ *has won 15 euros. So,*

$$0 = \mathbf{E}[X_T] = \mathbf{E}[15 - 1 + 3 - 1 - (T - 4)] = -\mathbf{E}[T] + 20,$$

   therefore $\mathbf{E}[T] = 20$*. It is interesting to note that it can be expected that, for example, the pattern* `ZZKK` *can already occur after* 16 *coin tosses using a similar calculation.*

13.4 Martingale convergence results

Again, $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, I countable (here it is also allowed that I is dense in $[0, \infty)$) and $(\mathcal{F}_t)_{t\in I}$ is a filtration. We are familiar with convergence theorems, such as the strong law of of large numbers. Martingales converge under relatively weak conditions.

We start in Proposition 13.26 with Doob's inequalities. These make statements about the distribution of $\sup_{s\leq t} X_s$ if $\mathcal{X} = (X_t)_{t\in I}$ is a (sub, super)-martingale.

Lemma 13.25 (maximum inequality)**.** *If I is at most countable and $\mathcal{X} = (X_t)_{t\in I}$ is a sub-martingale, then for $\lambda > 0$*

$$\lambda \mathbf{P}[\sup_{s\leq t} X_s \geq \lambda] \leq \mathbf{E}[X_t, \sup_{s\leq t} X_s \geq \lambda] \leq \mathbf{E}[|X_t|, \sup_{s\leq t} X_s \geq \lambda].$$

Proof. The second inequality is trivial. For the first one, we note that due to monotonic convergence (by choosing finer and finer index sets in index sets in $[0, t]$) it is sufficient to consider the discrete case, e.g. $I = \{0, 1, 2, ...\}$. We recall the definition of T_B from Definition 13.17, which is given after Remark 13.18.2 is a stopping time and set

$$T = t \wedge T_{[\lambda;\infty)}.$$

According to the Optional Sampling Theorem 13.22 is

$$\begin{aligned}\mathbf{E}[X_t] \geq \mathbf{E}[X_T] &= \mathbf{E}[X_T; \sup_{s\leq t} X_s \geq \lambda] + \mathbf{E}[X_T; \sup_{s\leq t} X_s < \lambda] \\ &\geq \lambda \mathbf{P}[\sup_{s\leq t} X_s \geq \lambda] + \mathbf{E}[X_t; \sup_{s\leq t} X_s < \lambda].\end{aligned}$$

Subtracting the last term gives the inequality. □

Proposition 13.26 (Doob's L^p inequality)**.** *Let I be at most countable and $\mathcal{X} = (X_t)_{t\in I}$ be a martingale or a positive sub-martingale.*

1. *For $p \geq 1$ and $\lambda > 0$ is*

$$\lambda^p \mathbf{P}[\sup_{s\leq t} |X_s| \geq \lambda] \leq \mathbf{E}[|X_t|^p].$$

2. *For $p > 1$ is*

$$\mathbf{E}[|X_t|^p] \leq \mathbf{E}[\sup_{s\leq t} |X_s|^p] \leq \Big(\frac{p}{p-1}\Big)^p \mathbf{E}[|X_t|^p].$$

Proof. Again, it suffices – due to monotonic convergence – to consider the case $I = \{0, 1, 2, ...\}$ to consider.

1 According to proposition 13.7, $(|X_t|^p)_{t\in I}$ is a sub-martingale and the assertion follows from Lemma 13.25.

2 The first inequality is clear. For the second inequality, note that according to Lemma 13.25 it holds that

$$\lambda \mathbf{P}\{\sup_{s\leq t} |X_s| \geq \lambda\} \leq \mathbf{E}[|X_s|; \sup_{s\leq t} |X_s| \geq \lambda].$$

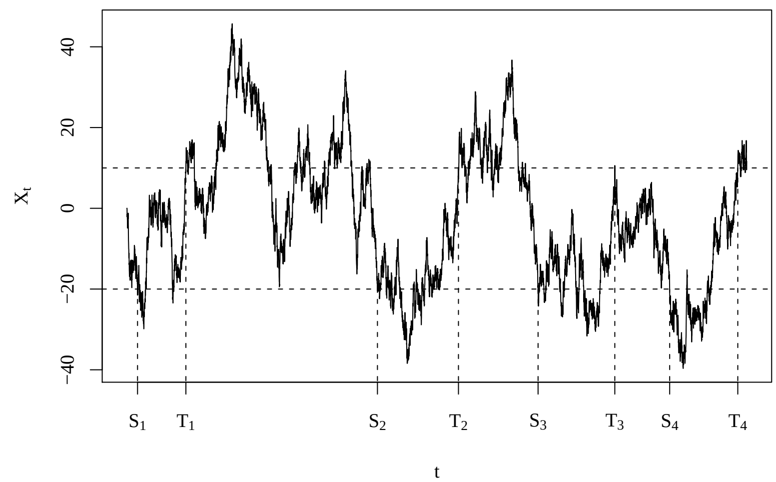


Figure 6:
An illustration of the stopping times $S_1, T_1, S_2, T_2, ...$ from definition 13.27

Therefore, for $K > 0$

$$\begin{aligned}
\mathbf{E}[\sup_{s\leq t}(|X_s| \wedge K)^p] &= \mathbf{E}\Big[\int_0^{\sup_{s\leq t}|X_s|\wedge K} p\lambda^{p-1}d\lambda\Big] \\
&= \mathbf{E}\Big[\int_0^K p\lambda^{p-1}1_{\{\lambda<\sup_{s\leq t}|X_s|\}}d\lambda\Big] \\
&= \int_0^K p\lambda^{p-1}\mathbf{P}(\sup_{s\leq t}|X_s| \geq \lambda)d\lambda \\
&\leq \int_0^K p\lambda^{p-2}\mathbf{E}[|X_t|, \sup_{s\leq t}|X_s| \geq \lambda]d\lambda \\
&= p\mathbf{E}\Big[|X_t|\int_0^{\sup_{s\leq t}|X_s|\wedge K} \lambda^{p-2}d\lambda\Big] \\
&= \frac{p}{p-1}\mathbf{E}[|X_t|(\sup_{s\leq t}|X_s| \wedge K)^{p-1}] \\
&\leq \frac{p}{p-1}\mathbf{E}[\sup_{s\leq t}(|X_s| \wedge K)^p]^{(p-1)/p} \cdot \mathbf{E}[|X_t|^p]^{1/p},
\end{aligned}$$

where we used the Hölder inequality in the last step. If you exponentiate both sides by p and then divide by $\mathbf{E}[\sup_{s\leq t}(|X_s| \wedge K)^p]^{p-1}$, the it follows

$$\mathbf{E}[\sup_{s\leq t}(|X_s|)^p] = \lim_{K\to\infty}\mathbf{E}[\sup_{s\leq t}(|X_s| \wedge K)^p] \leq \Big(\frac{p}{p-1}\Big)^p\mathbf{E}[|X_t|^p].$$

□

For the martingale convergence theorems, the upcrissong lemma 13.28 is central. Figure 6 illustrates the definition of an upcrossing.

Definition 13.27. *Let I be at most countable and $\mathcal{X} = (X_t)_{t\in I}$ a real-valued stochastic process. For $a < b$ an upcrossing is a piece of path $(X_r)_{s\le r\le s'}$ with $X_s \le a$ and $X_{s'} \ge b$. To count the number of such upcrossings, we carry out stopping times $0 =: T_0 < S_1 < T_1 < S_2 < T_2 < ...$*

$$S_k := \inf\{t \ge T_{k-1} : X_t \le a\},$$
$$T_k := \inf\{t \ge S_k : X_t \ge b\}$$

with $\inf \emptyset = \infty$. The k-th intersection between a and b is here between S_k and T_k. Further is

$$U^t_{a,b} := \sup\{k : T_k \le t\}$$

is the number of crossings between a and b up to time t.

Lemma 13.28 (Upcrossing lemma)**.** *Let I be at most countable and $\mathcal{X} = (X_t)_{t\in I}$ a sub-martingale. Then*

$$\mathbf{E}[U^t_{a,b}] \le \frac{\mathbf{E}[(X_t - a)^+]}{b - a}.$$

Proof. Again, we can assume – due to monotonic convergence – that $I = \{0, 1, 2, ...\}$. Since according to proposition 13.7 with $\mathcal{X}$ $((X_t - a)^+)_{t\in I}$ is also a sub-martingale and the upcrossings between a and b of $\mathcal{X}$ are the same as the upcrossings of $((X_t - a)^+)_{t\in I}$ between 0 and $b - a$, we can wlog assume that $\mathcal{X} \ge 0$ and $a = 0$. We define the process $\mathcal{H} = (H_t)_{t\in I}$ by

$$H_t := \sum_{k\ge 1} 1_{\{S_k < t \le T_k\}},$$

i.e. $H_t = 1$ exactly when t lies in an upcrossing. Since

$$\{H_t = 1\} = \bigcup_{k\ge 1} \{S_k \le t - 1\} \cap \{T_k > t - 1\},$$

H is previsible.

Given $T_k < \infty$ is obviously $X_{T_k} - X_{S_k} \ge b$. Further, in this case

$$(\mathcal{H} \cdot \mathcal{X})_{T_k} = \sum_{i=1}^{k} \sum_{s=S_i+1}^{T_i} (X_s - X_{s-1}) = \sum_{i=1}^{k} (X_{T_i} - X_{S_i}) \ge kb.$$

For $t \in \{T_k, ..., S_{k+1}\}$ is $(\mathcal{H} \cdot \mathcal{X})_t = (\mathcal{H} \cdot \mathcal{X})_{T_k}$ and for $t \in \{S_k + 1, ..., T_k\}$ is $(\mathcal{H} \cdot \mathcal{X})_t \ge (\mathcal{H} \cdot \mathcal{X})_{S_k} = (\mathcal{H} \cdot \mathcal{X})_{T_{k-1}}$. Therefore, $(\mathcal{H} \cdot \mathcal{X})_t \ge bU^t_{0,b}$. From Proposition 13.14 it follows that $((1 - \mathcal{H}) \cdot \mathcal{X})$ is a sub-martingale, in particular $\mathbf{E}[((1 - \mathcal{H}) \cdot \mathcal{X})_t] \ge 0$. With $X_t - X_0 = (1 \cdot \mathcal{X})_t = (\mathcal{H} \cdot \mathcal{X})_t + ((1 - \mathcal{H}) \cdot \mathcal{X})_t$ applies

$$\mathbf{E}[X_t] \ge \mathbf{E}[X_t - X_0] \ge \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t] \ge b\mathbf{E}[U^t_{0,b}].$$

□

Theorem 13.29 (martingale convergence theorem for sub-martingales)**.** *Let $I \subseteq [0, \infty)$ be countable, $\sup I = u \in (0, \infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$ and $\mathcal{X} = (X_t)_{t\in I}$ a sub-martingale with $\sup_{t\in I} \mathbf{E}[X_t^+] < \infty$. Then there is a null set N such that $\mathcal{X}$ converges outside of N along every ascending or descending sequence in I.*

In particular, if $I = \{0, 1, 2, ...\}$, $\mathcal{X}$ is a sub-martingale with $\sup_{t\in I} \mathbf{E}[X_t^+] < \infty$, then there exists a $\mathcal{F}_\infty$-measurable, integrable random variable X_∞ and $X_t \xrightarrow{t\to\infty}_{fs} X_\infty$.

Proof. Because of Lemma 13.28, $\mathbf{P}(U_{a,b}^t < \infty) = 1$ for all a, b, t. Therefore

$$N := \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \{\sup_{t\in I} U_{a,b}^t = \infty\}$$

is a null set. Assuming that there is an ascending or descending sequence $t_1, t_2, ... \in I$ exists such that $\mathbf{P}(\liminf_{n\to\infty} X_{t_n} < \limsup_{n\to\infty} X_{t_n}) > 0$. For $a, b \in \mathbb{Q}$ let

$$B(a,b) := \{\liminf_{n\to\infty} X_{t_n} < a < b < \limsup_{n\to\infty} X_{t_n}\}.$$

Since $\{\liminf_{n\to\infty} X_{t_n} < \limsup_{n\to\infty} X_{t_n}\} = \bigcup_{a,b\in\mathbb{Q}} B(a,b)$, there exist $a, b \in \mathbb{Q}$ with $\mathbf{P}(B(a,b)) > 0$. However, $\sup_t U_{a,b}^t = \infty$ applies to $B(a,b)$ in contradiction to the fact that N is a null set. Thus follows the almost sure convergence along every ascending or descending sequence.

Now let $I = \{0, 1, 2, ...\}$. Since all X_t are $\mathcal{F}_\infty$-measurable, X_∞ is also $\mathcal{F}_\infty$-measurable. It remains to show that X_∞ is integrable. According to Fatou's Lemma,

$$\mathbf{E}[X_\infty^+] \le \sup_{t\in I} \mathbf{E}[X_t^+] < \infty.$$

Moreover, since $\mathcal{X}$ is a sub-martingale, again using Fatou's lemma,

$$\mathbf{E}[X_\infty^-] \le \liminf_{t\to\infty} \mathbf{E}[X_t^-] = \liminf_{t\to\infty} \big(\mathbf{E}[X_t^+] - \mathbf{E}[X_t]\big) \le \sup_{t\in I} \mathbf{E}[X_t^+] - \mathbf{E}[X_0] < \infty.$$

□

Corollary 13.30 (martingale convergence theorem for positive super martingales)**.** *Let $I \subseteq [0,\infty)$ be at most countable, $\sup I = u \in (0,\infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$ and $\mathcal{X} = (X_t)_{t\in I}$ a non-negative super martingale. Then there exists a $\mathcal{F}_u$-measurable, integrable random variable X_u with $\mathbf{E}[X_u] \le \mathbf{E}[X_0]$ and $X_t \xrightarrow{t\to u}_{fs} X_u$.*

Proof. Theorem 13.29, applied to the sub-martingale $-\mathcal{X}$ provides the almost sure limit. With the Lemma of Fatou also

$$\mathbf{E}[X_u] \le \liminf_{t\to u} \mathbf{E}[X_t] \le \mathbf{E}[X_0].$$

□

Example 13.31 (Convergence of branching processes)**.** *Let us consider a critical or sub-critical branching process $\mathcal{Z} = (Z_t)_{t=0,1,2,...}$ from Example 13.5 (where the offspring distribution is not degenerate, i.e. $X_i^{(t)} = 1$ is not almost certain). These are non-negative super-martingales, so they must converge according to Corollary 13.30 almost surely against a random variable Z_∞. In this case, $\mathbf{P}(Z_\infty > 0) = 0$ must apply, otherwise the almost sure convergence is violated. (A population with a positive number of individuals has a positive probability of changing its size in one generation.) Therefore,*

$$Z_t \xrightarrow{t\to\infty} Z_\infty := 0$$

is almost certain.

In the case of the critical branching process, it is important to realize that $(Z_t)_{t=0,1,2,...,\infty}$ is not a martingale, because $\mathbf{E}[Z_\infty|\mathcal{F}_t] = \mathbf{E}[0|\mathcal{F}_t] \neq Z_t$ applies with positive probability.

If $\mathcal{Z}$ is supercritical, then $(Z_t/\mu^t)_{t=0,1,2,...}$ is a non-negative martingale that also converges almost surely according to the above corollary.

Theorem 13.32 (Convergence theorem for uniformly integrable martingales)**.** *Let I be countable with $\sup I = u \in (0, \infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$ and $\mathcal{X} = (X_t)_{t\in I}$ a (super, sub)-martingale. Then the following statements are equivalent:*

1. *$\mathcal{X}$ is uniformly integrable.*
2. *There exists a $\mathcal{F}_u$-measurable random variable X_u such that $(X_t)_{t\in I\cup u}$ is a (super, sub)martingale.*
3. *There exists a $\mathcal{F}_u$-measurable random variable X_u with $X_t \xrightarrow{t\to u}_{fs,L^1} X_u$.*

Proof. 2. $\Rightarrow$ 1 follows directly from Lemma 11.5.
1. $\Rightarrow$ 3. By Lemma 7.9, $\sup_{t\in I} \mathbf{E}[|X_t|] < \infty$. The almost certain convergence follows from theorem 13.29 and the L^1-convergence thus from theorem 7.11.
3. $\Rightarrow$ 2.: As for the proof that $(X_t)_{t\in I\cup\{u\}}$ is a (super, sub)-martingale, we only give the argument for sub-martingales, i.e. $\mathbf{E}[\mathbf{E}[X_u|\mathcal{F}_s]; A] \geq \mathbf{E}[X_s; A]$ for $A \in \mathcal{F}_s$ and $s \in I$. Because of the L^1 convergence according to Theorem 11.2.3, $\mathbf{E}[|\mathbf{E}[X_t|\mathcal{F}_s] - \mathbf{E}[X_u|\mathcal{F}_s]|] \xrightarrow{t\to u} 0$ and thus for $A \in \mathcal{F}_s$, so

$$\mathbf{E}[\mathbf{E}[X_u|\mathcal{F}_s]; A] = \lim_{t\to\infty} \mathbf{E}[\mathbf{E}[X_t|\mathcal{F}_s]; A] \geq \mathbf{E}[X_s; A],$$

i.e. $\mathbf{E}[X_u|\mathcal{F}_s] \geq X_s$ almost surely. □

Theorem 13.33 (Martingale convergence theorem for L^p-bounded martingales)**.** *Let I be countable with $\sup I = u \in [0, \infty)$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$, $p > 1$ and $\mathcal{X} = (X_t)_{t\in I}$ an L^p-bounded martingale. Then there is a $\mathcal{F}_u$-measurable random variable X_u with $\mathbf{E}[|X_u|^p] < \infty$, $X_t \xrightarrow{t\uparrow u}_{fs,L^p} X_u$. Furthermore, $(|X_t|^p)_{t\in I}$ is uniformly integrable.*

Proof. Because of Lemma 7.9, $\mathcal{X}$ is uniformly integrable. According to Theorem 13.32 there is thus the limit X_u with $X_t \xrightarrow{t\uparrow u}_{fs,L^1} X_u$. According to Doob's inequality from Proposition 13.26, for $t \in I$

$$\mathbf{E}[\sup_{t\in I} |X_t|^p] = \lim_{t\uparrow u} \mathbf{E}[\sup_{s\leq t} |X_s|^p] \leq \lim_{t\uparrow u} \Big(\frac{p}{p-1}\Big)^p \mathbf{E}[|X_t|^p] < \infty.$$

Thus $(|X_t|^p)_{t\in I}$ is uniformly integrable according to Example 7.8.3 According to Fatou's Lemma and Lemma 7.9, $\mathbf{E}[|X_u|^p] \leq \sup_{t\in I} \mathbf{E}[|X_t|^p] < \infty$ and Theorem 7.11 provides the convergence in L^p. □

Example 13.34 (Branching process)**.** *Let $\mathcal{Z}$ be a branching process as in Example 13.5 and Example 13.31 with $Z_0 = k$. The quadratic variation of $\mathcal{Y} = (Y_t)_{t=0,1,2,\dots}$, given $Y_t = Z_t/\mu^t$ is according to Proposition 13.11 is given as*

$$\begin{aligned}
\langle\mathcal{Y}\rangle_t &= \sum_{s=1}^t \frac{1}{\mu^{2s}} \mathbf{E}\Big[\Big(\sum_{i=1}^{Z_{s-1}} X_i^{(s-1)} - \mu Z_{s-1}\Big)^2 |\mathcal{F}_{s-1}\Big] \\
&= \sum_{s=1}^t \frac{1}{\mu^{2s}} \mathbf{V}\Big[\sum_{i=1}^{Z_{s-1}} X_i^{(s-1)} |Z_{s-1}\Big] \\
&= \sum_{s=1}^t \frac{1}{\mu^{2s}} Z_{s-1} \cdot \mathbf{V}[X_1^{(1)}].
\end{aligned}$$

In particular, the offspring distribution has a second moment, therefore $\mathbf{V}[X_1^{(1)}] =: \sigma^2 < \infty$, *so*

$$\mathbf{V}[Y_t] = \sum_{s=1}^{t} \frac{1}{\mu^{2s}} \mathbf{E}[Z_s] \cdot \sigma^2 = k\sigma^2 \sum_{s=1}^{t} \frac{1}{\mu^s}.$$

If $\mu \leq 1$, *then* $\mathcal{Y}$ *is not* L^2*-bounded, but for* $\mu > 1$ $\sup_{t=0,1,2,\ldots} \mathbf{V}[Y_t] < \infty$. *This means that there is a* $\mathcal{F}_\infty$*-measurable, square-integrable random variable* Y_∞, *so that* $(Y_t)_{t=0,1,2,\ldots,\infty}$ *is a martingale.*

Example 13.35 (product of random variables). *Let* $I = \{1, 2, ...\}$, $X_1, X_2, ...$ *be non-negative, independent, integrable random variable with* $\mathbf{E}[X_t] = 1, t \in I$ *and* $S_t := \prod_{s=1}^t X_s$ *according to Example 13.4.2 a martingale. According to the corollary 13.30 there is thus a* S_∞, *so that* $S_t \xrightarrow{t\to\infty}_{fs} S_\infty$. *Define*

$$a_t := \mathbf{E}[\sqrt{X_t}].$$

We now show:

$$\{S_t : t \in I\} \text{ uniformly integrable} \qquad \Longleftrightarrow \qquad \prod_{t=1}^{\infty} a_t > 0.$$

In particular, then also $S_t \xrightarrow{t\to\infty}_{L^1} S_\infty$. *In the proof we set for* $t = 1, 2, ..$

$$W_t := \prod_{s=1}^{t} \frac{\sqrt{X_s}}{a_s}.$$

This means that $(W_t)_{t=1,2,\ldots}$ *is a martingale. Here, too, it follows that there is a* W_∞ *with* $W_t \xrightarrow{t\to\infty}_{fs} W_\infty$.

'$\Leftarrow$*': Because of Jensen's inequality* $a_t^2 = (\mathbf{E}[\sqrt{X_t}])^2 \leq \mathbf{E}[X_t] = 1$, *thus* $a_t \leq 1$. *The following applies*

$$\sup_{t\in I} \mathbf{E}[W_t^2] = \sup_{t\in I} \mathbf{E}\Big[\prod_{s=1}^{t} \frac{X_s}{a_s^2}\Big] = \sup_{t\in I} \prod_{s=1}^{t} \frac{\mathbf{E}[X_s]}{a_s^2} \leq \frac{1}{\Big(\prod_{s=1}^{\infty} a_s\Big)^2} < \infty.$$

Thus $(W_t)_{t\in I}$ *is an* L^2*-bounded martingale, according to Theorem 13.33,* $\{W_t^2 : t \in I\}$ *is uniformly integrable. From this also the uniform integrability of* $\{S_t : t \in I\}$ *follows.*
'$\Rightarrow$*': Let us assume that* $\prod_{s=1}^\infty a_s = 0$. *Since* W_t *has an almost certain finite limit,* $S_t = \prod_{s=1}^t X_s \xrightarrow{t\to\infty}_{fs} 0$ *must hold. If* $\{S_t : t \in I\}$ *were uniformly integrable,* $0 = \mathbf{E}[S_\infty] = \lim_{t\to\infty} \mathbf{E}[S_t] = 1$, *i.e. a contradiction.*

Theorem 13.36 (Convergence of conditional expected values).

1. *Let* $I \subseteq [0, \infty)$ *be countable with* $\sup I = u \in (0, \infty]$, $(\mathcal{F}_t)_{t\in I}$ *a filtration and* $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$. *Then the following applies for* $X \in \mathcal{L}^1$ *that*

$$\mathbf{E}[X|\mathcal{F}_t] \xrightarrow{t\uparrow u}_{fs,L^1} \mathbf{E}[X|\mathcal{F}_u].$$

2. *Let* $I \subseteq (-\infty, \infty)$ *be countable with* $\inf I = u \in [-\infty, \infty)$, $(\mathcal{F}_t)_{t\in I}$ *a filtration and* $\mathcal{F}_u = \bigcap_{t\in I} \mathcal{F}_t$. *Then the following applies for* $X \in \mathcal{L}^1$ *that*

$$\mathbf{E}[X|\mathcal{F}_t] \xrightarrow{t\downarrow u}_{fs,L^1} \mathbf{E}[X|\mathcal{F}_u].$$

Proof. We only show 1. since the proof of 2. proceeds analogously. With $\mathbf{E}[|\mathbf{E}[X|\mathcal{F}_t]|] \leq \mathbf{E}[|X|] < \infty$ converges according to Theorem 13.29 the martingale $(\mathbf{E}[X|\mathcal{F}_t])_{t\in I}$ converges almost surely. The L^1-convergence follows with Theorem 13.32 and Lemma 11.5. The limit value X_u can be chosen $\mathcal{F}_u$-measurable can be chosen. We will now show that $X_u = \mathbf{E}[X|\mathcal{F}_u]$, from which the assertion follows.

It is clear that $\mathbf{E}[\mathbf{E}[X|\mathcal{F}_t], A] = \mathbf{E}[X, A]$ applies to all $A \in \mathcal{F}_s$ and $s \leq t$. With $t \uparrow u$ is therefore $\mathbf{E}[X_u, A] = \mathbf{E}[X, A]$ for all $A \in \mathcal{F}_s$ and with $s \uparrow u$ also $\mathbf{E}[X_u, A] = \mathbf{E}[X, A]$ for all $A \in \mathcal{F}_u$. Since X_u is measurable with respect to $\mathcal{F}_u$-measurable, this means that $X_u = \mathbf{E}[X|\mathcal{F}_u]$. □

We now come to backward martingales, which are martingales with an index set downward unlimited index set $I \subseteq (-\infty, 0]$. These converge under very weak conditions.

Theorem 13.37 (Martingale convergence theorem for backward martingales)**.** *Let $I \subseteq (\infty, 0]$ be discrete, $\inf I = u \in (-\infty, 0]$, $\mathcal{F}_u = \bigcap_{t\in I} \mathcal{F}_t$ and $\mathcal{X} = (X_t)_{t\in I}$ a sub-martingale. Then are equivalent*

1. *There is a $\mathcal{F}_u$-measurable, integrable random variable X_u with $X_t \xrightarrow{t\downarrow u}_{fs,\ L^1} X_u$*
2. $\inf_{t\in I} \mathbf{E}[X_t] > -\infty$.

Then $(X_t)_{t\in I\cup\{u\}}$ is also a sub-martingale. In particular, every backward martingale converges almost surely and in L^1.

Proof. Wlog, let $I = \{..., -2, -1, 0\}$ and $u = -\infty$.
'1. ⇒ 2.': From the convergence in the mean follows

$$\inf_{t\in I} \mathbf{E}[X_t] = \lim_{t\to-\infty} \mathbf{E}[X_t] = \mathbf{E}[X_{-\infty}] > -\infty.$$

'2. ⇒ 1.': The almost sure convergence follows as in the proof of Theorem 13.29, where the condition $\sup_{t\in I} \mathbf{E}[X_t^+] < \infty$ because of $I \subseteq (-\infty, 0]$ must be replaced by $\inf_{t\in I} \mathbf{E}[X_t^-] < \infty$. We further define for $t = ..., -2, -1, 0$

$$Y_t := \mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}] \geq 0.$$

Then,

$$\mathbf{E}\Big[\sum_{t=0}^{-\infty} Y_t\Big] = \mathbf{E}[X_0] - \inf_{t\in I} \mathbf{E}[X_t] < \infty.$$

Thus, $\sum_{t=0}^{-\infty} Y_t < \infty$ almost surely, and we define

$$A_t = \sum_{s\leq t} Y_s, \qquad M_t = X_t - A_t$$

Now $(A_t)_{t\in I}$ is uniformly integrable because $\mathbf{E}[A_0] < \infty$, and $(M_t)_{t\in I}$ is integrable because it is uniformly integrable by Lemma 11.5. Thus, $\mathcal{X}$ is uniformly integrable, and the L^1-convergence follows. The proof that $(X_t)_{t\in I\cup\{-\infty\}}$ is a sub-martingale proceeds analogous to the proof in 13.32. □

Example 13.38 (The strong law of large numbers). *Let $X_1, X_2, ... \in L^1$ be independently identically distributed. For $t \in \{..., -2, -1\}$ we set as in the Example 13.4.2,*

$$S_t := \frac{1}{|t|}\sum_{s=1}^{|t|} X_s$$

and $\mathcal{F}_t = \sigma(..., S_{t-1}, S_t) = \sigma(S_t, X_{t+1}, X_{t+2}, ...)$. Then $(S_t)_{t\in I}$ is a backward martingale with $S_t = \mathbf{E}[X_1|\mathcal{F}_t]$. According to Theorem 13.37, S_t converges almost surely and in L^1 against a random variable $S_{-\infty}$. This is measurable with respect to $\mathcal{F}_{-\infty}$, but also with respect to $\mathcal{T}(X_1, X_2, ...)$, the terminal σ-algebra of the family $\{X_1, X_2, ...\}$. Since this σ-algebra is trivial according to Kolmogoroff's 0-1-law, $S_{-\infty}$ is almost certainly constant. Since $(S_t)_{t\in I\cup\{-\infty\}}$ is a martingale, it follows that

$$\frac{1}{|t|}\sum_{s=1}^{|t|} X_s = S_t \xrightarrow{t\to-\infty}_{fs,L^1} S_{-\infty} = \mathbf{E}[S_{-\infty}] = \mathbf{E}[S_{-1}] = \mathbf{E}[X_1].$$

However, the almost sure convergence is exactly the statement of the law of large numbers.

We now come to an application of the martingale convergence theorems, an improvement of the Borel-Cantelli lemma, Theorem 8.8. For this we need a lemma.

Lemma 13.39 (Convergence and increasing process). *Let $\mathcal{M} = (M_t)_{t=0,1,2,...}$ be an L^2-integrable martingale, where $|M_t - M_{t-1}| \le K$ for some K and all $t = 1, 2, ...$ holds. Then there is a nullset N such that*

$$\begin{aligned}\{\langle\mathcal{M}\rangle_\infty < \infty\} &\subseteq \{\lim_{t\to\infty} M_t \text{ exists}\} \cup N,\\ \{\langle\mathcal{M}\rangle_\infty = \infty\} &\subseteq \{\lim_{t\to\infty} M_t/\langle\mathcal{M}\rangle_t = 0\} \cup N.\end{aligned}$$

Proof. We start with the first statement. First, for each $k = 1, 2, 3, ...$ the random time

$$T_k := \inf\{t : \langle\mathcal{M}\rangle_t > k\}$$

is a stopping time. From this already follows

$$\{\langle\mathcal{M}\rangle_\infty < \infty\} = \bigcup_{k=1}^\infty \{T_k = \infty\}. \tag{13.4}$$

Furthermore, the stopped process $(\langle\mathcal{M}\rangle_{t\wedge T_k})_{t=0,1,2,...}$ is previsible, because for $A \in \mathcal{B}(\mathbb{R})$,

$$\{\langle\mathcal{M}\rangle_{t\wedge T_k} \in A\} = \big(\{T_k > t-1\} \cap \{\langle\mathcal{M}\rangle_t \in B\}\big) \cup \bigcup_{s=0}^{t-1}\{T_k = s, \langle\mathcal{M}\rangle_s \in A\} \in \mathcal{F}_{t-1}.$$

Let us now consider the martingale $(\mathcal{M}^{T_k})^2 - \langle\mathcal{M}\rangle^{T_k} = (\mathcal{M}^2 - \langle\mathcal{M}\rangle)^{T_k}$ for $k = 1, 2, ...$ It is $\langle\mathcal{M}^{T_k}\rangle = \langle\mathcal{M}\rangle^{T_k}$ and $\langle\mathcal{M}\rangle^{T_k}$ is bounded by $k + K^2$. Thus $\mathcal{M}^{T_k}$ is bounded in L^2 and thus converges almost surely. However, on the set $\{T_k = \infty\}$, the process $\mathcal{M}^{T_k}$ converges if and only if $\mathcal{M}$ converges. Together with (13.4) the statement follows.

For the second statement, we consider the martingale $\mathcal{X} := (1 + \langle\mathcal{M}\rangle)^{-1} \cdot \mathcal{M}$. Since $(1 + \langle\mathcal{M}\rangle)^{-1}$ is bounded and $\mathcal{M}$ is an L^2-integrable martingale, $\mathcal{X}$ is an L^2-integrable martingale. Furthermore, according to Example 13.15,

$$\begin{aligned}\langle\mathcal{X}\rangle_t &= \Big(\frac{1}{(1+\langle\mathcal{M}\rangle)^2} \cdot \langle\mathcal{M}\rangle\Big)_t = \sum_{s=1}^t \frac{1}{(1+\langle\mathcal{M}\rangle_s)^2}(\langle\mathcal{M}\rangle_s - \langle\mathcal{M}\rangle_{s-1}) \\ &\le \sum_{s=1}^t \frac{1}{(1+\langle\mathcal{M}\rangle_s)(1+\langle\mathcal{M}\rangle_{s-1})}(\langle\mathcal{M}\rangle_s - \langle\mathcal{M}\rangle_{s-1}) = \sum_{s=1}^t \frac{1}{1+\langle\mathcal{M}\rangle_{s-1}} - \frac{1}{1+\langle\mathcal{M}\rangle_s} \\ &= 1 - \frac{1}{1+\langle\mathcal{M}\rangle_t}.\end{aligned}$$

This means that the martingale $\mathcal{X}$ converges after 1. i.e. in particular

$$\sum_{s=1}^\infty \frac{M_s - M_{s-1}}{1+\langle\mathcal{M}\rangle_s} < \infty.$$

Now the Kronecker lemma 8.24 provides that

$$\frac{\sum_{s=1}^t M_s - M_{s-1}}{\langle\mathcal{M}\rangle_t} \xrightarrow{t\to\infty} 0$$

on $\{\langle\mathcal{M}\rangle_\infty = \infty\}$. □

Theorem 13.40 (Extension of the Borel-Cantelli lemma)**.** *Let $A_t \in \mathcal{F}_t$, $t = 0, 1, 2, ...$ and*

$$X_s := \mathbf{P}(A_s|\mathcal{F}_{s-1}).$$

1. *On $\sum_{t=1}^\infty X_t < \infty$ only a finite number of the A_t occur, i.e.*

$$\Big\{\sum_{t=1}^\infty X_t < \infty\Big\} \subseteq \Big\{\sum_{t=1}^\infty 1_{A_t} < \infty\Big\}.$$

2. *On $\sum_{t=1}^\infty X_t = \infty$ applies $\sum_{t=1}^\infty 1_{A_t} / \sum_{t=1}^\infty X_t = 1$, thus*

$$\Big\{\sum_{t=1}^\infty X_t = \infty\Big\} \subseteq \Big\{\sum_{t=1}^\infty 1_{A_t} / \sum_{t=1}^\infty X_t = 1\Big\} \subseteq \Big\{\sum_{t=1}^\infty 1_{A_t} = \infty\Big\}.$$

Remark 13.41 (Extension)**.** *The Borel-Cantelli Lemma from theorem 8.8 can now be easily be derived. Namely, if*

$$\mathbf{E}\Big[\sum_{t=1}^\infty X_t\Big] = \sum_{t=1}^\infty \mathbf{P}(A_t) < \infty,$$

then $\sum_{t=1}^\infty X_t < \infty$ almost certainly applies. The statement now gives that at most a finite number of the A_n occur. If further $A_1, A_2, ...$ are independent, then we set $\mathcal{F}_t = \sigma(A_1, ..., A_t)$ and thus $X_s = \mathbf{E}[1_{A_s}|\mathcal{F}_{s-1}] = \mathbf{P}(A_s)$. Now, $\sum_{t=1}^\infty \mathbf{P}(A_t) = \infty$, infinitely many of the A_n's occur.

Proof. We consider the martingale $\mathcal{M}$ with

$$M_t = \sum_{s=1}^{t} 1_{A_s} - X_s.$$

Then,

$$\langle \mathcal{M} \rangle_t = \sum_{s=1}^{t} \mathbf{E}[1_{A_s}^2 X_s^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^{t} X_s(1 - X_s) \leq \sum_{s=1}^{t} X_s.$$

If now $\sum_{t=1}^{\infty} X_t < \infty$, then $\mathcal{M}$ converges according to Lemma 13.39.1. therefore also $\sum_{t=1}^{\infty} 1_{A_t} < \infty$.
If now $\sum_{t=1}^{\infty} X_t = \infty$ and $\langle \mathcal{M} \rangle_\infty < \infty$, then $\mathcal{M}$ converges and the assertion is clear.
If now $\sum_{t=1}^{\infty} X_t = \infty$ and $\langle \mathcal{M} \rangle_\infty = \infty$, then $M_t / \langle \mathcal{M} \rangle_t \xrightarrow{t\to\infty} 0$ according to Lemma 13.39.2 From this,

$$\Big| \frac{\sum_{s=1}^{t} 1_{A_s}}{\sum_{s=1}^{t} X_s} - 1 \Big| = \Big| \frac{M_t}{\sum_{s=1}^{t} X_s} \Big| \leq \Big| \frac{M_t}{\langle \mathcal{M} \rangle_t} \Big| \xrightarrow{t\to\infty} 0.$$

□

13.5 The Central Limit Theorem for martingales

The Central Limit Theorem from Section 10.2 states the convergence of a sum of *independent* random variables – suitably transformed – to a normally distributed random variable. Now we treat the case of a sequence of martingales $\mathcal{M}^1 = (M_t^1)_{t=0,1,2,...}, \mathcal{M}^2 = (M_t^2)_{t=0,1,2,...}, ...$, each started in 0, which are given by $X_t^n := M_t^n - M_{t-1}^n, t = 1, 2, ...$ as a sum through $M_t^n = X_1^n + \cdots + X_t^n$ now applies. Now note that the family $X_1^n, X_2^n, ...$ do not have to be independent. Nevertheless, we can – under suitable conditions – still prove convergence in distribution against a normally distributed random variable.

Theorem 13.42 (Central limit theorem for martingales)**.** *Let $I^n = \{0, 1, 2, ..., t_n\}$ and $\mathcal{M}^n = (M_t^n)_{t\in I^n}$ a martingale with $M_0^n = 0$ with respect to a filtration $\mathcal{F}^n = (\mathcal{F}_t^n)_{t\in\mathcal{I}^n}$, $n = 1, 2, ...$ For $X_t^n := M_t^n - M_t^{n-1}$ (with $t = 1, ..., t_n$) the following applies*

$$\mathbf{E}[\max_{1\leq s\leq t_n} |X_s^n|] \xrightarrow{n\to\infty} 0, \tag{13.5}$$

$$\sum_{s=1}^{t_n} (X_s^n)^2 \xrightarrow{n\to\infty}_p \sigma^2 > 0. \tag{13.6}$$

Then $M_{t_n}^n \xrightarrow{n\to\infty} X$ with $X \sim N(0, \sigma^2)$.

We need two lemmas in the proof of the theorem.

Lemma 13.43 (Convergence of products of random variables)**.** *Let $U_1, U_2, ..., T_1, T_2, ...$ be random variables that satisfy the following conditions:*

1. *$U_n \xrightarrow{n\to\infty}_p u$,*
2. *$(T_n)_{n=1,2,...}$ and $(T_nU_n)_{n=1,2,...}$ are uniformly integrable,*
3. *$\mathbf{E}[T_n] \xrightarrow{n\to\infty} 1$.*

Then $\mathbf{E}[T_n U_n] \xrightarrow{n\to\infty} u$.

Proof. Because of 3. it suffices to show that $\mathbf{E}[T_n(U_n - u)] \xrightarrow{n\to\infty} 0$. To do this, we simply show $T_n(U_n - u) \xrightarrow{n\to\infty}_p 0$, which implies the L^1-convergence due to 2. In particular, then $\mathbf{E}[T_n(U_n - u)] \xrightarrow{n\to\infty} 0$. Let $\varepsilon > 0$ and K be large enough so that $\sup_n \mathbf{P}(|T_n| >> K) \leq \varepsilon$. (Such a K exists because of the uniform integrability of $T_1, T_2, ...$) Now we write (note that for $x, y \geq 0$ and $\delta, \varepsilon > 0$ it always holds that $xy > \delta\varepsilon \to x > \delta$ or $y > \varepsilon$)

$$\limsup_{n\to\infty} \mathbf{P}(|T_n(U_n - u)| > \varepsilon) \leq \limsup_{n\to\infty} \mathbf{P}(|U_n - u| > \varepsilon/K) + \mathbf{P}(|T_n| > K) \leq \varepsilon.$$

The assertion follows from this. □

Lemma 13.44 (Estimation of the exponential function). *1. There is a $C > 0$ and a function r with $|r(x)| \leq C|x^3|$ such that*

$$\exp(ix) = (1 + ix)\exp(-x^2/2 + r(x))$$

for all $x \in \mathbb{R}$ is valid.

2. $|1 + ix| \leq e^{x^2/2}$ applies to all $x \in \mathbb{R}$.

Proof. 1. It is sufficient to show the assertion for small $|x|$, since it is trivial for large $|x|$. With the help of lemma 10.12, we write

$$\begin{aligned}
&\Big|\exp(ix) - (1 + ix)\exp(-x^2/2)\Big| \\
&\qquad = \Big|\exp(ix) - 1 - ix + x^2/2 - (1 + ix)(\exp(-x^2/2) - 1 + x^2/2) + ix^3/3\Big| \\
&\qquad \leq \Big|\exp(ix) - 1 - ix + x^2/2\Big| + |1 + ix| \cdot \Big|\exp(-x^2/2) - 1 + x^2/2\Big| + \Big|x^3/3\Big| \\
&\qquad \leq \tfrac{|x^3|}{6} + |1 + ix| \cdot \Big(\tfrac{|x^2|}{2} \wedge \tfrac{|x^4|}{8}\Big) + \tfrac{|x^3|}{3} \leq |x^3|
\end{aligned}$$

for all x. From this follows the assertion for small $|x|$, and thus 1. is proven. For 2. it is sufficient to use $|1 + ix|^2 = 1 + x^2 \leq e^{x^2}$ and take the root. □

Proof of Theorem 13.42. First we define

$$Z_s^n := X_s^n 1_{\sum_{r=1}^{s-1}(X_r^n)^2 \leq 2\sigma^2}$$

and $N_t^n := \sum_{s=1}^t Z_s^n$. Then $(N_t^n)_{t=1,2,\dots}$ is a $(\mathcal{F}_t^n)_{t\in I^n}$ martingale, because

$$\mathbf{E}[N_t^n - N_{t-1}^n | \mathcal{F}_{t-1}^n] = \mathbf{E}[Z_t^n | \mathcal{F}_{t-1}^n] = 1_{\sum_{r=1}^{s-1}(X_r^n)^2 \leq 2\sigma^2} \cdot \mathbf{E}[X_t^n | \mathcal{F}_{t-1}^n] = 0,$$

since $M_t^n = X_1^n + \cdots + X_t^n$. Now,

$$\begin{aligned}
\mathbf{P}(\max_{t=1,\dots,t_n} |M_t^n - N_t^n| > 0) &= \mathbf{P}(M_t^n \neq N_t^n \text{ for one } t \in I^n) \\
&= \mathbf{P}(X_t^n \neq Z_t^n \text{ for a } t \in I^n) \\
&= \mathbf{P}\Big(\sum_{s=1}^{t_n} (X_s^n)^2 > 2\sigma^2\Big) \xrightarrow{n\to\infty} 0,
\end{aligned} \tag{13.7}$$

where the convergence follows from (13.6). Now the following applies $M^n_{t_n} - N^n_{t_n} \xrightarrow{n\to\infty}_p 0$, so it suffices according to Slutzky's theorem, Corollary 9.9, $N^n_{t_n} \xRightarrow{n\to\infty} X \sim N(0,\sigma^2)$ to show. For this we will for any $\lambda \in \mathbb{R}$

$$\mathbf{E}[e^{i\lambda N^n_{t_n}}] \xrightarrow{n\to\infty} e^{-i\lambda^2\sigma^2/2}$$

show. With the function r from Lemma 13.44 now applies

$$\mathbf{E}[e^{i\lambda N^n_{t_n}}] = \prod_{s=1}^{t_n}(1+i\lambda Z^n_s)\cdot \exp\Big(-\tfrac{\lambda^2}{2}\sum_{s=1}^{t_n}(Z^n_s)^2 + \sum_{s=1}^{t_n} r(\lambda Z^n_s)\Big).$$

We now set

$$T_n := \prod_{s=1}^{t_n}(1+i\lambda Z^n_s), \qquad U_n := \exp\Big(-\tfrac{\lambda^2}{2}\sum_{s=1}^{t_n}(Z^n_s)^2 + \sum_{s=1}^{t_n} r(\lambda Z^n_s)\Big)$$

and show that for these random variables the conditions of Lemma 13.43 apply to these random variables (with $u = e^{-\lambda^2\sigma^2/2}$). For 1. first because of (13.7)

$$\lim_{n\to\infty}\sum_{s=1}^{t_n}(Z^n_s)^2 = \lim_{n\to\infty}\sum_{s=1}^{t_n}(X^n_s)^2 = \sigma^2.$$

Further, with C from Lemma 13.44

$$\begin{aligned}\Big|\sum_{s=1}^{t_n} r(\lambda Z^n_s)\Big| &\le C\cdot|\lambda^3|\cdot\sum_{s=1}^{t_n}|Z^n_s|^3 \le C\cdot|\lambda^3|\cdot\sum_{s=1}^{t_n}|X^n_s|^3\\ &\le C\cdot|\lambda^3|\cdot\max_{1\le s\le t_n}|X^n_s|\cdot\sum_{s=1}^{t_n}|X^n_s|^2 \xrightarrow{n\to\infty} 0,\end{aligned}$$

where the convergence follows from (13.5) and (13.6).
For 2. $|T_nU_n| = |e^{i\lambda N^n_{t_n}}| = 1$, from which the uniform integrability of $(T_nU_n)_{n\in I^n}$ already follows. For the uniform integrability of $(T_n)_{n\in I^n}$ we define

$$J_n := \inf\Big\{s \le t_n : \sum_{r=1}^{s}(X^n_r)^2 > 2\sigma^2\Big\}\wedge t_n$$

and write

$$\begin{aligned}|T_n| &= \prod_{s=1}^{J_n-1}|1+i\lambda Z^n_s|\cdot|1+i\lambda Z^n_{J_n}| \le \exp\Big(\tfrac{\lambda^2}{2}\sum_{s=1}^{J_n-1}(X^n_s)^2\Big)(1+\lambda|X^n_{J_n}|)\\ &\le \exp(\lambda^2\sigma^2)\cdot(1+|\lambda|\cdot\max_{1\le s\le t_n}|X^n_s|).\end{aligned}$$

Since $\max_{1\le s\le t_n}|X^n_s| \xrightarrow{n\to\infty}_{L^1} 0$, in particular the family $(\max_{1\le s\le t_n}|X^n_s|)_{n=1,2,\dots}$ is uniformly integrable, from which the uniform integrability of $(T_n)_{n=1,2,\dots}$ follows.

We now come to 3. by showing $\mathbf{E}[T_n] = 1$. Since $\mathbf{E}[Z^n_s|\mathcal{F}^n_{s-1}] = 0$ for all $s = 1,\dots,t_n$,

$$\begin{aligned}\mathbf{E}[T_n] &= \mathbf{E}\Big[\prod_{s=1}^{t_n}(1+i\lambda Z^n_s)\Big]\\ &= \mathbf{E}\Big[(1+i\lambda Z^n_1)\cdot\mathbf{E}[(1+i\lambda Z^n_2)\cdots\mathbf{E}[1+\lambda Z^n_{t_n}|\mathcal{F}^n_{t_n-1}]\cdots|\mathcal{F}^n_1]\Big] = 1.\end{aligned}$$

Now the assertion follows directly with Lemma 13.44. □

Example 13.45. *1. Let $X_1, X_2, \dots$ be independent, identically distributed, real-valued random variable with $\mathbf{E}[X_1] = 0$ and finite variance $\mathbf{V}[X_1] = \sigma^2$. It is then known that $\mathcal{M}^n = (M_t^n)_{t=0,1,2,\dots}$ with*

$$M_t^n = \tfrac{1}{\sqrt{n}} \sum_{s=1}^t X_t$$

is a martingale and

$$M_n^n \xRightarrow{n\to\infty} X \sim N(0, \sigma^2).$$

This can also be realized by means of Theorem 13.42: first we establish that $\int_0^\infty t\mathbf{P}(|X_1| > t)dt < \infty$ because of the finite second moment. This means that $\mathbf{P}(|X_1| > t) = o(1/t^2)$ for $t \to \infty$, can therefore be written as $\mathbf{P}(|X_1| > t) = a(t)/t^2$ with $a(t) \xrightarrow{t\to\infty} 0$. From this,

$$\begin{aligned}\mathbf{E}[\max_{1\leq s\leq n} |X_s|/\sqrt{n}] &= \int_0^\infty \mathbf{P}(\max_{1\leq s\leq n} |X_s| > t\sqrt{n})dt = \int_0^\infty 1 - (1 - \mathbf{P}(|X_1| > t\sqrt{n}))^n dt \\ &= \int_0^\infty 1 - \Big(1 - \frac{a(t\sqrt{n})}{t^2 n}\Big)^n dt \xrightarrow{n\to\infty} 0\end{aligned}$$

due to dominated convergence. Furthermore, with the law of large numbers,

$$\frac{1}{n}\sum_{s=1}^n X_s^2 \xrightarrow{n\to\infty}_{fs} \sigma^2.$$

So, the conditions of theorem 13.42 are fulfilled.

2. *We bring another example of a sequence of martingales that lead to sums of dependent random variables. For this, we recall the stochastic integral from Definition 13.13. Let $Y_1, Y_2, \dots$ be independent, identically distributed, restricted random variables with $\mathbf{E}[Y_1] = 0$ and $\mathbf{V}[Y_1] = 1$ and $\mathcal{H} = (H_t)_{t=0,1,2,\dots}$ and $\mathcal{M}^n = (M_t^n)_{t=0,1,2,\dots}$ given as*

$$H_s = \frac{1}{s-1}(Y_1^2 + \cdots + Y_{s-1}^2), \qquad M_t^n = \frac{1}{\sqrt{n}}\sum_{s=1}^t Y_s.$$

Then,

$$(\mathcal{H}\cdot\mathcal{M}^n)_t = \frac{1}{\sqrt{n}}\sum_{s=1}^t Y_s \frac{1}{s-1}\sum_{r=1}^{s-1} Y_r^2$$

is a martingale with

$$X_t^n := (\mathcal{H}\cdot\mathcal{M}^n)_t - (\mathcal{H}\cdot\mathcal{M}^n)_{t-1} = \frac{1}{\sqrt{n}} Y_t \frac{1}{t-1}\sum_{r=1}^{t-1} Y_r^2.$$

(Note that $(X_1^n, X_2^n, \dots)$ is not an independent family). Now, (13.5) applies by the boundedness of $Y_1, Y_2, \dots$ We further calculate

$$\sum_{s=1}^n (X_s^n)^2 = \frac{1}{n}\sum_{s=1}^n Y_s^2\Big(\frac{1}{s-1}\sum_{r=1}^{s-1} Y_r^2\Big)^2 \xrightarrow{n\to\infty} 1,$$

from which now $(\mathcal{H}\cdot\mathcal{M}^n)_n \xRightarrow{n\to\infty} X \sim N(0,1)$ follows.

13.6 Properties of martingales in continuous time

Example 13.46 (Martingales derived from the Poisson process)**.** *Let $I = [0, \infty)$, $\mathcal{X} = (X_t)_{t \in I}$ be a Poisson process with intensity λ and $\mathcal{F}_t = \sigma(X_s : s \leq t)$. Then,*

$$(X_t - \lambda t)_{t \in I} \qquad \text{and} \qquad \Big(X_t^2 - \lambda \int_0^t (2X_r + 1) dr\Big)_{t \in I}$$

is a martingale. The following applies for $0 \leq s \leq t$

$$\mathbf{E}[X_t - \lambda t | \mathcal{F}_s] = \mathbf{E}[X_s + X_t - X_s - \lambda t | \mathcal{F}_s] = X_s + \lambda(t-s) - \lambda t = X_s - \lambda s,$$

$$\begin{aligned} \mathbf{E}\Big[X_t^2 - X_s^2 - \lambda \int_s^t (2X_r + 1) dr | \mathcal{F}_s\Big] & \\ &= \mathbf{E}\Big[(X_t - X_s)^2 + 2(X_t - X_s)X_s - \lambda((2X_s + 1)(t-s) + 2\int_s^t (X_r - X_s) dr) | \mathcal{F}_s] \\ &= \lambda(t-s) + \lambda^2(t-s)^2 + 2\lambda(t-s)X_s - \lambda((2X_s + 1)(t-s) - \lambda^2(t-s)^2 = 0. \end{aligned}$$

Example 13.47 (Martingales derived from Brownian motion)**.** *Let $I = [0, \infty)$, $\mathcal{X} = (X_t)_{t \in I}$ be a Brownian motion, $\mathcal{F}_t = \sigma(X_s : s \leq t)$ and $\alpha \in \mathbb{R}$.*

1. *The processes*

$$(\alpha X_t)_{t \in I}, \qquad (\alpha X_t^2 - \alpha t)_{t \in I} \qquad \text{and} \qquad \big(\exp(\alpha X_t - \alpha^2 t/2)\big)_{t \in I} \tag{13.8}$$

are martingales. The following applies for $0 \leq s \leq t$

$$\begin{aligned} \mathbf{E}[\alpha X_t | \mathcal{F}_s] &= \mathbf{E}[\alpha X_s + \alpha(X_t - X_s) | \mathcal{F}_s] = \alpha X_s, \\ \mathbf{E}\Big[\alpha X_t^2 - \alpha t | \mathcal{F}_s] &= \alpha \mathbf{E}[(X_t - X_s)^2 + 2(X_t - X_s)X_s + X_s^2 - t | \mathcal{F}_s] \\ &= \alpha(t-s) + \alpha X_s^2 - \alpha t = \alpha X_s^2 - \alpha s, \\ \mathbf{E}\Big[\exp(\alpha X_t - \alpha^2 t/2) | \mathcal{F}_s] &= \exp(\alpha X_s - \alpha^2 t/2) \cdot \mathbf{E}[\exp(\alpha(X_t - X_s))] \\ &= \exp(\alpha X_s - \alpha^2 t/2 + \alpha^2(t-s)/2) = \exp(\alpha X_s - \alpha^2 s/2) \end{aligned}$$

according to Example 6.13.3.
Since the process $\big(\exp(\alpha X_t - \alpha^2 t/2)\big)_{t \in I}$ is a non-negative martingale with $\mathbf{E}[\exp(\alpha X_t - \alpha t/2)] = 1$, it represents a density. Therefore, for $\tau > 0$,

$$\mathbf{Q}_\tau : \begin{cases} \mathcal{B}(\mathbb{R})^{[0,\tau]} & \to [0,1] \\ A & \mapsto \mathbf{E}[\exp(\alpha X_\tau - \alpha^2 \tau/2), A] \end{cases}$$

is another probability measure on $\mathcal{B}(\mathbb{R})^{[0,\tau]}$, which leads to a probability measure $\mathbf{Q}$ on $\mathcal{B}(\mathbb{R})^I$ since

$$\mathbf{Q}|_{\mathcal{F}_\tau} = \mathbf{Q}_\tau \tag{13.9}$$

can be continued.

2. *For $\mu \in \mathbb{R}$ the process $(X_t + \mu t)_{t \in [0,\infty)}$ is called* Brownian motion with drift μ. *This is a martingale if and only if $\mu = 0$. For $\mu > 0$ it is a sub-martingale and for $\mu < 0$ it is a super-martingale.*

There is a close connection between the Brownian motion with drift and the martingale $\big(\exp(\mu X_t - \mu^2 t/2)\big)_{t\in I}$ from (13.8).

Proposition 13.48 (Brownian motion with drift and change of measure)**.** *Let $I = [0,\infty)$ and $\mathcal{X} = (X_t)_{t\in I}$ be a Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Further, let $\mathcal{Y} = (Y_t)_{t\in I}$ with $Y_t = X_t + \mu t$ for a $\mu \in \mathbb{R}$ and $\mathbf{Q}$ from (13.9). Then,*

$$\mathcal{X}_*\mathbf{Q} = \mathcal{Y}_*\mathbf{P} \qquad \text{and} \qquad \mathcal{Y}_*\mathbf{Q} = \mathcal{X}_*\mathbf{P},$$

i.e. the distribution of $\mathcal{Y}$ under the measure $\mathbf{Q}$ is that of a Brownian motion with drift μ under $\mathbf{P}$. In particular, is a martingale under $\mathbf{Q}$.

Proof. First, let f be continuous and bounded, and $0 \le s \le t$. Then,

$$\begin{aligned}
\mathbf{E_Q}[f(X_t)|\mathcal{F}_s] &= \mathbf{E_P}[f(X_t)e^{\mu X_t - \mu^2 t/2}|\mathcal{F}_s] \\
&= \frac{1}{\sqrt{2\pi(t-s)}} e^{\mu X_s - \mu^2 t/2} \int f(X_s + y) e^{\mu y} e^{-y^2/(2(t-s))} dy \\
&= \frac{1}{\sqrt{2\pi(t-s)}} e^{\mu X_s - \mu^2 t/2 + \mu^2 (t-s)/2} \int f(X_s + y) e^{-(y-\mu(t-s))^2/(2(t-s))} dy \\
&= \frac{1}{\sqrt{2\pi(t-s)}} e^{\mu X_s - \mu^2 s/2} \int f(X_s + y + \mu(t-s)) e^{-y^2/(2(t-s))} dy \\
&= \mathbf{E_P}[f(X_t + \mu(t-s))|\mathcal{F}_s] \cdot e^{\mu X_s - \mu^2 s/2}.
\end{aligned}$$

Now let $0 \le t_1 \le \cdots \le t_n$ and $f_1, ..., f_n$ be continuous and bounded. Then,

$$\begin{aligned}
\mathbf{E_Q}[f_1(X_{t_1})\cdots f_n(X_{t_n})] &= \mathbf{E_P}[f_1(X_{t_1})\cdots f_{n-1}(X_{t_{n-1}})\mathbf{E_P}[f_n(X_{t_n})e^{\mu X_{t_n} - \mu^2 t_n/2}|\mathcal{F}_{t_{n-1}}]] \\
&= \mathbf{E_P}[f_1(X_{t_1})\cdots f_{n-2}(X_{t_{n-2}})\cdot \\
&\qquad \mathbf{E_P}[f_{n-1}(X_{t_{n-1}})\mathbf{E_P}[f_n(X_{t_n} + \mu(t_n - t_{n-1})|\mathcal{F}_{t_{n-1}}]e^{\mu X_{t_{n-1}} - \mu^2 t_{n-1}/2}]|\mathcal{F}_{t_{n-2}}]] \\
&= \cdots = \mathbf{E_P}[f_1(X_{t_1} + \mu t_1)\cdots f_n(X_{t_n} + \mu t_n)] = \mathbf{E_P}[f_1(Y_{t_1})\cdots f_n(Y_{t_n})].
\end{aligned}$$

Since $f_1, ..., f_n$ were arbitrary, the finite-dimensional distributions of $\mathcal{X}_*\mathbf{Q}$ and $\mathcal{Y}_*\mathbf{P}$ are identical. The statement now follows from Proposition 12.6.1. □

We will now apply the results of martingales with a countable index set to the case of an uncountable index set, $I = [0,\infty)$. Central to this is Theorem 13.49, in which we will see that there is a right-continuous modification for very many sub-martingales.

Theorem 13.49 (Regularization of martingales in continuous time)**.** *Let $I = [0,\infty)$ and $\mathcal{X} = (X_t)_{t\in I}$ be a sub-martingale. Further, $\mathcal{Y} = (Y_t)_{t\in I\cap\mathbb{Q}}$ with $Y_t = X_t$ for $t \in I \cap \mathbb{Q}$. Then, with $(\mathcal{G}_t)_{t\in I}$ from Lemma 12.25, the following holds:*

1. *There is a null set N such that $Y_t^+ := \lim_{s\downarrow t} Y_t$ for all $t \in I$ outside N exists. The process $\mathcal{Z} = (Z_t)_{t\in I}$ with $Z_t = 1_{N^c} Y_t^+$ is a $(\mathcal{G}_t)_{t\in I}$ sub-martingale.*

2. *If $(\mathcal{F}_t)_{t\in I}$ is right-continuous, then $\mathcal{X}$ has a modification with paths in $\mathcal{D}_{\mathbb{R}}([0,\infty))$ if $t \mapsto \mathbf{E}[X_t]$ is right-continuous.*

Proof. Since $(|\mathcal{Y}_t|)_{t\in I\cap\mathbb{Q}}$ is a sub-martingale, $\sup_{t\leq\tau}\mathbf{E}[|Y_t|]<\infty$ for $\tau<\infty$. Thus, according to Theorem 13.29, for each $t\in I$ the limits $Y_{t\pm}, t\in I$ outside a nullset N. This means that $(Z_t)_{t\in I}$ with $Z_t=1_{N^c}Y_t^+$ is right-continuous with left-sided limits. Furthermore, Z_t is measurable with respect to $\sigma(\mathcal{F}_t,\mathcal{N})^+, t\in I$.

We now show that $(Z_t)_{t\in I}$ is a sub-martingale. Let $s<t$ and $s_n\downarrow s$, as well as $t_n\downarrow t$ (and $s_n\leq t$, $n=1,2,...$). Then obviously $Y_{s_m}\leq\mathbf{E}[Y_{t_n}|\mathcal{F}_{s_m}]$ for all m,n. This means that $Z_s\leq\mathbf{E}[Y_{t_n}|\mathcal{F}_{s+}]$ according to Theorem 13.36. Since $\sup_n\mathbf{E}[Y_{t_n}]<\infty$, the sub-martingale $(Y_{t_n})_{n=1,2,...}$ is according to Theorem 13.37 uniformly integrable with $Y_{t_n}\xrightarrow{n\to\infty}_{fs,L^1}Z_t$, and thus $\mathbf{E}[Y_{t_n}|\mathcal{F}_{s+}]\xrightarrow{n\to\infty}_{fs,L^1}\mathbf{E}[Z_t|\mathcal{F}_{s+}]$. From this, $Z_s\leq\mathbf{E}[Z_t|\mathcal{F}_{s+}]=\mathbf{E}[Z_t|\mathcal{G}_s]$.

2. With the same notation, for $t\in I$ and $t_n\downarrow t$ with $t_1,t_2,...\in\mathbb{Q}$,

$$\mathbf{E}[X_{t_n}]=\mathbf{E}[Y_{t_n}],\qquad X_t\leq\mathbf{E}[Y_{t_n}|\mathcal{F}_t].$$

Because of $t_n\downarrow t$, $\lim_{s\downarrow t}\mathbf{E}[X_s]=\mathbf{E}[Z_t]$. Furthermore, due to the right-continuity of $(\mathcal{F}_t)_{t\in I}$ and Theorem 13.37 $X_t\leq\mathbf{E}[Z_t|\mathcal{F}_t]=Z_t$. If $\mathcal{X}$ has a right-continuous modification, then $Z_t=X_t$ is almost certain, and thus $\lim_{s\downarrow t}\mathbf{E}[X_s]=\mathbf{E}[X_t]$, therefore $t\mapsto\mathbf{E}[X_t]$ right-handed. On the other hand, if $t\mapsto\mathbf{E}[X_t]$ is right-handed, then $\mathbf{E}[|Z_t-X_t|]=0$, and thus $Z_t=X_t$ almost surely. Thus $(Z_t)_{t\in I}$ is a right-continuous modification of $\mathcal{X}$. □

Remark 13.50 (Usual conditions). *Let $I=[0,\infty)$. In the following, we will always assume that the filtration $(\mathcal{F}_t)_{t\geq 0}$ is right-continuous and complete. Furthermore, Theorem 13.49 shows that, under these assumptions, for each sub-martingale $\mathcal{X}$ there is a modification with paths in $\mathcal{D}_{\mathbb{R}}([0,\infty))$ if $t\mapsto\mathbf{E}[X_t]$ is right-continuous. We also want to assume this modification of each sub-martingale has paths in $\mathcal{D}_{\mathbb{R}}([0,\infty))$. All this we will summarize and say that the* usual conditions *hold.*

Theorem 13.51 (Martingale convergence theorems for continuous I). *Let $I\subseteq[0,\infty)$ be an interval. Under the usual conditions, the statements of Lemma 13.25, Proposition 13.26, Lemma 13.28, Theorem 13.29, Corollary 13.30, Theorem 13.32, Theorem 13.33, Theorem 13.36 and Theorem 13.37 apply accordingly.*

Proof. Note that all statements already apply in the case of countable index set, e.g. $I\cap\mathbb{Q}$, have already been shown. All statements follow in the continuous case, because under the usual conditions, the process $\mathcal{X}=(X_t)_{t\in I}$, as well as all its limit values, can be uniquely constructed from $(X_t)_{t\in I\cap\mathbb{Q}}$ and its limits can be constructed. □

All martingale convergence theorems are now also shown for the case of continuous index set. The following are the statements of the Optional Sampling (Theorem 13.22) and Optional Stopping Theorem (Proposition 13.19) in the continuous case.

Theorem 13.52 (Optional Sampling Theorem in the continuous case). *Let $I\subseteq[0,\infty)$ be an interval, $S\leq T$ almost surely finite stopping times and $\mathcal{X}=(X_t)_{t\in I}$ a sub-martingale. If either T is bounded or $\mathcal{X}$ is uniformly integrable, then X_T is integrable and $X_S\geq\mathbf{E}[X_T|\mathcal{F}_S]$. Furthermore, Lemma 13.23 is also valid for $I=[0,\infty)$.*

Proof. Without restriction, $I=[0,\infty)$. Let $S_n:=2^{-n}[2^nS+1]$ and $T_n:=2^{-n}[2^nT+1]$ such that $S_n\downarrow S$ and $T_n\downarrow T$ as in Proposition 12.28. With Theorem 13.22 follows $X_{S_m}\leq\mathbf{E}[X_{T_n}|\mathcal{F}_{S_m}]$ for all $m\geq n$. With $m\to\infty$ and Theorem 13.36.2,

$$X_S\leq\mathbf{E}[X_{T_n}|\mathcal{F}_S]. \tag{13.10}$$

If T is almost surely bounded, then $..., X_{T_2}, X_{T_1}$ is a sub-martingale with $\inf_n \mathbf{E}[X_{T_n}] > -\infty$. Therefore, according to Theorem 13.37 it is a uniformly integrable, almost surely and in L^1 convergent sub-martingale with limit X_T. Now follows the statement from (13.10) with $m \to \infty$.

If $\mathcal{X}$ is uniformly integrable, then it converges according to Theorem 13.32 (or Theorem 13.51) as $X_t \xrightarrow{t\to\infty}_{fs,L^1} X_\infty$ with integrable limit X_∞. and $X_s \leq \mathbf{E}[X_\infty|\mathcal{F}_s]$ applies.

As above, first $X_S \leq \mathbf{E}[X_{T_n}|\mathcal{F}_S]$, and the sub-martingale $..., X_{T_2}, X_{T_1}$ converges almost surely and in L^1 against X_T. So the statement applies again because of (13.10).

The proof of Lemma 13.23 applies unchanged. □

Corollary 13.53 (Optional stopping in the continuous case)**.** *Let $I \subseteq [0, \infty)$ be an interval and $\mathcal{X} = (X_t)_{t\in I}$ a (sub, super)-martingale and T an almost surely finite stopping time. Then $\mathcal{X}^T = (X_{T\wedge t})_{t\in I}$ is a (sub, super) martingale.*

Proof. The corollary follows with the Optional Sampling Theorem, since $T \wedge s \leq T \wedge t$, thus $X_{T\wedge s} \leq \mathbf{E}[X_{T\wedge t}|\mathcal{F}_{T\wedge s}] \leq \mathbf{E}[X_{T\wedge t}|\mathcal{F}_s]$. □

14 Markov processes

The simplest stochastic processes $\mathcal{X} = (X_t)_{t\in I}$ are those in which $\mathcal{X}$ is an independent family. We now come to the second simplest dependency structure that occurs in stochastic processes. By a Markov process $\mathcal{X}$ we understand a process in which at time t the future $(X_u)_{u>t}$ depends only on X_t, but not on $(X_s)_{s<t}$. In other words: $(X_s)_{s>t}$ and $(X_s)_{s<t}$ are given independently X_t.

Many of the stochastic processes already introduced are Markov processes and will serve as examples in this section. Throughout this section, let (E, r) be a complete and separable metric space.

14.1 Definition and examples

In this section, we will introduce the notion of conditional independence from Section 11.4 will be needed. Finally, Markov processes are those in which the future – given the present – does not depend on the past. After the introduction of Markov processes and some examples, we will determine in Theorem 14.5, when Gaussian processes are Markov. A central notion will be Markov kernels $\mu^{\mathcal{X}}_{s,t}$, which represent just the transition probabilities between two points in time s and t. Formally equivalent, we introduce operators $T^{\mathcal{X}}_{s,t}$, which indicate how expected values of functions $f(X_t)$ change over time.

Definition 14.1 (Markov process)**.** *Let $(\mathcal{F}_t)_{t\in I}$ be a filtration and $\mathcal{X} = (X_t)_{t\in I}$ an adapted stochastic process.*

1. *The process $\mathcal{X}$ is called* Markov process *if $\mathcal{F}_s$ is independent of X_t given X_s, $s \leq t$. This means that for $A \in \mathcal{B}(E)$ (see Proposition 11.18)*
$$\mathbf{P}(X_t \in A|\mathcal{F}_s) = \mathbf{P}(X_t \in A|X_s) \tag{14.1}$$
or equivalently
$$\mathbf{E}(f(X_t)|\mathcal{F}_s) = \mathbf{E}(f(X_t)|X_s)$$
for all measurable and bounded $f : E \to \mathbb{R}$.

2. *The Markov kernels (or transition kernels)* $\mu^{\mathcal{X}}_{s,t}$ *(from* E *to* E*) of* $\mathcal{X}$ *are given by*

$$\mu^{\mathcal{X}}_{s,t}(X_s, B) = \mathbf{P}(X_t \in B|X_s) = \mathbf{P}(X_t \in B|\mathcal{F}_s).$$

3. *Let* $\mathcal{B}(E)$ *be (not only the Borel's* σ*-algebra on* E*, but also) the set of bounded, measurable functions* $f : E \to \mathbb{R}$*. Then we define for* $s \leq t$ *the transition operator*

$$T^{\mathcal{X}}_{s,t} : \begin{cases} \mathcal{B}(E) & \to \mathcal{B}(E) \\ f & \mapsto x \mapsto \mathbf{E}[f(X_t)|X_s = x] = \int \mu^{\mathcal{X}}_{s,t}(x, dy)f(y). \end{cases}$$

4. *For Markov kernels* μ, ν *from* E *to* E *we define a Markov kernel from* E *to* E^2 *by*

$$(\mu \otimes \nu)(x, A \times B) = \int \mu(x, dy)\nu(y, dz)1_{y\in A, z\in B}$$

and a Markov kernel from E *to* E *by*

$$(\mu\nu)(x, A) = (\mu \otimes \nu)(x, E \times A).$$

Remark 14.2 (Interpretations)**.** 1. *Just as with martingales, the Markov property is formulated with respect to a filtration* $(\mathcal{F}_t)_{t\in I}$*. In the following, however, we will always use* $\mathcal{F}_t = \sigma((X_s)_{s\leq t})$*,* $t \in I$*.*

2. *We want the transition kernels* $(\mu^{\mathcal{X}}_{s,t})_{s\leq t}$ *as regular versions of the conditional expectation of* X_t *given* X_s*. This is possible because* E *is Polish and according to Theorem 11.23, then the regular version of the conditional distribution exists.*

3. *The transition operator* $T^{\mathcal{X}}_{s,t}$ *is best interpreted as follows: Given a function* f *and* X_s*), then* $(T^{\mathcal{X}}_{s,t}f)(X_s)$ *is the expectation of* $f(X_t)$ *at the start in* X_s*. This naturally depends on the value* X_s *so* $T^{\mathcal{X}}_{s,t}f$ *is a function of* X_s*.*

4. *To interpret the Markov kernels* $\mu^{\mathcal{X}}_{s,t} \otimes \mu^{\mathcal{X}}_{t,u}$ *and* $\mu^{\mathcal{X}}_{s,t}\mu^{\mathcal{X}}_{t,u}$ *for* $s \leq t \leq u$ *note the following: It is* $\mu^{\mathcal{X}}_{s,t}\otimes\mu^{\mathcal{X}}_{t,u}(x, A\times B)$ *is the probability, given* $X_s = x$*, that is both* $X_t \in A$ *and* $X_u \in B$*. In addition, under* $\mu^{\mathcal{X}}_{s,t}\mu^{\mathcal{X}}_{t,u}$ *the state at time* t *is integrated out, i.e.* $\mu^{\mathcal{X}}_{s,t}\mu^{\mathcal{X}}_{t,u}(x, B)$ *is the probability, given* $X_s = x$*, that* $X_u \in B$*. (Of course, in the case of a of a Markov process must be equal to* $\mu^{\mathcal{X}}_{s,u}(x, B)$*; see also the Chapman-Kolmogorov equations in Corollary 14.16.)*

Example 14.3 (Markov chains)**.** *(See also example 5.10.) Markov processes* $\mathcal{X} = (X_t)_{t\in I}$ *with at most countable state space* E *are called* Markov chains*. Furthermore, if* $I = \{0, 1, 2, ...\}$*, then the transition kernel* $\mu^{\mathcal{X}}_{t,t+1}$ *is represented by a matrix* $P_{t,t+1} = (p_{t,t+1}(x, y))_{x,y\in E}$ *so that*

$$p_{t,t+1}(x, y) = \mathbf{P}(X_{t+1} = y|X_t = x)$$

and

$$\mu^{\mathcal{X}}_{t,t+1}(x, A) = \sum_{y\in A} p_{t,t+1}(x, y).$$

Further here is

$$(\mu^{\mathcal{X}}_{t,t+1}\otimes\mu^{\mathcal{X}}_{t+1,t+2})(x,A\times B)=\sum_{y\in A,z\in B}p_{t,t+1}(x,y)p_{t+1,t+2}(y,z)$$

and

$$(\mu^{\mathcal{X}}_{t,t+1}\mu^{\mathcal{X}}_{t+1,t+2})(x,A)=\sum_{y\in E,z\in A}p_{t,t+1}(x,y)p_{t+1,t+2}(y,z).$$

For the transition operator $(T^{\mathcal{X}}_{s,t})_{s\leq t}$ can be written $f:E\to\mathbb{R}$ can be written as a restricted vector, namely as $f=(f(x))_{x\in E}$ and thus

$$(T^{\mathcal{X}}_{t,t+1}f)(x)=\sum_{y\in E}\mu^{\mathcal{X}}_{t,t+1}(x,dy)f(y)=\sum_{y\in E}p_{t,t+1}(x,y)f(y),$$

so the application of $T^{\mathcal{X}}_{t,t+1}$ to f corresponds to a multiplication of the matrix $p_{t,t+1}$ with the vector f.

Example 14.4 (Sums and products of independent random variables etc.)**.**

1. *Let $X_1,X_2,...$ be real-valued, almost certainly finite and independent. Then $\mathcal{S}=(S_t)_{t=0,1,2,...}$ with $S_t=\sum_{s=1}^t X_s$ and also $\mathcal{S}=(S_t)_{t=0,1,2,...}$ with $S_t=\prod_{s=1}^t X_s$ Markov processes. The following applies for example for $A\in\mathcal{B}(\mathbb{R})$*

$$\begin{aligned}\mathbf{P}(S_{t+1}\in A|\mathcal{F}_t)&=\int\mathbf{P}(S_t\in A-x,X_{t+1}\in dx|\mathcal{F}_t)\\&=\int 1_{S_t\in A-x}\mathbf{P}(X_{t+1}\in dx)=\mathbf{P}(S_{t+1}\in A|S_t).\end{aligned}$$

   In this case

$$\mu^{\mathcal{S}}_{t,t+1}(x,A)=\mathbf{P}(X_{t+1}\in A-x)$$

   and

$$(T^{\mathcal{S}}_{t,t+1}f)(x)=\mathbf{E}[f(x+X_{t+1})].$$

2. *Let $\mathcal{X}=(X_t)_{t\geq0}$ be a Poisson process with intensity λ. Then $(X_t)_{t\geq0}$ and $(X_{f(t)})_{t\geq0}$ for each growing function f Markov processes, just like $(X_t-\lambda t)_{t\geq0}$. However, $(X_t^2-\lambda\int_0^t(2X_r+1)dr)_{t\geq0}$ is not a Markov process; see also Example 13.46. (Note for the last process: assuming $X_t^2-\lambda\int_0^t(2X_r+1)dr=x$, the process decreases linearly with slope $\lambda(2X_t+1)$. However this slope is not a function of x).*

   Let's look at the Poisson process $\mathcal{X}$. Here the Markov kernels for $x\in\{0,1,2,...\}$ are given as

$$\mu^{\mathcal{X}}_{s,t}(x,A)=\sum_{k\in A\cap\{x,x+1,...\}}e^{-\lambda(t-s)}\frac{(\lambda(t-s))^{k-x}}{(k-x)!},$$

   and the transition operator for $f:\{0,1,2,...\}\to\mathbb{R}$ is bounded

$$(T^{\mathcal{X}}_{s,t}f)(x)=\sum_{k=0}^\infty e^{-\lambda(t-s)}\frac{(\lambda(t-s))^k}{k!}f(x+k)=\mathbf{E}[f(x+P)],$$

   where P is a Poisson distributed random variable with parameters $\lambda(t-s)$.

3. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. Then both $(\mu X_t)_{t\geq 0}$ and $(\mu X_t^2 - \mu t)_{t\geq 0}$ as well as $(\exp(\mu X_t - \mu^2 t/2))_{t\geq 0}$ for $\mu \in \mathbb{R}$ Markov processes (as well as martingales according to example 13.47). For example*

$$\begin{aligned}\mathbf{P}[X_u^2 - u \leq x|\mathcal{F}_t] &= \mathbf{P}[(X_u - X_t)^2 + 2(X_u - X_t)X_t + X_t^2 \leq u + x|\mathcal{F}_t]\\ &= \mathbf{P}[(X_u - X_t)^2 + 2(X_u - X_t)X_t + X_t^2 \leq u + x|X_t] = \mathbf{P}[X_u^2 - u \leq x|X_t].\end{aligned}$$

Let us consider the Brownian motion $\mathcal{X}$. Its Markov kernels is given by

$$\mu_{s,t}^{\mathcal{X}}(x, A) = \frac{1}{\sqrt{2\pi(t-s)}} \int_A \exp\Big(-\frac{(y-x)^2}{2(t-s)}\Big)dy$$

and the transition operator for $f \in \mathcal{B}(\mathbb{R})$

$$(T_{s,t}^{\mathcal{X}} f)(x) = \frac{1}{\sqrt{2\pi(t-s)}} \int \exp\Big(-\frac{y^2}{2(t-s)}\Big)f(x+y)dy = \mathbf{E}[f(x + \sqrt{t-s}Z)],$$

where Z is a $N(0,1)$-distributed random variable.

Theorem 14.5 (Gaussian Markov processes)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Gaussian process. Then, $\mathcal{X}$ is Markov iff*

$$\mathbf{COV}(X_s, X_u) \cdot \mathbf{V}(X_t) = \mathbf{COV}(X_s, X_t) \cdot \mathbf{COV}(X_t, X_u) \tag{14.2}$$

for all $s \leq t \leq u$.

Proof. By subtracting the expected values, we can assume wlog that $\mathbf{E}[X_t] = 0$ holds for all $t \geq 0$. We note that (if $\mathbf{V}(X_t) > 0$) with

$$X_u' = X_u - \frac{\mathbf{COV}(X_t, X_u)}{\mathbf{V}(X_t)} X_t$$

holds that $\mathbf{COV}(X_u', X_t) = 0$. Therefore, X_u' and X_t are independent (and the joint distribution is a normal distribution). In the case $\mathbf{V}(X_t) = 0$ we set $X_u' = X_u$ from which the same follows.

First, let $\mathcal{X}$ be Markov and $s \leq t \leq u$. Then X_s is independent of X_u given X_t, so X_s is also independent of X_u' given X_t. Since X_t and X_u' are independent, we find

$$\begin{aligned}\mathbf{P}(X_s \in A, X_u' \in B) &= \mathbf{E}[\mathbf{P}(X_s \in A|X_t) \cdot \mathbf{P}(X_u' \in B|X_t)]\\ &= \mathbf{E}[\mathbf{P}(X_s \in A|X_t) \cdot \mathbf{P}(X_u' \in B)] = \mathbf{P}(X_s \in A) \cdot \mathbf{P}(X_u' \in B)\end{aligned}$$

and therefore X_s and X_u' are independent. This means that

$$0 = \mathbf{COV}(X_s, X_u') = \mathbf{COV}(X_s, X_u) - \frac{\mathbf{COV}(X_t, X_u)}{\mathbf{V}(X_t)}\mathbf{COV}(X_s, X_t)$$

and (14.2) follows.

Conversely, let $\mathcal{X}$ fulfill (14.2). Then (with the same calculation as above), X_s is independent of X_u' for all $s \leq t$. This means that X_u' is independent of $\mathcal{F}_t = \sigma((X_s)_{s\leq t})$ and

$$\begin{aligned}\mathbf{P}(X_u \in A|\mathcal{F}_t) &= \int \mathbf{P}\Big(X_u' \in dx, \tfrac{\mathbf{COV}(X_t,X_u)}{\mathbf{V}(X_t)} X_t \in A - x|\mathcal{F}_t\Big)\\ &= \int \mathbf{P}\Big(X_u' \in dx, \tfrac{\mathbf{COV}(X_t,X_u)}{\mathbf{V}(X_t)} X_t \in A - x|X_t\Big)\\ &= \mathbf{P}(X_u \in A|X_t).\end{aligned}$$

□

Example 14.6 (Examples of Gaussian Markov processes). *1. We have already shown that a Brownian motion $\mathcal{X}$ is a Markov process. To be on the safe side, also note that in this case for $s \leq t \leq u$*

$$\mathbf{COV}(X_s, X_u) \cdot \mathbf{V}(X_t) = s \cdot t = \mathbf{COV}(X_s, X_t) \cdot \mathbf{COV}(X_t, X_u).$$

2. A fractional Brownian motion with Hurst parameter h is a Gaussian process $\mathcal{X} = (X_t)_{t \geq 0}$ with $\mathbf{E}[X_t] = 0$, $t \geq 0$ and

$$\mathbf{COV}(X_s, X_t) = \frac{1}{2}(t^{2h} + s^{2h} - (t-s)^{2h}).$$

As you can easily calculate, this is only for $h = \frac{1}{2}$ a Markov process. Then $\mathcal{X}$ is the Brownian motion.

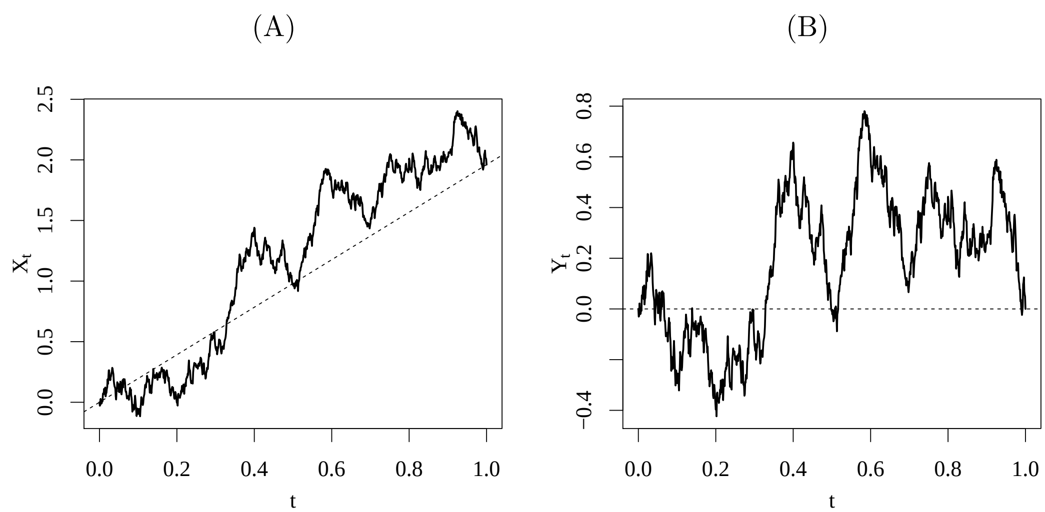


Figure 7: (A) The path of a Brownian motion $\mathcal{X} = (X_t)_{t \in [0,1]}$. (B) The corresponding path of the Brownian bridge $\mathcal{Y} = (Y_t)_{t \in [0,1]}$ with $Y_t = X_t - tX_1$.

The verbal description of Markov processes states that the future of the process is independent of the past, given the present. However, in Definition (14.1) it is only required that individual time points in the future are independent of the past, given the present. The fact that this in fact corresponds to with the verbal description is now shown.

Lemma 14.7 (Extended Markov property). *Let $\mathcal{X} = (X_t)_{t \in I}$ be a Markov process. Then $(X_u)_{u \geq t}$ is independent of $\mathcal{F}_t$ given X_t*

Proof. Let $t = t_0 < t_1 < ... < t_n \in I$ and $A_0, ..., A_n \in E$. Then the following applies

$$\begin{aligned}
\mathbf{P}(X_{t_0} \in A_0, ..., X_{t_n} \in A_n | \mathcal{F}_t) &= \mathbf{E}[1_{X_{t_0} \in A_0}, ..., 1_{X_{t_{n-1}} \in A_{n-1}} \cdot \mathbf{E}[1_{X_{t_n} \in A_n} | \mathcal{F}_{t_{n-1}}] | \mathcal{F}_t] \\
&= \mathbf{E}[1_{X_{t_0} \in A_0}, ..., 1_{X_{t_{n-1}} \in A_{n-1}} \cdot \mathbf{E}[1_{X_{t_n} \in A_n} | X_{t_{n-1}}] | \mathcal{F}_t] \\
&= \mathbf{E}[1_{X_{t_0} \in A_0}, ..., 1_{X_{t_{n-2}} \in A_{n-2}} \cdot \underbrace{\mathbf{E}[1_{X_{t_{n-1}} \in A_{n-1}} \mathbf{E}[1_{X_{t_n} \in A_n} | X_{t_{n-1}}] | \mathcal{F}_{t_{n-2}}]}_{\substack{= \mathbf{E}[1_{X_{t_{n-1}} \in A_{n-1}} \mathbf{E}[1_{X_{t_n} \in A_n} | X_{t_{n-1}}, X_{t_{n-2}}] | X_{t_{n-2}}] \\ = \mathbf{E}[1_{X_{t_{n-1}} \in A_{n-1}} \cdot 1_{X_{t_n} \in A_n} | X_{t_{n-2}}]}} \ \mathcal{F}_t] \\
&= \cdots = \mathbf{E}[1_{X_{t_0} \in A_0} \mathbf{E}[1_{X_1 \in A_1}, ..., 1_{X_{t_n} \in A_n} | X_{t_0}] | \mathcal{F}_t] \\
&= \mathbf{E}[1_{X_{t_0} \in A_0}, ..., 1_{X_{t_n} \in A_n} | X_t] = \mathbf{P}[X_{t_0} \in A_0, ..., X_{t_n} \in A_n | X_t].
\end{aligned}$$

where we have used Proposition 11.18. This shows that $(X_{t_0}, ..., X_{t_n})$ is independent of $\mathcal{F}_t$ given X_t, i.e. the independence on cylinder sets $\{X_{t_0} \in A_0, ..., X_{t_n} \in A_n\}$. This is extended by means of an argument with a Dynkin system to all sets in $\sigma((X_u)_{u \geq t})$. □

A special case is that of a Markov process that is spatially homogeneous. This always behaves in the same way, regardless of its current value is. We have already become familiar with such processes via the Brownian motion and the Poisson process.Equivalent to this is that the process has independent increments, as Lemma 14.9 shows.

Definition 14.8 (Spatially homogeneous Markov process)**.**
Let E be an Abelian group.

1. *A Markov kernel from E to E is called* homogeneous *if $\mu(x, B) = \mu(0, B - x)$ for all $x \in E$ and $B \in \mathcal{B}(E)$. (Here $B - x = \{y - x : y \in B\}$.)*
2. *A Markov process $\mathcal{X}$ is called* spatially homogeneous*, if the Markov kernels $\mu^{\mathcal{X}}_{s,t}$ are homogeneous, $s \leq t$.*
3. *A Markov process $\mathcal{X} = (X_t)_{t \geq 0}$ has independent increments if $X_t - X_s$ is independent of $\mathcal{F}_s$, $s \leq t$.*

Lemma 14.9 (Homogeneity and independent increments)**.** *Let $\mathcal{X} = (X_t)_{t \in I}$ be a Markov process with state space E, where E is an Abelian group. The process $\mathcal{X}$ has independent increments if and only if $\mathcal{X}$ is spatially homogeneous. In this case, the completion of the filtration $(\mathcal{F}_t)_{t \geq 0}$ with $\mathcal{F}_t = \sigma((X_s)_{s \leq t})$ is right-continuous.*

Proof. First, let $\mathcal{X}$ be a spatially homogeneous Markov process, i.e. $\mu^{\mathcal{X}}_{s,t}(x, B) = \mu^{\mathcal{X}}_{s,t}(0, B - x)$ for all $x \in E$ and $B \in \mathcal{B}(E)$. Then,

$$\mathbf{P}(X_t - X_s \in B | \mathcal{F}_s) = \mathbf{P}(X_t \in X_s + B | \mathcal{F}_s) = \mu_{s,t}(X_s, X_s + B) = \mu^{\mathcal{X}}_{s,t}(0, B).$$

Thus $X_t - X_s$ is according to Lemma 11.13 independent of $\mathcal{F}_s$, so $\mathcal{X}$ has independent increments.

Conversely, $\mathcal{X}$ has independent increments. Then $(X_t - X_s)_{t \geq s}$ is also a Markov process with the same Markov kernels and

$$\mu^{\mathcal{X}}_{s,t}(X_s, B) = \mathbf{P}(X_t \in B | \mathcal{F}_s) = \mathbf{P}(X_t - X_s \in B - X_s | \mathcal{F}_s) = \mu^{\mathcal{X}}_{s,t}(0, B - X_s).$$

We now come to the second part of the statement, the right continuity of the filtration generated by $\mathcal{X}$. Let $t \in I$ and $u_1, u_2, ... \in I$ with $u_n \downarrow t$. Wlog we assume that $\mathcal{F}_t$ is complete. We must show that $\mathcal{F}_t^+ = \bigcap_n \mathcal{F}_{u_n} = \mathcal{F}_t$. First of all, $(\mathcal{F}_t, \mathcal{G}_1, \mathcal{G}_2, ...)$ is with $\mathcal{G}_n = \sigma(X_{u_{n-1}} - X_{u_n})$ an independent family. It is $\mathcal{F}_t^+$ independent of $(\mathcal{G}_1, ..., \mathcal{G}_n)$ for each n. Let $A \in \mathcal{F}_t^+$ be. Then, according to Proposition 11.18,

$$\mathbf{P}(A|\mathcal{F}_t) = \mathbf{P}(A|\mathcal{F}_t, \mathcal{G}_1, ..., \mathcal{G}_n) \xrightarrow{n\to\infty} 1_A$$

almost surely by Theorem 13.36 and because 1_A is measurable with respect to $\sigma(\mathcal{F}_t, \mathcal{G}_1, .\mathcal{G}_2, ...)$. In particular, since $\mathcal{F}_t$ is complete, $\mathcal{F}_t^+ \subseteq \mathcal{F}_t \subseteq \mathcal{F}_t^+$. □

14.2 Strong Markov processes

With martingales, we have become familiar with the procedure that a property that applies for fixed times (e.g. $X_s = \mathbf{E}[X_t|\mathcal{F}_s]$) is transferred to stopping times. (This led to the Optional Sampling Theorem, i.e. $X_S = \mathbf{E}[X_T|\mathcal{F}_S]$ for almost surely bounded stopping times $S \leq T$).

The Markov property is initially again a property for fixed points in time, which can be written, for example, as

$$\mathbf{P}(X_{s+t} \in A|\mathcal{F}_s) = \mu^{\mathcal{X}}_{s,s+t}(X_s, A).$$

Replacing the fixed time s in the last equation with a stopping time S leads to strong Markov processes. Most of the processes discussed here belong to this class, however Example 14.14 is an exception.

Definition 14.10 (Strong Markov process). *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Markov process with generated filtration $(\mathcal{F}_t)_{t\in I}$ and progressively measurable. Further let S be a $(\mathcal{F}_t)_{t\in I}$ stopping time. Then $\mathcal{X}$ has the* strong Markov property at S *if*

$$\mathbf{P}(X_{S+t} \in A|\mathcal{F}_S) = \mu^{\mathcal{X}}_{S,S+t}(X_S, A)$$

for $A \in \mathcal{B}(E)$ or equivalent to this

$$\mathbf{E}[f(X_{S+t})|\mathcal{F}_S] = (T^{\mathcal{X}}_{S,S+t} f)(X_S)$$

applies to $f \in \mathcal{B}(E)$. Further, $\mathcal{X}$ is a strong Markov process *if $\mathcal{X}$ has the strong Markov property at all almost surely finite stopping times.*

Proposition 14.11 (Strong Markov at discrete stopping times). *Let $\mathcal{X} = (X_t)_{t\in I}$ be a progressively measurable Markov process with generated filtration $(\mathcal{F}_t)_{t\in I}$. Further let S be an almost surely finite $(\mathcal{F}_t)_{t\in I}$-stopping time, which only assumes discrete (i.e. in particular only countably many) values. Then $\mathcal{X}$ has the strong Markov property for S.*

If I in particular is discrete, then every Markov process $\mathcal{X}$ also has the strong Markov property.

Proof. Let $\{s_1, s_2, ...\}$ be the range of values of S and $f \in \mathcal{B}(E)$ and $A \in \mathcal{F}_S$. Then (since the range of values of S is discrete) $A \cap \{S = s_i\} \in \mathcal{F}_{s_i}$ and

$$\begin{aligned}
\mathbf{E}[f(X_{S+t}), A] &= \sum_i \mathbf{E}[f(X_{S+t}), A \cap \{S = s_i\}] \\
&= \sum_i \mathbf{E}[f(X_{s_i+t}), A \cap \{S = s_i\}] \\
&= \sum_i \mathbf{E}\big[\mathbf{E}[f(X_{s_i+t})|X_{s_i}], A \cap \{S = s_i\}\big] \\
&= \sum_i \mathbf{E}[(T_{s_i,s_i+t}f)(X_{s_i}), A \cap \{S = s_i\}] \\
&= \sum_i \mathbf{E}[(T_{S,S+t}f)(X_S), A \cap \{S = s_i\}] \\
&= \mathbf{E}[(T_{S,S+t}f)(X_S), A].
\end{aligned}$$

Since $(T_{S,S+t}f)(X_S)$ is measurable according to $\mathcal{F}_S$, the assertion follows. □

Theorem 14.12 (Strong Markov with continuous transition operator). *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Markov process with generated filtration $(\mathcal{F}_t)_{t\in I}$ with right-continuous paths. If $T^{\mathcal{X}}_{s,t}f$ is continuous for $f \in \mathcal{C}_b(E)$ and $s \mapsto T^{\mathcal{X}}_{s,s+t}f$ continuous for all $f \in \mathcal{C}_b(E)$ (with respect to the supremum norm on $\mathcal{C}_b(E)$), then $\mathcal{X}$ is a strong Markov process.*

Proof. First, according to Lemma 12.32, the process $\mathcal{X}$ is progressively measurable. Let S be an almost surely finite stopping time, which, according to Proposition 12.28, we replace by stopping times $S_1, S_2, ...$ with $S_n \downarrow S$ so that S_n only takes on discrete values, $n = 1, 2, ...$ Then, because of the right continuity of the paths of $\mathcal{X}$ that $X_{S_n} \xrightarrow{n\to\infty} X_S$ is almost certain and for $f \in \mathcal{C}_b(E)$ is

$$\begin{aligned}
\mathbf{E}[f(X_{S+t})|\mathcal{F}_S] &= \lim_{n\to\infty} \mathbf{E}[\mathbf{E}[f(X_{S_n+t})|\mathcal{F}_{S_n}]|\mathcal{F}_S] \\
&= \lim_{n\to\infty} \mathbf{E}[(T^{\mathcal{X}}_{S_n,S_n+t}f)(X_{S_n})|\mathcal{F}_S] \\
&= \mathbf{E}[(T^{\mathcal{X}}_{S,S+t}f)(X_S)|\mathcal{F}_S] = (T^{\mathcal{X}}_{S,S+t}f)(X_S),
\end{aligned}$$

where the continuity conditions are used in in the third equality. □

Example 14.13 (Poisson process and Brownian motion are strong Markov).

1. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Poisson process with intensity $\lambda \geq 0$. Then $\mathcal{X}$ is strong Markov, because:*
   According to Example 14.4.2, $(T^{\mathcal{X}}_{s,t}f)(x) = \mathbf{E}[f(x + P)]$, where $P \sim Poi(\lambda(t - s))$. Thus $s \mapsto T^{\mathcal{X}}_{s,s+t}f$ is constant. Further, $x \mapsto (T^{\mathcal{X}}_{s,s+t}f)(x)$ is measurable and due to the discrete topology on $\{0, 1, 2, ...\}$ also continuous. The strong Markov property thus follows from Theorem 14.12.

2. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. Then $\mathcal{X}$ is strong Markov, because:*
   According to Example 14.4.3 is $(T^{\mathcal{X}}_{s,t}f)(x) = \mathbf{E}[f(x + \sqrt{t - s}Z)]$, where $Z \sim N(0, 1)$. This means that $s \mapsto T^{\mathcal{X}}_{s,s+t}f$ is constant and $x \mapsto (T^{\mathcal{X}}_{s,s+t}f)(x)$ is constant. Again, the strong Markov property follows from Theorem 14.12.

It is not so easy to specify non-strong Markov processes. However, here is an example.

Example 14.14 (A non-strong Markov process)**.** *Let $T \sim \exp(1)$ be distributed. We further define the stochastic process $\mathcal{X} = (X_t)_{t\geq 0}$ with*

$$X_t = (t-T)^+$$

and completion of the canonical filtration $(\mathcal{F}_t)_{t\geq 0}$. Then for $f \in \mathcal{B}(\mathbb{R})$

$$\mathbf{E}[f(X_{s+t})|\mathcal{F}_s] = \begin{cases} \mathbf{E}[f((t-T)^+)], & \text{if } X_s = 0, \\ f(x+t), & \text{if } X_s > 0. \end{cases}$$

In particular, the right-hand side only depends on X_s and therefore $\mathcal{X}$ is a Markov process with transition operator

$$(T^{\mathcal{X}}_{s,s+t})f(x) = 1_{x=0}\mathbf{E}[f((t-T)^+)] + 1_{x>0}f(x+t).$$

Now consider the random time $S = \inf\{t : X_t > 0\}$ (i.e. $S = T$). According to Proposition 12.30.2, T is an option time and thus, since $\{T = t\}$ is a nullset and $\mathcal{F}_t$ is complete, $\{T \leq t\} = \{T < t\} \cup \{T = t\} \in \mathcal{F}_t$. Thus T is $(\mathcal{F}_t)_{t\geq 0}$ a stopping time. Now,

$$\mathbf{E}[f(X_{S+t})|\mathcal{F}_S] = f(t),$$

since S is measurable according to $\mathcal{F}_S$ and $X_{S+t} = t$ almost surely. On the other hand, $X_S = 0$ and therefore

$$(T^{\mathcal{X}}_{S,S+t}f)(X_S) = (T^{\mathcal{X}}_{S,S+t}f)(0) = \mathbf{E}[f((t-T)^+)].$$

Since the right-hand sides of the last two equations for many $f \in \mathcal{B}(E)$ do not match, $\mathcal{X}$ is not a strong Markov process.

14.3 The distribution of a Markov process

For a Markov process $\mathcal{X}$, the Markov kernels $\mu^{\mathcal{X}}_{s,t}$ and the transition operators $T^{\mathcal{X}}_{s,t}$ are important tools. We will discuss in Theorem 14.17 that a consistency condition (the Chapman-Kolmogorov equations, see Corollary 14.16) is not only necessary but also sufficient for a family of Markov kernels to be Markov kernels for a Markov process.

Lemma 14.15 (Finite-dimensional distributions)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Markov process with $X_t \sim \nu^{\mathcal{X}}_t$ for distributions $\nu^{\mathcal{X}}_t$ on E and Markov kernels $(\mu^{\mathcal{X}}_{s,t})_{s\leq t}$. Then, for $t_0 < \cdots < t_n$*

$$(X_{t_0}, ...X_{t_n}) \sim \nu^{\mathcal{X}}_{t_0} \otimes \mu^{\mathcal{X}}_{t_0,t_1} \otimes \cdots \otimes \mu^{\mathcal{X}}_{t_{n-1},t_n}$$

and

$$\mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot|\mathcal{F}_{t_0}) = (\mu^{\mathcal{X}}_{t_0,t_1} \otimes \cdots \otimes \mu^{\mathcal{X}}_{t_{n-1},t_n})(X_{t_0}, \cdot)$$

Proof. The proof of the first formula is done by induction. For $n = 0$ the statement is clear. If it applies for n, then the following applies for $f \in \mathcal{C}_b(E^{n+2})$

$$\begin{aligned} \mathbf{E}[f(X_{t_0},..., X_{t_{n+1}})] &= \mathbf{E}[\mathbf{E}[f(X_{t_0}, ..., X_{t_{n+1}})|\mathcal{F}_{t_n}]] \\ &= \mathbf{E}\Big[\int f(X_{t_0}, ..., X_{t_n}, x_{n+1})\mu^{\mathcal{X}}_{t_n,t_{n+1}}(X_{t_n}, dx_{n+1})\Big] \\ &= \int \nu^{\mathcal{X}}_{t_0} \otimes \mu^{\mathcal{X}}_{t_0,t_1} \otimes \cdots \otimes \mu^{\mathcal{X}}_{t_n,t_{n+1}}(dx_0, ..., dx_{n+1})f(x_0, ..., x_{n+1}) \end{aligned}$$

so the first formula applies to $n+1$. For the second formula, we note that the right-hand side X_{t_0} is measurable. Furthermore, with Lemma 14.7

$$\mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot | \mathcal{F}_{t_0}) = \mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot | X_{t_0})$$

and for $A \in \mathcal{B}(E)$ and $B \in \mathcal{B}(E^n)$ with the first formula

$$\begin{aligned} &\mathbf{E}[1_{(X_{t_1},...,X_{t_n}) \in B}, X_{t_0} \in A] = \mathbf{P}((X_{t_0}, ..., X_{t_n}) \in A \times B) \\ &\quad = \int_A \nu_{t_0}^{\mathcal{X}}(dx)(\mu_{t_0,t_1}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n,t_{n+1}}^{\mathcal{X}}(x, B) = \mathbf{E}[(\mu_{t_0,t_1}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n,t_{n+1}}^{\mathcal{X}})(X_{t_0}, B), X_{t_0} \in A], \end{aligned}$$

from which the assertion follows. □

Corollary 14.16 (Chapman-Kolmogorov equations)**.** *Let $\mathcal{X}$ be a Markov process with $X_t \sim \nu_t^{\mathcal{X}}$ for distributions $\nu_t^{\mathcal{X}}$ on E, Markov kernels $(\mu_{s,t}^{\mathcal{X}})_{s \leq t}$ and transition operators $(T_{s,t}^{\mathcal{X}})_{s \leq t}$. Then, for $s \leq t \leq u$*

$$\mu_{s,t}^{\mathcal{X}} \mu_{t,u}^{\mathcal{X}} = \mu_{s,u}^{\mathcal{X}}, \tag{14.3}$$

and for $f \in \mathcal{B}(E)$

$$(T_{s,t}^{\mathcal{X}}(T_{t,u}^{\mathcal{X}} f))(X_s) = (T_{s,u}^{\mathcal{X}} f)(X_s) \tag{14.4}$$

$\nu_s^{\mathcal{X}}$-almost surely.

Proof. According to Proposition 14.15, for $\nu_s^{\mathcal{X}}$-almost all X_s for $A \in \mathcal{B}(E)$

$$\begin{aligned} \mu_{s,u}^{\mathcal{X}}(X_s, A) &= \mathbf{P}(X_u \in A | \mathcal{F}_s) = \mathbf{P}((X_t, X_u) \in E \times A | \mathcal{F}_s) \\ &= (\mu_{s,t}^{\mathcal{X}} \otimes \mu_{t,u}^{\mathcal{X}})(X_s, E \times A) = (\mu_{s,t}^{\mathcal{X}} \mu_{t,u}^{\mathcal{X}})(X_s, A) \end{aligned}$$

and for $f \in \mathcal{B}(E)$.

$$\begin{aligned} (T_{s,u}^{\mathcal{X}} f)(X_s) &= \mathbf{E}[f(X_u) | \mathcal{F}_s] = \mathbf{E}\big[\mathbf{E}[f(X_u) | \mathcal{F}_t] | \mathcal{F}_s\big] \\ &= \mathbf{E}[(T_{t,u}^{\mathcal{X}} f)(X_t) | \mathcal{F}_s] = (T_{s,t}^{\mathcal{X}}(T_{t,u}^{\mathcal{X}} f))(X_s). \end{aligned}$$

□

It is clear that for each Markov process there are the Markov kernels $(\mu_{s,t}^{\mathcal{X}})_{s \leq t}$ exist. Conversely, we now show that for every family of Markov kernels $(\mu_{s,t})_{s \leq t}$, which satisfies the Chapman-Kolmogorov equations, there is a Markov process.

Theorem 14.17 (Existence of Markov processes)**.**
Let I be an index set with $\min I = 0$, ν_0 a probability measure on E. Then the following applies:

1. *If $(\mu_{s,t})_{s \leq t}$ is a family of Markov kernels with $\mu_{s,t}\mu_{t,u} = \mu_{s,u}$ for all $s \leq t \leq u$. Then there is a Markov process with starting distribution ν_0 and transition kernels $(\mu_{s,t})_{s \leq t}$.*

2. *If $(T_{s,t})_{s \leq t}$ is a family of transition operators with $T_{s,t}T_{t,u} = T_{s,u}$ for all $s \leq t \leq u$. Then there is a Markov process with starting distribution ν_0 and transition operators $(T_{s,t})_{s \leq t}$.*

Proof. Given $(\mu_{s,t})_{s\leq t}$, it is easy to calculate that

$$(T_{s,t}f)(x) := \int \mu_{s,t}(x, dy)f(y)$$

with $f \in \mathcal{B}(E)$ a family of transition operators $(T_{s,t})_{s\leq t}$, which exactly then (14.4)is fulfilled if $(\mu_{s,t})_{s\leq t}$ fulfills the conditions (14.3) are fulfilled. If the other way around $(T_{s,t})_{s\leq t}$ is given, then defines

$$\mu_{s,t}(x, A) = (T_{s,t}1_A)(x)$$

a family of Markov kernels that (14.3) is fulfilled iff $(T_{s,t})_{s\leq t}$ fulfills the condition (14.4). It is therefore sufficient to show 1.For this we first define the measures for $t_1 < ... < t_n$ with $\{t_1, ..., t_n\} \subseteq_f I$

$$\nu_{t_1,...,t_n} = \nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \otimes \mu_{t_{n-1},t_n}.$$

To show that $(\nu_{t_1,...,t_n})_{\{t_1,...,t_n\}\subseteq_f I}$ a projective family is $J = \{t_1, ..., t_n\}$ and $H = \{t_1, ..., t_{k-1}, t_{k+1}, ..., t_n\}$. Then for $B = B_1 \times \cdots \times B_{k-1} \times B_{k+1} \times \cdots \times B_n \in \mathcal{B}(E^H)$

$$\begin{aligned}
(\pi^J_H)_*\nu_J(B) &= \nu_J((\pi^J_H)^{-1}(B)) \\
&= (\nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \mu_{t_{n-1},t_n})(B_1 \times \cdots \times B_{k-1} \times E \times B_{k+1} \times \cdots \times B_n) \\
&= (\nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \mu_{t_{k-1},t_k}\mu_{t_k,t_{k+1}} \otimes \mu_{t_{k+1},t_{k+2}} \otimes \cdots \mu_{t_{n-1},t_n})(B) \\
&= (\nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \mu_{t_{k-1},t_{k+1}} \otimes \mu_{t_{k+1},t_{k+2}} \otimes \cdots \mu_{t_{n-1},t_n})(B) \\
&= \nu_H(B).
\end{aligned}$$

According to Theorem 5.24 there is a process $\mathcal{X} = (X_t)_{t\in I}$ with the finite-dimensional distributions $(\nu_J)_{J\subseteq_f I}$ and starting distribution ν_0. It remains to show that $\mathcal{X}$ is a Markov process. For this, let $A \in \mathcal{B}(E^J)$ for a $J \subset I$ and $\max J = s \leq t$ and $B \in \mathcal{B}(E)$. Then,

$$\mathbf{P}((X_r)_{r\in J} \in A, X_t \in B) = \nu_{J\cup\{t\}}(A \times B) = \mathbf{E}[\mu_{s,t}(X_s, B), (X_r)_{r\in J} \in A].$$

If $(\mathcal{F}_t)_{t\in I}$ is the filtration generated by $\mathcal{X}$, then the following applies to $A \in \mathcal{F}_s$

$$\mathbf{P}(X_t \in B, A) = \mathbf{E}[\mu_{s,t}(X_s, B), A].$$

From the definition of the conditional expectation, we can read that $\mathbf{P}(X_s \in B|\mathcal{F}_s) = \mu_{s,t}(X_s, B) = \mathbf{P}(X_s \in B|X_s)$. From this the assertion follows. □

Corollary 14.18 (Distribution of Markov processes)**.** *Let ν and $(\mu_{s,t})_{s\leq t}$ be as in Theorem 14.17. Then there is a probability distribution $\mathbf{P}_\nu$ on $\mathcal{B}(E)^I$, such that $\mathbf{P}_\nu$ is the distribution of the Markov process with transition kernels $(\mu_{s,t})_{s\leq t}$ and initial distribution ν. Furthermore, $x \mapsto \mathbf{P}_x := \mathbf{P}_{\delta_x}$ defines a transition kernel from E to $\mathcal{B}(E)^I$ and*

$$\mathbf{P}_\nu = \int \nu(dx)\mathbf{P}_x.$$

Proof. It is easy to calculate that $\mathbf{P}_\nu(A) = \int \nu(dx)\mathbf{P}_x(A)$ applies to cylinder sets A. As usual, one extends this statement to all $A \in \mathcal{B}(E)^I$. □

14.4 Semigroups and generators

time-homogeneous Markov processes play a special role. With these, $\mu_{s,t}^{\mathcal{X}}$ depends only on the time difference $t-s$.

Definition 14.19 (Time-homogeneous Markov processes and their semigroups)**.** *Let I be closed under addition. A Markov process $\mathcal{X}$ is called* time-homogeneous *if there is a family of Markov kernels $(\mu_t)_{t\in I}$ with $\mu_{s,t}^{\mathcal{X}} = \mu_{t-s}$. Then we also write $\mu_t^{\mathcal{X}} = \mu_t$ and denote $(\mu_t^{\mathcal{X}})_{t\in I}$ as* transitionsemigroup[20].

This is (of course) exactly the case if there is a family of transition operators $(T_t)_{t\in I}$ with $T_{s,t}^{\mathcal{X}} = T_{t-s}$. In this case, we write $T_t^{\mathcal{X}} = T_t$ and denote $(T_t^{\mathcal{X}})_{t\in I}$ as operator semigroup.

Remark 14.20 (Transfer to time-homogeneous Markov processes)**.** *Let $\mathcal{X}$ be a time-homogeneous Markov process with transition and operator semigroup $(\mu_t^{\mathcal{X}})_{t\in I}$ and $(T_t^{\mathcal{X}})_{t\in I}$. Then, according to the results from Section 14.3,*

$$(X_{t_0}, ...X_{t_n}) \sim \nu_{t_0}^{\mathcal{X}} \otimes \mu_{t_1-t_0}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n-t_{n-1}}^{\mathcal{X}}$$

and

$$\mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot|\mathcal{F}_{t_0}) = (\mu_{t_1-t_0}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n-t_{n-1}}^{\mathcal{X}})(X_{t_0}, \cdot).$$

In addition, the Chapman-Kolmogorov equations become

$$\begin{aligned} \mu_s^{\mathcal{X}} \mu_t^{\mathcal{X}} &= \mu_{s+t}^{\mathcal{X}}, \\ T_s^{\mathcal{X}} T_t^{\mathcal{X}} &= T_{s+t}^{\mathcal{X}} \end{aligned}$$

for all $s, t \in I$. The strong Markov property is in this case

$$\begin{aligned} \mathbf{P}[X_{S+t} \in A|\mathcal{F}_S] &= \mu_t(X_S, A), \\ \mathbf{E}[f(X_{S+t})|\mathcal{F}_S] &= (T_t f)(X_S) \end{aligned}$$

for all almost surely finite stopping times S, $A \in \mathcal{B}(E)$ or $f \in \mathcal{B}(E)$.

Remark 14.21 (Semigroup property)**.** *Let $(\mu_t^{\mathcal{X}})_{t\in I}$ be the transition semigroup and $(T_t^{\mathcal{X}})_{t\in I}$ the operator semigroup of a time-homogeneous Markov process $\mathcal{X}$. Then, by the Chapman-Kolmogorov equations*

$$\begin{aligned} \mu_s^{\mathcal{X}} \mu_t^{\mathcal{X}} &= \mu_{s+t}^{\mathcal{X}}, \\ T_s^{\mathcal{X}} T_t^{\mathcal{X}} &= T_{s+t}^{\mathcal{X}} \end{aligned}$$

for all $s, t \in I$. For this reason, one speaks of (commutative) transition and operator semigroups.

Certain properties of operator semigroups often facilitate proofs. This leads to the concept of the Feller semigroup. To save us save paperwork, we use the distributions $\mathbf{P}_x$ from Corollary 14.18 and denote the expected value with respect to this distribution with $\mathbf{E}_x$.

Definition 14.22 (Feller semigroup, Feller process)**.** *Let $I = \mathbb{R}_+$.*

[20]A semigroup is a pair $(I, *)$, where $*$ is an associative map $I \times I \to I$

1. *Let $(T_t)_{t\in I}$ be a family of operators with $T_t : \mathbf{B}(E) \to \mathcal{B}(E)$. This is called an* operator semigroup *if $T_t(T_s f) = T_{t+s} f$ for all $f \in \mathcal{B}(E)$. Such a semigroup is called*
   (a) positive *if $T_t f \geq 0$ if $f \geq 0$ for all $t \in I$,*
   (b) contraction *if $0 \leq T_t f \leq 1$ for $0 \leq f \leq 1$ for a*
   (c) conservative *if $T_t 1 = 1$ for all $t \in I$,*
   (d) strongly continuous *if $||T_t f - f||_\infty \xrightarrow{t\to 0} 0$ for all $f \in \mathcal{C}_b(E)$.*
   (e) Feller semigroup *if $T_t f(x) \xrightarrow{t\to 0} f(x)$ for $x \in E$ and $f \in \mathcal{C}_b(E)$ and $T_t f \in \mathcal{C}_b(E)$ for all $f \in \mathcal{C}_b(E)$ and $t \in I$.*
2. *A time-homogeneous Markov process $\mathcal{X} = (X_t)_{t\in\mathcal{I}}$ is called* Feller process *if its operator semigroup $(T_t^{\mathcal{X}})_{t\in I}$ is a Feller semigroup.*

Remark 14.23 (Probabilistic properties of Feller processes)**.** *Let $I = \mathbb{R}_+$ and $(T_t^{\mathcal{X}})_{t\in I}$ be the operator semigroup of a Markov process $\mathcal{X} = (X_t)_{t\in I}$.*

1. *The semigroup $(T_t^{\mathcal{X}})_{t\in I}$ is conservative and a positive contraction.*
   Indeed: Of course, $T_t^{\mathcal{X}} 1(x) = \mathbf{E}_x[1] = 1$, which shows the conservativeness of $(T_t^{\mathcal{X}})_{t\in I}$. Similarly, one writes for $f \in \mathcal{B}(E)$ with $0 \leq f \leq 1$
   $$T_t^{\mathcal{X}} f(x) = \mathbf{E}_x[f(x)] \leq \mathbf{E}_x[1] = 1$$
   and thus $(T_t^{\mathcal{X}})_{t\in I}$ is a contraction.
2. *Let $X_0 = x$. Then $T_t^{\mathcal{X}} f(x) \xrightarrow{t\to 0} f(x)$ for all $f \in \mathcal{C}_b(E)$ if and only if $X_t \xrightarrow{t\to 0}_p x$.*
   Indeed: '$\Rightarrow$': It follows with $g(y) := r(x, y) \wedge 1$ that $\mathbf{E}_x[r(x, X_t) \wedge 1] = T_t^{\mathcal{X}} g(x) \xrightarrow{t\to 0} g(x) = 0$, which shows the claimed convergence. '$\Leftarrow$': $X_t \xRightarrow{t\to 0} x$ applies and thus according to the definition of weak convergence for $f \in \mathcal{C}_b(E)$ in particular $T_t^{\mathcal{X}} f(x) = \mathbf{E}_x[f(X_t)] \xrightarrow{t\to\infty} \mathbf{E}_x[f(x)] = f(x)$.

Lemma 14.24 (Poisson process and Brownian motion are Feller)**.** *Both the Poisson process (with rate $\lambda \geq 0$) and the Brownian motion are Feller processes.*

Proof. Let $\mathcal{X}^x = (X_t^x)_{t\geq 0}$ be a Poisson process and $\mathcal{Y}^y = (Y_t^y)_{t\geq 0}$ a Brownian motion, each started in $x \in \mathbb{R}$ and $y \in \mathbb{R}$. The following applies $\mathcal{X}^x \stackrel{d}{=} x + \mathcal{X}^0$ and $\mathcal{Y}^y \stackrel{d}{=} y + \mathcal{Y}^0$. Then $X_t^x \sim N(x, t)$ and $Y_t^y \sim y + \text{Poi}(t\lambda)$. In particular, obviously $X_t \xrightarrow{t\to 0}_p x, Y_t \xrightarrow{t\to 0}_p$. Therefore, $T_t^{\mathcal{X}} f(x) \xrightarrow{t\to 0} f(x)$ and $T_t^{\mathcal{Y}} f(y) \xrightarrow{t\to 0} f(y)$ for $f \in \mathcal{C}_b(\mathbb{R})$ according to remark 14.23.2 Further,

$$T_t^{\mathcal{X}} f(x) = \mathbf{E}_x[f(X_t)] = \mathbf{E}_0[f(x + X_t)] \xrightarrow{x\to x'} \mathbf{E}_0[f(x' + X_t)] = T_t^{\mathcal{X}} f(x')$$

and analogously for the process $\mathcal{Y}$. From this follow all assertions. □

For concrete Markov processes, semigroups are usually difficult to specify. (However, see the exceptions of the Poisson process and the Brownian motion from example 14.4). It is easier to define what happens in an infinitesimally short time. This is described by the generator of the operator semigroup.

Definition 14.25 (Generator of a semigroup)**.** *Let $I = [0, \infty)$, $\mathcal{X} = (X_t)_{t\in I}$ be a time-homogeneous Markov process with operator semigroup $(T_t^{\mathcal{X}})_{t\in I}$. Then the* generator *of $\mathcal{X}$ (or of its operator semigroup) is defined as*

$$(G^{\mathcal{X}} f)(x) = \lim_{t\to 0} \frac{\mathbf{E}_x[f(X_t) - f(x)]}{t} = \lim_{t\to 0} \frac{1}{t}((T_t^{\mathcal{X}} f)(x) - f(x)),$$

for all f for which the limit value exists. The set of functions f for which $(G^{\mathcal{X}} f)(x)$ exists for all $x \in E$ exists is the domain of $G^{\mathcal{X}}$ *and is denoted by $\mathcal{D}(G^{\mathcal{X}})$.*

Example 14.26 (Generator for Poisson process and Brownian motion)**.**

1. *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Poisson process with parameter λ and $G^{\mathcal{X}}$ its generator. Then,*

$$(G^{\mathcal{X}} f)(x) = \lambda(f(x+1) - f(x)$$

for $x \in \mathbb{N}$ and $f \in \mathcal{B}(\mathbb{N})$.

Because we calculate, if P_t is a Poisson distributed random variable with parameter λt

$$\begin{aligned}(G^{\mathcal{X}} f)(x) &= \lim_{t\to 0} \frac{1}{t}(\mathbf{E}_x[f(x+P_t) - f(x)]) = \lim_{t\to 0} \frac{1}{t} \sum_{k=1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!}(f(x+k) - f(x)) \\ &= \lim_{t\to 0} \lambda \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{(k+1)!}(f(x+1+k) - f(x)) \\ &= \lambda(f(x+1) - f(x))\end{aligned}$$

due to dominated convergence.

2. *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Brownian motion and $G^{\mathcal{X}}$ its generator. Then*

$$(G^{\mathcal{X}} f)(x) = \tfrac{1}{2} f''(x)$$

for $x \in \mathbb{R}$ and $f \in \mathcal{C}_b^2(\mathbb{R})$, the set of bounded, twice continuously differentiable functions with with bounded derivatives.

Because we calculate, if Z is a $N(0,1)$-distributed random variable with the Taylor approximation and a random variable Y with $|Y| \le |Z|$

$$\begin{aligned}(G^{\mathcal{X}} f)(x) &= \lim_{t\to 0} \frac{1}{t}(\mathbf{E}_x[f(x+\sqrt{t}Z) - f(x)]) \\ &= \lim_{t\to 0} \frac{1}{t}(\mathbf{E}_x[f'(x)\sqrt{t}Z + \tfrac{1}{2} f''(x) t Z^2 + \tfrac{1}{2}(f''(x+\sqrt{t}Y) - f''(x)) t Z^2]) \\ &= \tfrac{1}{2} f''(x) + \lim_{t\to\infty} \mathbf{E}[\tfrac{1}{2}(f''(x+\sqrt{t}Y) - f''(x)) Z^2] = \tfrac{1}{2} f''(x)\end{aligned} \tag{14.5}$$

by dominated convergence.

We calculate analogously: If $\mathcal{X} = (X_t)_{t\in I}$ with $X_t = (X_t^1, ..., X_t^d)$ is a d-dimensional Brownian motion. Then,

$$(G^{\mathcal{X}} f)(x) = \frac{1}{2} \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2}(x)$$

for $x \in \mathbb{R}^d$ and $f \in \mathcal{C}_b^2(\mathbb{R}^2)$.

Remark 14.27 (Feller semigroups and strong continuity)**.** *If E is at least locally compact, one can – if one replaces $\mathcal{C}_b(E)$ by $\mathcal{C}_0(E)$, the continuous functions vanishing at infinity – at least show that every Feller semigroup is strongly continuous. This makes it easier in some proofs to verifying the uniform convergence for strong continuity. In particular, according to Lemma 14.24 the (Feller) semigroups of the Poisson process and Brownian motion are strongly continuous.*

Lemma 14.28 (Relationship between operator semigroup and generator)**.** *Let $\mathcal{X}$ be a Feller process with operator semigroup $(T_t^{\mathcal{X}})_{t\in I}$. Further, let $G^{\mathcal{X}}$ be the generator of $\mathcal{X}$ and $\mathcal{D} \subseteq \mathcal{D}(G^{\mathcal{X}})$ with $G^{\mathcal{X}}(\mathcal{D}) \subseteq \mathcal{C}_b(E)$. For $f \in \mathcal{C}_b(E)$ is then $\int_0^t (T_s^{\mathcal{X}} f)ds \in \mathcal{D}(G^{\mathcal{X}})$ with*

$$(T_t^{\mathcal{X}} f)(x) - f(x) = \Big(G^{\mathcal{X}}\Big(\int_0^t (T_s^{\mathcal{X}} f)ds\Big)\Big)(x) \tag{14.6}$$

and for $f \in \mathcal{D}$ and $t \geq 0$ is also $T_t^{\mathcal{X}} f \in \mathcal{D}(G^{\mathcal{X}})$ and the following applies

$$\begin{aligned} G^{\mathcal{X}}(T_t^{\mathcal{X}} f) &= T_t^{\mathcal{X}}(G^{\mathcal{X}} f), \\ (T_t^{\mathcal{X}} f)(x) - f(x) &= \int_0^t (T_s^{\mathcal{X}}(G^{\mathcal{X}} f))(x)ds, \end{aligned} \tag{14.7}$$

thus

$$\mathbf{E}_x[f(X_t)] = f(x) + \int_0^t \mathbf{E}[(G^{\mathcal{X}} f)(X_s)]ds.$$

Proof. For $x \in E$ and $f \in \mathcal{C}_b(E)$, $t \mapsto (T_t^{\mathcal{X}} f)(x)$ is continuous. Because of the Feller property of $(T_t^{\mathcal{X}})_{t\in I}$,

$$(T_{t+h}^{\mathcal{X}} f)(x) = (T_t^{\mathcal{X}}(T_h^{\mathcal{X}} f))(x) = (T_h^{\mathcal{X}}(T_t^{\mathcal{X}} f))(x).$$

For the first equation,

$$\begin{aligned} \frac{1}{h}\mathbf{E}_x\Big[\int_0^t (T_s^{\mathcal{X}} f)(X_h) - (T_s^{\mathcal{X}} f)(x)ds\Big] &= \frac{1}{h}\Big(\int_0^t (T_{s+h}^{\mathcal{X}} f)(x) - (T_s^{\mathcal{X}} f)(x)ds\Big) \\ &= \frac{1}{h}\Big(\int_h^{t+h} (T_s^{\mathcal{X}} f)(x)ds - \int_0^t (T_s^{\mathcal{X}} f)(x)ds\Big) \\ &= \frac{1}{h}\int_t^{t+h} (T_s^{\mathcal{X}} f)(x)ds - \frac{1}{h}\int_0^h (T_s^{\mathcal{X}} f)(x)ds \\ &\xrightarrow{h\to 0} (T_t^{\mathcal{X}} f)(x) - f(x). \end{aligned}$$

For the other statements, first of all

$$\begin{aligned} \frac{d}{dt}\mathbf{E}_x[f(X_t)] &= \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[f(X_{t+h}) - f(X_t)] \\ &= (T_t^{\mathcal{X}} \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[f(X_h) - f(x)] = (T_t^{\mathcal{X}}(G^{\mathcal{X}} f))(x), \end{aligned}$$

but also

$$\begin{aligned} \frac{d}{dt}\mathbf{E}_x[f(X_t)] &= \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[f(X_{t+h}) - f(X_t)] \\ &= \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[(T_t^{\mathcal{X}} f)(X_h) - (T_t^{\mathcal{X}} f)(x)] = (G^{\mathcal{X}}(T_t^{\mathcal{X}} f))(x), \end{aligned}$$

which shows the first equation. For the second equation, we note that $t \mapsto (T_t^{\mathcal{X}}(G^{\mathcal{X}} f))(x)$ is continuous according to the condition, so

$$(T_t^{\mathcal{X}} f)(x) - f(x) = \int_0^t \frac{d}{ds}\mathbf{E}_x[f(X_s)]ds = \int_0^t (T_s^{\mathcal{X}}(G^{\mathcal{X}} f))(x)ds.$$

□

Corollary 14.29 (Domain is dense)**.** *Let $\mathcal{X}$, $(T_t^{\mathcal{X}})_{t\in I}$ and $G^{\mathcal{X}}$ as in Lemma 14.28 and the conditions in Lemma 14.28 apply with $\mathcal{D} = \mathcal{C}_b(E)$. Furthermore, let $(T_t^{\mathcal{X}})_{t\in I}$ be strongly continuous. Then $\mathcal{D}(G^{\mathcal{X}})$ is dense in $\mathcal{C}_b(E)$ with respect to the supremum norm, i.e. each $f \in \mathcal{C}_b(E)$ can be approximated by functions from $\mathcal{D}(G^{\mathcal{X}})$.*

Proof. For each $f \in \mathcal{C}_b(E)$ the following applies according to the condition

$$\tfrac{1}{t}\int_0^t (T_s^{\mathcal{X}} f)ds \xrightarrow{t\to 0} f$$

with respect to the supremum norm. Since the function on the left-hand side according to (14.6) lie in $\mathcal{D}(G^{\mathcal{X}})$, the assertion is shown. □

Theorem 14.30 (Martingales derived from Markov processes)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Feller process with generator $G^{\mathcal{X}}$ and domain $\mathcal{D}(G^{\mathcal{X}})$. Further let $\mathcal{D} \subseteq \mathcal{D}(G^{\mathcal{X}})$ be such that $G^{\mathcal{X}}(\mathcal{D}) \subseteq \mathcal{C}_b(E)$. Then, for $f \in \mathcal{D}$ both*

$$\Big(f(X_t) - \int_0^t (G^{\mathcal{X}} f)(X_s)ds\Big)_{t\in I}$$

as well as, in the case of $(G^{\mathcal{X}} f)/f \in L$

$$\Big(f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}} f)(X_s)}{f(X_s)}ds\Big)\Big)_{t\in I}$$

are martingales.

Proof. Let $t \geq s$. For the first process, we note

$$\begin{aligned}
&\mathbf{E}\Big[f(X_t) - f(X_s) - \int_s^t (G^{\mathcal{X}} f)(X_r)dr\Big|\mathcal{F}_s\Big] \\
&\qquad = \mathbf{E}\Big[f(X_t) - f(X_s) - \int_s^t (G^{\mathcal{X}} f)(X_r)dr\Big|X_s\Big] \\
&\qquad = (T_{t-s}f)(X_s) - f(X_s) - \int_s^t (T_{r-s}(G^{\mathcal{X}} f))(X_s)dr = 0
\end{aligned}$$

according to Lemma 14.28. Furthermore,

$$\begin{aligned}
&\mathbf{E}_x\Big[f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}} f)(X_r)}{f(X_r)}dr\Big) - f(X_s)\exp\Big(-\int_0^s \frac{(G^{\mathcal{X}} f)(X_r)}{f(X_r)}dr\Big)\Big|\mathcal{F}_s\Big] \\
&\quad = \mathbf{E}_x\Big[f(X_t)\exp\Big(-\int_s^t \frac{(G^{\mathcal{X}} f)(X_r)}{f(X_r)}dr\Big) - f(X_s)\Big|X_s\Big]\cdot\exp\Big(-\int_0^s \frac{(G^{\mathcal{X}} f)(X_r)}{f(X_r)}dr\Big)
\end{aligned}$$

and

$$\frac{d}{dt}\mathbf{E}_{X_s}\Big[f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)\Big]$$
$$= \mathbf{E}_{X_s}\Big[(G^{\mathcal{X}}f)(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)$$
$$- f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)\frac{(G^{\mathcal{X}}f)(X_t)}{f(X_t)}\Big] = 0.$$

Again, integration from s to t provides the assertion. □

Example 14.31 (Ordinary differential equation)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ with values in $\mathbb{R}^d$ which is a solution of the ordinary differential equation*

$$\frac{d}{dt}X_t = g(X_t)$$

where $g = (g_i)_{i=1,\dots,d} : \mathbb{R}^d \to \mathbb{R}^d$ is a Lipshitz function. Then $\mathcal{X}$ is deterministic, but can also be regarded as homogeneous in time (because g does not additionally depend on t) Markov process. The generator of $\mathcal{X}$ is calculated for $f \in \mathcal{C}_b^1(\mathbb{R}^d)$ and $X_0 = x$ as

$$(G^{\mathcal{X}}f)(x) = \lim_{t\to 0}\tfrac{1}{t}(f(X_t) - f(x)) = \frac{d}{dt}(f(X_t))\Big|_{t=0} = \sum_{i=1}^d \frac{\partial f}{\partial x_i}(g(x))\cdot g_i(x) = (\nabla f)(g(x))\cdot g(x).$$

Example 14.32 (Poisson process and Brownian motion)**.** *In the following, let $f_n(x) = xe^{-x/n}$, i.e. $f_n \in \mathcal{C}_b(\mathbb{R}_+)$ and $g_n(x) = x^2e^{-x/n}$ such that $f_n(x) \xrightarrow{n\to\infty} f(x)$ and $g_n(x) \xrightarrow{n\to\infty} g(x)$ with $f(x) = x$ and $g(x) = x^2$.*

1. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Poisson process with rate λ. Thus, according to theorem 14.30 and Example 14.26*

$$\big(X_t \wedge n - \int_0^t \lambda 1_{X_s\leq n-1}ds\big)_{t\geq 0}$$

is a martingale. Since X_t is integrable, it follows from dominated convergence that

$$\big(X_t - \lambda t\big)_{t\geq 0}$$

is a martingale. Analogously, one concludes (from the integrability of X_t^2 that

$$\big(X_t^2 - \lambda\int_0^t (X_s+1)^2 - X_s^2 ds\big)_{t\geq 0}$$

is a martingale. See also example 13.46.

2. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. From the integrability of X_t, X_t^2 and $e^{\mu X_t}$, one concludes from Theorem 14.30 that because of $G^{\mathcal{X}}h(x) = \frac{1}{2}h''(x)$*

$$\Big(X_t - \tfrac{1}{2}\int_0^t id''(X_s)ds\Big)_{t\geq 0} = (X_t)_{t\geq 0},$$
$$\Big(X_t^2 - \tfrac{1}{2}\int_0^t (id^2)''(X_s)ds\Big)_{t\geq 0} = (X_t^2 - t)_{t\geq 0},$$
$$\Big(\exp\Big(\mu X_t - \tfrac{1}{2}\int_0^t \frac{(e^{\mu\cdot})''(X_s)}{e^{\mu X_s}}ds\Big)\Big)_{t\geq 0} = \Big(\exp\Big(\mu X_t - \tfrac{1}{2}\mu^2 t\Big)\Big)_{t\geq 0}$$

are all martingales. See also example 13.47.

Example 14.33 (Jump processes)**.** *The simplest Markov processes are piecewise constant processes. We now describe the following process: Given $X_s = x$, the process jumps after an exponentially distributed time with rate $\lambda(x)$. The process jumps according to the Markov kernel $\mu(X_s, .)$, i.e. it jumps with probability $\mu(X_s, dy)$ to y.*

Let $\lambda \in \mathcal{B}(E)$ be given with $0 \le \lambda \le \lambda^*$ and the Markov kernel μ from E to E. Further, let $(Y_k)_{k=0,1,2,...}$ be a Markov chain in discrete time with $\mathbf{P}(Y_{k+1} \in A|Y_k) = \mu(Y_k, A)$ for all $A \in \mathcal{B}(E)$. Furthermore, let $T_1, T_2, ...$ be independent and $\exp(1)$-distributed. (We note that this means that T_k/λ according to $\exp(\lambda)$ is distributed). We define the jump process $(X_t)_{t\ge 0}$ by

$$X_t = \begin{cases} Y_0, & t < \dfrac{T_0}{\lambda(Y_0)}, \\ Y_k, & \displaystyle\sum_{j=0}^{k-1} \frac{T_j}{\lambda(Y_j)} \le t < \sum_{j=0}^{k} \frac{T_j}{\lambda(Y_j)}. \end{cases} \tag{14.8}$$

This is a Markov process since it is memoryless by the exponential distribution. To calculate the generator of $\mathcal{X}$, we note that the probability that in time t more than than 2 jumps take place is at most $1 - e^{-\lambda^* t}(1 + \frac{1}{2}\lambda^* t) = \mathcal{O}(t^2)$. So the following applies for $f \in \mathcal{C}_b(E)$

$$\begin{aligned} (G^{\mathcal{X}} f)(x) &= \lim_{t\to 0} \frac{\mathbf{E}_x[f(X_t) - f(x)]}{t} \\ &= \lim_{t\to 0} \tfrac{1}{t}\Big((e^{-\lambda(x)t} - 1) f(x) + \lambda(x) t e^{-\lambda(x)t} \int \mu(x, dy) f(y) \Big) \\ &= \lambda(x) \int \mu(x, dy)\big(f(y) - f(x)\big) dy. \end{aligned} \tag{14.9}$$

We now give some more examples of Markov jump processes on countable state spaces.

Example 14.34 (Master equation)**.** *Let $\mathcal{X} = (X_t)_{t\ge 0}$ be a jump process on a countable state space E, given as in the last example by the functions λ and the Markov kernel $\mu(,.)$. We now set $\lambda(x, y) := \lambda(x)\mu(x, y)$ and denote this quantity as the jump rate from x to y, i.e.*

$$Gf(x) = \sum_{y\in E} \lambda(x, y)(f(y) - f(x))$$

is the generator of $\mathcal{X}$. If you insert the following into this equation function $f(y) = 1_{y=x}$ (for a fixed x) into this equation, you get

$$\begin{aligned} \frac{d}{dt}\mathbf{P}(X_t = x) &= \frac{d}{dt}\mathbf{E}[f(X_t)] = \mathbf{E}[(Gf)(X_t)] \\ &= \mathbf{E}\Big[\sum_{y\in E} \lambda(X_t, y)(1_{y=x} - 1_{X_t=x})\Big] \\ &= \sum_{z\in E} \mathbf{P}(X_t = z) \sum_{y\in E} \lambda(z, y)(1_{x=y} - 1_{x=z}) \\ &= \sum_{z\in E} \lambda(z, x)\mathbf{P}(X_t = z) - \lambda(x, z)\mathbf{P}(X_t = x). \end{aligned} \tag{14.10}$$

This equation is therefore a differential equation for $(\mathbf{P}(X_t = x))_{x\in E}$. The solution of this equation thus provides the exact distribution of X_t. This equation is also known in physics as the master equation*.*

We will now also replace the generator equation with

$$\mathbf{E}_x[f(X_h)] = f(x) + hGf(x) + o(h).$$

write.

Example 14.35 (Branching processes in continuous time)**.** *In a continuous-time branching process (with state space $\mathbb{Z}_+$), each individual dies at rate 1 and is replaced by a random number of random number of offspring (with distribution μ). Here the generator results in*

$$Gf(x) = x\sum_{n=0}^{\infty}\mu(n)(f(x-1+n) - f(x)).$$

For example, for $f_r(x) = r^x$,

$$Gf_r(x) = xr^{x-1}\sum_{n=0}^{\infty}\mu(n)(r^n - r) = xr^{x-1}(g_\mu(r) - r) = (g_\mu(r) - r)\frac{d}{dr}f_r(x).$$

From this you calculate

$$\mathbf{E}_x[r^{X_t}] = r^x + (g_\mu(r) - r)\int_0^t \frac{d}{dr}\mathbf{E}_x[r^{X_s}]ds,$$

so the function $u : (t, r) \mapsto \mathbf{E}_x[r^{X_t}]$ solves the equation

$$\frac{d}{dt}u(t,r) = (g_\mu(r) - r)\frac{d}{dr}u(t,r) \tag{14.11}$$

with the boundary conditions $u(0, r) = r^x, u(t, 1) = 1$.

Example 14.36 (Yule process)**.** *The simplest branching process is the Yule process, in which each individual is replaced by two offspring. In this case $\mu = \delta_2$ and thus $g_\mu(r) = r^2$, so here in* (14.11)

$$\frac{d}{dt}u(t,r) = -r(1-r)\frac{d}{dr}u(t,r)$$

apply. We now claim that this equation in the case $x = 1$ is given by

$$u(t,r) = \frac{e^{-t}r}{1 - r(1 - e^{-t})}.$$

Indeed,

$$(1 - r(1 - e^{-t}))^2\frac{d}{dt}u(t,r) = -(1 - r(1 - e^{-t}))re^{-t} + e^{-2t}r^2 = -r(1-r)e^{-t}$$
$$(1 - r(1 - e^{-t}))^2\frac{d}{dr}u(t,r) = (1 - r(1 - e^{-t}))e^{-t} + e^{-t}r(1 - e^{-t}) = e^{-t}.$$

Since the generating function of the geometric distribution is just

$$g_{geo(p)}(r) = \sum_{n=1}^{\infty}(1-p)^{n-1}pr^n = \frac{pr}{1 - r(1-p)}$$

we have shown that in this case $X_t \sim geo(e^{-t})$. This can also be shown using the master equation

$$\frac{d}{dt}\mathbf{P}(X_t = x) = (x-1)\mathbf{P}(X_t = x-1) - x\mathbf{P}(X_t = x).$$

This is because for $\mathbf{P}(X_t = x) = (1-e^{-t})^{x-1}e^{-t}$

$$\begin{aligned}\frac{d}{dt}(1-e^{-t})^{x-1}e^{-t} &= (x-1)(1-e^{-t})^{x-2}e^{-2t} - (1-e^{-t})^{x-1}e^{-t} \\ &= (1-e^{-t})^{x-2}e^{-t}((x-1)e^{-t} - (1-e^{-t})) = (1-e^{-t})^{x-2}e^{-t}(xe^{-t}-1)\end{aligned}$$

and

$$\begin{aligned}(x-1)\mathbf{P}(X_t = x-1) - x\mathbf{P}(X_t = x) &= (1-e^{-t})^{x-2}e^{-t}(x-1-x(1-e^{-t})) \\ &= (1-e^{-t})^{x-2}e^{-t}(xe^{-t}-1).\end{aligned}$$

Example 14.37 (Probability of extinction of a branching process)**.** *Let $T = T_0$ be the extinction time of a branching process. Then obviously $\mathbf{P}_x(T<\infty) = \mathbf{P}_1(T<\infty)^x$ and*

$$\mathbf{P}_1(T<\infty) = (1-h)\mathbf{P}_1(T<\infty) + h\sum_{n=0}^{\infty}\mu(n)\mathbf{P}_1(T<\infty)^n + o(h)$$

therefore, for $r := \mathbf{P}_1(T<\infty)$ just

$$r = g_\mu(r) \tag{14.12}$$

apply. This equation trivially has the solution $r=1$. In the case $\sum_n n\mu(n) \leq 1$ this is the only solution, which shows that the extinction probability in this case is 1. (This we have already seen through martingale theory). In the case $\mu = p\delta_0 + q\delta_2$ with $q>p$ (i.e. $\sum_n n\mu(n) > 1$) you calculate that (14.12) *applies exactly when $0 = qr^2 - r + p$, i.e. for*

$$r = \frac{1 \pm \sqrt{1-4pq}}{2q} = \frac{1 \pm 2q - 1}{2q}.$$

Since the extinction probability must be less than 1 is therefore just $(p/q)\wedge 1$.

Example 14.38 (Hitting times)**.** *Let $E' \subseteq E$ and $T := T_{E'}$ be the hitting time of E'. We want to calculate the mapping $u : x \mapsto \mathbf{E}_x[T]$. Obviously, $u(x) = \mathbf{E}_x[T] = 0$ for $x \in E'$, so with $\lambda(x) = \sum_y \lambda(x,y)$*

$$\begin{aligned}\mathbf{E}_x[T] &= (1-h\lambda(x))\mathbf{E}_x[T+h] + \sum_y \mathbf{E}_x[T|X_h = y]\cdot\mathbf{P}(X_h = y) \\ &= \mathbf{E}_x[T] + h(1-\lambda(x)\mathbf{E}_x[T] + \sum_y \lambda(x,y)\mathbf{E}_y[T] + O(h^2) \\ &= \mathbf{E}_x[T] + h(1+G\mathbf{E}_\bullet[T]) + O(h^2).\end{aligned}$$

Therefore, the function u must fulfill the equation

$$\begin{aligned}Gu(x) &= -1, \qquad x \notin E', \\ u(x) &= 0, \qquad x \in E'.\end{aligned}$$

Example 14.39 (Birth-death processes)**.** *Markov processes with* $E = \mathbb{Z}_+$ *and transition rate* $\lambda(x, y) = 0$ *for* $|x - y| > 1$ *are called birth-death processes. Typically one denotes*

$$\lambda(n, n+1) =: \lambda_n, \qquad \lambda(n, n-1) =: \mu_n,$$

and thus the generator is given by

$$Gf(n) = \lambda_n(f(n+1) - f(n)) + \mu_n(f(n-1) - f(n)).$$

For the expected hitting times of 0*, i.e.* $u(n) := \mathbf{E}_n[T_0]$*, we now show that*

$$\begin{aligned} u(0) &= 0, \\ u(n) &= \sum_{k=1}^{n} \frac{1}{\mu_k \pi_k} \sum_{j=k}^{\infty} \pi_j \end{aligned}$$

with $\pi_1 = 1$ *and*

$$\pi_i = \prod_{j=2}^{i} \frac{\lambda_{j-1}}{\mu_j}.$$

Then,

$$\begin{aligned} Gu(n) &= \lambda_n \frac{1}{\mu_{n+1}\pi_{n+1}} \sum_{j=n+1}^{\infty} \pi_j - \mu_n \frac{1}{\mu_n \pi_n} \sum_{j=n}^{\infty} \pi_j \\ &= \frac{1}{\pi_n} \sum_{j=n+1}^{\infty} \pi_j - \frac{1}{\pi_n} \sum_{j=n}^{\infty} \pi_j = -1. \end{aligned}$$

15 Properties of Brownian motion

Although we have already introduced Brownian motion in Chapter 12.3, there is still a lot of properties we have not covered yet. We already know that Brownian motion is a martingale, a Gaussian process and a strong Markov process with independent and identically distributed increments and has continuous paths. From this, we can deduce new properties, for example, Blumenthal's 0-1 law, which is an addition to the Kolmogorov's 0-1 law. We will assume that filtrations are complete.

Theorem 15.1 (Blumenthal's 0-1 law)**.** *Let* $\mathcal{X} = (X_t)_{t\geq 0}$ *be a Brownian motion, defined on a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, started in* $x \in \mathbb{R}$*, and* $\mathcal{F}_{0+} := \bigcap_{t>0} \sigma(X_s : s \leq t)$*. Then* $\mathcal{F}_{0+}$ $\mathbf{P}$*-trivial, i.e.* $\mathbf{P}(A) \in \{0, 1\}$ *for* $A \in \mathcal{F}_{0+}$*.*

Let further $\mathcal{T} := \bigcap_{s\geq 0} \sigma(X_t : t \geq s)$ *be the terminal* σ*-algebra of* $\mathcal{X}$*. Then* $\mathcal{T}$ *is* $\mathbf{P}$*-trivial.*

Proof. According to Lemma 14.9, the filtration $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t = \sigma(X_s : s \leq t)$ is right-continuous. From the right continuity in 0 follows $\mathcal{F}_{0+} = \sigma(X_0)$. Since $X_0 = x$ is constant, $\mathcal{F}_{0+}$ must therefore be a $\mathbf{P}$-trivial σ-algebra.

Furthermore, with $\mathcal{X}$ according to theorem 12.19 also $\mathcal{X}' = (X'_t)_{t\geq 0}$ with $X'_t = tX_{1/t}$ a Brownian motion started in 0. With what has just been shown, $\bigcap_{t\geq 0} \sigma(X'_s : s \leq t)$ is $\mathbf{P}$-trivial. It follows, however, that

$$\bigcap_{s\geq 0} \sigma(X_t : t \geq s) = \bigcap_{s\geq 0} \sigma(tX_{1/t} : t \leq s) = \bigcap_{s\geq 0} \sigma(X'_t : t \leq s)$$

is $\mathbf{P}$-trivial, i.e. the assertion. □

Remark 15.2. *Although Blumenthal's 0-1 law looks simple, it may nevertheless be surprising. As we will show later, the Brownian motion – in a suitable sense – can be thought of as the limit of random walks. If we start a random walk in 0, then this random walk either jumps upwards first or downwards first. In particular, for small times they spend either more time in the positive or in the negative.*

Let us define analogously for Brownian motion

$$A_t := \Big\{ \int_0^t 1_{X_s>0} ds \geq \int_0^t 1_{X_s<0} ds \Big\}$$

the set of Brownian paths that have spent more time in the positive by time t and and $A := \bigcap_{t>0} \bigcap_{0<s\leq t} A_s$, which is the set of paths that have spent more time in the positive up to some small time t. Then $A \in \mathcal{F}_{0+}$, so for reasons of symmetry $\mathbf{P}(A) = 0$ must apply. So there is almost certainly no Brownian path that has spent more time in the positive, for very small times.

However, this law is only the prelude to a series of further properties. Here we examine the quadratic variation in Section 15.1, the reflection principle based on the strong Markov property in Section 15.2, the law of the iterated logarithm in Section 15.3, the convergence of random walks against Brownian motion in Section 15.4 and a further connection with random walks in Section 15.5.

15.1 Quadratic variation

The paths of Brownian motion in Figures 5 and 7 look – albeit steady – very *rough*. This property should now be specified.

Definition 15.3 (Variation and quadratic variation)**.** *Let $f \in \mathcal{D}_{\mathbb{R}}([0, \infty))$, $t \geq 0$ and for $n = 1, 2, ...$ let $0 = t_{n,0} < t_{n,1} < \cdots < t_{n,k_n} = t$ be given. We denote $\zeta_n := \{t_{n,0}, ..., t_{n,k_n}\}$ as n-th partition (of $[0,t]$). Assuming $\max_k(t_{n,k} - t_{n,k-1}) \xrightarrow{n\to\infty} 0$, i.e. the partitions exploit the interval $[0,t]$ for $n \to \infty$ better and better. Then we define the ℓ-variation of f with respect to $\zeta = (\zeta_n)_{n=1,2,...}$ as*

$$\nu_{\ell,t,\zeta}(f) := \lim_{n\to\infty} \nu^n_{\ell,t,\zeta}(f)$$

with

$$\nu^n_{\ell,t,\zeta}(f) = \sum_{k=1}^{k_n} |f(t_{n,k}) - f(t_{n,k-1})|^\ell.$$

If the limit value is independent of ζ, we call this the ℓ-variation and denote it by $\nu_{\ell,t}(f)$. The 1-variation is also called variation and the 2-variation is also called quadratic variation.

In addition, ζ is called ascending *if $\zeta_n \subseteq \zeta_{n+1}$ holds for all $n = 1, 2, ...$.*

Lemma 15.4 (Elementary properties of the (quadratic) variation)**.** *Let f be continuous and $t \geq 0$. Then the following applies to ζ as in Definition 15.3*

$$\nu_{\ell,t,\zeta}(f) < \infty \Rightarrow \nu_{\ell+1,t,\zeta}(f) = 0,$$
$$\nu_{\ell+1,t,\zeta}(f) > 0 \Rightarrow \nu_{\ell,t,\zeta}(f) = \infty.$$

Proof. It is sufficient to show the first property. We write

$$\begin{aligned} 0 &\le \lim_{n\to\infty} \sum_{k=1}^{k_n} |f(t_{n,k}) - f(t_{n,k-1})|^{\ell+1} \\ &\le \lim_{n\to\infty} \sup_k |f(t_{n,k}) - f(t_{n,k-1})| \cdot \lim_{n\to\infty} \sum_{k=1}^{k_n} |f(t_{n,k}) - f(t_{n,k-1})|^{\ell} = 0 \end{aligned}$$

since f is uniformly continuous on $[0,t]$. □

Proposition 15.5 (Quadratic variation of Brownian motion)**.** *Let $\mathcal{X} = (X_t)_{t\ge 0}$ be a Brownian motion. Then for ζ as in Definition 15.3,*

$$\nu^n_{2,t,\zeta}(\mathcal{X}) \xrightarrow{n\to\infty}_{L^2} t.$$

If ζ is ascending, then also

$$\nu^n_{2,t,\zeta}(\mathcal{X}) \xrightarrow{n\to\infty}_{fs} t.$$

In particular, the variation of $\mathcal{X}$ is almost surely infinite.

Proof. We write $\nu^n_{2,t,\zeta} := \nu^n_{2,t,\zeta}(\mathcal{X})$. First to the L^2-convergence. It is known that $X_t - X_s \sim \sqrt{t-s}X_1$ is valid for $s \le t$. Therefore

$$\mathbf{E}[\nu^n_{2,\zeta}] = \sum_{k=1}^{k_n} \mathbf{E}[(X_{t_{n,k}} - X_{t_{n,k-1}})^2] = \sum_{k=1}^{k_n} (t_{n,k} - t_{n,k-1})\mathbf{E}[X_1^2] = \sum_{k=1}^{k_n} (t_{n,k} - t_{n,k-1}) = t$$

as well as

$$\mathbf{E}[(\nu^n_{2,\zeta} - t)^2] = \mathbf{V}[\nu^n_{2,\zeta}] = \sum_{k=1}^{k_n} \mathbf{V}[(X_{n,k} - X_{n,k-1})^2] = \sum_{k=1}^{k_n} (t_{n,k} - t_{n,k-1})^2 \mathbf{E}[X_1^4] \xrightarrow{n\to\infty} 0.$$

For the almost sure convergence, we first assume wlog that there is $0 \le t_1, t_2, ... \le t$, so that $\zeta_n = \{t_1, ..., t_n\}$. We will further show that $(\nu^{-n}_{2,\zeta})_{n=...,-2,-1}$ is a (backward) martingale, so that

$$\mathbf{E}[\nu^{n-1}_{2,\zeta} - \nu^n_{2,\zeta} | \nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...] = 0$$

applies to all n. If t'_n and t''_n are the points in time directly before and after t_n in ζ_n,

$$\begin{aligned} \nu^{n-1}_{2,\zeta} - \nu^n_{2,\zeta} &= (X_{t''_n} - X_{t'_n})^2 - (X_{t''_n} - X_{t_n})^2 - (X_{t_n} - X_{t'_n})^2 \\ &= 2(X_{t''_n} - X_{t_n})(X_{t_n} - X_{t'_n}). \end{aligned}$$

We define a second Brownian motion $(\widetilde{X}_t)_{t\ge 0}$ by an independent random variable Y with $\mathbf{P}(Y=1) = \mathbf{P}(Y=-1) = \frac{1}{2}$ and

$$\widetilde{X}_s = X_{s\wedge t_n} + Y(X_s - X_{s\wedge t_n}).$$

This means that $(\widetilde{X}_s)_{0\le s\le t}$ after t_n at X_{t_n} is mirrored. In particular, $(X_{t_n} - X_{t'_n}) = (\widetilde{X}_{t_n} - \widetilde{X}_{t'_n})$ and $(X_{t''_n} - X_{t_n}) = -(\widetilde{X}_{t''_n} - \widetilde{X}_{t_n})$. It is $\nu^k_{2,t,\zeta}(\mathcal{X}) = \nu^k_{2,t,\zeta}(\widetilde{\mathcal{X}})$ for $k = n, n+1, ...$ and thus

$$\mathbf{E}[\nu^{n-1}_{2,t,\zeta}(\mathcal{X}) - \nu^n_{2,t,\zeta}(\mathcal{X}) | \nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...] = \mathbf{E}[\nu^{n-1}_{2,t,\zeta}(\widetilde{\mathcal{X}}) - \nu^n_{2,t,\zeta}(\widetilde{\mathcal{X}}) | \nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...],$$

thus

$$\begin{aligned}
\mathbf{E}[\nu_{2,t,\zeta}^{n-1}(\mathcal{X}) - \nu_{2,t,\zeta}^{n}(\mathcal{X})|\nu_{2,\zeta}^{n}, \nu_{2,\zeta}^{n+1}, ...]
&= \tfrac{1}{2}\big(\mathbf{E}[\nu_{2,t,\zeta}^{n-1}(\mathcal{X}) - \nu_{2,t,\zeta}^{n}(\mathcal{X})|\nu_{2,\zeta}^{n}, \nu_{2,\zeta}^{n+1}, ...] + \mathbf{E}[\nu_{2,t,\zeta}^{n-1}(\widetilde{\mathcal{X}}) - \nu_{2,t,\zeta}^{n}(\widetilde{\mathcal{X}})|\nu_{2,\zeta}^{n}, \nu_{2,\zeta}^{n+1}, ...]\big) \\
&= \mathbf{E}[(X_{t_n''} - X_{t_n})(X_{t_n} - X_{t_n'}) + (\widetilde{X}_{t_n''} - \widetilde{X}_{t_n})(\widetilde{X}_{t_n} - \widetilde{X}_{t_n'})|\nu_{2,\zeta}^{n}, \nu_{2,\zeta}^{n+1}, ...] = 0,
\end{aligned}$$

which shows the desired martingale property. According to Theorem 13.37, $(\nu_{2,t,\zeta}^{n})_{n=1,2,...}$ converges almost surely towards t. □

Corollary 15.6 (Brownian motion has nowhere differentiable paths)**.** *A Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ almost certainly has nowhere differentiable paths. This means that*

$$\mathbf{P}\Big(\lim_{h\to 0} \frac{X_{t+h} - X_t}{h} \text{ exists for some } t > 0\Big) = 0.$$

Proof. It is sufficient to consider the set of paths of Brownian motion whose quadratic variation in time $[0, t]$ is exactly t. (The set of these paths has probability 1, as Proposition 15.5 shows). Each path in this set has positive quadratic variation in every small time interval, i.e. according to Lemma 15.4 infinite variation. Since differentiability requires at least a finite variation in a small time interval the assertion follows. □

15.2 Strong Markov property and reflection principle

In Example 14.13 we saw that Brownian motion is a strong Markov process. This has some useful consequences, as we will now see. The reflection principle is illustrated in Figure 8.

Lemma 15.7 (Reflection principle)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion and T is a stopping time. Then the reflected process is $\mathcal{X}' = (X_t')_{t\geq 0}$ with*

$$X_t' := X_{t\wedge T} - (X_t - X_{t\wedge T}) = \begin{cases} X_t, & t \leq T, \\ 2X_T - X_t, & t > T \end{cases}$$

is also a Brownian motion.

Proof. First of all, it is clear from the construction that $\mathcal{X}'$ has continuous paths. Wlog, we assume that $T < \infty$ holds. We define $\mathcal{Y} = (Y_t)_{t\geq 0}$ by $Y_t := X_{t\wedge T}$ and $\mathcal{Z} = (Z_t)_{t\geq 0}$ by $Z_t := X_{T+t} - X_T$. Then $\mathcal{Z}$ is a Brownian motion, since by the strong Markov property, $(T, \mathcal{Y})$ is independent. This means that $(T, \mathcal{Y}, \mathcal{Z}) \stackrel{d}{=} (T, \mathcal{Y}, -\mathcal{Z})$, since $\mathcal{Z} \stackrel{d}{=} -\mathcal{Z}$. It also follows that $(\mathcal{Y}, \mathcal{Z}^T) \stackrel{d}{=} (\mathcal{Y}, -\mathcal{Z}^T)$ with $\mathcal{Z}^T := (Z_t^T)_{t\geq 0}$, $Z_t^T := Z_{(t-T)^+}$. From this,

$$\mathcal{X} = \mathcal{Y} + \mathcal{Z}^T \stackrel{d}{=} \mathcal{Y} - \mathcal{Z}^T = \mathcal{X}'.$$

This shows the assertion. □

As an application of the reflection principle, we now calculate the distribution of the maximum of a Brownian motion up to a time t. First, however, we note that from Doob's L^p inequality, Proposition 13.26, estimates about the distribution of the maximum. Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion and $\mathcal{M} = (M_t)_{t\geq 0}$ with $M_t = \sup_{s\leq t} X_s$ the maximum process. Then

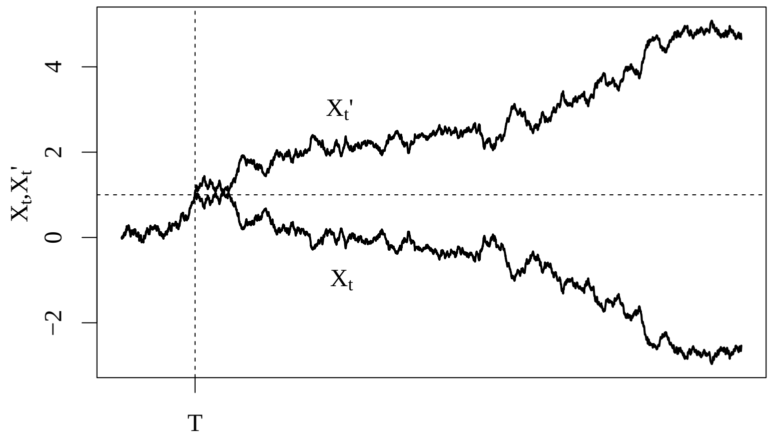


Figure 8:
The reflection principle of Brownian motion states that for a Brownian motion $(X_t)_{t\geq 0}$ the process reflected to T at $x = X_T$ $(X'_t)_{t\geq 0}$ is also a Brownian motion.

it follows from Proposition 13.26 (or the extension to continuous-time processes from Theorem 13.51) with $p = 2$

$$\mathbf{P}(M_t \geq x) \leq \tfrac{1}{x^2}\mathbf{E}[X_t^2] = \tfrac{t}{x^2}.$$

However, especially for large x, this probability is in fact much smaller, as the next result shows.

Theorem 15.8 (Maximum of Brownian motion)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion started in $X_0 = 0$. We define the maximum process $\mathcal{M} = (M_t)_{t\geq 0}$ by $M_t = \sup_{0\leq s\leq t} X_s$. Then,*

$$M_t \stackrel{d}{=} M_t - X_t \stackrel{d}{=} |X_t|.$$

All three random variables have the density

$$x \mapsto \sqrt{\frac{2}{\pi t}} \exp\Big(-\frac{x^2}{2t}\Big) 1_{x\geq 0}.$$

Proof. Let $\varphi_t(x) = \frac{1}{\sqrt{2\pi t}} \exp\big(-\frac{x^2}{2t}\big)$ the density of Brownian motion at time t. Then the density of $|X_t|$ is given by $2\varphi_t(x)1_{x\geq 0}$. So it remains to show that both M_t and $M_t - X_t$ have exactly this density. For this we set $T := T_x = \inf\{s \geq 0 : X_s = x\}$. For $0, y \leq x$, because of Lemma 15.7, if $(X'_t)_{t\geq 0}$ is the process mirrored at T,

$$\mathbf{P}(M_t \geq x, X_t \leq y) = \mathbf{P}(X'_t \geq 2x - y) = \int_{2x-y}^{\infty} \varphi_t(z)dz$$

and thus for $x \geq 0$

$$\begin{aligned}\mathbf{P}(M_t \geq x) &= \mathbf{P}(M_t \geq x, X_t \leq x) + \mathbf{P}(X_t \geq x)\\ &= 2\int_x^{\infty} \varphi_t(z)dz\end{aligned}$$

from which it follows that $M_t \overset{d}{=} |X_t|$. We further calculate

$$\begin{aligned}
\mathbf{P}(M_t - X_t \geq x) &= \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \int_0^\infty \mathbf{P}(z \leq M_t \leq z + \varepsilon, X_t \leq z - x) dz \\
&= \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \int_0^\infty \mathbf{P}(M_t \geq z, X_t \leq z - x) - \mathbf{P}(M_t \geq z + \varepsilon, X_t \leq z - x) dz \\
&= \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \int_0^\infty 2\varphi_t(z + x) dz = \int_x^\infty 2\varphi(z) dz.
\end{aligned}$$

Again, $M_t - X_t \overset{d}{=} |X_t|$ applies. □

Remark 15.9 (The path-valued reflection principle)**.** *The reflection principle only shows the equality of the distributions of $|X_t|, M_t, M_t - X_t$ at a fixed time t. It now remains open whether $(|X_t|)_{t\geq 0} \sim (M_t - X_t)_{t\geq 0}$ is also valid. Even if we do not show this here, this assertion turns out to be correct. (By the way: Surely $(M_t)_{t\geq 0}$ is distributed differently than $(|X_t|)_{t\geq 0}$ or $(M_t - X_t)_{t\geq 0}$, since the last two processes can also decrease, but $(M_t)_{t\geq 0}$ not).*

15.3 The Law of the Iterated Logarithm

We want to determine how a Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ maximally *grows*. This means that we have a function $t \mapsto h_t$ so that

$$0 < \limsup_{t \to \infty} \frac{X_t}{h_t} < \infty. \tag{15.1}$$

We already know from the law of large numbers that $\frac{X_t}{t} \xrightarrow{t\to\infty} 0$. The following also applies

$$\limsup_{t \to \infty} \frac{X_t}{\sqrt{t}} = \infty. \tag{15.2}$$

Indeed: Certainly $\limsup_{t\to\infty} \frac{X_t}{\sqrt{t}}$ is measurable with respect to the terminal σ-algebra of $\mathcal{X}$, i.e. according to Theorem 15.1 almost certainly constant. Suppose, $\limsup_{t\to\infty} \frac{X_t}{\sqrt{t}} \xrightarrow{t\to\infty} \gamma$ for a $0 < \gamma < \infty$. Then it would apply in particular that $\mathbf{P}(\frac{X_t}{\sqrt{t}} > 2\gamma) \xrightarrow{t\to\infty} 0$, in contradiction to the central limit theorem.

The task now is to find a function $t \mapsto h_t$ with $\sqrt{t} \leq h_t \leq t$ so that (15.1) applies. This is determined by the *iterated logarithm* as follows:

Theorem 15.10 (Iterated logarithm for Brownian motion)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. Then*

$$\limsup_{t \to \infty} \frac{X_t}{\sqrt{2t \log\log t}} = \limsup_{t \to 0} \frac{X_t}{\sqrt{2t \log\log 1/t}} = 1, \tag{15.3}$$

almost surely.

Remark 15.11. *For reasons of symmetry, i.e. because $-\mathcal{X}$ is also a Brownian motion,*

$$\liminf_{t \to \infty} \frac{X_t}{\sqrt{2t \log\log t}} = \liminf_{t \to 0} \frac{X_t}{\sqrt{2t \log\log 1/t}} = -1$$

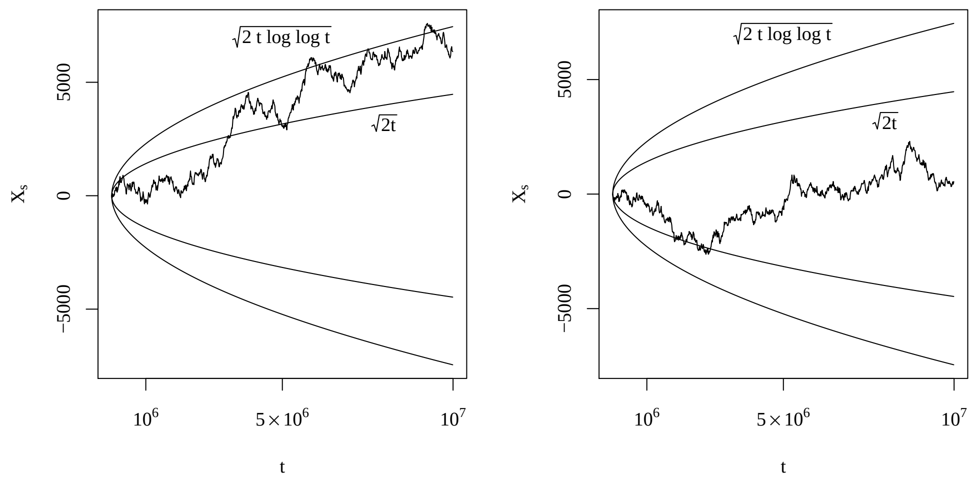


Figure 9:
Here are two paths of a Brownian movement are given. As you can see, the two paths leave the curves $t \mapsto \pm\sqrt{2t}$ much more frequently than the curve $t \mapsto \pm h_t$.

almost surely. For illustration see Figure 9. The fact that $h_t := \sqrt{2t \log\log t}$ is the correct function means that almost every path of the Brownian motion is only finitely often outside the two curves $t \mapsto \pm h_t$ but infinitely often outside the two curves $t \mapsto \pm(1-\varepsilon)h_t$, where $0 < \varepsilon < 1$ is arbitrary.

Proof. First of all, we note that with Theorem 12.19 also $(tX_{1/t})_{t\geq 0}$ is also a Brownian motion. If we apply the statement for the $t \to \infty$ limit, it follows that

$$\limsup_{t\to 0} \frac{X_t}{\sqrt{2t\log\log 1/t}} = \limsup_{t\to\infty} \frac{X_{1/t}}{\sqrt{2\frac{1}{t}\log\log t}} = \limsup_{t\to\infty} \frac{tX_{1/t}}{\sqrt{2t\log\log t}} = 1$$

almost surely. In addition, we write $h_t := h(t) := \sqrt{2t\log\log t}$. The proof for $t \to \infty$ requires a few estimations. We divide the proof into three steps.

Step 1: Estimation of the normal distribution: Let $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ the density of X_1. Then

$$\mathbf{P}(X_1 > x) \leq \tfrac{1}{x}\varphi(x), \tag{15.4}$$
$$\mathbf{P}(X_1 > x) \geq \tfrac{x}{1+x^2}\varphi(x), \tag{15.5}$$

Indeed: $\varphi'(y) = -y\varphi(y)$ and therefore

$$\varphi(x) = \int_x^\infty y\varphi(y)dy \geq x\int_x^\infty \varphi(y)dy = x \cdot \mathbf{P}(X > x),$$

which shows (15.4). For (15.5) we write, quite similarly, $\left(\frac{\varphi(y)}{y}\right)' = -\frac{1+y^2}{y^2}\varphi(y)$, and thus

$$\frac{\varphi(x)}{x} = \int_x^\infty \frac{1+y^2}{y^2}\varphi(y)dy \leq \frac{1+x^2}{x^2}\int_x^\infty \varphi(y)dy = \frac{1+x^2}{x^2}\cdot\mathbf{P}(X > x).$$

In the following we write $a(x) \overset{x\to\infty}{\approx} b(x)$, if $\frac{a(x)}{b(x)} \xrightarrow{x\to\infty} 1$ applies. So, for example, according to what has just been shown

$$\mathbf{P}(X_t > x\sqrt{t}) \overset{x\to\infty}{\approx} \tfrac{1}{x}\varphi(x).$$

2nd step: upper estimate: According to Theorem 15.8 is for $x > 0$

$$\mathbf{P}(\sup_{0\le s\le t} X_s > x\sqrt{t}) = 2 \cdot \mathbf{P}(X_t > x\sqrt{t}) \overset{x\to\infty}{\approx} \tfrac{2}{x}\varphi(x).$$

Now let $r > 1$. We first notice

$$h(r^{n-1}) = \sqrt{\frac{2(\log(n-1) + \log\log r)}{r}}\sqrt{r^n} \overset{n\to\infty}{\approx} \sqrt{\frac{2\log n}{r}}\sqrt{r^n}$$

Now for $c > 0$ with the last two estimates

$$\begin{aligned}\mathbf{P}\big(\sup_{0\le t\le r^n} X_t > ch(r^{n-1})\big) &\overset{n\to\infty}{\approx} 2 \cdot \mathbf{P}\Big(X_{r^n} > c\sqrt{\frac{2\log n}{r}}\sqrt{r^n}\Big)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\sqrt{\frac{2r}{\log n}}\varphi(c\sqrt{2\log n^{1/r}}) \qquad (15.6)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\sqrt{\frac{r}{\pi\log n}}\frac{1}{n^{c^2/r}}.\end{aligned}$$

Therefore, for $c > 1$ and $1 < r < c^2$, the right-hand side of the last equation is summable, so the following holds due to the Borel-Cantelli lemma

$$\mathbf{P}\Big(\limsup_{t\to\infty}\frac{X_t}{h_t} \ge c\Big) \le \mathbf{P}\big(\sup_{0\le t\le r^n} X_t > ch_{r^{n-1}} \text{ for infinitely many } n\big) = 0.$$

Thus '$\le$' follows in (15.3).

3rd step: lower estimate: Let $r > 1$ (typically large) and $c > 0$ (typically close to 1). Define the events

$$A_n := \{X_{r^n} - X_{r^{n-1}} > ch(r^n - r^{n-1})\}.$$

Since $X_{r^n} - X_{r^{n-1}} \sim N(0, r^n - r^{n-1})$, the following applies according to Step 1

$$\begin{aligned}\mathbf{P}(A_n) &= \mathbf{P}\Big(\frac{X_{r^n} - X_{r^{n-1}}}{\sqrt{r^n - r^{n-1}}} > c\frac{h(r^n - r^{n-1})}{\sqrt{r^n - r^{n-1}}}\Big)\\ &= \mathbf{P}\big(X_1 > c\sqrt{2\log\log(r^n - r^{n-1})}\big)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\frac{1}{\sqrt{4\pi\log\log(r^n - r^{n-1})}}\exp\big(-c^2\log\log(r^n - r^{n-1})\big)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\frac{1}{\sqrt{4\pi\log n}}\frac{1}{n^{c^2}}\end{aligned}$$

If $c < 1$, these probabilities cannot be summed up in n. Since the events $A_1, A_2, ...$ are independent, according to the Borel-Cantelli lemma, an infinite number of A_n occur. Thus, for an infinite number of n, if $c < 1$

$$X_{r^n} > ch(r^n - r^{n-1}) + X_{r^{n-1}}.$$

According to the '$\leq$' direction, $X_{r^{n-1}} > -2h(r^{n-1})$ for almost all n, i.e. $\liminf_{n\to\infty} \frac{X_{r^{n-1}}}{h(r^n)} \geq -\liminf_{n\to\infty} \frac{h(r^{n-1})}{h(r^n)} = -\frac{2}{\sqrt{r}}$ is almost certain. Further, $h(r^n - r^{n-1})/h(r^n) \xrightarrow{n\to\infty} 1$ and thus

$$\limsup_{t\to\infty} \frac{X_t}{h_t} \geq \limsup_{n\to\infty} \frac{X_{r^n}}{h(r^n)} \geq \limsup_{n\to\infty} \frac{X_{r^n} - X_{r^{n-1}}}{h(r^n - r^{n-1})} - \frac{2}{\sqrt{r}} \geq c - \frac{2}{\sqrt{r}}.$$

Since $0 < c < 1$ and $r > 0$ were arbitrary, '$\geq$' follows in (15.3). □

15.4 Donsker's Theorem

Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ is a stochastic process with continuous paths. Paul Lévy considered approximated Brownian motion as the path of a random walk, where

$$X_{t+dt} - X_t = \pm\sqrt{dt}, \text{ each with probability } \tfrac{1}{2}.$$

(Of course, this can only be a formal representation, after all it is unclear what $\sqrt{dt}$ is supposed to be). Donsker's Theorem presented here makes the connection between random walks and Brownian motion. It asserts the convergence of random walks against Brownian motion in distribution.

Remark 15.12 (Random walks and Brownian motion)**.** *In this section, $Y_1, Y_2, ...$ are independent and identically distributed random variables with $\mathbf{E}[Y_1] = 0$ and $\mathbf{V}[Y_1] = \sigma^2$ and $\widetilde{X}_{n,t} := \frac{Y_1 + \cdots + Y_{\lfloor nt \rfloor}}{\sqrt{n\sigma^2}}$ for $t \geq 0$ and $\widetilde{\mathcal{X}}_n = (\widetilde{X}_{n,t})_{t\geq 0}$. We know from the central limit theorem that for $t > 0$*

$$\widetilde{X}_{n,t} \xRightarrow{n\to\infty} X_t,$$

where $X_t \sim N(0,t)$ is distributed. Analogously, for $0 < t_1 < \cdots < t_k < \infty$

$$\big(\widetilde{X}_{n,t_1}, \widetilde{X}_{n,t_2} - \widetilde{X}_{n,t_1}, ..., \widetilde{X}_{n,t_k} - \widetilde{X}_{n,t_{k-1}}\big) \xRightarrow{n\to\infty} (X_{t_1}, X_{t_2} - X_{t_1}, ..., X_{t_k} - X_{t_{k-1}}),$$

if $(X_{t_1}, ..., X_{t_k})$ is Brownian motion $\mathcal{X}$ at the points in time $t_1, .., t_k$. Does this now mean already the convergence of the random walks against the Brownian motion, therefore $\mathcal{X}_n \xRightarrow{n\to\infty} \mathcal{X}$? No! For this convergence, we must use both $\mathcal{X}_n$ and $\mathcal{X}$ as random variables with values in a topological space – let's call it $\mathcal{C}$ – where the convergence in distribution is based on the convergence of expected values with respect to continuous, bounded functions $f : \mathcal{C} \to \mathbb{R}$. However, for the uncountable product space, the σ-algebra $\mathcal{B}(\mathbb{R})^{\otimes[0,\infty)}$ is not the Borel σ-algebra on the product space, and we have developed the theory of weak convergence only for the case of probability measures on a Borel's σ-algebra.

In order to formulate the convergence in distribution against the Brownian motion we first need a suitable state space. This is defined as $\mathcal{C} := \mathcal{C}_{\mathbb{R}}([0,\infty))$, provided with the topology of compact convergence (see Definition 15.13). For convergence in this space, we define the linear interpolation of the processes $\widetilde{\mathcal{X}}_n$ so that their paths are also continuous. For this we set

$$X_{n,t} := \widetilde{X}_{n,t} + (nt - \lfloor nt \rfloor)\frac{Y_{\lfloor nt \rfloor + 1}}{\sqrt{n\sigma^2}}. \tag{15.7}$$

and $\mathcal{X}_n = (X_{n,t})_{t\geq 0}$. Now it makes sense to ask whether

$$\mathcal{X}_n \xRightarrow{n\to\infty} \mathcal{X}$$

applies, whereby here the weak convergence with respect to the distributions on $\mathcal{B}(\mathcal{C}_{\mathbb{R}}([0,\infty)))$ is meant here.

is

Definition 15.13 (Uniform convergence on compacta). *Let (E,r) be a metric space. For $f, f_1, f_2, ... \in \mathcal{C}_E([0,\infty))$ let $f_n \xrightarrow{n\to\infty} f$* uniform on compacta *if and only if* $\sup_{0\le s\le t} r(f_n(s), f(s)) \xrightarrow{n\to\infty} 0$ *for all $t > 0$.*

Lemma 15.14 ($\mathcal{C}_E([0,\infty))$ is Polish). *Let E be Polish with complete metric r. Then the topology of uniform convergence on compacta on $\mathcal{C}_E([0,\infty))$ is separable. Moreover,*

$$r_{\mathcal{C}}(f,g) := \int_0^\infty e^{-t}\cdot(1\wedge \sup_{0\le s\le t}|r(f(s),g(s))|)dt$$

is a complete metric on $\mathcal{C}_E([0,\infty))$, which induces this topology. In particular, $\mathcal{C}_E([0,\infty))$ is Polish.

Proof. To show separability, it is sufficient to name a countable class of functions that every function in $\mathcal{C}_E([0,\infty))$ can be locally approximated by such functions on compacta. For this purpose, let $D \subseteq E$ be dense and countable. For every finite sequence $x_1, ..., x_n \in D$ and $t_1, ..., t_n$ let $f = f_{x_1,...,x_n,t_1,...,t_n}$ be a continuous function with $f(t_i) = x_i$. Then $\bigcup_n \{f_{x_1,...,x_n,t_1,...,t_n} : 1x_1, ..., x_n \in D, t_1, ..., t_n \ge 0\}$ is countable and dense in $\mathcal{C}_E([\infty))$.

Now to the metric. Since $t \mapsto \sup_{0\le s\le t} r(f(s), g(s)) \wedge 1$ is monotonically increasing, $r_{\mathcal{C}}(f_n, f) \xrightarrow{t\to\infty} 0$ holds if and only if $\sup_{0\le s\le t} r(f_n(s), f(s)) \xrightarrow{n\to\infty} 0$ for all t is valid. But this is exactly the compact convergence. Let further $f_1, f_2, ...$ be a Cauchy sequence with respect to $r_{\mathcal{C}}$. Then for every $t > 0$ the sequence $f_1, f_2, ...$, restricted to $[0,t]$ is a Cauchy sequence with respect to the supremum norm on $[0,t]$, i.e. uniformly convergent on $[0,t]$. The assertion now follows by means of of a diagonal sequence argument. □

First, we define two types of convergence of stochastic processes that we have just learned about.

Definition 15.15 (Convergence of stochastic processes). *Let $\mathcal{X} = (X_t)_{t\ge0}, \mathcal{X}^1 = (X_t^1)_{t\ge0}, \mathcal{X}^2 = (X_t^2)_{t\ge0}, ...$ stochastic processes with state space E.*

1. *For each choice of $t_1, ..., t_k, k = 1, 2, ...,$ it holds that*

$$(X^n_{t_1}, ..., X^n_{t_k}) \xRightarrow{n\to\infty} (X_{t_1}, ..., X_{t_k}),$$

   we say that the finite-dimensional distributions of $\mathcal{X}^1, \mathcal{X}^2, ...$ converge to those of $\mathcal{X}$ converge and write

$$\mathcal{X}^n \underset{fdd}{\xRightarrow{n\to\infty}} \mathcal{X}.$$

   (Here fdd *stands for* finite dimensional distributions*).*

2. *If the processes $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ have paths in $\mathcal{C}_E([0,\infty))$ and*

$$\mathcal{X}^n \xRightarrow{n\to\infty} \mathcal{X},$$

   where we use $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ as the random variable in $\mathcal{C}_E([0,\infty))$, we say that $\mathcal{X}^1, \mathcal{X}^2, ...$ converges in distribution against $\mathcal{X}$.

The fdd convergence is weaker than the weak convergence of processes. However, if the processes are tight (see Definition 9.14), both terms coincide.

Proposition 15.16 (Weak and fdd convergence)**.** *Let $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ be random variables with values in $\mathcal{C}_E([0,\infty))$. Then are equivalent*

1. $\mathcal{X}^n \xRightarrow{n\to\infty} \mathcal{X}$.
2. $\mathcal{X}^n \xRightarrow[fdd]{n\to\infty} \mathcal{X}$ *and* $\{\mathcal{X}^n : n = 1, 2, ...\}$ *is tight in* $\mathcal{C}_E([0,\infty))$.

Proof. '1.⇒2.': First, from the weak convergence, according to Corollary 9.18 the tightness of $\{\mathcal{X}^n : n = 1, 2, ...\}$ follows. Furthermore, the mappings $f \mapsto (f(t_1), ..., f(t_k))$ are continuous for $t_1, ..., t_k \in [0,\infty)$, so the fdd convergence follows according to Theorem 9.10.

'2.⇒1.': We define the function class

$$\mathcal{M} := \{f \mapsto \varphi(f(t_1), ..., f(t_k)) : t_1, ..., t_k \in [0,\infty), \varphi \in \mathcal{C}_b(E^k)\} \subseteq \mathcal{C}_b(\mathcal{C}_E([0,\infty)))).$$

It is clear that the fdd convergence $\mathcal{X}^n \xRightarrow[fdd]{n\to\infty} \mathcal{X}$ is equivalent to $\mathbf{E}[\varphi(\mathcal{X}^n)] \xrightarrow{n\to\infty} \mathbf{E}[\varphi(\mathcal{X})]$ for all $\varphi \in \mathcal{M}$. Furthermore $\mathcal{M}$ is an algebra and separates points, according to Theorem 9.24 is therefore separating. Now follows the weak convergence follows from Proposition 9.27. □

To show the convergence of processes, after Proposition 15.16 both the convergence of the finite-dimensional distributions as well as the tightness must be shown. In applications, the verification of tightness is usually non-trivial. In particular, one needs to understand how (relatively) compact subsets of $\mathcal{C}_E([0,\infty))$ can be characterized. This is done using the theorem of Arzela-Ascoli's theorem, which is based on the modulus of continuity.

Definition 15.17 (Modulus of continuity)**.** *For $f \in \mathcal{C}_E([0,\infty))$ we define the* modulus of continuity

$$w(f, \tau, h) := \sup\{r(f(s), f(t)) : s, t \le \tau, |t - s| \le h\}.$$

Theorem 15.18 (Arzela-Ascoli)**.** *A set $A \subseteq \mathcal{C}_E([0,\infty))$ is relatively compact if and only if $\{f(t) : f \in A\}$ for all $t \in \mathbb{Q}_+ := [0,\infty) \cap \mathbb{Q}$ is relatively compact in E and for all $\tau > 0$*

$$\lim_{h\to 0} \sup_{f\in A} w(f, \tau, h) = 0. \tag{15.8}$$

Proof. First, let A be relatively compact. Then $\{f(t) : t \in A\}$ must be relatively compact for all $t \ge 0$, otherwise it would be easy to construct a divergent sequence. Furthermore, A is according to Proposition A.9 totally bounded. Further, let $\tau > 0$, $\varepsilon > 0$ and $f_1, ..., f_N$, so that $A \subseteq \bigcup_{i=1}^N B_{\varepsilon/3}(f_i)$. Since $f_1, ..., f_N$ is based on $[0, \tau]$ are uniformly continuous, there is an $h > 0$ with

$$0 \le s, t \le \tau, |t - s| < h \qquad \Longrightarrow \qquad r(f_i(t), f_i(s)) \le \varepsilon/3, \qquad i = 1, ..., N.$$

So, for every $f \in A$ and $s, t \le \tau, |t - s| \le h$, that

$$r(f(s), f(t)) \le \min_{i=1,...,N} r(f(s), f_i(s)) + r(f_i(s), f_i(t)) + r(f_i(t), f(t)) \le \varepsilon$$

and thus

$$w(f, \tau, h) = \sup\{r(f(t), f(s)) : s, t \le \tau, |t-s| \le h\} \le \varepsilon,$$

independent of f. From this follows (15.8).

Conversely, (15.8) applies. It suffices to show that every sequence in A has a subsequence that is Cauchy. By the relative compactness of $\{f(t) : f \in A\}$ for $t \in \mathbb{Q}_+$ it is clear that for each sequence there is a subsequence $f_1, f_2, ...$ such that $f_1(t_i), f_2(t_i), ...$ for all $t_i \in \mathbb{Q}_+$ is a Cauchy sequence (i.e. convergent). Now let $\varepsilon > 0$. According to the condition there is an $h > 0$, so that from $|t-s| \le h$ and $f \in A$ it follows that $r(f(s), f(t)) \le \varepsilon/3$ holds. Further, let $M = \lceil \tau/h \rceil$ and $0 = t_1, ..., t_M \in \mathbb{Q}_+$, so that $|t_{i+1} - t_i| \le h$, $i = 1, ..., M-1$ and $t_M \ge \tau$. Further there is an N such that from $n, m > N$ it follows that $\sup_{t=t_1,...,t_M} r(f_n(t), f_m(t)) \le \varepsilon/3$. Thus, for $0 \le s \le t$

$$r(f_n(s), f_m(s)) \le r(f_n(s), f_n(t_{\lceil s/h \rceil})) + r(f_n(t_{\lceil s/h \rceil}), f_m(t_{\lceil s/h \rceil})) + r(f_m(t_{\lceil s/h \rceil}), f_m(s)) \le \varepsilon.$$

It follows that $f_1, f_2, ...$ is a Cauchy sequence with respect to compact convergence on $[0, t]$, i.e. it converges on this range converges uniformly. A diagonal sequence argument extends this statement to compact convergence. □

Theorem 15.19 (Tightness in $\mathcal{C}_{\mathbb{E}}([0, \infty))$)**.** *Let $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ be random variables with values in $\mathcal{C}_E([0, \infty))$. Then $\mathcal{X}^n \xRightarrow{n\to\infty} \mathcal{X}$ iff $\mathcal{X}^n \xRightarrow[fdd]{n\to\infty} \mathcal{X}$ and*

$$\lim_{h\to 0} \limsup_{n\to\infty} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1] = 0 \tag{15.9}$$

for all $\tau > 0$.

Proof. According to Proposition 15.16 it suffices to show that (15.9) is equivalent to the tightness of the family $(\mathcal{X}^n)_{n=1,2,...}$.

First, let $(\mathcal{X}^n)_{n=1,2,...}$ be tight and $\varepsilon > 0$. Then there is a compact set $K \subseteq \mathcal{C}_E([0, \infty))$ such that $\limsup_{n\to\infty} \mathbf{P}(\mathcal{X}^n \notin K) \le \varepsilon$. For $\tau > 0$ you can use the Arzela-Ascoli Theorem h can be chosen small enough so that $w(f, \tau, h) \le \varepsilon$ applies to $f \in K$. This means that

$$\limsup_{n\to\infty} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1] \le \varepsilon + \sup_{n=1,2,...} \mathbf{P}[w(\mathcal{X}^n, \tau, h) > \varepsilon] \le 2\varepsilon,$$

from which (15.9) follows.

Conversely, (15.9) and $\mathcal{X}^n \xRightarrow[fdd]{n\to\infty} \mathcal{X}$. The mapping w is increasing in h and $w(\mathcal{X}^n, \tau, h) \xrightarrow{h\to 0} 0$ almost certainly for $n = 1, 2, ...$ So $\lim_{h\to 0} \sup_{n=1,2,...} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1] = \lim_{h\to 0} \sup_{n=k,k+1,...} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1]$ for all k, i.e. also $\lim_{h\to 0} \sup_{n=1,2,...} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1] = \lim_{h\to 0} \limsup_{n\to\infty} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1]$. So (15.9) is equivalent to

$$\lim_{h\to 0} \sup_{n=1,2,...} \mathbf{P}[w(\mathcal{X}^n, \tau, h) > \varepsilon] = 0$$

for all $\varepsilon > 0$ and $\tau > 0$. Let $\tau_k = k$ and $\varepsilon > 0$. Then there exist $h_1, h_2, ... > 0$ such that

$$\sup_{n=1,2,...} \mathbf{P}(w(\mathcal{X}^n, \tau_k, h_k) > 2^{-k}) \le 2^{-(k+1)}\varepsilon.$$

Further, let $t_1, t_2, ...$ be a count of $\mathbb{Q}_+$ and $C_1, C_2, ... \subseteq \mathbb{R}$ compact such that

$$\sup_{n=1,2,...} \mathbf{P}(X^n(t_k) \notin C_k) \leq 2^{-(k+1)}\varepsilon.$$

Now we define

$$B := \bigcap_{k=1}^{\infty} \big\{ f \in \mathcal{C}_E([0,\infty)) : f(t_k) \in C_k, w(f, \tau_k, h_k) \leq 2^{-k} \big\}.$$

According to Arzela-Ascoli's theorem, $B \subseteq \mathcal{C}_E([0,\infty))$ is relatively compact. Furthermore,

$$\begin{aligned}
\sup_{n=1,2,...} \mathbf{P}(\mathcal{X}^n \notin B) &\leq \sup_{n=1,2,...} \sum_{k=1}^{\infty} \mathbf{P}(X^n(t_k) \notin C_k) + \mathbf{P}(w(\mathcal{X}^n, \tau_k, h_k) > 2^{-k}) \\
&\leq \sum_{k=1}^{\infty} 2^{-(k+1)}\varepsilon + 2^{-(k+1)}\varepsilon = \varepsilon.
\end{aligned}$$

It follows that $(\mathcal{X}^n)_{n=1,2,...}$ is tight. □

We want to apply the last result to prove the convergence of the random walk against Brownian motion. For this we need one more lemma.

Lemma 15.20. *Let $Y_1, Y_2, ...$ be independent and identically distributed random variables with $\mathbf{E}[Y_1] = 0$ and $\mathbf{V}[Y_1] = \sigma^2 > 0$ and $S_n := Y_1 + \cdots + Y_n$. Then the following applies for $r > 1$*

$$\mathbf{P}(\max_{1 \leq k \leq n} S_k > 2r\sqrt{n}) \leq \frac{\mathbf{P}(|S_n| > r\sqrt{n})}{1 - \sigma^2 r^{-2}}.$$

Proof. We define $T := \inf\{k : |S_k| > 2r\sqrt{n}\}$. Then, since $(S_n)_{n=1,2,...}$ is strongly Markov,

$$\begin{aligned}
\mathbf{P}(|S_n| > r\sqrt{n}) &\geq \mathbf{P}(|S_n| > r\sqrt{n}, \max_{1 \leq k \leq n} S_k > 2r\sqrt{n}) \\
&\geq \mathbf{P}(T \leq n, |S_n - S_T| \leq r\sqrt{n}) \\
&\geq \mathbf{P}(\max_{1 \leq k \leq n} S_k > 2r\sqrt{n}) \cdot \min_{1 \leq k \leq n} \mathbf{P}(|S_k| \leq r\sqrt{n}).
\end{aligned}$$

From Chebychev's inequality,

$$\min_{1 \leq k \leq n} \mathbf{P}(|S_k| \leq r\sqrt{n}) \geq \min_{1 \leq k \leq n} 1 - \frac{\sigma^2 k}{r^2 n} = 1 - \frac{\sigma^2}{r^2}.$$

□

Theorem 15.21 (Donsker's theorem)**.** *Let $Y_1, Y_2, ...$ be independent, identically distributed random variables with $\mathbf{E}[Y_1] = 0$ and $\mathbf{V}[Y_1] = \sigma^2 > 0$, and $\mathcal{X}_n = (X_{n,t})_{t \geq 0}$ given by*

$$X_{n,t} := \frac{1}{\sqrt{n\sigma^2}} \big(Y_1 + \cdots + Y_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor) Y_{\lfloor nt \rfloor + 1} \big)$$

and $\mathcal{X} = (X_t)_{t \geq 0}$ a Brownian motion. Then,

$$\mathcal{X}_n \xrightarrow{n \to \infty} \mathcal{X}.$$

Proof. Let wlog $\sigma^2 = 1$. As stated in Remark 15.12 it holds that $\mathcal{X}_n \xrightarrow[fdd]{n\to\infty} \mathcal{X}$. Therefore, according to Proposition 15.16, the tightness of the family $\{\mathcal{X}_n : n \in \mathbb{N}\}$, so (15.9) from Theorem 15.19, must be proven. We write $S_n := Y_1 + \cdots + Y_n$ in the following. With Lemma 15.20,

$$\begin{aligned}
&\lim_{h\to 0} \frac{1}{h} \limsup_{n\to\infty} \mathbf{P}\Big(\sup_{0\le s\le h} |X_{n,t+s} - X_{n,t}| > \varepsilon\Big) \\
&\quad\le \lim_{h\to 0} \frac{1}{h} \limsup_{n\to\infty} \mathbf{P}\Big(\sup_{k=1,\dots,\lceil nh\rceil} |S_k| > \frac{\varepsilon}{\sqrt{h}}\sqrt{nh}\Big) \\
&\quad\le \lim_{h\to 0} \frac{1}{h} \limsup_{n\to\infty} \mathbf{P}\Big(\frac{|S_{\lceil nh\rceil}|}{\sqrt{nh}} > \frac{\varepsilon}{2\sqrt{h}}\Big) \\
&\quad\le \lim_{h\to 0} \frac{2}{h} \int_{\varepsilon/(2\sqrt{h})}^{\infty} \varphi(x)dx \\
&\quad= \lim_{h\to 0} \frac{2}{h} \frac{2\sqrt{h}}{\varepsilon} \varphi(\varepsilon/(2\sqrt{h})) = 0
\end{aligned}$$

by (15.5), where φ is the density of the $N(0,1)$ distribution. Now let $\delta > 0$ and h be small enough for

$$\limsup_{n\to\infty} \mathbf{P}\Big(\sup_{0\le s\le h} |X_{n,t+s} - X_{n,t}| > \varepsilon\Big) \le \delta h.$$

With this we can write

$$\begin{aligned}
\limsup_{n\to\infty} \mathbf{P}(w(\mathcal{X}_n, \tau, h) > 2\varepsilon) &= \limsup_{n\to\infty} \mathbf{P}(\sup_{0\le t\le \tau-h, 0\le s\le h} |X_{n,t+s} - X_{n,t}| > 2\varepsilon) \\
&\le \limsup_{n\to\infty} \mathbf{P}(\sup\{|X_{n,kh+s} - X_{n,kh}| : k = 0, 1, ..., [\tau/h], 0 \le s \le h\} > \varepsilon) \\
&\le \sum_{k=0}^{[\tau/h]} \limsup_{n\to\infty} \mathbf{P}(\sup\{|X_{n,kh+s} - X_{n,kh}| : 0 \le s \le h\} > \varepsilon) \\
&\le [\tau/h]\delta h \xrightarrow{h\to 0} \tau\delta.
\end{aligned}$$

Since $\delta > 0$ was arbitrary, the result follows (15.9). □

We end this section with a tightness criterion that is often is applicable. It builds on theorem 12.8.

Theorem 15.22 (Kolmogorov-Chentsov criterion for tightness)**.** *Let $\mathcal{X}_1 = (X_1(t))_{t\ge 0}, \mathcal{X}_2 = (X_2(t))_{t\ge 0}, \dots$ are stochastic processes with continuous paths. Assuming $\{X_n(0) : n \in \mathbb{N}\}$ is tight and for every each $\tau > 0$ there are numbers $\alpha, \beta, C > 0$ with*

$$\sup_n \mathbf{E}[r(X_n(s), X_n(t))^\alpha] \le C|t - s|^{1+\beta}$$

for all $0 \le s, t \le \tau$. Then $\{\mathcal{X}_n : n \in \mathbb{N}\}$ is tight in $\mathcal{C}_E([\infty))$.

Proof. Let $0 < \gamma < \beta/\alpha$ be arbitrary. We use the notation from the proof of theorem 12.8, e.g. $\xi_{nk} := \max\{r(X_n(s), X_n(t)) : s, t \in D_k, |t - s| = 2^{-k}\}$. Wlog let $\tau = 1$. Just as in (12.1) we calculate

$$\sum_{k=0}^{\infty} 2^{\alpha\gamma k} \mathbf{E}[\xi_{nk}^\alpha] \le C \sum_{k=0}^{\infty} 2^{(\alpha\gamma-\beta)k}.$$

Since the right-hand side does not depend on n, there is a C' with $\sup_n \mathbf{E}[\xi_{nk}^\alpha] \leq C' 2^{-\alpha\gamma k}$. It is important to realize that $w(\mathcal{X}_n, 1, 2^{-m}) \leq \sum_{k=m}^\infty \xi_{nk}$. From this,

$$\sup_n \mathbf{E}[w(\mathcal{X}_n, 1, 2^{-m})^\alpha \wedge 1] \leq \sup_n \mathbf{E}\Big[\Big(\sum_{k=m}^\infty \xi_{nk}\Big)^\alpha\Big] \leq \sup_n \Big(\sum_{k=m}^\infty \mathbf{E}[\xi_{nk}^\alpha]^{1/\alpha}\Big)^\alpha$$
$$\leq C'\Big(\sum_{k=m}^\infty 2^{-\gamma k}\Big)^{1/\alpha} \xrightarrow{m\to\infty} 0,$$

from which the assertion follows. □

15.5 The Skorohod embedding Theorem

The name Skorohod was already mentioned in the connection between weak and almost sure convergence, see Theorem 9.11. Simply spoken, a sequence of random variables converges weakly iff it converges almost surely in a suitable probability space. If we look again at Donsker's theorem, we can ask ourselves the question as to what the probability space should look like, on which the random walks converge almost surely against a Brownian motion. In other words: how must one define the random walks and the Brownian motion so that both *always* are close together. This is answered by Skorohod's embedding theorem, Theorem 15.26. It allows further conclusions to be drawn about the error, such as the law of the iterated logarithm, Theorem 15.29. The following lemma is fundamental:

Lemma 15.23 (Randomization). *For $w < 0 < z$ let $Y_{w,z}$ be a random variable with state space $\{w, z\}$ with*

$$\mathbf{P}(Y_{w,z} = w) = \frac{z}{z + |w|}$$

and $Y_{w,z} = 0$ for $w, z = 0$. Further, let Y be a real-valued random variable with $\mathbf{E}[Y] = 0$. Then there is a pair of random variables (W, Z) with $W \leq 0, Z \geq 0$, so that Y has the distribution $Y_{W,Z}$.

Proof. We set $c = \mathbf{E}[Y^+] = \mathbf{E}[Y^-]$. Further, let $f : \mathbb{R} \to \mathbb{R}_+$ is measurable with $f(0) = 0$. Then, if $Y \sim \mu$,

$$\begin{aligned}
c \cdot \mathbf{E}[f(Y)] &= \mathbf{E}[Y^+] \cdot \mathbf{E}[f(-Y^-)] + \mathbf{E}[Y^-] \cdot \mathbf{E}[f(Y^+)] \\
&= \int\int (zf(w) + |w|f(z)) 1_{z\geq 0} 1_{w\leq 0} \mu(dw)\mu(dz) \\
&= \int\int (z + |w|)\mathbf{E}[f(Y_{w,z})] 1_{z\geq 0} 1_{w\leq 0} \mu(dw)\mu(dz).
\end{aligned}$$

This means that we define (W, Z) as a random variable with a joint distribution

$$\mu_{W,Z}(dw, dz) = \mu(0)\delta_{0,0}(dw, dz) + \tfrac{1}{c}(z + |w|) 1_{w\leq 0} 1_{z\geq 0} \mu(dw)\mu(dz)$$

can be selected. (It is easy to check that the total mass of this measure is 1). Then,

$$c\mathbf{E}[f(Y_{W,Z})] = c\mathbf{E}[\mathbf{E}[f(W_{W,Z})|(W, Z)]] = \int\int (z + |w|)\mathbf{E}[f(Y_{w,z})] 1_{w\leq 0} 1_{z\geq 0} \mu(dw)\mu(dz)$$

and the assertion is shown, since f was arbitrary. □

Remark 15.24 (strong embedding)**.** *The lemma initially only asserts equality in distribution, $Y \sim Y_{W,Z}$. Furthermore, it is also possible to define the probability space on which Y is defined by adding random variables (W, Z) and $Y_{W,Z}$, so that $Y = Y_{W,Z}$ is almost certain.*

Lemma 15.25 (Embedding of a random variable in a Brownian motion)**.** *Let Y be a real-valued random variable with $\mathbf{E}[Y] = 0$. Further, let (W, Z) be distributed as in Lemma 15.23, and $\mathcal{X} = (X_t)_{t\geq 0}$ is an independent Brownian motion. Then*

$$T_{W,Z} = \inf\{t \geq 0 : X_t \in \{W, Z\}\}$$

is a stopping time with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t = \sigma(W, Z; X_s : s \leq t)$. In addition,

$$X_{T_{W,Z}} \sim Y, \qquad \mathbf{E}[T_{W,Z}] = \mathbf{E}[Y^2].$$

Proof. The Brownian motion $\mathcal{X}$ is adapted to $(\mathcal{F}_t)_{t\geq 0}$. Therefore, $T_{W,Z}$ according to Proposition 12.30 is a stopping time. Clearly, for $w < 0 \leq z$ the random variable $X_{T_{w,z}}$ only takes the values w and z. According to Proposition 13.19, $(X_{T_{w,z}\wedge t})_{t\geq 0}$ is is a martingale which, according to Theorem 13.22 converges in L^1 against $X_{T_{w,z}}$. Therefore

$$0 = \mathbf{E}[X_{T_{w,z}}] = w\mathbf{P}(X_{T_{w,z}} = w) + z(1 - \mathbf{P}(X_{T_{w,z}} = w)),$$

also

$$\mathbf{P}(X_{T_{w,z}} = w) = \frac{z}{z + |w|}.$$

So $X_{T_{w,z}}$ has the same distribution as $Y_{w,z}$ from Lemma 15.23 and is independent of X. According to the lemma it follows that $X_{T_{W,Z}} \sim Y_{W,Z} \sim Y$. Further, $(X_t^2 - t)_{t\geq 0}$ is a martingale and for $y < 0 \leq z$, the process $(X^2_{T_{w,z}\wedge t} - T_{w,z} \wedge t)_{t\geq 0}$ is a martingale. This means that with monotone and dominated convergence,

$$\begin{aligned} \mathbf{E}[T_{W,Z}] &= \mathbf{E}[\mathbf{E}[T_{W,Z}|W, Z]] = \mathbf{E}[\lim_{t\to\infty} \mathbf{E}[T_{W,Z} \wedge t]|W, Z] \\ &= \mathbf{E}[\mathbf{E}[X^2_{T_{W,Z}}|W, Z]] = \mathbf{E}[X^2_{T_{W,Z}}] = \mathbf{E}[Y^2]. \end{aligned}$$

□

Theorem 15.26 (Skorohod's embedding theorem)**.** *Let $Y_1, Y_2, ...$ be independent and identically distributed with $\mathbf{E}[Y_1] = 0$, and $S_n = Y_1 + \cdots + Y_n$. Then there is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with filtration $(\mathcal{F}_t)_{t\geq 0}$, as well as a Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ on this probability space, which is a $(\mathcal{F}_t)_{t\geq 0}$ martingale and stopping times $T_1, T_2, ...$, so that:*

1. *$(X_{T_1}, X_{T_2}, ...) \sim S_1, S_2, ...$ and*
2. *$(T_{n+1} - T_n)_{n=0,1,2,...}$ are independent with $\mathbf{E}[T_{n+1} - T_n] = \mathbf{V}[Y_1]$ for $n = 1, 2, ...$*

Remark 15.27 (Strong embedding)**.**

1. *As in remark 15.24, it is possible to define the probability space on which $Y_1, Y_2, ...$ are defined so that $(X_{T_1}, X_{T_2}, ...) = S_1, S_2, ...$ almost certainly holds.*
2. *Without the restriction of the integrability of $T_{n+1} - T_n$ the statement of the theorem would be trivial. Then you could simply recursively $0 = T_0 \leq T_1, ...$ by means of*

$$T_n = \inf\{t \geq T_{n-1} : X_t = S_n\}.$$

However, these waiting times cannot be integrated.

Proof of theorem 15.26. Let the pairs $(W_1, Z_1), (W_2, Z_2), ...$ be distributed exactly as in Lemma 15.23. We extend the probability space by an independent Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$. We recursively define $0 = T_0 \leq T_1 \leq T_2 ...$ by

$$T_n := \inf\{t \geq T_{n-1} : X_t - X_{T_{n-1}} \in \{W_n, Z_n\}\}.$$

Thus $T_1, T_2, ...$ are stopping times with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t = \sigma(W_1, Z_1, W_2, Z_2, ...; X_s : s \leq t)$ and $\mathcal{X}$ is a martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. Furthermore, the pairs $(T_{n+1} - T_n, X_{T_{n+1}} - X_{T_n})_{n=0,1,2,...}$ because of the strong Markov property of the Brownian motion are independent of each other. Therefore, it follows from Lemma 15.25 that

$$(X_{T_1}, X_{T_2} - X_{T_1}, ...) \sim (Y_1, Y_2, ...),$$

also

$$(X_{T_1}, X_{T_2}, ...) \sim (S_1, S_2, ...),$$

and $\mathbf{E}[T_{n+1} - T_n] = \mathbf{E}[Y_n]$. □

Since, thanks to the last theorem, the relationship between the random walks and Brownian motion is shown, it is obvious to formulate another extension of Donsker's theorem, Theorem 15.21.

Corollary 15.28 (Stochastic convergence of the random walks)**.** *Let $Y_1, Y_2, ...$ be real-valued, independent, identically distributed random variables with $\mathbf{E}[Y_1] = 0$, $\mathbf{V}[Y_1] = 1$ and $S_n = Y_1 + \cdots + Y_n$. Then you can expand the probability space so that there is a Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ with*

$$\sup_{0\leq s\leq t} \Big| \frac{1}{\sqrt{n}} S_{[sn]} - \frac{1}{\sqrt{n}} X_{sn} \Big| \xrightarrow{n\to\infty}_p 0 \tag{15.10}$$

for all $t > 0$.

Proof. We use the construction from Theorem 15.26 and Remark 15.27. Since $T_{n+1} - T_n$ are independent and identically distributed with $\mathbf{E}[T_{n+1} - T_n] = 1$ and $T_n/n \xrightarrow{n\to\infty} 1$ according to the law of large numbers. This means that $\frac{1}{n}\sup_{s\leq t} |T_{[sn]} - sn| \xrightarrow{n\to\infty}_{fs} 0$. (To see this, we consider the set $\{\frac{1}{n}\sup_{s\leq t} |T_{[sn]} - sn| > \varepsilon\}$ for a $\varepsilon > 0$. On this set there are $s_1, s_2, ... \leq t$ with $|T_{[s_n n]} - s_n n| > \varepsilon n$. However, this contradicts $\lim_{n\to\infty} T_{[s_n n]}/[s_n n] = \lim_{n\to\infty} T_n/n = 1$.)

We recall the definition of the continuity modulus w from Definition 15.17. With the scaling property of the Brownian motion from Theorem 12.19, it follows that $S_{[sn]} = X_{T_{[sn]}}$,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbf{P}\Big(\frac{1}{\sqrt{n}} \sup_{0\leq s\leq t} |S_{[sn]} - X_{sn}| > \varepsilon\Big) \\
&\leq \inf_h \limsup_{n\to\infty} \mathbf{P}(w(\mathcal{X}, (t+h)n, nh) > \varepsilon\sqrt{n}) + \mathbf{P}(\sup_{s\leq t} |T_{[sn]} - sn| > nh) \\
&= \inf_h \mathbf{P}(w(\mathcal{X}, t+h, h) > \varepsilon) = 0.
\end{aligned}$$

□

Now that the random walks and Brownian motion are directly related to each other, it makes sense to transfer the properties of Brownian motion to the random walks. We do this for the Law of the iterated logarithm.

Theorem 15.29 (Law of the iterated logarithm for random walks)**.** *Let $Y_1, Y_2, \ldots$ be real-valued, independent, identically distributed random variable with $\mathbf{E}[Y_1] = 0$, $\mathbf{V}[Y_1] = 1$ and $S_n = Y_1 + \cdots + Y_n$. Then,*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1$$

almost surely.

Proof. We only show that the probability space can be extended in such a way such that there is a Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ exists with

$$\frac{S_{[t]} - X_t}{\sqrt{2t\log\log t}} \xrightarrow{t\to\infty}_{fs} 0. \tag{15.11}$$

Then the statement follows from the law of the iterated logarithm for Brownian motion, theorem 15.10.

According to Theorem 15.26 there is an extension of the probability space and stopping times $0 = T_0, T_1, \ldots$, so that $X_{T_n} = S_n$. Again, $T_n/n \xrightarrow{n\to\infty} 1$ applies according to the law of large numbers, which is also $T_{[t]}/t \xrightarrow{t\to\infty} 1$ implies. Now let $r > 1$, $c^2 > r - 1$ and $h(t) = \sqrt{2t\log\log t}$. Then, with a similar calculation as in (15.6)

$$\begin{aligned}
&\mathbf{P}\big(\sup_{r^{n-1}\leq t\leq r^n} |X_t - X_{r^{n-1}}| > ch(r^{n-1})\big) = \mathbf{P}\big(\sup_{0\leq t\leq r^n - r^{n-1}} |X_t| > ch(r^{n-1})\big)\\
&= 2\mathbf{P}(X_{r^n - r^{n-1}} > ch(r^{n-1})) = 2\mathbf{P}(X_1 > ch(r^{n-1})/\sqrt{r^n - r^{n-1}})\\
&\overset{n\to\infty}{\approx} \frac{1}{c}\sqrt{\frac{(r-1)}{\pi\log n}} n^{-c^2/(r-1)},
\end{aligned}$$

since $h(r^{n-1})/\sqrt{r^n - r^{n-1}} \overset{n\to\infty}{\approx} \sqrt{(2\log n)/(r-1)}$. The right-hand side is summable, so with the Borel-Cantelli lemma and $X_{T_{[t]}} = S_{[t]}$

$$\begin{aligned}
\mathbf{P}\Big(\limsup_{t\to\infty} \frac{|S_{[t]} - X_t|}{h(t)} = 0\Big) &\geq \mathbf{P}\Big(\lim_{r\downarrow 1} \limsup_{t\to\infty} \sup_{t\leq u\leq rt} \frac{|X_u - X_t|}{h(t)} = 0\Big)\\
&\geq \mathbf{P}\Big(\lim_{r\downarrow 1} \limsup_{n\to\infty} \sup_{r^{n-1}\leq t\leq r^n} \frac{|X_t - X_{r^{n-1}}|}{h(r^{n-1})} = 0\Big)\\
&= \inf_{c>0} \mathbf{P}\Big(\lim_{r\downarrow 1, r<c^2+1} \limsup_{n\to\infty} \sup_{r^{n-1}\leq t\leq r^n} \frac{|X_t - X_{r^{n-1}}|}{h(r^{n-1})} \leq c\Big) = 1.
\end{aligned}$$

Therefore follows (15.11). □

Part IV
Stochastic Analysis

In Definition 13.13 we already got to know the discrete stochastic integral. For $I = \{0, 1, 2, ...\}$ and a filtration $(\mathcal{F}_t)_{t\in I}$ was $\mathcal{H} = (H_t)_{t\in I}$ a real-valued and predictable stochastic process (i.e. H_t is $\mathcal{F}_{t-1}$-measurable) and $\mathcal{X} = (X_t)_{t\in I}$ was real-valued and adapted. The stochastic integral $\mathcal{H} \cdot \mathcal{X} = ((\mathcal{H} \cdot \mathcal{X})_t)_{t\in I}$ was defined as the adapted process

$$(\mathcal{H} \cdot \mathcal{X})_t := \sum_{s=1}^{t} H_s(X_s - X_{s-1}).$$

Intuitively, this means that the stochastic integral $\mathcal{H} \cdot \mathcal{X}$ measures the process $\mathcal{H}$ relative to the changes of $\mathcal{X}$. Furthermore, we have already learned about important properties of this integral. For example, $\mathcal{H} \cdot \mathcal{X}$ is a martingale if $\mathcal{X}$ is one (see Proposition 13.14 and Table 4, which summarizes the martingale properties of $\mathcal{H} \cdot \mathcal{X}$). Furthermore, it is noticeable that we have defined quadratic variation for both discrete-time stochastic integrals (Example 13.15) and for Brownian motion (see Section 15.1). Again, cross-connections are to be expected. The aim of this chapter is to extend the theory of stochastic integrals to processes in continuous time. The discrete-time case will serve as a template.

In the following, $(\Omega, \mathcal{F}, \mathbf{P})$ will always be a probability space on which all stochastic processes are defined, and $(\mathcal{F}_t)_{t\geq 0}$ will always be a filtration. Unless otherwise stated, all stochastic processes are real-valued.

16 Introduction

Integrals (based, for example, on σ-finite measures on $\mathcal{B}(\mathbb{R})$) have already been treated in detail in measure theory. Later, in probability theory, the integral corresponded to the expectation value of a random variable; see remark 6.1.5. The situation is different with stochastic integration. Here, the integral itself is meant to is a random variable.

In the following, we use

$$\mathcal{H} \cdot \mathcal{X} = \int H_s dX_s.$$

as a shorthand for stochastic integrals. The process $\mathcal{H} = (H_t)_{t\geq 0}$ is called the integrand and $\mathcal{X} = (X_t)_{t\geq 0}$ the integrator. The class of possible integrators in the general theory of stochastic integration are semi-martingales (that is, the sum of a local martingale – see Definition 16.23 – and a process of finite variation). Just as in the discrete stochastic integral, the integrands are predictable processes, which in particular includes the adapted, left-continuous processes.

After laying some foundations in Section 16.1, we introduce stochastic integrals in Section 16.2 first for processes with finite variation, and then – after an introduction to local martingales in Section 16.3 – for local martingales with continuous paths in Section 16.4. Table 3 summarizes the most important steps in the construction of stochastic integrals. We will only briefly discuss the general theory with possibly discontinuous semimartingales as integrators later on.

16.1 Basic

Just as in integration theory, which we encountered in measure theory, we first consider stochastic integrals of simple integrands. The process $\mathcal{I}$ from (16.2) will then serve as a

$\mathcal{H} \cdot \mathcal{X}$	$\mathcal{H} \in \mathbb{S}$	$\mathcal{H}$ hat Pfade in $\mathcal{G}_{\mathbb{R}}([0,\infty))$	$\mathcal{H}$ progressive	$\mathcal{X}$ locally bounded variation	$\mathcal{X} \in \mathcal{M}^2$ (continuous paths)	$\mathcal{X}$ local martingale, continuous paths	defining property
Definition 16.1, lemma 16.12	•			•			$(\mathcal{H} \cdot \mathcal{X})_t =$ $\sum G_i(X_{T_{i+1}\wedge t} - X_{T_i \wedge t})$
Definition 16.7.2			•	•			Lebesgue-Stieltjes Integral
Definition 16.19	•				•		$(\mathcal{H} \cdot \mathcal{X})_t =$ $\sum G_i(X_{T_{i+1}\wedge t} - X_{T_i \wedge t})$
Proposition16.20, Definition 16.21		•			•		$\mathcal{M}^2$-Grenzwert
Definition 16.34, Lemma 16.35	•					•	$(\mathcal{H} \cdot \mathcal{X})_t =$ $\sum G_i(X_{T_{i+1}\wedge t} - X_{T_i \wedge t})$
Theorem 16.36			•			•	$[\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}] = \mathcal{H} \cdot [\mathcal{X},\ ConY]$

Table 3: The stochastic integral $\mathcal{H} \cdot \mathcal{X}$ is introduced for various classes of processes $\mathcal{H}$ and $\mathcal{X}$. In chronological order, the corresponding definitions and results are listed here. Note that the defining property of the stochastic integral for $\mathcal{H} \in \mathbb{S}$ (simple, predictable processes) is always the same.

	$\mathcal{X}$ martingale	$\mathcal{X}$ martingale, bounded variation, $\sup_{t\geq 0}\mathbf{E}[X_t^2] < \infty$	$\mathcal{X}$ local martingale, continuous paths	
$\mathcal{H} \in \mathbb{S}, G_1, ..., G_n \in L^1$	$\mathcal{H} \cdot \mathcal{X}$ martingale, Proposition 16.2			
$\mathcal{H}$ beschränkt, Pfade in $\mathcal{G}_{\mathbb{R}}([0,\infty))$		$\mathcal{H} \cdot \mathcal{X}$ Martingal, Proposition 16.15	$\mathcal{H} \cdot \mathcal{X} \in Con_2$, Prop. 16.20	
$\mathcal{H}$ progressive, $\mathbf{E}[(\mathcal{H}^2 \cdot [\mathcal{X}])_t] < \infty$				$\mathcal{H} \cdot \mathcal{X}$ local martingale, Theorem 16.36

Table 4: The martingale properties of the stochastic integral $\mathcal{H} \cdot \mathcal{X}$ are summarized here.

definition of the stochastic integral. After we shown important martingale properties of this process in Proposition 16.2, the introduction of previsible processes follows.

Definition 16.1 (Simple predictable processes)**.** *A simple predictable process $\mathcal{H} = (H_t)_{t\geq 0}$ is of the form*

$$H_t = \sum_{i=1}^{n} G_i 1_{(T_i, T_{i+1}]}(t) \tag{16.1}$$

for stopping times $T_1 \leq ... \leq T_{n+1}$ and $G_1, ..., G_n$ such that G_i is $\mathcal{F}_{T_i}$-measurable, $i = 1, ..., n$. The set of all simple, predictable processes is denoted by $\mathbb{S}$.

The following result is central to the general definition of the stochastic integral. It is important to note that no preconditions are placed on the martingale $\mathcal{X}$. This suggests defining the stochastic integral with respect to simple predictable functions of the form (16.1) by (16.2) and then extending the resulting integral concept in a suitable way extended.

Proposition 16.2 (Martingale Property of the Stochastic Integral)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a martingale and $\mathcal{H} = (H_t)_{t\geq 0} \in \mathbb{S}$ a simple predictable process as in* (16.1) *with $G_1, ..., G_n \in \mathcal{L}^1$, and $\mathcal{I} = (I_t)_{t\geq 0}$, given by*

$$I_t = \sum_{i=1}^{n} G_i (X_{T_{i+1}\wedge t} - X_{T_i \wedge t}). \tag{16.2}$$

It holds that:

1. *The process $\mathcal{I}$ is a martingale.*
2. *If, in addition, $\mathcal{H} \leq \ell$ is bounded, then*

$$\sup_{t\geq 0} ||I_t||_2 \leq \ell \cdot \sup_{t\geq 0} ||X_t||_2.$$

Proof. First, a brief preliminary consideration based on the Optional Sampling Theorem (Theorem 13.22). If S, T are two bounded stop times (and not necessarily $S \leq T$) and $\mathcal{X} = (X_t)_{t\geq 0}$ is a martingale, then

$$\begin{aligned}\mathbf{E}[X_T|\mathcal{F}_S] &= \mathbf{E}[(1_{S\leq T} + 1_{S>T})X_T|\mathcal{F}_S] = 1_{S\leq T}\mathbf{E}[X_{S\vee T}|\mathcal{F}_S] + 1_{S>T}X_T \\ &= 1_{S\leq T}X_S + 1_{S>T}X_T = X_{S\wedge T}.\end{aligned}$$

Now for the proof of 1. Because of the linearity of the conditional expectation, it is sufficient to prove the assertion in the case of $H_t = G \cdot 1_{(S,T]}(t)$, i.e.

$$I_t = G(X_{T\wedge t} - X_{S\wedge t})$$

for G measurable with respect to $\mathcal{F}_S$ and $S \le T$. For $s \le t$ we write

$$\begin{aligned}
\mathbf{E}[I_t|\mathcal{F}_s] &= \mathbf{E}[(1_{T,S>s} + 1_{S\le s<T} + 1_{T\le s})G(X_{T\wedge t} - X_{S\wedge t})|\mathcal{F}_s] \\
&= \mathbf{E}[1_{T,S>s}G \cdot \mathbf{E}[X_{T\wedge t} - X_{S\wedge t}|\mathcal{F}_{S\vee s}]|\mathcal{F}_s] \\
&\qquad + 1_{S\le s<T}G \cdot \mathbf{E}[X_{T\wedge t} - X_{S\wedge t}|\mathcal{F}_s] \\
&\qquad + \mathbf{E}[1_{T\le s}G(X_{T\wedge s} - X_{S\wedge s})|\mathcal{F}_s] \\
&= \mathbf{E}[1_{T,S>s}G \cdot (X_{T\wedge t\wedge(S\vee s)} - X_{S\wedge t\wedge(S\vee s)})|\mathcal{F}_s] \\
&\qquad + 1_{S\le s<T}G \cdot (X_{T\wedge s} - X_{S\wedge s}) \\
&\qquad + 1_{T\le s}G \cdot (X_{T\wedge s} - X_{S\wedge s}) \\
&= (1_{S\le s<T} + 1_{T\le s})G \cdot (X_{T\wedge s} - X_{S\wedge s}) = G \cdot (X_{T\wedge s} - X_{S\wedge s}) = I_s,
\end{aligned}$$

where we have used the fact that $1_{S\le s<T}G$ is $\mathcal{F}_s$-measurable. This proves the claim.
2. If $\mathcal{H} \le \ell$, then we write

$$\begin{aligned}
\mathbf{E}[I_t^2] &\le \mathbf{E}\Big[\sum_{i=1}^n \sum_{j=1}^n G_i G_j (X_{T_{i+1}\wedge t} - X_{T_i\wedge t})(X_{T_{j+1}\wedge t} - X_{T_j\wedge t})\Big] \\
&= \mathbf{E}\Big[\sum_{i=1}^n G_i^2 (X_{T_{i+1}\wedge t} - X_{T_i\wedge t})^2\Big] \\
&\qquad\qquad + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^n \mathbf{E}\big[G_i G_j (X_{T_{i+1}\wedge t} - X_{T_i\wedge t})\mathbf{E}[X_{T_{j+1}\wedge t} - X_{T_j\wedge t}|\mathcal{F}_{T_j\wedge t}]\big] \\
&= \mathbf{E}\Big[\sum_{i=1}^n G_i^2 (X_{T_{i+1}\wedge t} - X_{T_i\wedge t})^2\Big] \\
&\le \ell^2 \sum_{i=1}^n \mathbf{E}\big[(X_{T_{i+1}} - X_{T_i})^2\big] \\
&= \ell^2 \sum_{i=1}^n \mathbf{E}\big[X_{T_{i+1}}^2 - 2\mathbf{E}[X_{T_{i+1}}|\mathcal{F}_{T_i}]X_{T_i} + X_{T_i}^2\big] \\
&= \ell^2 \sum_{i=1}^n \mathbf{E}\big[X_{T_{i+1}}^2 - X_{T_i}^2\big] = \ell^2 \cdot \sup_{t\ge 0} \mathbf{E}[X_t^2].
\end{aligned}$$

□

The process $\mathcal{I}$ just defined inherits the continuity properties of the process $\mathcal{X}$.

Lemma 16.3 (Continuity properties of the stochastic integral)**.** *Let $\mathcal{X}$ be a stochastic process and $\mathcal{H}$ a simple predictable process. Then the following holds for the process $\mathcal{I}$ from Proposition 16.2:*

1. *If $\mathcal{X}$ has continuous paths, then so does $\mathcal{I}$.*
2. *If $\mathcal{X}$ has right-continuous paths, then so does $\mathcal{I}$.*

Proof. Clear. □

The simple predictable integrands from the last proposition do not yet form a particularly large class of stochastic processes. Therefore, we will try to extend the concept of the integral by approximating more general processes by simple predictable stochastic processes. This leads us to the predictable stochastic processes, which we will now introduce.

Definition 16.4 (Left-continuous processes and predictable σ-algebra)**.** *1. We denote by $\mathcal{G}_E([0,\infty))$ the set of left-continuous functions $f : [0,\infty) \to E$ with right-sided limits.*

2. The predictable sigma$-$*algebra $\mathcal{V}$ on $[0,\infty)\times\Omega$ is the smallest σ-algebra with respect to which all processes with paths in are measurable in $\mathcal{G}_E([0,\infty))$. (This means that for each $t \mapsto X_t(\omega)$ for $\mathcal{X} = (X_t)_{t\geq 0}$ is measurable with paths in $\mathcal{G}_E([0,\infty))$ with respect to $\mathcal{V}/\mathcal{B}(E)$.) An adapted process $\mathcal{X}$ is predictable if it is measurable with respect to $\mathcal{V}$.*

Proposition 16.5 (Predictable Processes)**.** *1. Every adapted process with paths in $\mathcal{G}_E([0,\infty))$ is predictable. (In particular, processes with continuous paths are predictable.)*

2. Every predictable process is progressively measurable.

3. If $\mathcal{X}$ has paths in $\mathcal{D}_E([0,\infty))$, then $\mathcal{X}_- := (X_{t-})_{t\geq 0}$ is predictable.

Proof. 1. Obvious from the definition of the predictable σ-algebra.
2. Just as in the proof of lemma **??**, one deduces that processes with paths in $\mathcal{G}_E([0,\infty))$ are progressively measurable. Since $\mathcal{V}$ is generated by the processes with paths in $\mathcal{G}_E([0,\infty))$ and these processes are progressively measurable, the statement follows.
3. It is clear that $\mathcal{X}_-$ has paths in $\mathcal{G}_E([0,\infty))$. Thus, the statement follows from 1. □

The class of left-continuous stochastic processes is quite large. We now show that such processes can be approximated very well by simple predictable processes.

Lemma 16.6 (Approximations of Processes with Paths in $\boldsymbol{\mathcal{G}_E([0,\infty))}$)**.** *Let $\mathcal{Y} = (Y_t)_{t\geq 0}$ be a bounded process with paths in $\mathcal{G}_E([0,\infty))$ and $Y_0 = 0$. Then, for every $t > 0$ a sequence $\mathcal{Y}^1 = (Y_t^1)_{t\geq 0}, \mathcal{Y}^2 = (Y_t^2)_{t\geq 0}, ... \in \mathbb{S}$ of simple, predictable processes and $\varepsilon_1, \varepsilon_2, ... > 0$ with $\varepsilon_n \downarrow 0$ and*

$$\sup_{0\leq s\leq t} |Y_s - Y_s^n| \leq \varepsilon_n$$

almost surely.

Proof. We note that $(Y_{t+})_{t\geq 0}$ has paths in $\mathcal{D}_E([0,\infty))$. Let $\varepsilon > 0$. We define recursively $T_0^\varepsilon = 0$,

$$T_{n+1}^\varepsilon := \inf\{t > T_n^\varepsilon : |Y_t - Y_{T_n^\varepsilon +}| > \varepsilon\}.$$

Then $T_{n+1}^\varepsilon \uparrow \infty$. Further, we set

$$Y_t^{n,\varepsilon} := \sum_{i=1}^n Y_{T_i^\varepsilon +} \cdot 1_{(T_i^\varepsilon \wedge n, T_{i+1}^\varepsilon \wedge n,]}(t),$$

so that $\sup_{0\leq s\leq t\wedge n} |Y_s - Y_s^{n,\varepsilon}| \leq \varepsilon$ by definition almost surely holds. The assertion now follows if we consider a sequence $\varepsilon_n \downarrow 0$ and $\mathcal{Y}^n = (Y_t^{n,\varepsilon_n})_{t\geq 0}$. □

16.2 (Stochastic) Stieltjes Integrals

The construction of stochastic integrals is done different cases separately. First, we construct the stochastic integral in the case of integrators that are stochastic processes with paths of finite variation (see Definition 15.3). This is done via the Stieltjes integral, which is a (simple) extension of the Lebesgue integral.

We recall Proposition 2.19: A σ-finite measure on $\mathbb{R}$ is exactly described by a non-decreasing, right-continuous function $g : \mathbb{R} \to \mathbb{R}$ with $g(0) = 0$. We first introduce the corresponding integral term for this measure .

Definition 16.7 (Lebesgue-Stieltjes integral for non-decreasing functions)**.** *Let $g : \mathbb{R}_+ \to \mathbb{R}_+$ be right-continuous and non-decreasing. Then g uniquely defines a measure μ_g. Furthermore, if $f : \mathbb{R}_+ \to \mathbb{R}$ is Borel-measurable, then, if it exists, we set*

$$f \cdot g := \int f dg := f \cdot \mu_g = \int f d\mu_g.$$

Here, $\int f dg$ is called the Lebesgue-Stieltjes integral of f with respect to g. (The notation $f \cdot \mu_g$ was introduced in Definition **??***.)*

Example 16.8 (Lebesgue integral, integration with respect to a Poisson process)**.**

1. *We denote the (one-dimensional) Lebesgue measure. If $g(x) = x$ in the above definition, then*
$$f \cdot g = \int f(x)\lambda(dx).$$

2. *As is well known, the Poisson process $\mathcal{X} = (X_t)_{t\geq 0}$ has monotonically non-decreasing, right-continuous paths. Therefore, for each path $(X_t(\omega))_{t\geq 0}$ of a Poisson process a σ-finite measure on $\mathbb{R}_+$. (This is in particular a counting measure.) Let $T_1, T_2, ...$ be the jump times of a Poisson process and $\mathcal{H} = (H_t)_{t\geq 0}$ a stochastic process. Then we can write*
$$(\mathcal{H} \cdot \mathcal{X})_t = \int_0^t H_s dX_s = \sum_{k:T_k \leq t} H_{T_k}.$$
write. (To see the last equality, note that the measure $\mu_{\mathcal{X}}$ puts atoms of size 1 on the times $T_1, T_2, ...$ If one integrates with respect to such a measure, only these atoms play a role. In particular, the Lebesgue-Stieltjes integral already explains the integration with respect to the Poisson process.

Monotonically non-decreasing functions (and thus also stochastic processes with such paths) are rare. Much more common are those with locally finite variation. A function $f : [0, \infty] \to \mathbb{R}$ is locally of finite variation if

$$\nu_{1,t}(f) := \sup_{n, 0\leq t_0 < ... < t_n < t} \sum_{k=1}^n |f(t_k) - f(t_{k-1})| < \infty,$$

holds for all $t > 0$. (See also Definition 15.3.) There is the following connection between non-decreasing functions and functions with finite variation:

Lemma 16.9 (Darstellung von Funktionen von lokal endlicher Variation)**.** *A function $g : [0,\infty) \to \mathbb{R}$ is locally of finite variation if and only if it can be represented as the difference of two monotonically non-decreasing functions.*

Proof. '$\Leftarrow$': First, let $g = a - b$, where a and b are non-decreasing. Then $\nu_{1,t}(a) = a(t) - a(0)$ and $\nu_{1,t}(b) = b(t) - b(0)$. Furthermore,

$$\begin{aligned}\nu_{1,t}(g) &= \sup_{n,0\le t_0<\ldots<t_n<t} \sum_{k=1}^{n} |g(t_k) - g(t_{k-1})| \\ &\le \sup_{n,0\le t_0<\ldots<t_n<t} \sum_{k=1}^{n} |a(t_k) - a(t_{k-1})| + |b(t_k) - b(t_{k-1})| \\ &\le \nu_{1,t}(a) + \nu_{1,t}(b) < \infty.\end{aligned}$$

'$\Rightarrow$': It is clear that both $t \mapsto \nu_{1,t}(g)$ and $t \mapsto \nu_{1,t}(g) - g(t)$ are non-decreasing. Therefore, $g(t) = \nu_{1,t}(g) - (\nu_{1,t}(g) - g(t))$ already fulfills the desired property. □

Definition 16.10 (Stochastic Lebesgue-Stieltjes Integral)**.**

1. *Let $g : \mathbb{R}_+ \to \mathbb{R}$ be right-continuous and of finite variation. Further, let $g = a - b$, where a, b are non-decreasing. Then, for any measurable f, if it exists, we set*

$$f \cdot g := \int f da - \int f db.$$

2. *Let $\mathcal{X}$ be a stochastic process with right-continuous paths of finite variation and $\mathcal{H}$ be progressively measurable. Then we denote by $\mathcal{H} \cdot \mathcal{X}$ the* stochastic integral of $\mathcal{H}$ with respect to $\mathcal{X}$ *as the stochastic process with*

$$(\mathcal{H} \cdot \mathcal{X})_t(\omega) := (H_s(\omega))_{0\le s\le t} \cdot (X_s(\omega))_{0\le s\le t}.$$

Remark 16.11 (Well-definedness and signed measures)**.**

1. *The integral $f \cdot g$ for a function g of finite variation (and thus also the stochastic integral) is well-defined: namely, let $g = a - b = a' - b'$ be two representations of the function g. Then $h := a + b' = a' + b$, so also $f \cdot a + f \cdot b' = f \cdot a' + f \cdot b$. Hence the assertion follows.*

2. *In the section Measure Theory, we have seen σ-finite measures on $\mathcal{B}([0,\infty))$ as mappings $\mathcal{B}([0,\infty)) \to [0,\infty]$. This notion can be generalized to σ-finite,* signed measures. *These are σ-finite mappings $\mathcal{B}([0,\infty)) \to (-\infty,\infty]$ or $\mathcal{B}([0,\infty)) \to [-\infty,\infty)$. (Sets can therefore also have negative measure.) is that the difference of two σ-finite measures $\mu^+ - \mu^-$ of which at least one is finite, is a signed measure. The reverse is also true and is known as Jordan's decomposition theorem. In particular, we could write the integral $f \cdot g$ for a function g of finite variation as an integral with respect to a finite signed measure,*

Just as in the integration theory that we encountered in measure theory, we now consider stochastic integrals of simple integrands.

Lemma 16.12 (Stochastic Integration of Simple Functions)**.** *Let $\mathcal{X} = (X_t)_{t\ge0}$ be a process with paths of locally finite variation and let $\mathcal{H} = (H_t)_{t\ge0}$ be a simple previsible process as in* (16.1)*. Then*

$$(\mathcal{H} \cdot \mathcal{X})_t = \sum_{i=1}^{n} G_i(X_{T_{i+1}\wedge t} - X_{T_i\wedge t}). \tag{16.3}$$

Proof. There are non-decreasing processes $\mathcal{Y} = (Y_t)_{t\geq 0}$ and $\mathcal{Z} = (Z_t)_{t\geq 0}$ with $\mathcal{X} = \mathcal{Y} - \mathcal{Z}$. Furthermore, $(1_{(T_i,T_{i+1}]}\cdot\mathcal{Y})_t = Y_{T_{i+1}\wedge t} - Y_{T_i\wedge t}$ and $(1_{(T_i,T_{i+1}]}\cdot\mathcal{Z})_t = Z_{T_{i+1}\wedge t} - Z_{T_i\wedge t}$ by definition of the Stieltjes integral, $i = 1, ..., n$. The statement now follows because of the linearity of the integral. □

We now turn to the special case of integrators of bounded variation. In this case, the Lebensgue-Stieltjes integral can also be interpreted as a Riemann integral. Similar to in Proposition 3.23, the following applies:

Proposition 16.13 (Riemann Integrability). *Let $g : \mathbb{R}_+ \to \mathbb{R}$ have locally finite variation, i.e., for each $t > 0$ there exist $0 = t_{n,0} \leq ... \leq t_{n,k_n} = t$ with $\max_k |t_{n,k} - t_{n,k-1}| \xrightarrow{n\to\infty} 0$ is $\sum_{i=1}^{k_n} |g(t_{n,k}) - g(t_{n,k-1})| < \infty$. Then $f \cdot g$ exists for a continuous function f if and only if*

$$(f \cdot g)_t = \lim_{n\to\infty} \sum_{k=1}^{\infty} f(s_{n,k})(g(t_{n,k}) - g(t_{n,k-1})) \tag{16.4}$$

holds for arbitrary $t_{n,k-1} \leq s_{n,k} \leq t_{n,k}$.

Proof. Analogous to the proof of Proposition 3.23. □

If the integrand has continuous paths of finite variation, there is a transformation formula. We note here that similar transformations in the case of unbounded variation require an additional term.

Theorem 16.14 (Transformation formula). *Let $X = (X_t)_{t\geq 0}$ be a process with continuous paths of locally finite variation and $f \in \mathcal{C}^1(\mathbb{R})$. Then*

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)dX_s.$$

Proof. Since $(f'(X_t))_{t\geq 0}$ is continuous, the right-hand side exists. Furthermore, let $0 = t_{n,1} \leq \cdots \leq t_{n,k_n} = t$ with $\max_k t_{n,k} - t_{n,k-1} \xrightarrow{n\to\infty} 0$. Then, for suitable random variables $t_{n,k-1} \leq S_{n,k} \leq t_{n,k}$ according to Proposition 16.13 and the mean value theorem

$$\begin{aligned} f(X_t) - f(X_0) &= \sum_{i=1}^{k_n} f(X_{t_{n,k}}) - f(X_{t_{n,k-1}}) \\ &= \sum_{i=1}^{k_n} f'(X_{S_{n,k}})(X_{t_{n,k}} - X_{t_{n,k-1}}) \xrightarrow{n\to\infty} \int_0^t f'(X_s)dX_s. \end{aligned}$$

□

Proposition 16.15 (Martingale property of the stochastic integral). *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a martingale with paths of bounded variation, $\sup_{t\geq 0} \mathbf{E}[X_t^2] < \infty$ and $\mathcal{H}$ is a bounded adapted process with paths in $\mathcal{G}_{\mathbb{R}}([0,\infty))$. Then $\mathcal{H} \cdot \mathcal{X}$ is a martingale with $\sup_{t\geq 0} \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t^2] < \infty$.*

Proof. According to Lemma 16.6, there are $\mathcal{H}^1 = (H_t^1)_{t\geq 0}, \mathcal{H}^2 = (H_t^2)_{t\geq 0}, ... \in \mathbb{S}$ and $\varepsilon_1, \varepsilon_2, ... > 0$ with $\varepsilon_2 \downarrow 0$ and $\sup_{0\leq s\leq t} |H_s - H_s^n| \leq \varepsilon_n$ almost surely. For every $\varepsilon > 0$, let $K > 0$ be such that $P(\nu_{1,t}(\mathcal{X}) > K) < \varepsilon/2$. Furthermore, let N be large Then

$$\mathbf{P}(\lim_{n\to\infty} \sup_{0\leq s\leq t} |((\mathcal{H}^n - \mathcal{H}) \cdot \mathcal{X})_s| = 0) \leq \mathbf{P}(\lim_{n\to\infty} \varepsilon_n \nu_{1,s}(\mathcal{X}) = 0) = 1. \tag{16.5}$$

Therefore, $(\mathcal{H}^n \cdot \mathcal{X})_t \xrightarrow{n\to\infty}_{fs} (\mathcal{H}^n \cdot \mathcal{X})_t$, and in fact uniformly on compact. According to the Fatou lemma, if $\mathcal{H}, \mathcal{H}^1, \mathcal{H}^2, ... < \ell$, then, according to Proposition 16.2.2,

$$\mathbf{E}[((\mathcal{H} \cdot \mathcal{X})_t)^2] \leq \liminf_{n\to\infty} \mathbf{E}[((\mathcal{H}^n \cdot \mathcal{X})_t)^2] \leq \ell^2 \cdot \sup_{t\geq 0} \mathbf{E}[X_t^2] < \infty.$$

Finally, we verify the martingale property of $\mathcal{H} \cdot \mathcal{X}$. Let $s \leq t$. By (16.5) and the L^2-boundedness of $\mathcal{H}^n \cdot \mathcal{X}$, it holds that $(\mathcal{H}^n \cdot \mathcal{X})_t \xrightarrow{n\to\infty}_{L^1} (\mathcal{H} \cdot \mathcal{X})_t$. Then

$$\mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t|\mathcal{F}_s] = \lim_{n\to\infty} \mathbf{E}[(\mathcal{H}^n \cdot \mathcal{X})_t|\mathcal{F}_s] = \lim_{n\to\infty} (\mathcal{H}^n \cdot \mathcal{X})_s = (\mathcal{H} \cdot \mathcal{X})_s$$

and all assertions are shown. □

Example 16.16 (Integration with respect to a Poisson process). *Let $\mathcal{Y} = (Y_t)_{t\geq 0}$ be a Poisson process with intensity $\lambda > 0$ and $\mathcal{X} = (X_t)_{t\geq 0}$ with $X_t = Y_t - \lambda t$. Then, by example 13.46, $\mathcal{X}$ is a martingale and has paths of locally bounded variation. Further, let $f, g : \mathbb{N}_0 \to \mathbb{R}$ be bounded and such that $g(i) = f(i+1) - f(i)$ for $i = 1, 2, ...$ Since $\mathcal{Y}_- = (Y_{t-})_{t\geq 0}$ is a process with paths in $\mathcal{G}_{\mathbb{R}}([0,\infty))$, it follows, if $T_1, T_2, ...$ are the jump times of $\mathcal{Y}$, that*

$$\begin{aligned}(f(Y_{t-}+1) - f(Y_{t-})) \cdot \mathcal{X} &= g(\mathcal{Y}_-) \cdot \mathcal{X} \\ &= \Big(\sum_{i:T_i\leq t} g(i-1) - \lambda \int_0^t g(Y_{s-})ds\Big)_{t\geq 0} \\ &= \Big(f(Y_t) - f(0) - \lambda \int_0^t f(Y_s+1) - f(Y_s)ds\Big)_{t\geq 0}\end{aligned}$$

is a martingale. Indeed, this martingale property also follows from Theorem 14.30 together with Example 14.26.1.

16.3 L^2-bounded continuous martingales as integrators

Although the definition of the stochastic integral with processes of locally bounded variation was quite straightforward, we are not yet able to integrate with respect to continuous martingales (such as the Brownian motion) with respect to the path. Indeed, as we have seen in Proposition 15.5, the variation of the paths of a Brownian motion is almost surely infinite (and only the quadratic variation is almost surely finite and positive).

The approach so far was based on the fact that a function of limited variation can be understood as a signed measure. For this, it was essential that a non-decreasing process uniquely defines a σ-finite measure on $\mathcal{B}(\mathbb{R})$. This is not the case for functions with unlimited variation.

Nevertheless, in order to allow for continuous martingales with unlimited variation as integrators, we recall the definition of the stochastic integral with respect to simple predictable processes from Proposition 16.2. We can use (16.2) as definition of the stochastic integral with respect to simple predictable processes and then extend the integral concept using the martingale property of the stochastic integral from Proposition 16.2. The integrators we consider here are continuous L^2-bounded martingales.

Definition 16.17 (L^2-bounded, continuous martingale). *We denote by $\mathcal{M}^2$ the set of continuous, L^2-bounded martingales $\mathcal{X} = (X_t)_{t\geq 0}$ with $X_0 = 0$. (This means that $\mathcal{X}$ has continuous paths and $\sup_{t\geq 0} \mathbf{E}[X_t^2] < \infty$). By Theorem 13.51 (and Theorems 13.32 and 13.33) there exists for each $\mathcal{X} = (X_t)_{t\geq 0} \in \mathcal{M}^2$ an X_∞ such that $(X_t)_{0\leq t\leq\infty}$ is a martingale. We define the norm $||\mathcal{X}|| := ||X_\infty||_2$ on $\mathcal{M}^2$ and recall that $||\sup_{t\geq 0} X_t^2||_2 \leq 2||\mathcal{X}||$ by Proposition 13.26.*

To approximate stochastic integrals for general integrands, we will define them as L^2 limits. This works because the space L^2 – and thus also the space $\mathcal{M}^2$ – is complete. The completeness of $\mathcal{M}^2$ will now be shown.

Lemma 16.18. *The space $\mathcal{M}^2$ is a Hilbert space with a scalar product $\langle \mathcal{X}, \mathcal{Y} \rangle := \mathbf{E}[X_\infty Y_\infty]$.*

Proof. We have to show that $\mathcal{M}^2$ is complete. To do this, let $\mathcal{X}^1 = (X_t^1)_{0 \le t \le \infty}, \mathcal{X}^2 = (X_t^2)_{0 \le t \le \infty}, ...$ be a Cauchy sequence in $\mathcal{M}^2$. According to the definition of the norm in $\mathcal{M}^2$, $X_\infty^1, X_\infty^2, ...$ is a Cauchy sequence in L^2. Therefore, the limit $X_\infty^n \xrightarrow{n\to\infty} X_\infty$ exists. We now define $\mathcal{X} = (X_t)_{t \ge 0}$ by $X_t = \mathbf{E}[X_\infty | \mathcal{F}_t]$ and note that $X_t \xrightarrow{t\to\infty}_{fs, L^2} X_\infty$. Furthermore,

$$|| \sup_{t \ge 0} (X_t^n - X_t) ||_2 \le 2 || X_\infty^n - X_\infty || \xrightarrow{n\to\infty} 0$$

according to Propotision 13.26. By passing to a subsequence, this shows that $\mathcal{X}^n$ converges uniformly to $\mathcal{X}$. In particular, $\mathcal{X}$ has continuous paths and $X_0 = 0$. □

Definition 16.19 (Stochastic integral of simple predictable processes)**.** *Let $\mathcal{X} \in \mathcal{M}^2$ and $\mathcal{H}$ be a simple predictable process as in definition 16.1. Then the stochastic integral $\mathcal{H} \cdot \mathcal{X}$ is defined by*

$$(\mathcal{H} \cdot \mathcal{X})_t := \sum_{i=1}^n G_i (X_{T_{i+1} \wedge t} - X_{T_i \wedge t}),$$

i.e. as in (16.3)*. If $G_1, ..., G_n \in \mathcal{L}^1$, this is, by Proposition 16.2, once more a martingale and additionally has continuous paths.*

Proposition 16.20 (Integration of processes with paths in $\boldsymbol{\mathcal{G}_\mathbb{R}([0, \infty))}$)**.** *Let $\mathcal{X} = (X_t)_{t \ge 0} \in \mathcal{M}^2$ and $\mathcal{H}$ a bounded adapted process with paths in $\mathcal{G}_\mathbb{R}([0, \infty))$. Further, let $\mathcal{H}^1, \mathcal{H}^2, ...$ is a sequence of simple, predictable processes that converge uniformly on compact against $\mathcal{H}$ (as in Lemma 16.6), then for every $\tau > 0$, the sequence $((\mathcal{H}^n \cdot \mathcal{X})_{t \wedge \tau})_{t \ge 0}$ converges to a martingale in $\mathcal{M}^2$.*

Definition 16.21 (Stochastic integral of processes with paths in $\mathcal{G}_\mathbb{R}([0, \infty))$)**.** *Let $\mathcal{X} \in \mathcal{M}^2$ and $\mathcal{H}$ be a bounded, adapted stochastic process with paths in $\mathcal{G}_\mathbb{R}([0, \infty))$. Then we define $\mathcal{H} \cdot \mathcal{X}$ as the stochastic process for which $(\mathcal{H} \cdot \mathcal{X})_t$ is the limit from Proposition 16.20 for (an arbitrary) $\tau \ge t$.*

Proof of Proposition 16.20. It suffices to consider martingales with compact index set $t \in [0, \tau]$. For the approximating sequence $\mathcal{H}^1, \mathcal{H}^2, ...$ there are $\varepsilon_1, \varepsilon_2, ...$ with $\varepsilon_n \downarrow 0$ and $\sup_{0 \le t \le \tau} |H_t^n - H_t| \le \varepsilon_n$ almost surely by Lemma 16.6. Furthermore, because $\sup_{0 \le t \le \tau} |H_t^n - H_t^m| \le \varepsilon_m + \varepsilon_n$, because of Proposition 16.2,

$$||\mathcal{H}^n \cdot \mathcal{X} - \mathcal{H}^m \cdot \mathcal{X}|| = ||(\mathcal{H}^n - \mathcal{H}^m) \cdot \ Constance|| \le (\varepsilon_m + \varepsilon_n) \cdot || \ Constance|| \xrightarrow{m,n\to\infty} 0.$$

Therefore, $\int_n H \cdot X$ is a Cauchy sequence in $\mathcal{M}^2$ that, because of the completeness, converges to an element $\int H \cdot X \in \mathcal{M}^2$. □

Example 16.22 (Integral with respect to Brownian motion)**.** *Using the last proposition, we can now define integrals with respect to Brownian motion. For this purpose, let $\mathcal{X} = (X_t)_{t \ge 0}$ be a Brownian motion with $X_0 = 0$ and $f \in \mathcal{C}_b^2(\mathbb{R})$ is bounded. We will now show that*

$$\int_0^t f'(X_s) dX_s = f(X_t) - f(X_0) - \tfrac{1}{2} \int_0^t f''(X_s) ds.$$

For this purpose, we define recursively $T_0^\varepsilon = 0$, $T_{i+1}^\varepsilon = \inf\{t > T_i^\varepsilon : |X_t - X_{T_i^\varepsilon}| = \varepsilon\}$. Then, for suitable $C, C' > 0$

$$\mathbf{E}[T_{i+1}^\varepsilon - T_i^\varepsilon] = \mathbf{E}[T_1^\varepsilon] = C\varepsilon^2,$$
$$\mathbf{E}[(T_{i+1}^\varepsilon - T_i^\varepsilon)^2] = \mathbf{E}[(T_1^\varepsilon)^2] = C'\varepsilon^4$$

and thus

$$\begin{aligned}
&\mathbf{E}\Big[\Big(\sum_{i=0}^\infty f'(X_{T_i^\varepsilon\wedge t})(X_{T_{i+1}^\varepsilon\wedge t} - X_{T_i^\varepsilon\wedge t}) - \Big(f(X_t) - f(X_0) - \tfrac12\int_0^t f''(X_s)ds\Big)\Big)^2\Big] \\
&= \mathbf{E}\Bigg[\Bigg(\sum_{i=0}^\infty f'(X_{T_i^\varepsilon\wedge t})(X_{T_{i+1}^\varepsilon\wedge t} - X_{T_i^\varepsilon\wedge t}) - \Big(f(X_{T_{i+1}^\varepsilon\wedge t}) - f(X_{T_i^\varepsilon\wedge t}) - \tfrac12\int_{T_i^\varepsilon\wedge t}^{T_{i+1}^\varepsilon\wedge t} f''(X_s)ds\Big)\Bigg)^2\Bigg] \\
&= \sum_{i=0}^\infty \mathbf{E}\Big[\Big(f'(X_{T_i^\varepsilon\wedge t})(X_{T_{i+1}^\varepsilon\wedge t} - X_{T_i^\varepsilon\wedge t}) - \Big(f(X_{T_{i+1}^\varepsilon\wedge t}) - f(X_{T_i^\varepsilon\wedge t}) - \tfrac12\int_{T_i^\varepsilon\wedge t}^{T_{i+1}^\varepsilon\wedge t} f''(X_s)ds\Big)\Big)^2\Big] \\
&= \tfrac14\sum_{i=0}^\infty \mathbf{E}\Big[\Big(f''(X_{S_i^\varepsilon\wedge t})(X_{T_{i+1}^\varepsilon\wedge t} - X_{T_i^\varepsilon\wedge t})^2 - f''(X_{\tilde S_i^\varepsilon\wedge t})(T_{i+1}^\varepsilon\wedge t - T_i^\varepsilon\wedge t)\Big)^2\Big] \\
&\le \tfrac12\sum_{i=0}^\infty \mathbf{E}\Big[\Big(f''(X_{T_i^\varepsilon\wedge t})((X_{T_{i+1}^\varepsilon\wedge t} - X_{T_i^\varepsilon\wedge t})^2 - (T_{i+1}^\varepsilon\wedge t - T_i^\varepsilon\wedge t))\Big)^2\Big] \\
&\qquad + \mathbf{E}[(f''(X_{T_i^\varepsilon\wedge t}) - f''(X_{S_i^\varepsilon\wedge t}))^2\varepsilon^4] + \mathbf{E}[(f''(X_{T_i^\varepsilon\wedge t}) - f''(X_{\tilde S_i^\varepsilon\wedge t}))^2(T_{i+1}^\varepsilon\wedge t - T_i^\varepsilon\wedge t)^2] \\
&\xrightarrow{\varepsilon\to 0} 0
\end{aligned}$$

for random variables $T_i^\varepsilon \le S_i^\varepsilon, \tilde S_i^\varepsilon \le T_{i+1}^\varepsilon$ according to the Taylor formula and the mean value theorem. (In the last inequality sign, we used the simple estimate $(ab)^2 \le 2(a-a')^2b^2 + 2(a')^2b^2$). On the one hand, this means that $\sum_{i=0}^\infty f'(X_{T_i^\varepsilon\wedge t})(X_{T_{i+1}^\varepsilon\wedge t} - X_{T_i^\varepsilon\wedge t})$ converges in L^2 to $\int_0^t f'(X_s)dX_s$ by definition of the stochastic integral, but also to $f(X_t) - f(X_0) - \frac12\int_0^t f''(X_s)ds$. Due to the uniqueness of the L^2 limit, the assertion follows.

It seems clear that this calculation method for stochastic integrals is feasible, but not particularly elegant. Therefore, we will use the Itô formula (Theorem 16.51) to learn a simpler method for calculating in similar cases.

16.4 Local Martingales as Integrators

In calculations, it would often be nice to know that X_s for a martingale $\mathcal{X} = (X_t)_{t\ge 0}$ cannot be *too large*. If $\mathcal{X}$ has continuous paths, this is possible by transitioning to a sequence $\mathcal{X}^1, \mathcal{X}^2, ...$ of stopped martingales such that $\mathcal{X}^n$ is stopped whenever $|X_t| = n$. On the other hand, continuous processes for which such stopping times exist are not necessarily martingales. Therefore, we need a class of stochastic processes that is larger than the class of martingales.

Definition 16.23 (Local Martingale and Stopped Process). *1. A real-valued stochastic process $\mathcal{X} = (X_t)_{t\ge 0}$ is called a* local martingale *if there are stop times $T_1, T_2, ...$ with $T_n \uparrow \infty$ such that $(X_{t\wedge T_n})_{t\ge 0}$ is a martingale for all n. Here, $T_1, T_2, ...$ is called a localizing sequence of stopping times.*

2. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a stochastic process and T a stop time. Then we denote by $\mathcal{X}^T := (X_{T\wedge t})_{t\geq 0}$ the process stopped at T.*

In other words, the above definition implies that $\mathcal{X}$ is a local martingal if and only if $\mathcal{X}^{T_n}$ is a martingale for an appropriate sequence of stopping times $T_n \uparrow \infty$.

Remark 16.24 (Properties of Local Martingales)**.** 1. *According to the Optional Stopping Theorem, Proposition 13.19 and Corollary 13.53, every martingale is a local martingale.*

2. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a continuous local martingale. Then $T_n = \inf\{t : |X_t| = n \vee |X_0|\}$ is a localizing sequence of stopping times.*

   Because: Since $\mathcal{X}$ is real-valued and continuous, every path on $[0, t]$ takes its supremum. It follows that $T_n \uparrow \infty$ applies. Furthermore, let $S_m \uparrow \infty$ be a localizing sequence of stopping times for $\mathcal{X}$. For $s \leq t$,

$$\mathbf{E}[X_{t\wedge T_n}|\mathcal{F}_s] = \lim_{m\to\infty} \mathbf{E}[X_{t\wedge T_n\wedge S_m}|\mathcal{F}_s] = \lim_{m\to\infty} X_{s\wedge T_n\wedge S_m} = X_{s\wedge T_n},$$

   since $X_{t\wedge T_n}$ is bounded.

3. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a bounded local martingale. Then $\mathcal{X}$ is a martingale.*

   Because: Let $T_1, T_2, ...$ be a localizing sequence of stop times. According to majorized convergence, $\mathbf{E}[X_t|\mathcal{F}_s] = \lim_{n\to\infty} \mathbf{E}[X_{t\wedge T_n}|\mathcal{F}_s] = \lim_{n\to\infty} X_{s\wedge T_n} = X_s$ for $s \leq t$.

Example 16.25 (A genuine local martingale)**.** *Let $\mathcal{W} = (X_t, Y_t, Z_t)_{t\geq 0}$ be a three-dimensional Brownian motion started at $(x, y, z) \neq 0$. We now consider the process $\mathcal{V} = (V_t)_{t\geq 0}$, given by*

$$V_t = \frac{1}{\sqrt{X_t^2 + Y_t^2 + Z_t^2}}$$

and claim that, although $\mathcal{V}$ is a local martingale, it is not a martingale.
Because: The generator of the Brownian motion is well-known

$$(Gf)(w) = \frac{1}{2}\Big(\frac{\partial^2 f(w)}{\partial x^2} + \frac{\partial^2 f(w)}{\partial y^2} + \frac{\partial^2 f(w)}{\partial z^2}\Big).$$

Since for $f(w) = 1/\sqrt{x^2 + y^2 + z^2}$

$$\frac{1}{2}\frac{\partial^2 f(w)}{\partial x^2} f(w) = -\frac{\partial}{\partial x}\frac{x}{(x^2 + y^2 + z^2)^{3/2}} = -\frac{1}{(x^2 + y^2 + z^2)^{3/2}} + 3\frac{x^2}{(x^2 + y^2 + z^2)^{5/2}},$$

is for $w \neq 0$

$$(Gf)(w) = 0.$$

Let $B_\varepsilon(0)$ be the ball around 0 with radius ε and $g^\varepsilon \in \mathcal{C}_b^2(\mathbb{R}^3)$ such that $g^\varepsilon|_{B_\varepsilon(0)^c} = f|_{B_\varepsilon(0)^c}$ and $T_\varepsilon := \inf\{t > 0 : ||W_t||_2 = \varepsilon\}$.

According to Theorem 14.30, $\Big(g^\varepsilon(V_t) - \int_0^t (Gg^\varepsilon)(V_s)ds\Big)_{t\geq 0}$ is a martingale, so by the Optional Stopping Theorem – Proposition 13.19 – also

$$(f(V_{t\wedge T_\varepsilon}))_{t\geq 0} = (g^\varepsilon(V_{t\wedge T_\varepsilon}))_{t\geq 0} = \Big(g^\varepsilon(V_{t\wedge T_\varepsilon}) - \int_0^{T_\varepsilon\wedge t} (Gg^\varepsilon)(V_s)ds\Big)_{t\geq 0}$$

a martingale. Now let $T = T_\varepsilon \wedge T_R$ *for* $\varepsilon < ||W_0||_2 < R$. *Then* $f(V_{t\wedge T}))_{t\geq 0}$ *is also a martingale, hence*

$$1/||V_0||_2 = f(V_0) = g^\varepsilon(V_0) = \mathbf{E}[g^\varepsilon(V_T)] = \mathbf{P}(T_\varepsilon > T_R)\frac{1}{R} + (1 - \mathbf{P}(T_\varepsilon > T_R))\frac{1}{\varepsilon}.$$

It follows that

$$\mathbf{P}(T_\varepsilon > T_R) = \frac{1/\varepsilon - 1/||V_0||_2}{1/\varepsilon - 1/R} \xrightarrow{\varepsilon\to 0} 1.$$

Since $T_R \xrightarrow{R\to\infty} \infty$ *due to the continuity of the paths of the Brownian motion,* $T_\varepsilon \xrightarrow{\varepsilon\to 0} \infty$.

We have now shown that $(f(V_t))_{t\geq 0}$ *is a local martingale. However, one calculates*

$$\mathbf{E}[f(V_t)] = \mathbf{E}\Big[\frac{1}{||V_t||_2}\Big] = \frac{1}{\sqrt{t}} \cdot \mathbf{E}\Big[\frac{1}{||V_1||_2}\Big] \xrightarrow{t\to\infty} 0.$$

Since the expectation of a martingale is constant, $(f(V_t))_{t\geq 0}$ *cannot be a martingale.*

We already have stochastic integrals with respect to processes of bounded variation and with respect to continuous martingales. We now show that we have not treated any cases twice.

Theorem 16.26 (Continuous local martingales of bounded variation are constant)**.** *Let* $\mathcal{X} = (X_t)_{t\geq 0}$ *be a local martingale with continuous paths and bounded variation. Then* $t \mapsto X_t$ *is almost surely constant.*

Proof. Without restriction, assume that $X_0 = 0$ and $\mathcal{X}$ is a martingale. (If $\mathcal{X}$ is only a local martingale, we can show that all stopped martingales are constant along the localizing sequence of stopping times, which implies the statement.) Let $t \mapsto \nu_{1,t}(\mathcal{X})$ be the variation of $\mathcal{X}$. We define the stopping times

$$T_N := \inf\{t \geq 0 : \nu_{1,t}(\mathcal{X}) \geq N\}$$

and note that $(X_{t\wedge T_N})_{t\geq 0}$ is a martingale whose variation is bounded by N. Furthermore, $t \mapsto X_t$ is constant if and only if $t \mapsto X_{t\wedge T_N}$ is constant for some N. By assumption, $T_N \uparrow \infty$. Thus, it suffices to show the assertion in the case that the variation of $\mathcal{X}$ is bounded by N. For $t > 0$ we set $t_{n,k} := tk/n$ and define

$$\begin{aligned} Z_n &:= \sum_{k=1}^n (X_{t_{n,k}} - X_{t_{n,k-1}})^2 \leq \max_{1\leq k\leq n} |X_{t_{n,k}} - X_{t_{n,k-1}}| \cdot \sum_{k=1}^n |X_{t_{n,k}} - X_{t_{n,k-1}}| \\ &\leq \max_{1\leq k\leq n} |X_{t_{n,k}} - X_{t_{n,k-1}}| \cdot \nu_{1,t}(\mathcal{X}) \xrightarrow{n\to\infty} 0, \end{aligned}$$

since $\mathcal{X}$ has continuous paths. By definition, $Z_n \leq \nu_{1,t}(\mathcal{X})^2 \leq N^2$. Therefore, we conclude with majorized convergence that

$$\mathbf{E}[X_t^2] = \mathbf{E}\Big[\Big(\sum_{k=1}^n X_{t_{n,k}} - X_{t_{n,k-1}}\Big)^2\Big] = \mathbf{E}[Z_n] \xrightarrow{n\to\infty} 0.$$

Thus, $X_t = 0$ is almost sure. ☐

We continue with a frequently used characterization of local martingales.

Lemma 16.27 (Characterizations of local martingales)**.** *Let $T_n \uparrow \infty$ and $\mathcal{X}$ be a real-valued stochastic process. Then $\mathcal{X}$ is a local martingale if and only if $\mathcal{X}^{T_n}$ is a local martingale for all n.*

Proof. First, assume that $\mathcal{X}$ is a local martingale and $S_1, S_2, ...$ is a localizing sequence of stopping times. Then, for each stopping time T, the process $(\mathcal{X}^{S_n})^T = (\mathcal{X}^T)^{S_n}$ is a martingale according to the Optional Stopping Theorem. This means that $\mathcal{X}^T$ is a local martingale (with the localizing sequence of stopping times $S_1, S_2, ...$)

On the other hand, for each n, the process $\mathcal{X}^{T_n}$ is a local martingale with a localizing sequence of stopping times $S_1^n, S_2^n, ...$ Since $S_k^N \uparrow \infty$ for $k \to \infty$, we choose k_n with

$$\mathbf{P}(S_{k_n}^n < T_n \wedge n) \leq 2^{-n}.$$

According to the Borel-Cantelli lemma, $T_n' := S_{k_n}^n \wedge T_n \uparrow \infty$. To obtain an increasing sequence of stopping times, we define $T_n'' := \inf_{m \geq n} T_n'$. Furthermore, $\mathcal{X}^{T_n''} = (\mathcal{X}^{T_n'})^{T_n''} = ((X^{T_n})^{S_{k_n}^n})^{T_n''}$. By assumption, $(\mathcal{X}^{T_n})^{S_{k_n}^n}$ is a martingale, hence so is $\mathcal{X}^{T_n''}$. In particular, $\mathcal{X}$ is a local martingale. □

For the Brownian motion $\mathcal{X}$, we had already seen that the quadratic variation is given by $[\mathcal{X}]_t = t$. Since quadratic variation (and co-variation between two processes) will play a crucial role for stochastic integration with respect to local martingals (see Theorem 16.36), we will now construct it.

Proposition 16.28 (Quadratic Variation for bounded, continuous martingales)**.** *Let X be a bounded martingale with continuous paths and $X_0 = 0$. Further, let recursively $T_0^n = 0$ and*

$$T_{k+1}^n := \inf\{t > T_k^n : |X_t - X_{T_k^n}| > 2^{-n}\}.$$

Then there is an almost surely unique process $[\mathcal{X}] = ([\mathcal{X}]_t)_{t\geq 0}$ with $[\mathcal{X}]_0 = 0$ and non-decreasing paths of finite variation, such that for

$$\mathcal{Q}^n = (Q_t^n)_{t\geq 0} \text{ with } Q_t^n := \sum_{k=0}^{\infty} (X_{t\wedge T_{k+1}^n} - X_{t\wedge T_k^n})^2$$

holds that $\sup_{t\geq 0} |Q_t^n - [\mathcal{X}]_t| \xrightarrow{n\to\infty}_p 0$. *Furthermore, $\mathcal{X}^2 - [\mathcal{X}]$ is a martingale.*

Proof. The almost sure uniqueness follows from Theorem 16.26. (Assuming that there are two processes $[\mathcal{X}]$ and $[\mathcal{X}]'$ with the required properties. Then $[\mathcal{X}] - [\mathcal{X}]'$ would be a martingale with paths of finite variation, so almost surely constant. Since $[\mathcal{X}]_0 = [\mathcal{X}']_0$, $[\mathcal{X}] = [\mathcal{X}]'$ would be almost surely constant.) We define the process $\mathcal{H}^n = (H_t^n)_{t\geq 0} \in \mathbb{S}$ by

$$H_t^n = \sum_{k=0}^{\infty} X_{T_k^n} 1_{(T_k^n, T_{k+1}^n]}(t).$$

Then $||\mathcal{H}^n - \mathcal{X}||_2 \leq 2^{-n}$. According to Proposition 16.2.2, this also implies that

$$||\mathcal{H}^n \cdot \mathcal{X} - \mathcal{H}^m \cdot \mathcal{X}||_2 \xrightarrow{m,n\to\infty} 0.$$

Thus, $(\mathcal{H}^n \cdot \mathcal{X})_{n=1,2,\dots}$ is a M^2-Cauchy sequence that converges to a martingale $\mathcal{N} = (N_t)_{t\geq 0}$. Further, we write

$$\begin{aligned}
(\mathcal{H}^n \cdot \mathcal{X})_t &= \sum_{k=0}^{\infty} X_{T_k^n}(X_{T_{k+1}^n \wedge t} - X_{T_k^n \wedge t}), \\
X_t^2 &= \sum_{k=0}^{\infty} X_{T_{k+1}^n \wedge t}^2 - X_{T_k^n \wedge t}^2 \\
&= \sum_{k=0}^{\infty} (X_{T_{k+1}^n \wedge t} - X_{T_k^n \wedge t})^2 + 2X_{T_k^n \wedge t}(X_{T_{k+1}^n \wedge t} - X_{T_k^n \wedge t}) \\
&= Q_t^n + 2(\mathcal{H}^n \cdot \mathcal{X})_t.
\end{aligned}$$

We now define $[\mathcal{X}] := \mathcal{X}^2 - 2\mathcal{N}$ (so so that $\mathcal{X}^2 - [\mathcal{X}]$ is automatically a martingale). Then

$$\sup_{t\geq 0} |Q_t^n - [\mathcal{X}]_t| = \sup_{t\geq 0} |X_t^2 - 2(\mathcal{H}^n \cdot \mathcal{X})_t - X_t^2 + 2N_t = \sup_{t\geq 0} |2N_t - 2(\mathcal{H}^n \cdot \ ConX)_t| \xrightarrow{n\to\infty}_p 0.$$

Since $\mathcal{Q}^n$ has non-decreasing paths, the same is true for $[\mathcal{X}]$. □

We now extend the last result from bounded martingales to local martingales.

Theorem 16.29 (Quadratische Variation von lokalen Martingalen)**.** *Let $\mathcal{X}$ be a local martingale with continuous paths. Then there exists an almost surely unique process $[\mathcal{X}] = ([\mathcal{X}]_t)_{t\geq 0}$ with $[\mathcal{X}]_0 = 0$ and non-decreasing paths of finite variation, such that $\mathcal{X}^2 - [\mathcal{X}]$ is a local martingale. Furthermore, $[\mathcal{X}^T] = [\mathcal{X}]^T$ for every stopping time T.*

Proof. Again, the almost sure uniqueness follows from Theorem 16.26. Moreover, if $\mathcal{X}^2 - [\mathcal{X}]$ is a local martingale and T is an almost surely finite stopping time, then $(\mathcal{X}^T)^2 - [\mathcal{X}]^T = (\mathcal{X}^2 - [\mathcal{X}])^T$ is a local martingale, hence $[\mathcal{X}^T] = [\mathcal{X}]^T$.

For the existence of $[\mathcal{X}]$ we define $T_n := \inf\{t \geq 0 : |X_t| = n\}$, whereby $T_n \uparrow \infty$ holds. Further, $\mathcal{X}^{T_n}$ is a (by n) bounded martingale with continuous paths. Thus, $[\mathcal{X}^{T_n}]$ exists according to Proposition 16.28. Since for $m \geq n$ $(\mathcal{X}^{T_m})^{T_n} = \mathcal{X}^{T_n}$, $\mathcal{X}^{T_m} = \mathcal{X}^{T_n}$ on $[0, T_n]$, hence $[\mathcal{X}^{T_n}] = [\ \mathcal{X}^{T_m}]$ on $[0, T_n]$. We define $[\ \mathcal{X}]_t := \lim_{n\to\infty}[\ \mathcal{X}^{T_n}]_t$, where the convergence follows from $T_n \uparrow \infty$. Furthermore, it is clear that $(\mathcal{X}^2 - [\mathcal{X}])^{T_n} = (\mathcal{X}^{T_n})^2 - [\mathcal{X}^{T_n}]$ for each n is a martingale, so $\mathcal{X}^2 - [\mathcal{X}]$ is a local martingale. □

Example 16.30 (Martingales derived from the Brownian motion)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. As is well known (see example 13.47), $\mathcal{Y} = (Y_t)_{t\geq 0}$ with $Y_t = X_t^2 - t$ is a martingale. We show that*

$$[\mathcal{Y}]_t = 4 \int_0^t X_s^2 ds.$$

Because: According to example 14.26.2,

$$\Big(X_t^4 - 6 \int_0^t X_s^2 ds \Big)_{t\geq 0}$$

is a martingale. Further, we consider the process $(t, X_t)_{t\geq 0}$, i.e. the Markov process with state space $\mathbb{R}_+ \times \mathbb{R}$ and generator

$$(Gf)(t,x) = \frac{\partial f(t,x)}{\partial t} + \tfrac{1}{2} \frac{\partial^2 f(t,x)}{\partial x^2}.$$

With $f(t,x) = tx^2$, it follows from Theorem 14.30 that

$$\Big(tX_t^2 - \int_0^t X_s^2 + sds\Big)_{t\geq 0} = \Big(tX_t^2 - \tfrac{1}{2}t^2 - \int_0^t X_s^2 ds\Big)_{t\geq 0}$$

is a martingale. Thus,

$$\Big(X_t^4 - 6\int_0^t X_s^2 ds - 2tX_t + t^2 + 2\int_0^t X_s^2 ds\Big)_{t\geq 0} = \Big((X_t^2 - t)^2 - 4\int_0^t X_s^2 ds\Big)_{t\geq 0}$$

is also a martingale. Hence the assertion follows.

Theorem 16.31 (Covariation of local martingales)**.** *Let $\mathcal{X}, \mathcal{Y}$ be local martingales. Then there exists a near-unique process $[\mathcal{X}, \mathcal{Y}]$ of locally bounded variation and $[\mathcal{X}, \mathcal{X}]_0$ such that $\mathcal{X}\mathcal{Y} - [\mathcal{X}, \mathcal{Y}]$ is symmetrical and bilinear with*

$$[\mathcal{X}, \mathcal{Y}]^T = [\ Con,\ Con^T] = [\ Con^T,\ Con] = [\ Con^T,\ Con^T]$$

for every stopping time T.

Proof. We define

$$[\mathcal{X}, \mathcal{Y}] = \tfrac{1}{4}([\mathcal{X} + \mathcal{Y}] - [\mathcal{X} - \mathcal{Y}]).$$

Then

$$4(\mathcal{X}\mathcal{Y} - [\mathcal{X},\ Constance]) = (\ X + \mathcal{Y})^2 - (\mathcal{X} - \mathcal{Y})^2 - [\mathcal{X} + \mathcal{Y}] + [\ \mathcal{X} -\ \mathcal{Y}],$$

from which all assertions follow using Theorem 16.29. □

Proposition 16.32 (Continuity of quadratic variation)**.**
Let $\mathcal{X}^1 = (X_t^1)_{t\geq 0}, \mathcal{X}^2 = (X_t^2)_{t\geq 0}, ...$ be a sequence of local martingales with continuous paths starting from 0. Then, $\sup_{t\geq 0}|X_t^n| \xrightarrow{n\to\infty}_p 0$ if and only if $[\mathcal{X}^n]_\infty \xrightarrow{n\to\infty}_p 0$. In particular, the map $\mathcal{X} \mapsto [\mathcal{X}]$ is continuous on the space of local martingales with continuous paths. The same is true for the covariation.

Proof. First, let $\sup_{t\geq 0}|X_t^n| \xrightarrow{n\to\infty}_p 0$ and $\varepsilon > 0$. We define $T_n := \inf\{t \geq 0 : |X_t^n| > \varepsilon$, $n = 1, 2, ...$ Then for $\mathcal{Y}^n := (\mathcal{X}^n)^2 - [\mathcal{X}^n]$, the process $(\mathcal{Y}^n)^{T_n}$ is a martingale, which, according to Theorem 13.33 (and Theorem 13.51), can be extended to a martingale with index set $[0, \infty]$. Since $\mathbf{E}[Y_t^n] = 0$ and $X_{t\wedge T_n}^n \leq \varepsilon$, we have $\mathbf{E}[[\mathcal{X}^n]_{T_n}] \leq \varepsilon^2$. From the Markov inequality, we conclude

$$\mathbf{P}([\mathcal{X}^n]_\infty > \varepsilon) \leq \mathbf{P}(T_n < \infty) + \frac{1}{\varepsilon}\mathbf{E}[[\mathcal{X}^n]_{T_n}] \leq \mathbf{P}(\sup_{t\geq 0}|X_t^n| > \varepsilon) + \varepsilon.$$

By assumption, the right-hand side converges to ε, from which $[\mathcal{X}^n]_\infty \xrightarrow{n\to\infty}_p 0$ follows.

If, on the other hand, $[\mathcal{X}^n]_\infty \xrightarrow{n\to\infty}_p 0$, then we write $T_n := \inf\{t : [\mathcal{X}^n]_t > \varepsilon\}$

$$\begin{aligned}\mathbf{P}(\sup_{t\geq 0}|X_t^n| > \varepsilon) &\leq \mathbf{P}(T_n < \infty) + \mathbf{P}(\sup_{t\geq 0}|X_{t\wedge T_n}^n| > \varepsilon)\\ &\leq \mathbf{P}(T_n < \infty) + \frac{\mathbf{E}[[\mathcal{X}^n]_\infty^{T_n}]}{\varepsilon^2}\\ &= \mathbf{P}(T_n < \infty) + \frac{\mathbf{E}[[\mathcal{X}^n]_\infty \wedge \varepsilon]}{\varepsilon^2} \xrightarrow{n\to\infty} 0.\end{aligned}$$

□

Proposition 16.33 (Cauchy-Schwartz inequalities for covariation)**.** *Let $\mathcal{X}, \mathcal{Y}$ be local martingales and $\mathcal{H} = (H_t)_{t\geq 0}$ progressively measurable. Then*

$$[\mathcal{X}, \mathcal{Y}]^2 \leq [\mathcal{X}][\mathcal{Y}]$$

as well as

$$(\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])_t^2 \leq (\mathcal{H}^2 \cdot [\mathcal{X}])_t [\mathcal{Y}]_t$$

almost surely.

Proof. First, we note the following: Let $A, B \geq 0$. If $a^2 A + 2abC + b^2 B \geq 0$ for all a, b, then $C^2 \leq AB$. Namely, if we set $a = \pm 1/\sqrt{A}, b = 1/\sqrt{B}$, then $\pm 2C/\sqrt{AB} \geq -2$, thus $|C| \leq \sqrt{AB}$ or $C^2 \leq AB$.

For all a, b, because of the linearity of covariation,

$$0 \leq [a\mathcal{X} + b\mathcal{Y}] = a^2[\mathcal{X}] + 2ab[\mathcal{X}, \mathcal{Y}] + b^2[\mathcal{Y}]$$

almost certainly. (We can even choose the exception null set A independently of a, b. To do this, we note that the union of the exception null sets for $a, b \in \mathbb{Q}$ is again a null set. This must simultaneously be the exception null set for all $a, b \in \mathbb{R}$, since the quadratic variation is continuous; see Proposition 16.32.) From the initial remark, it now follows immediately that $[\mathcal{X}, \mathcal{Y}]_t^2 \leq [\mathcal{X}]_t[\mathcal{Y}]_t$ for all $t \geq 0$. The first assertion follows from the fact that the right-hand side is a non-decreasing function in t.

To show the second statement, we assume without further ado that $\mathcal{H} \geq 0$. Furthermore, we note that the first statement is analogously proven for subintervals, so $([\mathcal{X}, \mathcal{Y}]_t - [\mathcal{X}, \ ConY]_s)^2 \leq ([\ ConX]_t - [\ ConX]_s)([\mathcal{Y}]_t - [\ Con_Y]_s)$. Now let $\mathcal{H} = \sum_{i=1}^n G_i 1_{I_i}$ for disjoint open intervals $I_i = (V_i, W_i)$ and $G_1, ..., G_n \geq 0$. Then on A^c we have

$$\begin{aligned}
(\mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}])_t &= \sum_{i=1}^n G_i([\mathcal{X}, \mathcal{Y}]_{t\wedge W_i} - [\mathcal{X}, \mathcal{Y}]_{t \wedge V_i}) \\
&\leq \sum_{i=1}^n G_i \sqrt{[\mathcal{X}]_{t\wedge W_i} - [\mathcal{X}]_{t\wedge V_i}} \sqrt{[\mathcal{Y}]_{t\wedge W_i} - [\mathcal{Y}]_{t\wedge V_i}} \\
&\leq \sqrt{\sum_{i=1}^n G_i^2([\mathcal{X}]_{t\wedge W_i} - [\mathcal{X}]_{t\wedge V_i})} \sqrt{\sum_{k=1}^n [\mathcal{Y}]_{t\wedge W_k} - [\mathcal{Y}]_{t\wedge V_k}} \\
&= (\mathcal{H}^2 \cdot [\mathcal{X}])_t^{1/2} [\mathcal{Y}]_t^{1/2}.
\end{aligned}$$

The case of progressively measurable processes $\mathcal{H}$ follows from the last calculation, initially by approximating measurable sets instead of intervals $I_1, ..., I_n$. Subsequently, monotone convergence yields an approximation of $\mathcal{H}$, from which the statement follows by monotone convergence. □

Now we are dealing with the stochastic integral with local martingales as integrators.

Definition 16.34 (Stochastic integral of predictable, simple processes)**.** *Let $\mathcal{X}$ be a local martingale and $\mathcal{H}$ a simple, predictable process as in Definition 16.1. Then we define the stochastic integral $\mathcal{H} \cdot \mathcal{X}$ in the same way as in Definition 16.19.*

Lemma 16.35 (Stochastic integral for simple predictable processes)**.** *Let $\mathcal{X}, \mathcal{Y}$ be local martingales with continuous paths, and $\mathcal{H} \in \mathbb{S}$. Then $\mathcal{H} \cdot \mathcal{X}$ is a local martingale again and it holds that*

$$[\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}] = \mathcal{H} \cdot [\ ConX,\ ConY]. \tag{16.6}$$

Proof. First, $\mathcal{H} \cdot \mathcal{X}$ is a local martingale (with the same localizing sequence of stopping times as for $\mathcal{X}$. We have to show that $(\mathcal{H} \cdot \mathcal{X})\mathcal{Y} - \mathcal{H} \cdot [\ \mathcal{X},\ \mathcal{Y}]$ is a local martingale . Due to linearity, it suffices to consider $H_t = G(1_{T\wedge t} - 1_{S\wedge t})$ for G measurable with respect to $\mathcal{F}_S$. Since $\mathcal{X}\mathcal{Y} - [\mathcal{X}, \mathcal{Y}]$ is a local martingale, we conclude with repeated application of the Optional Sampling Theorem

$$\begin{aligned}
&\mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t Y_t - (\mathcal{H} \cdot [\ \mathcal{X},\ \mathcal{Y}])_t | \mathcal{F}_s] \\
&\quad = \mathbf{E}\big[G(X_{T\wedge t}Y_t - [\ \mathcal{X},\ \mathcal{Y}]_{T\wedge t} - (X_{S\wedge t}Y_t - [\mathcal{X},\ ConY]_{S\wedge t}))|\mathcal{F}_s\big] \\
&\quad = G1_{T\leq s}\big(X_{T\wedge s}Y_s - [\mathcal{X},\ ConY]_{T\wedge s} - (X_{S\wedge s}Y_s - [\mathcal{X}, \mathcal{Y}]_{S\wedge s})\big) \\
&\qquad + G1_{S\leq s\leq T}\mathbf{E}[X_{T\wedge t}Y_{T\wedge t} - [\mathcal{X}, \mathcal{Y}]_{T\wedge t} - (X_{S\wedge t}Y_s - [\mathcal{X}, \mathcal{Y}]_{S\wedge t}))|\mathcal{F}_s] \\
&\qquad + \mathbf{E}\big[G1_{s<S\leq t}\mathbf{E}[X_{T\wedge t}Y_{T\wedge t} - [\ Con_X,\ Con_Y]_{T\wedge t} - (X_{S\wedge t}Y_t - [\ Con_X,\ Con_Y]_{S\wedge t}))|\mathcal{F}_{S\vee s}]|\mathcal{F}_s\big] \\
&\quad = G1_{S\leq s}\big(X_{T\wedge s}Y_s - [\mathcal{X},\ ConY]_{T\wedge s} - (X_{S\wedge s}Y_s - [\ ConX,\ ConY]_{S\wedge s})\big) \\
&\quad = G\big((X_{T\wedge s} - X_{S\wedge s})Y_s - ([\mathcal{X}, \mathcal{Y}]_{T\wedge s} - [\mathcal{X},\ ConY]_{S\wedge s})\big) \\
&\quad = (\ ConH \cdot \mathcal{X})_s Y_s - (\ ConH \cdot [\ ConX,\ ConY])_s.
\end{aligned}$$

Using lemma 13.23, the statement follows. □

To extend the stochastic integral of integrands in $\mathbb{S}$ to progressive stochastic processes, we take (16.6) as the defining property.

Theorem 16.36 (General Stochastic Integral)**.** *Let $\mathcal{X}$ be a local martingale with continuous paths and $\mathcal{H}$ a progressively measurable stochastic process with $(\mathcal{H}^2 \cdot [\mathcal{X}])_t < \infty$ for all $t \geq 0$. Then there is an almost surely unique local martingale $\mathcal{H} \cdot \mathcal{X}$ with $(\mathcal{H} \cdot \mathcal{X})_0 = 0$, such that for every local martingale $\mathcal{Y}$ with continuous paths*

$$[\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}] = \mathcal{H} \cdot [\mathcal{X}, \mathcal{Y}]$$

holds.

Proof. First, we show uniqueness. If the process $\mathcal{H} \cdot \mathcal{X}$ were not unique, there would be two processes $\mathcal{Z}', \mathcal{Z}''$ such that $[\mathcal{Z}', \mathcal{Y}] = [\mathcal{Z}'', \mathcal{Y}] = [\mathcal{H} \cdot \mathcal{X}, \mathcal{Y}]$ for all local martingale $\mathcal{Y}$ with continuous paths. Then, due to the linearity of the covariation with $\mathcal{Y} = \mathcal{Z}' - \mathcal{Z}''$, it holds that $[\mathcal{Z}' - \mathcal{Z}''] = 0$. From Proposition 16.28 it follows that $\mathcal{Z}' = \mathcal{Z}''$.

For the existence of $\mathcal{H} \cdot \mathcal{X}$, it suffices, as in the proof of Theorem 16.29, to consider the case $\mathbf{E}[(\mathcal{H}^2 \cdot [\mathcal{X}])_\infty] < \infty$ (otherwise, we define $T_n := \inf\{t \geq 0 : (\mathcal{H}^2 \cdot \mathcal{X})_t \leq n\}$. Then, by assumption, $T_n \uparrow \infty$. There exist continuous local martingales $\mathcal{H} \cdot \mathcal{X}^{T_n}$ such that $[\mathcal{H} \cdot \mathcal{X}^{T_n}, \mathcal{Y}] = \mathcal{H} \cdot [\mathcal{X}^{T_n}, \mathcal{Y}]$ for all continuous local martingales $\mathcal{Y}$ and $n = 1, 2, ...$ Furthermore,

$\mathcal{H}\cdot\mathcal{X}^{T_n} = \mathcal{H}\cdot\mathcal{X}^{T_m}$ must hold for $m \geq n$ on $[0, T_n]$, since on the same set also $\mathcal{X}^{T_n} = \mathcal{X}^{T_m}$ holds. Now we can define $\mathcal{H}\cdot\mathcal{X} = \lim_{n\to\infty}\mathcal{H}\cdot\mathcal{X}^{T_n}$. Then $(\mathcal{H}\cdot\mathcal{X})^{T_n} = \mathcal{H}\cdot\mathcal{X}^{T_n}$. Since the latter is a local martingale for every n, it follows from lemma 16.27 that also $\mathcal{H}\cdot\mathcal{X}$ is a local martingale.)

We now have to show that $(\mathcal{H}\cdot\mathcal{X})\mathcal{Y} - \mathcal{H}\cdot[\mathcal{X},\mathcal{Y}]$ is a local martingale. First, using Proposition 16.33 and the Cauchy–Schwartz inequality, we have that for $\mathcal{Y}\in\mathcal{M}^2$

$$|\mathbf{E}[(\mathcal{H}\cdot[\mathcal{X},\mathcal{Y}])_\infty]| \leq \mathbf{E}[|\mathcal{H}\cdot[\mathcal{X},\mathcal{Y}])|_\infty] \leq (\mathbf{E}[(\mathcal{H}^2\cdot[\mathcal{X}])_\infty])^{1/2}(\mathbf{E}[[\mathcal{Y}]_\infty])^{1/2} < \infty.$$

Thus, $\mathcal{Y}\mapsto\mathbf{E}[(\mathcal{H}\cdot[\mathcal{X},\mathcal{Y}])_\infty]$ is a continuous, linear functional on $\mathcal{M}^2$. According to the Riesz-Fréchet theorem[21] (applied on the Hilbert space $\mathcal{M}^2$) there is a unique process in $\mathcal{M}^2$ that we call $\mathcal{H}\cdot\mathcal{X}$ such that

$$\mathbf{E}[(\mathcal{H}\cdot[\mathcal{X},\mathcal{Y}])_\infty] = \mathbf{E}[(\mathcal{H}\cdot\mathcal{X})_\infty\mathcal{Y}_\infty]$$

for all $\mathcal{Y}\in\mathcal{M}^2$. From this equation it also follows that $\mathcal{H}\mapsto\mathcal{H}\cdot\mathcal{X}$ is continuous, and the definition of $\mathcal{H}\cdot\mathcal{X}$ for $\mathcal{H}\in\mathbb{S}$ from Lemma 16.35 . Furthermore, by Theorem 16.31 for a stop time T

$$\mathbf{E}[(\mathcal{H}\cdot[\mathcal{X},\mathcal{Y}])_T] = \mathbf{E}[(\mathcal{H}\cdot[\mathcal{X},\ ConY^T])_\infty] = \mathbf{E}[(\mathcal{H}\cdot\mathcal{X})_\infty\ ConY^T_\infty] = \mathbf{E}[(\mathcal{H}\cdot\mathcal{X})_T\mathcal{Y}_T].$$

From lemma 13.23 it now follows that $(\mathcal{H}\cdot\mathcal{X})\mathcal{Y} - \mathcal{H}\cdot[\mathcal{X},\mathcal{Y}]$ is a martingale. From this the assertion follows. □

Corollary 16.37 (Kettenregel)**.** *Let H, X, Y be as in Theorem 16.36 and K progressively measurable with $(K^2\cdot[Y])_t < \infty$ for all $t \geq 0$. Then holds*

$$[\mathcal{H}\cdot\mathcal{X}, \mathcal{K}\cdot\mathcal{Y}] = (\mathcal{H}\mathcal{K})\cdot[\ \mathcal{X},\ \mathcal{Y}].$$

Proof. Since $\mathcal{K}\cdot\mathcal{Y}$ is a local martingale, it follows immediately from Theorem 16.36 that

$$[\mathcal{H}\cdot\mathcal{X}, \mathcal{K}\cdot\mathcal{Y}] = \mathcal{K}\cdot[\ \mathcal{H}\cdot\mathcal{X},\ \mathcal{Y}] =\ \mathcal{H}\cdot\mathcal{K}\cdot[\ \mathcal{X},\ \ \mathcal{Y}].$$

Since $[\ \mathcal{X}, xxx]$

rules of calculus as for the Lebesgue integral, in particular when calculating with densities. Now the assertion follows from Lemma **??**.2. □

Example 16.38 (Quadratische Variation von $X_t^2 - t$)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. According to example 16.30, for $\mathcal{Y} = (Y_t)_{t\geq 0}$ with $Y_t = X_t^2 - t$,*

$$[\mathcal{Y}]_t = 4\int_0^t X_s^2 ds.$$

We can verify this fact again with the help of the last result : According to example 16.22 with $f(x) = x^2$, it holds that

$$2\int_0^t X_s dX_s = X_t^2 - t$$

. Therefore, we write

$$[\mathcal{Y}] = [2\mathcal{X}\cdot\mathcal{X}] = 4\mathcal{X}^2\cdot[\mathcal{X}] = \Big(4\int_0^t X_s^2 ds\Big)_{t\geq 0}.$$

[21] Riesz-Fréchet theorem: Let H be a Hilbert space (with a scalar product $\langle .,\rangle$) over $\mathbb{R}$ and H' the space of linear, continuous mappings $H\to\mathbb{R}$. Then $x'\in H'$ can be written as: $x'(x) = \langle y, x\rangle$ for some suitable $y\in H$.

Finally, we prove a continuity property of the stochastic integral.

Proposition 16.39 (Continuity of the stochastic integral)**.** *Let $\mathcal{X}^1, \mathcal{X}^2, ...$ be continuous local martingales and $\mathcal{H}^1, \mathcal{H}^2, ...$ progressively measurable with $((\mathcal{H}^n)^2 \cdot [\mathcal{X}^n])_t < \infty$, $t \geq 0, n = 1, 2, ...$ Then $\sup_{t\geq 0} |(\mathcal{H}^n \cdot \mathcal{X}^n)_t| \xrightarrow{n\to\infty}_p 0$ holds only if $((\mathcal{H}^n)^2 \cdot [\mathcal{X}^n]) \xrightarrow{n\to\infty}_p 0$.*

Proof. By Corollary 16.37, $[\mathcal{H}^n, \mathcal{X}^n] = (\mathcal{H}^n)^2 \cdot [\mathcal{X}^n]$. Now it suffices to apply Proposition 16.32. □

16.5 Calculation Rules for Stochastic Integrals

The following are important rules for calculating with stochastic integrals. The most important ones are partial integration (Theorem 16.48) and the Itô formula (Theorem 16.51). However, in order to establish a framework that is as general as possible framework, we introduce the class of continuous semimartingales. A semimartingale $\mathcal{X}$ is the sum of a process of locally finite variation $\mathcal{A}$ and a local martingale $\mathcal{M}$. Integrals with respect to continuous semimartingales are then defined by the sum of the integrals with respect to $\mathcal{A}$ and $\mathcal{M}$.

Definition 16.40 (Semimartingale)**.**

1. *An adapted process $\mathcal{X}$ is called a semimartingale if it has right-continuous paths and can be written as $\mathcal{X} = \mathcal{A} + \mathcal{M}$, where $\mathcal{A}$ with $A_0 = 0$ is a process with locally finite variation and $\mathcal{M}$ a local martingale.*

2. *A continuous semimartingale $\mathcal{X}$ is a semimartingale for which $\mathcal{A}$ and $\mathcal{M}$ can be chosen as continuous processes. The decomposition $\mathcal{X} = \mathcal{A} + \mathcal{M}$ is then called the canonical decomposition.*

Lemma 16.41 (Canonical decomposition unique)**.** *The canonical decomposition of a continuous semimartingale is unique.*

Proof. Let $\mathcal{X} = \mathcal{A} + \mathcal{M} = \mathcal{A}' + \mathcal{M}'$ be two decompositions of the continuous semimartingale $\mathcal{X}$. Then we have $\mathcal{A} - \mathcal{A}' = \mathcal{M} - \mathcal{M}'$. Since the right-hand side is a continuous local martingale with $M_0 - M_0' = 0$, then by Theorem 16.26 we have that $\mathcal{M} = \mathcal{M}'$. Thus, $\mathcal{A} = \mathcal{A}'$. □

Example 16.42 (Decomposition of semimartingales not unique)**.** *The decomposition $\mathcal{X} = \mathcal{A} + \mathcal{M}$ is unique for continuous semimartingales, but not for semimartingales. As a simple counterexample, let $\mathcal{X} = (X_t)_{t\geq 0}$ be the Poisson process (with parameter λ). Then $\mathcal{X}$ can be written as*

$$X_t = X_t + 0 \qquad \text{and} \qquad X_t = t + (X_t - t).$$

Here $A_t = X_t, M_t = 0$ and $A_t' = t, M_t' = X_t - t$ are two different decompositions of $\mathcal{X}$.

Example 16.43 (Functionals of continuous Markov processes)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Markov process with state space E and continuous paths. The generator of $\mathcal{X}$ is $G^{\mathcal{X}}$ with $\mathcal{D}(G^{\mathcal{X}}) \subseteq \mathcal{C}_b(E)$. Then, for every $f \in \mathcal{D}(G)$, $(f(X_t))_{t\geq 0}$ is a semimartingale. Indeed, by Theorem 14.30*

$$f(X_t) = A_t + M_t \text{ with } A_t = \int_0^t (G^{\mathcal{X}} f)(X_s) ds, M_t = f(X_t) - \int_0^t (G^{\mathcal{X}} f)(X_s) ds$$

, which is a decomposition into the process with locally finite variation $\mathcal{A}$ and the (local) martingale $\mathcal{M}$.

Definition 16.44 (Continuous semimartingales as integrators)**.** *Let $\mathcal{X} = \mathcal{A} + \mathcal{M}$ be a continuous semimartingale. Define*

$$L(\mathcal{A}) := \{\mathcal{H} \text{ progressive} : \mathcal{H} \cdot \mathcal{A} \text{ exists}\}$$

as well as

$$L(\mathcal{M}) := \{\mathcal{H} : \mathcal{H}^2 \in L([\mathcal{M}])\}$$

and $L(\mathcal{X}) := L(\mathcal{A}) \cap L(\mathcal{M})$. For $\mathcal{H} \in L(\mathcal{X})$ we set

$$\mathcal{H} \cdot \mathcal{X} := \mathcal{H} \cdot \mathcal{A} + \mathcal{H} \cdot \mathcal{M}.$$

We define the quadratic variation of $\mathcal{X}$ as

$$[\mathcal{X}] := [\mathcal{M}].$$

If $\mathcal{Y} = \mathcal{B} + \mathcal{N}$ is the canonical decomposition of another continuous semimartingale, we set the covariation

$$[\mathcal{X}, \mathcal{Y}] := [\mathcal{M}, \mathcal{N}].$$

Lemma 16.45 (Covariation of the stochastic integral)**.** *Let $\mathcal{X} = \mathcal{A} + \mathcal{M}$ and $\mathcal{Y} = \mathcal{B} + \mathcal{N}$ the canonical decompositions of the continuous semimartingales $\mathcal{X}$ and $\mathcal{Y}$. Then*

$$4[\mathcal{X}, \mathcal{Y}] = [\mathcal{X} + \mathcal{Y}] - [\ Con -\ Con].$$

If $\mathcal{H} \in L(\mathcal{X})$, then $\mathcal{H} \cdot \mathcal{X} = \mathcal{H} \cdot \mathcal{A} + \mathcal{H} \cdot \mathcal{M}$ is the canonical decomposition of the semimartingale $\mathcal{H} \cdot \mathcal{M}$. Furthermore,

$$[\mathcal{H} \cdot \mathcal{M}, \mathcal{Y}] = \mathcal{H} \cdot [\mathcal{H}, \mathcal{Y}].$$

Proof. All statements follow directly from the definition of $\mathcal{H} \cdot \mathcal{X}$ and $[\mathcal{X}]$ as well as $[\mathcal{X}, \mathcal{Y}]$. □

We repeat a simple property of the integral with respect to a measure with density. If μ is a σ-finite measure, $\nu = g \cdot \mu$ is the measure with density g with respect to μ and f is a (bounded) measurable function, then $\nu[f] = g \cdot \mu[f] = \mu[fg]$. This statement has an analogue for stochastic integration with respect to processes of locally finite variation. Here is a simple example: If $\mathcal{H} = (H_t)_{t \geq 0}, \mathcal{K} = (K_t)_{t \geq 0}$ are progressively measurable processes, then

$$\mathcal{H} \cdot \Big(\int_0^t K_s ds \Big)_{t \geq 0} = \Big(\int_0^t H_s K_s ds \Big)_{t \geq 0}.$$

This example will now be significantly extended.

Proposition 16.46 (Chain Rule Theorem)**.** *Let $\mathcal{X} = (X_t)_{t \geq 0}$ be a continuous semimartingale and $\mathcal{K} \in L(\mathcal{X})$. Then $\mathcal{H} \in L(\mathcal{K} \cdot \mathcal{X})$ if and only if $\mathcal{H}\mathcal{K} \in L(\mathcal{X})$. In this case,*

$$\mathcal{H} \cdot (\mathcal{K} \cdot \mathcal{X}) = \ H\mathcal{K} \cdot \mathcal{X}$$

Proof. Let $\mathcal{X} = \mathcal{A} + \mathcal{M}$ be the canonical decomposition of $\mathcal{X}$. It is clear that $\mathcal{H}(\mathcal{K} \cdot \mathcal{A}) = (\mathcal{H}\mathcal{K}) \cdot \mathcal{A}$, which follows exactly as in the proof of Corollary 16.37 from Lemma **??**.2. if one of the two sides exists. Now we note that $\mathcal{H} \in L(\mathcal{K} \cdot \mathcal{M})$ if and only if $\mathcal{H}^2 \in L(\mathcal{K}^2 \cdot [\mathcal{M}])$, which is the case if and only if $\mathcal{H}^2\mathcal{K}^2 \in L([\mathcal{M}])$, hence $\mathcal{H} \cdot K \in L(\mathcal{M})$. Similarly, $\mathcal{H} \in L(\mathcal{K} \cdot \mathcal{A})$ if and only if $\mathcal{H} \cdot \mathcal{K} \in L(\mathcal{A})$, hence $\mathcal{H} \cdot \mathcal{K} \in L(\mathcal{A})$. Hence, the equivalence follows.

We now have to show that $\mathcal{H} \cdot (\mathcal{K} \cdot \mathcal{M}) = (\mathcal{H} \cdot \mathcal{K}) \cdot \mathcal{M}$. To do this, we note that for any local martingale $\mathcal{N}$

$$[(\mathcal{H} \cdot \mathcal{K}) \cdot \mathcal{M}, \mathcal{N}] = (\mathcal{H} \cdot \mathcal{K}) \cdot [\mathcal{M}, \mathcal{N}] = \mathcal{H} \cdot (\mathcal{K} \cdot [\mathcal{M}, \mathcal{N}]) = \mathcal{H} \cdot [\ \mathcal{K} \cdot \mathcal{M},\ N] = (\mathcal{H} \cdot \mathcal{K}) \cdot \mathcal{M}$$

is a local martingale with quadratic variation 0. Thus, the claim follows from Theorem 16.26. □

In integration theory, the theorem of majorized convergence played an important role. This now gets an analogue for stochastic integrals.

Proposition 16.47 (Majorized Convergence for Stochastic Integrals)**.** *Let $\mathcal{X}$ be a continuous semimartingale and $\mathcal{H}, \mathcal{K} = (K_t)_{t \geq 0}, \mathcal{K}^1 = (K_t^1)_{t \geq 0}, \mathcal{K}^2 = (K_t^2)_{t \geq 0}, ... \in L(\mathcal{X})$ with $|\mathcal{K}^n| \leq \mathcal{H}$ and $\sup_{t \geq 0} |K_n(t) - K(t)| \xrightarrow{n \to \infty}_{fs} 0$. Then $\sup_{0 \leq s \leq t} |(\mathcal{K}^n \cdot \mathcal{X} - \mathcal{K} \cdot \mathcal{X})_s| \xrightarrow{n \to \infty}_p 0$ for all $t \geq 0$.*

Proof. Let $\mathcal{X} = \mathcal{A} + \mathcal{M}$ be the canonical decomposition of $\mathcal{X}$. Since $\mathcal{H} \in L(\mathcal{X})$, it holds that $H \in L(\mathcal{A})$ and $H^2 \in L([\mathcal{M}])$. Since we can transfer the theorem on the majorized convergence to Stieltjes integrals , it follows that $((\mathcal{K}^n - \mathcal{K})^2 \cdot [\mathcal{M}])_t \xrightarrow{n \to \infty}_{fs} 0$ for $t \geq 0$. This implies (by stopping at t) according to Proposition 16.39 that $\sup_{0 \leq s \leq t} |(\mathcal{K}^n \cdot \mathcal{M} - \mathcal{K} \cdot \mathcal{M})_s| \xrightarrow{n \to \infty}_p 0$. Furthermore, $((\mathcal{K}^n - \mathcal{K}) \cdot \mathcal{A})_t \xrightarrow{n \to \infty}_{fs} 0$, again by the convergence in measure for Stieltjes integrals. From the last two convergences, the statement follows. □

The following is another rule of Lebesgue integration: If f is measurable and locally bounded and λ is the one-dimensional Lebesgue integral, then $f \cdot \lambda$ is of locally bounded variation. Fubini's theorem applies

$$((f \cdot \lambda) \cdot (f \cdot \lambda))_t = ((f(f \cdot \lambda)) \cdot \lambda)_t = \int_0^t f(s) \int_0^s f(r) dr ds = \tfrac{1}{2}(f \cdot \lambda)_t^2,$$

i.e. $2(f \cdot \lambda) \cdot (f \cdot \lambda) = (f \cdot \lambda)^2$. Analogously, we conclude that for a process with locally finite variation $\mathcal{A}$,

$$2\mathcal{A} \cdot \mathcal{A} = \mathcal{A}^2. \tag{16.7}$$

We will now generalize this result.

Theorem 16.48 (Partial Integration)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be continuous semimartingales. Then we have*

$$\mathcal{X} \cdot \mathcal{Y} = X_0 Y_0 + \mathcal{X} \cdot \mathcal{Y} + \mathcal{Y} \cdot \mathcal{X} + [\ \mathcal{X},\ \mathcal{Y}].$$

Proof. We first show only the case $\mathcal{X} = \mathcal{Y}$, i.e. $\mathcal{X}^2 = X_0^2 + 2\mathcal{X} \cdot \mathcal{X} + [\mathcal{X}]$. First, we consider $\mathcal{X} \in \mathcal{M}^2$ and set $T_0^n, T_1^n, ...$ and $\mathcal{Q}^n$ as in Proposition 16.28, and $\mathcal{H}^n = (H_t^n)_{t \geq 0} \in \mathbb{S}$ by

$$H_t^n = \sum_{k=0}^{\infty} X_{T_k^n} 1_{(T_k^n, T_{k+1}^n]}(t).$$

Just as in Propositin 16.28, $X_t^2 - X_0^2 = Q_t^n + 2(\mathcal{H}^n \cdot \mathcal{X})_t$. and $\sup_{t\geq 0} |H_t^n - X_t| \leq 2^{-n} \xrightarrow{n\to\infty 0}$.
Due to Proposition 16.47 we have

$$2(\mathcal{X} \cdot \mathcal{X})_t = \lim_{n\to\infty} 2(H^n \cdot \mathcal{X})_t = X_t^2 - X_0^2 - \lim_{n\to\infty} Q_t^n = X_t^2 - X_0^2 - [\mathcal{X}]_t.$$

If $\mathcal{X}$ is a local martingale, then the statement follows as for $\mathcal{X} \in \mathcal{M}^2$ by suitably stopping $\mathcal{X}$. On the other hand, if $\mathcal{X} = \mathcal{A}$ is a process with locally finite variation, then the statement follows as in (16.7).

Now consider the case of a semimartingale $\mathcal{X}$ with canonical decomposition $\mathcal{X} = \mathcal{A} + \mathcal{X}$. Then the assertion is equivalent to

$$\mathcal{M}^2 + 2\mathcal{M}\mathcal{A} + \mathcal{A}^2 = 2\mathcal{M} \cdot \mathcal{M} + 2\mathcal{M} \cdot \mathcal{A} + 2\mathcal{A} \cdot \mathcal{M} + 2\mathcal{A} \cdot \ ConA + [\ ConM].$$

So we have to show that

$$\mathcal{M}\mathcal{A} = \mathcal{A} \cdot \mathcal{M} + \mathcal{M} \cdot \mathcal{A}.$$

We now define for $t \geq 0$ and $n = 1, 2, ...$ the processes $\mathcal{A}^n = (A_s^n)_{0\leq s\leq t}$ and $\mathcal{M}^n = (M_s^n)_{0\leq s\leq t}$ by

$$A_s^n = A_{(k-1)t/n} and M_s^n = M_{kt/n} for s \in t(k-1, k]/n.$$

Then

$$\begin{aligned}(\mathcal{A}^n \cdot \mathcal{M})_t + (\mathcal{M}^n \cdot \mathcal{A})_t &= \sum_{k=1}^n A_{(k-1)t/n}(M_{kt/n} - M_{(k-1)t/n}) + M_{kt/n}(A_{kt/n} - A_{(k-1)t/n}) \\ &= \sum_{k=1}^n M_{kt/n}A_{kt/n} - M_{(k-1)t/n}A_{(k-1)t/n} = A_t M_t.\end{aligned}$$

From the majorized convergence for Stieltjes integrals (in the term $(\mathcal{M}^n \cdot \mathcal{A})_t$) and from Proposition 16.47 (in the term $(\mathcal{A}^n \cdot \mathcal{M})_t$) now follows the statement in the case $\mathcal{X} = \mathcal{Y}$. In the case $\mathcal{X} \neq \mathcal{Y}$, we write

$$\begin{aligned}4\mathcal{X}\mathcal{Y} &= (\mathcal{X} + \mathcal{Y})^2 - (\mathcal{X} - \mathcal{Y})^2 \\ &= (X_0 + Y_0)^2 + 2(\mathcal{X} + \mathcal{Y}) \cdot (\mathcal{X} + \mathcal{Y}) + [\mathcal{X} + \mathcal{Y}] - (X_0 - Y_0)^2 - 2(\mathcal{X} - \mathcal{Y}) \cdot (\mathcal{X} - \mathcal{Y}) - [\mathcal{X} - \mathcal{Y}] \\ &= 4X_0Y_0 + 4\mathcal{X} \cdot \mathcal{Y} + 4\mathcal{Y} \cdot \mathcal{X} + 4[\ \mathcal{X},\ \mathcal{Y}]\end{aligned}$$

and the statement is proven. □

Example 16.49 (Semimartingales derived from Brownian motion). *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. Then we saw in example 16.30 that $2\int_0^t X_s dX_s = X_t^2 - t$. This is also exactly the formula of partial integration (if you note that $[\mathcal{X}]_t = t$).*

In example 16.30, we saw that for a Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$, the process

$$\Big(tX_t^2 - \frac{1}{2}t^2 - \int_0^t X_s^2 ds\Big)_{t\geq 0}$$

is a martingale. This can also be seen by means of partial integration. Indeed, for $\mathcal{Y} = (Y_t)_{t\geq 0}$ *and* $Y_t = X_t^2 - t$, *we have*

$$\begin{aligned} tX_t^2 &= \Big(id \cdot \mathcal{X}^2 + \mathcal{X}^2 \cdot id\Big)_t \\ &= \Big(id \cdot \mathcal{Y} + id \cdot id + \mathcal{X}^2 \cdot id\Big)_t \\ &= \int_0^t s dY_s + \int_0^t s ds + \int_0^t X_s^2 ds \end{aligned}$$

Therefore, the above martingale is identical to $id \cdot \mathcal{Y}$.

Lemma 16.50 (The Covariation)**.** *Let* X *and* Y *be continuous semimartingales. For* $t > 0$, *let* $\zeta_n := \{t_{n,0}, ..., t_{n,k_n}\}$ *be a partition of* $[0, t]$. *If* $\max_k |t_{n,k} - t_{n,k-1}| \xrightarrow{n\to\infty} 0$, *then*

$$Z_n := \sum_{k=1}^{k_n} (X_{t_{n,k}} - X_{t_{n,k-1}})(Y_{t_{n,k}} - Y_{t_{n,k-1}}) \xrightarrow{n\to\infty}_p [\mathcal{X}, \mathcal{Y}]_t.$$

Proof. Similar to the last proof, we define

$$X_s^n = X_{t_{n,k-1}} \text{ and } Y_s^n = Y_{t_{n,k-1}} \text{ for } s \in (t_{n,k-1}, t_{n,k}].$$

Then

$$\begin{aligned} X_t Y_t &= \sum_{k=1}^{n} X_{t_{n,k}} Y_{t_{n,k}} - X_{t_{n,k-1}} Y_{t_{n,k-1}} \\ &= \sum_{k=1}^{n} X_{t_{n,k-1}}(Y_{t_{n,k}} - Y_{t_{n,k-1}}) + Y_{t_{n,k-1}}(X_{t_{n,k}} - X_{t_{n,k-1}}) \\ &\qquad\qquad + (X_{t_{n,k}} - X_{t_{n,k-1}})(Y_{t_{n,k}} - Y_{t_{n,k-1}}) \\ &= (\mathcal{X}^n \cdot \mathcal{Y})_t + (\mathcal{Y}^n \cdot \mathcal{X})_t + Z_n. \end{aligned}$$

Since $(\mathcal{X}^n \cdot \mathcal{Y})_t \xrightarrow{n\to\infty}_p (\mathcal{X} \cdot \mathcal{Y})_t$ and $(\mathcal{Y}^n \cdot \mathcal{X})_t \xrightarrow{n\to\infty}_p (\mathcal{Y} \cdot \mathcal{X})_t$ by Proposition 16.47, it follows with partial integration

$$\lim_{n\to\infty} Z_n = X_t Y_t - (\mathcal{X} \cdot \mathcal{Y})_t - (\mathcal{Y} \cdot \mathcal{X})_t = [\mathcal{X}, \mathcal{Y}]_t.$$

□

Theorem 16.51 (Itô-Formel)**.** *Let* $\mathcal{X}^1, ..., , \mathcal{X}^d$ *be continuous semimartingales and* $f \in \mathcal{C}^2(\mathbb{R}^d)$. *With* $\underline{\mathcal{X}} = (\mathcal{X}^1, ..., \mathcal{X}^d)$, $\underline{X}_t = (X_t^1, ..., X_t^d)$ *and* $f(\underline{\mathcal{X}}) = (f(\underline{X}_t))_{t\geq 0}$ *is*

$$f(\underline{\mathcal{X}}) = f(\underline{X}_0) + \sum_{i=1}^{d} f_i(\underline{\mathcal{X}}) \cdot \mathcal{X}^i + \tfrac{1}{2} \sum_{i,j=1}^{d} f_{ij}(\underline{\mathcal{X}}) \cdot [\mathcal{X}^i, \mathcal{X}^j]$$

where $f_i = \partial f / \partial x_i$ *and* $f_{ij} = \partial^2 f / \partial x_i \partial x_j$. *In particular, if* $d = 1$, *then*

$$f(\mathcal{X}) = f(X_0) + f'(\mathcal{X}) \cdot \mathcal{X} + \tfrac{1}{2} f''(\mathcal{X}) \cdot [\ Con_X]. \tag{16.8}$$

Proof. We only show the case $d = 1$. The general case can be proved analogously. Let $\mathcal{C} \subseteq \mathcal{C}_b(\mathbb{R})$ be the class of functions for which (16.8) holds. Then $\mathcal{C}$ is a vector space and id$\in \mathcal{C}$. We show that $\mathcal{C}$ is closed under multiplication. If namely $f, g \in \mathcal{C}$, then by Proposition 16.46 and Theorem 16.48,

$$\begin{aligned}(fg)(\mathcal{X}) - (fg)(X_0) &= f(\mathcal{X}) \cdot g(\mathcal{X}) + g(\mathcal{X}) \cdot f(\mathcal{X}) + [f(\mathcal{X}), g(\mathcal{X})] \\ &= f(\mathcal{X}) \cdot (g'(\mathcal{X}) \cdot \mathcal{X} + \tfrac{1}{2} g''(\mathcal{X}) \cdot [\mathcal{X}]) \\ &\qquad + g(\mathcal{X}) \cdot (f'(\mathcal{X}) \cdot \mathcal{X} + \tfrac{1}{2} f''(\mathcal{X} \cdot [\mathcal{X}]) + [f'(\mathcal{X}) \cdot \mathcal{X}), g'(\mathcal{X}) \cdot \mathcal{X}] \\ &= (fg' + f'g)(\ ConX) \cdot \mathcal{X} + \tfrac{1}{2}(f''g + 2f'g' + fg'')(\mathcal{X}) \cdot [\mathcal{X}] \\ &= (fg)'(\mathcal{X}) \cdot \ ConX + \tfrac{1}{2}(fg)''(\mathcal{X}) \cdot [\mathcal{X}].\end{aligned}$$

Let $f \in \mathcal{C}''(\mathbb{R})$ be arbitrary and $p_1, p_2, ... \in \mathcal{C}$ polynomials such that $\sup_{|x| \le c} |p_n(x) - f''(x)| \xrightarrow{n\to\infty} 0$ for each $c > 0$. By integration, polynomials $f_1, f_2, ...$ are obtained with

$$\sup_{|x|\le c} |f_n(x) - f(x)| \vee |f_n'(x) - f'(x)| \vee |f_n''(x) - f''(x)| \xrightarrow{n\to\infty} 0.$$

For the canonical decomposition $\mathcal{X} = \mathcal{A} + \mathcal{M}$, this means that, with majorized convergence for Stieltjes integrals,

$$(f_n'(\mathcal{X}) \cdot \mathcal{A} + \tfrac{1}{2} f_n''(\ ConX) \cdot [\ ConX] \xrightarrow{n\to\infty} (f'(\ ConX) \cdot \ ConA + \tfrac{1}{2} f''(\ ConX) \cdot [\ ConX]$$

uniformly on compact. Furthermore, it holds

$$((f_n(\mathcal{X}) - f(\mathcal{X}))^2 \cdot [\mathcal{X}])_t \xrightarrow{n\to\infty} 0,$$

so it follows from Proposition 16.39 that $f_n(\mathcal{X}) \cdot \mathcal{M} \xrightarrow{n\to\infty}_p f(\mathcal{X}) \cdot \mathcal{M}$ uniformly on compact. Hence the claim follows. □

Example 16.52 (Application to Brownian Motion)**.** *In example 16.22 we saw that for $f' \in \mathcal{C}_b^2(\mathbb{R})$ and a Brownian motion $\mathcal{X} = (X_t)_{t\ge0}$ holds that*

$$f'(\mathcal{X}) \cdot \mathcal{X} = f(\mathcal{X}) - f(X_0) - \tfrac{1}{2} f''(\mathcal{X}) \cdot [\mathcal{X}],$$

since $[\mathcal{X}]_t = t$. This is exactly the Itô formula applied to the semimartingale $\mathcal{X}$.

Remark 16.53 (Itô Formula as Taylor Expansion)**.** *The Itô formula can be written in different ways. If the stochastic integrals in* (16.8) *are written out, one obtains*

$$f(X_t) - f(X_0) = \int_0^t f'(X_s) dX_s + \tfrac{1}{2} \int_0^t f''(X_s) d[\mathcal{X}]_s.$$

Comparing this notation with Theorem 16.14 (which deals with processes of locally bounded variation), we see an important difference in calculating with continuous martingales compared to processes with locally bounded variation. The term $\frac{1}{2}\int_0^t f''(X_s)d[\mathcal{X}]_s$ is called the Itô-correction term. This is also obtained if $f(X_t)$is represented by the Taylor series up to the second term. In differential notation, this means

$$df(\mathcal{X}) = f'(\mathcal{X}) d\mathcal{X} + \tfrac{1}{2} f''(\mathcal{X}) d[\mathcal{X}].$$

Remark 16.54 (The Stratonovich integral)**.** *In addition to the stochastic Itô integral, stochastic integrals can also be introduced in the sense of R. Stratonovich. This is done by approximation using*

$$(\mathcal{H} \bullet \mathcal{X})_t = \lim_{n\to\infty} \sum_{i=1}^{k_n} \tfrac{1}{2}(H_{t_{i+1}^n} + H_{t_i^n})(X_{t_{i+1}^n} - X_{t_i^n}) \tag{16.9}$$

for partitions $0 = t_0^n < \cdots < t_{k_n}^n$ *with* $\max_i |t_{i+1}^n - t_i^n| \xrightarrow{n\to\infty} 0$ *is introduced. This construction is therefore similar to integration in the sense of Riemann. After constructing the stochastic integral, this leads to*

$$\mathcal{X} \bullet \mathcal{Y} = \mathcal{X} \cdot \mathcal{Y} + \tfrac{1}{2}[\mathcal{X}, \mathcal{Y}]$$

for semimartingales $\mathcal{X}$ *and* $\mathcal{Y}$*. Similarly, we get the formula*

$$f(\mathcal{X}) = f(X_0) + f'(\mathcal{X}) \bullet \mathcal{X}$$

for $f \in \mathcal{C}^2(\mathbb{R})$*, which is exactly the Itô formula without the Itô correction term. So it looks like the Stratonovich integral is the natural extension of the Lebesgue integral. However, in* (16.9) *one has to approximate the integration by non-adapted processes, which in turn seems unnatural. In stochastics, the Itô integral has become widely accepted.*

17 Applications of the Itô Formula

The Itô formula is considered the most important formula in stochastic integration theory. In this section, we will present some applications. In particular, we will focus on those that establish the relationship between general continuous local martingales and Brownian motion. For example, in Section 17.1 we will see that by means of a time transformation, every continuous local martingale can be transformed into a Brownian motion (Theorem 17.4), and in section 17.2 also as a stochastic integral with respect to a Brownian motion (Theorem 17.9). Furthermore, in section 17.3 we will use non-negative martingales to carry out a change of measure, whereby semimartingales can be converted into martingales can be converted into martingales (Theorem 17.14). Finally, in section **??**, the concept of the local time of a semimartingale is introduced, which allows an extension of Theorem 15.8 about the distribution of the maximum of a Brownian motion.

17.1 Transformations of Brownian Motion

The Itô formula provides a general transformation formula for continuous local martingales. Central to this is the notion of quadratic variation. We will first get to know Lévy's characterization of Brownian motion in Theorem 17.3 , which states that the Brownian motion is the only continuous local martingale $\mathcal{X}$ with $[\mathcal{X}]_t = t$. This not only emphasizes the importance of the Brownian motion, but also opens the door to representing general continuous local martingales in terms of the Brownian motion. For example, every continuous local martingale is a time-transformed Brownian motion (Theorem 17.4). The proof of Theorem 17.3 is easiest if we introduce $\mathbb{C}$-valued local martingales.

Remark 17.1 ($\mathbb{C}$-valued martingales)**.** *Let* $X = Y + iZ$ *be a stochastic process with values in* $\mathbb{C}$*, then we call it a (local) martingale if both* Y *and* Z *are (local) martingales.*

Lemma 17.2 (An exponential martingale)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a local martingale with $X_0 = 0$. Then $\mathcal{Z} = (Z_t)_{t\geq 0}$ with*

$$Z_t = \exp(iX_t + \tfrac{1}{2}[\mathcal{X}]_t)$$

is a $\mathbb{C}$-valued local martingale and

$$Z_t = 1 + i(\mathcal{Z} \cdot \mathcal{X})_t$$

almost surely.

Proof. We apply Itô's formula to the semimartingale $(iX_t + \frac{1}{2}[\mathcal{X}]_t)_{t\geq 0}$ and the function $f(z) = e^z$. This gives

$$dZ = Z(idX + \tfrac{1}{2}d[\mathcal{X}] + \tfrac{1}{2}d[i\mathcal{X}]) = iZdX$$

or $\mathcal{Z} = Z_0 + i(\mathcal{Z} \cdot \mathcal{X})$, from which the claim follows. □

Theorem 17.3 (Lévy's Characterization of Brownian Motion)**.** *Let $\underline{\mathcal{X}} = (\underline{X}_t)_{t\geq 0} = (\mathcal{X}^1, ..., \mathcal{X}^d)$, $\mathcal{X}^k = (X_t^k)_{t\geq 0}$ be an adapted stochastic process with $X_0^k = 0, k = 1, ..., d$. Then the following are equivalent:*

1. *$\underline{\mathcal{X}}$ is a Brownian motion, i.e. $X_t^k - X_s^k$ is distributed according to $N(0, t-s)$ for all k and $t - s$ and independent of $\mathcal{F}_s$ and of $(\mathcal{X}^l)_{l\neq k}$.*

2. *$\underline{\mathcal{X}}$ is a continuous local martingale and*

$$[\mathcal{X}^k, \mathcal{X}^l]_t = \delta_{kl} t.$$

Proof. 1.⇒2. is clear from example 13.47.
2.⇒1.: Let $\underline{\gamma} \in \mathbb{R}^d$. The process $(\langle \underline{\gamma}, \underline{X}_t \rangle)_{t\geq 0}$ is a local martingale with

$$[\langle \underline{\gamma}, \underline{\mathcal{X}} \rangle]_t = \sum_{k=1}^{d} \gamma_k^2 t = \langle \underline{\gamma}, \underline{\gamma} \rangle t.$$

From lemma 17.2 it follows that for each $\underline{\gamma} \in \mathbb{R}^d$ the process

$$(\exp(i\langle \underline{\gamma}, \underline{X}_t \rangle + \tfrac{1}{2}\langle \underline{\gamma}, \underline{\gamma} \rangle t))_{t\geq 0} \tag{17.1}$$

is a continuous martingale. From this it follows that

$$\mathbb{E}[\exp(i\langle \underline{\gamma}, \underline{X}_t - \underline{X}_s \rangle)|\mathcal{F}_s] = \exp(-\tfrac{1}{2}\langle \underline{\gamma}, \underline{\gamma} \rangle (t-s)),$$

from which all assertions follow. □

The importance of Brownian motion in the class of local martingales is now further emphasized. We now show that by means of a time transformation, we can transform every continuous local martingale into a Brownian motion.

Theorem 17.4 (Time Transformation, Dubins-Schwartz)**.** *Let $\mathcal{Y}$ be a continuous local martingale with $Y_0 = 0$, such that $[\mathcal{Y}]_t \uparrow \infty$ as $t \to \infty$. Define*

$$T_t := \inf\{u \geq 0 : [\mathcal{Y}]_u \geq t\}, \qquad \mathcal{G}_t := \mathcal{F}_{T_t}, \qquad X_t := Y_{T_t}.$$

Then $\mathcal{X} = (X_t)_{t\geq 0}$ is a $(\mathcal{G}_t)_{t\geq 0}$-adapted Brownian motion (thus in particular a martingale with respect to $(\mathcal{G}_t)_{t\geq 0}$). Furthermore, $[\mathcal{Y}]_t$ is a $(\mathcal{G}_t)_{t\geq 0}$ stopping time and it holds that

$$Y_t = X_{[\mathcal{Y}]_t}.$$

Remark 17.5 (Merkhilfe)**.** *The relationships $X_t := Y_{T_t}$ and $Y_t = X_{[\mathcal{Y}]_t}$ can be justified with the help of quadratic variation. After all, by definition of T_t and because $[\mathcal{X}]_t = t$, namely*

$$[\mathcal{X}]_t = t = [\mathcal{Y}]_{T_t}, \qquad [\mathcal{Y}]_t = [\ ConX]_{[\ ConY]_t}.$$

Example 17.6. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. We know from example 16.30 that $\mathcal{Y} = (Y_t)_{t\geq 0}$ with $Y_t = X_t^2 - t$ is a martingale and $[\mathcal{Y}]_t = 4\mathcal{X}^2 \cdot \lambda$. This means that $\mathcal{Y}$ can be transformed back into a Brownian motion by time change. See Figure 10.*

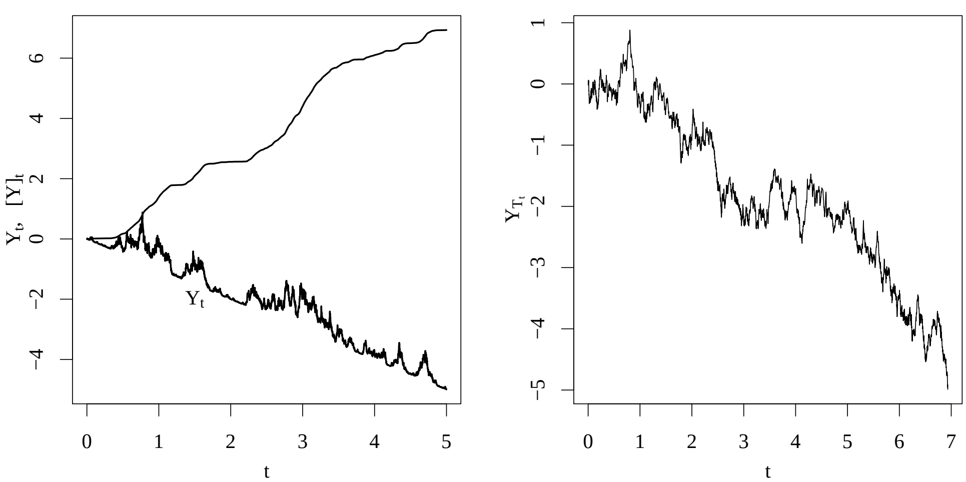


Figure 10: Sketch of the time transformation of the Brownian motion. Here, T_t is the first time for which the quadratic variation of $\mathcal{Y}$ reaches t .

Proof. We start the proof by showing that $\mathcal{X}$ has continuous paths. To do this, we need to show that the following almost certainly applies : If $t \mapsto T_t$ is constant on an interval, (i.e. $[\mathcal{Y}]_t$ is constant), then $\mathcal{Y}$ is also constant on the same interval. It suffices to assume that $\mathcal{X} \in \mathcal{M}^2$; otherwise we move on to stopped processes. Furthermore, it is sufficient to show the statement for intervals with rational endpoints. For

$$S_s := \inf\{t > s : [\mathcal{Y}]_t > [\ ConY]_s\}$$

we calculate with the help of the Optional Sampling Theorem

$$\mathbf{E}[(Y_{S_s})^2 - [\mathcal{Y}]_{S_s} | \mathcal{F}_s] = Y_s^2 - [\ ConstanceY]_s,$$

so because $[\ ConstanceY]_s = [\ ConstanceY]_{S_s}$ also

$$\mathbf{E}[(Y_{S_s} - Y_s)^2 | \mathcal{F}_s] = 0,$$

from which the continuity of $\mathcal{X}$ follows.

We now show that both $\mathcal{X}$ and $(X_t^2 - t)$ are continuous local martingales with respect to $(\mathcal{G}_t)_{t\geq 0}$. Then, by applying Theorem 17.3, the assertion follows. Now let

$$U_n := \inf\{t \geq 0 : |\mathcal{Y}_t| > n\}, \qquad V_n := [\mathcal{Y}]_{U_n}.$$

Then

$$T_{t\wedge V_n} = T_t \wedge U_n.$$

(If $t \leq V_n$, then the quadratic variation up to time U_n is already greater than t. Thus, the time up to which quadratic variation t is accumulated before U_n is $U_n \geq T_t$. If the other way around, $t \geq V_n$, then t has not yet been reached as quadratic variation up to $U_n. So, time passes before$ t is reached, so time still passes, so $T_t \geq U_n$. It is clear in this case $T_{V_n} = U_n$.) Since $\{V_n \leq t\} = \{U_n \leq T_t\} \in \mathcal{G}_t$, V_n is a $(\mathcal{G}_t)_{t\geq 0}$-stop time. Furthermore, for a $(\mathcal{G}_t)_{t\geq 0}$-stop time T, the following applies

$$\begin{aligned} \mathbf{E}[X_{t\wedge V_n \wedge T}] &= \mathbf{E}[Y_{T_{t\wedge T\wedge V_n}}] = \mathbf{E}[Y_{T_{t\wedge T}\wedge U_n}] = 0, \\ \mathbf{E}[X^2_{t\wedge V_n\wedge T} - t\wedge V_n \wedge T] &= \mathbf{E}[Y^2_{T_{t\wedge V_n \wedge T}} - t \wedge [\mathcal{Y}]_{U_n} \wedge T] = 0, \end{aligned}$$

Furthermore, for $s \leq t$

$$\mathbf{E}[X_{t\wedge V_n} | \mathcal{G}_s] = \mathbf{E}[Y_{T_{t\wedge V_n}} | \mathcal{G}_s] = \mathbf{E}[Y_{T_t \wedge U_n} | \mathcal{G}_s] = Y_{T_s\wedge U_n} = X_{s\wedge V_n}.$$

Thus, $\mathcal{X}$ is a local martingale with $V_1, V_2, ...$ as a localizing sequence of stopping times. Furthermore,

$$\begin{aligned} \mathbf{E}[X^2_{t\wedge V_n} - t\wedge V_n | \mathcal{G}_s] &= \mathbf{E}[Y^2_{T_t\wedge U_n} - [\mathcal{Y}]_{T_t\wedge U_n} + [\mathcal{Y}]_{T_{t\wedge V_n}} - t\wedge V_n | \mathcal{G}_s] \\ &= Y^2_{T_s\wedge U_n} - [\mathcal{Y}]_{T_s\wedge V_n} = X^2_{s\wedge V_n} - s \wedge V_n. \end{aligned}$$

Since $\mathcal{X}^2 - t$ is a local martingale, the claim follows. □

17.2 Martingale Representations

We now look at how to represent (continuous) local martingales can be represented as integrals with respect to Brownian motion. Let Y be a local martingale. Then a Brownian motion $\mathcal{X}$ is given and we are looking for a process $\mathcal{H}$ such that $\mathcal{Y} = \mathcal{H} \cdot \mathcal{X}$; see Theorem 17.9. On the other hand, the process $\mathcal{H}$ is given and a Brownian motion $\mathcal{X}$ is searched for, such that $\mathcal{Y} = \mathcal{H} \cdot \mathcal{X}$ holds; see Theorem 17.10.

Example 17.7. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion and $\mathcal{Y} = (Y_t)_{t\geq 0}$ with $Y_t = X_t^2 - t$.*

1. *We already know that $\mathcal{Y} = 2\mathcal{X} \cdot \mathcal{X}$ holds. This means that we have found a process $\mathcal{H} = 2\mathcal{X}$ such that $\mathcal{Y} = \mathcal{H} \cdot \mathcal{X}$. The general case of this is treated in Theorem 17.9.*

2. *Let $\mathcal{H} = 2\mathcal{X}$. We are looking for a Brownian motion $\mathcal{X}'$ such that $\mathcal{Y} = \mathcal{H}\cdot\mathcal{X}'$. Of course, in this example we could use the Brownian motion $\mathcal{X}' = \mathcal{X}$. But there is but there is another way: We note that $[\mathcal{Y}] = \mathcal{H}^2 \cdot \lambda$ according to example 16.30. Furthermore, we use (at least for times t with $X_t \neq 0$) the process $\mathcal{X}' = \frac{1}{2}\mathcal{X}^{-1} \cdot \mathcal{Y}$. This is a continuous local martingale with $[\mathcal{X}'] = \frac{1}{4}\mathcal{X}^{-2} \cdot [\mathcal{Y}] = \mathcal{X}^{-2}\mathcal{X}^2 \cdot \lambda = \lambda$ as in Example 16.30. With Theorem 17.3, it follows that $\mathcal{X}'$ is a Brownian motion. Furthermore, $\mathcal{H}\cdot\mathcal{X}' = \frac{1}{2}\mathcal{X}^{-1}2\mathcal{X}\cdot \mathcal{Y} = 1 \cdot \mathcal{Y} = \mathcal{Y}$ by Proposition 16.46. The general case is treated in Theorem 17.10.*

Proposition 17.8 (Representation of $Y \in L^2$ by a Brownian integral)**.**
Let $\underline{\mathcal{X}} = (\mathcal{X}^1, ..., \mathcal{X}^d)$ a d-dimensional Brownian motion, so that $(\mathcal{G}_t)_{t\geq 0}$ is the completed generated filtration of $\underline{\mathcal{X}}$ and $Y \in \mathcal{L}^2$ and measurable under $\mathcal{G}_\infty$. Then there exists an almost surely unique process $\underline{\mathcal{H}} = (\mathcal{H}^1, ..., \mathcal{H}^d)$ with $\mathbf{E}[\int_0^\infty ||H_s||_2^2 ds] < \infty$, such that

$$Y = \mathbf{E}[Y] + \sum_{i=1}^d (\mathcal{H}^i \cdot \mathcal{X}^i)_\infty.$$

Proof. Blumenthal's 0-1 law, Theorem 15.1, shows that $\mathbf{E}[Y|\mathcal{G}_0] = \mathbf{E}[Y]$ applies. Thus, it suffices to assume that $\mathbf{E}[Y] = 0$. Let $\mathcal{I}$ denote the Hilbert space of the $\mathcal{G}_\infty$-measurable random variables Z with $\mathbf{E}[Z] = 0$ and $\mathbf{E}[Z^2] < \infty$, and $\mathcal{J} \subseteq \mathcal{I}$ is the subspace of random variables that allows the desired representation $Z = \sum_{i=1}^d (\mathcal{H}^i \cdot \mathcal{X}^i)_\infty$. This representation is unique: namely, let $Z = \sum_{i=1}^d (\mathcal{H}^i \cdot \mathcal{X}^i)_\infty \sum_{i=1}^d (\mathcal{K}^i \cdot \mathcal{X}^i)_\infty$, then

$$0 = \mathbf{E}\Big[(\sum_{i=1}^d (\mathcal{H}^i - \mathcal{K}^i) \cdot \mathcal{X}^i)_\infty^2\Big] = \mathbf{E}\Big[\Big[\sum_{i=1}^d (\mathcal{H}^i - \mathcal{K}^i) \cdot \ ConX^i\Big]_\infty\Big] = \sum_{i=1}^d \mathbf{E}\Big[\int_0^\infty (H_s^i - K_s^i)^2 ds\Big],$$

so that $\mathcal{H}^i = \mathcal{K}^i$ is almost surely. Analogously, it follows that if $\mathcal{I}$ is complete, then so is $\mathcal{J}$, and that for $Z = \sum_{i=1}^d (\mathcal{H}^i \cdot \mathcal{X}^i)_\infty \in \mathcal{J}$

$$\mathbf{E}\Big[\int_0^\infty ||H_s||_2 ds\Big] = \mathbf{E}[Z^2] < \infty.$$

Thus, $\mathcal{J} \subseteq \mathcal{I}$ is a closed subspace and we have to show that for $Y \in \mathcal{I}$ with $Y \perp \mathcal{J}$ we always have $Z = 0$. Then, in fact, $\mathcal{J} = \mathcal{I}$.

Let $\underline{h} = (h^1, ..., h^d)$ be deterministic functions with $\int_0^\infty ||\underline{h}_s||_2^2 ds < \infty$. Then $\sum_{k=1}^d h^k \cdot \mathcal{X}^k$ is a local martingale with $[\sum_{k=1}^d h^k \cdot \mathcal{X}^k]_t = \int_0^t ||h_s||_2^2 ds$. By lemma 17.2 $\mathcal{Z} = (Z_t)_{t\geq}$ with $Z_t = \exp(i \sum_{i=1}^d (h^k \cdot \underline{\mathcal{X}}^k)_t + \frac{1}{2}\int_0^t ||\underline{h}_s||_2^2 ds)$ is a local martingale and $\mathcal{Z} - 1 = i(\mathcal{Z} \cdot \sum_{k=1}^d h^k \cdot \mathcal{X}^k) = i(\sum_{k=1}^d \mathcal{Z} h^k \cdot \mathcal{X}^k)$, where the last equality follows from Proposition 16.46. Thus, $\mathcal{Z} - 1 \in \mathcal{J}$ and $Y \perp \mathcal{J}$, that

$$0 = \mathbf{E}[Y(Z_\infty - 1)] = \mathbf{E}\Big[Y \exp(i \sum_{i=1}^d (h^k \cdot \underline{\mathcal{X}}^k)_\infty + \tfrac{1}{2}\int_0^\infty ||\underline{h}_s||_2^2 ds)\Big]$$

If you choose specifically step functions $h^1, ..., h^d$, it follows from the uniqueness of the characteristic function that

$$\mathbf{E}[Y, (X_{t_1}, ..., X_{t_n}) \in C] = 0$$

for $t_1, ..., t_n \in [0, \infty)$ and $C \in \mathcal{B}^n$. This statement can now be extended to $\mathbf{E}[Y, C] = 0$ for $C \in \mathcal{G}_\infty$. Since Y is measurable with respect to $\mathcal{G}_\infty$, it follows that $Y = 0$. □

Theorem 17.9 (Representation of local martingales by Brownian integrals).
Let $\underline{\mathcal{X}} = (\mathcal{X}^1, ..., \mathcal{X}^d)$ be a d-dimensional Brownian motion such that $(\mathcal{G}_t)_{t\geq 0}$ is the completed generated filtration of $\underline{\mathcal{X}}$, and $\mathcal{Y} = (Y_t)_{t\geq 0}$ is a local $(\mathcal{G}_t)_{t\geq 0}$ martingale. Then $\mathcal{Y}$ has almost surely continuous paths and there is an almost surely unique process $\underline{\mathcal{H}} = (\mathcal{H}^1, ..., \mathcal{H}^d)$ such that

$$\mathcal{Y} = Y_0 + \sum_{i=1}^{d} \mathcal{H}^i \cdot \mathcal{X}^i$$

almost surely.

Proof. The most important step is to show that $\mathcal{Y}$ almost surely has continuous paths. At first, it is sufficient to show this assertion if $\mathcal{Y}$ is uniformly integrable. (This can be achieved by a localizing sequence of stopping times.) So we write $Y_t = \mathbf{E}[Y_\infty|\mathcal{G}_t]$ for a $\mathcal{G}_\infty$-measurable random variable Y_∞. Further, we choose bounded random variables $Y_\infty^1, Y_\infty^2, ...$, such that $\mathbf{E}[|Y_\infty^n - Y_\infty|] \leq 3^{-n}$. Since $Y_\infty^n \in L^2$, we can Proposition 17.8 write Y_∞^n as an integral with respect to $\underline{\mathcal{X}}$. In particular, with $Y_t^n := \mathbf{E}[Y_\infty^n|\mathcal{G}_t]$, the martingale $(Y_t^n)_{t\geq 0}$ is continuous. Furthermore, it follows from Lemma 13.25 that

$$\mathbf{P}(\sup_{t\geq 0} |Y_t^n - Y_t| > 2^{-n}) \leq 2^n \mathbf{E}[|Y_\infty^n - Y_\infty|] \leq (2/3)^n.$$

Since the right-hand side is summable, the Borel-Cantelli lemma implies the uniform convergence of $\mathcal{Y}^n$ against $\mathcal{Y}$. Since the uniform limit of continuous functions is continuous, $\mathcal{Y}$ is continuous. Since $\mathcal{Y}$ is now continuous, there is a stopping time T such that $\mathcal{Y}^T$ is bounded. Applying Proposition 17.8 again shows that $\mathcal{Y}^T = (\sum_{k=1}^d \mathcal{H}^k \cdot \mathcal{X}^k)^T$ for suitable $\mathcal{H}^1, ..., \mathcal{H}^d$. □

In the next result, we need an extension of the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. This is a space $(\Omega', \mathcal{F}', \mathbf{P}')$ such that the embedding $\Omega \to \Omega'$ is measure-preserving. It is important to realize that further stochastic processes can be defined on Ω'.

Theorem 17.10 (Integral Representation of Local Martingales).
Let $\underline{\mathcal{Y}} = (\mathcal{Y}^1, ..., \mathcal{Y}^d)$ be a vector of continuous local martingales with $Y_0^i = 0, i = 1, ..., d$ and

$$[\mathcal{Y}^i, \mathcal{Y}^j]_t = \sum_{k=1}^{n} \int_0^t H_s^{i,k} H_s^{j,k} ds$$

for progressively measurable processes $\mathcal{H}^{i,k} = (H_t^{i,k})_{t\geq 0}$, $i = 1, ..., d$, $k = 1, ..., n$. Then there exists a extension of the probability space and a Brownian motion $\underline{\mathcal{X}} = (X^1, ..., X^n)$ with

$$Y^i = \sum_{k=1}^{n} H^{i,k} \cdot X^k, \quad i = 1, ..., d.$$

Proof. We consider the $d \times n$ matrix $(\mathcal{H}_t^{i,k})_{i=1,...,d,k=1,...,n}$. Let $N_t \subseteq \mathbb{R}^n$ be the kernel and $R_t \subseteq \mathbb{R}^d$ be the image of the corresponding linear map. Furthermore, let $N_t^\perp$ and $R_t^\perp$ be the orthogonal spaces. We denote the projections onto these subspaces by $\pi_{N_t}, \pi_{N_t^\perp}, \pi_{R_t}$ and $\pi_{R_t^\perp}$. Then $(\mathcal{H}_t^{i,k})_{i=1,...,d,k=1,...,n}$ is a bijection $N_t^\perp \to R_t$ and we denote its inverse with $((\mathcal{H}_t^{i,k})^{-1})_{i=1,...,d,k=1,...,n}$.

We extend the probability space by an independent Brownian motion $\underline{\mathcal{Z}} = (\mathcal{Z}^1, ..., \mathcal{Z}^n)$, and by $\mathcal{G}_t = \sigma(\mathcal{F}_t, (Z_s)_{s\leq t})$. Then $\underline{\mathcal{Y}}$ is a local martingale and $\underline{\mathcal{Z}}$ is a martingale with respect to $(\mathcal{G}_t)_{t\geq 0}$. We now define

$$\mathcal{X} = \mathcal{H}^{-1}\pi_R \cdot \mathcal{Y} + \pi_N \cdot \mathcal{Z}.$$

Thus

$$d\mathcal{X}_t^k = \begin{cases} \sum_{i=1}^d (H_t^{i,k})^{-1} dY_t^i, & H_t \in R_t \\ dZ_t, & H_t \notin R_t \end{cases},$$

$$d[\mathcal{X}^k, \mathcal{X}^l]_t = \begin{cases} \sum_{i,j=1}^d (H_t^{i,k})^{-1}(H_t^{j,l})^{-1} d[\mathcal{Y}^i,\ ConY^j]_t, & H_t \in R_t, \\ \delta_{kl} dt, & H_t \notin R_t \end{cases}$$

Since, by assumption,

$$\begin{aligned} \sum_{i,j=1}^d (H_t^{i,k})^{-1}(H_t^{j,l})^{-1} d[\mathcal{Y}^i, \mathcal{Y}^j]_t &= \sum_{m=1}^n \sum_{i,j=1}^d (H_t^{i,k})^{-1}(H_t^{j,l})^{-1} H_t^{i,m} H_t^{j,m} dt \\ &= \sum_{m=1}^n \delta_{km}\delta_{lm} dt = \delta_{kl} dt, \end{aligned}$$

so $d[\mathcal{X}^k, \mathcal{X}^l] = \delta_{kl} \cdot \lambda$. According to Theorem 17.3, $\underline{\mathcal{X}}$ is therefore a Brownian motion. Furthermore, Proposition 16.46 applies

$$\mathcal{Y}^i = 1 \cdot \mathcal{Y}^i = \sum_{j=1}^d \delta_{ij} \cdot \mathcal{Y}^j = \sum_{k=1}^n \sum_{j=1}^d (\mathcal{H}^{i,k} \cdot (\mathcal{H}^{j,k})^{-1} \cdot \mathcal{Y}^j = \sum_{k=1}^n \mathcal{H}^{i,k} \cdot \mathcal{X}^k.$$

□

17.3 Change of Measure and Transformations of the Drift

We now turn to two measures $\mathbf{P}$ and $\mathbf{Q}$ on the measure space $(\Omega, \mathcal{F})$ on which our stochastic processes are defined. We always assume that this space is equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ that is right-continuous and complete (with respect to $\mathbf{P}$); we write $\mathcal{F}_\infty := \sigma((\mathcal{F}_t)_{t\geq 0})$. Since the martingale property requires a conditional expectation , and this in turn requires a probability measure, we will now speak of $\mathbf{P}$-martingales and $\mathbf{Q}$-martingales, respectively. Furthermore, we write $\mathbf{E_P}[.]$ for expectation values with respect to $\mathbf{P}$ and $\mathbf{E_Q}[.]$ for expectation values with respect to $\mathbf{Q}$.

We know from the Radon–Nikodym theorem, Corollary 4.17, that in the case $\mathbf{Q} \ll \mathbf{P}$ (on $\mathcal{F}$) there is a random variable Z with $\mathbf{E_P}[Z] = 1$, so that $\mathbf{Q}(A) = \mathbf{E_P}[Z, A]$ holds. However, if only $\mathbf{Q} \ll \mathbf{P}$ on $\mathcal{F}_t$, i.e. $\mathbf{P}$-nullsets $N \in \mathcal{F}_t$ are also $\mathbf{Q}$-nullsets, there is at least a $\mathcal{F}_t$-measurable random variable Z_t such that $\mathbf{Q}(A) = \mathbf{E_P}[Z_t, A]$ for $A \in \mathcal{F}_t$. We now want to consider this case. We recall that $Z \cdot \mathbf{P}$ for a non-negative random variable Z with $\mathbf{E_P}[Z] = 1$ is the probability measure with density Z with respect to $\mathbf{P}$

Example 17.11 (Transformation of Brownian Motion from Proposition 13.48)**.** *We already know the basic example of a change of measure in Proposition 13.48. Let $\mathcal{X}$ be a Brownian motion, let $\mathbf{P}$ be its distribution, let $(\mathcal{F}_t)_{t\geq 0}$ be the generated filtration, and let $\mathcal{Z} = (Z_t)_{t\geq 0}$ with $Z_t = \exp(\mu X_t - \frac{1}{2}\mu^2 t)$ and $\mathbf{Q} = Z_t \cdot \mathbf{P}$ on $\mathcal{F}_t$. Then we showed that the distribution of $\mathcal{X}$*

under $\mathbf{Q}$ *is just that of a process* $\mathcal{Y}$ *with* $Y_t = X_t + \mu t$*. So, by changing the measure, we have changed the drift of the Brownian motion. Changing the distribution of a stochastic process by changing the measure is exactly the topic covered here.*

Lemma 17.12 (Measure transformations and martingales)**.** *Let* E *be a Polish space and* $\mathbf{P}$ *a probability measure on* $(E^I, (\mathcal{B}(E))^I)$ *for* $I = [0, \infty)$*. Further, let* $(\mathcal{F}_t)_{t\geq 0}$ *be a right-continuous and complete filtration.*

1. *If* $\mathcal{Z} = (Z_t)_{t\geq 0}$ *with* $Z_t \geq 0$ *and* $Z_0 = 1$ *is a* $\mathbf{P} - ((\mathcal{B}(E))_t)_{t\geq 0}$ *martingale, then there is a probability measure* $\mathbf{Q}$ *on* $(E^I, (\mathcal{B}(E))^I)$ *such that* $\mathbf{Q} = Z_t \cdot \mathbf{P}$ *on* $\mathcal{F}_t$*,* $t \geq 0$*.*
2. *Let* $\mathbf{Q}$ *be a probability measure such that* $\mathbf{Q} = Z_t \cdot \mathbf{P}$ *on* $\mathcal{F}_t$*,* $t \geq 0$*, for suitable* Z_t*. Then* $\mathcal{Z} = (Z_t)_{t\geq 0}$ *is a* $\mathbf{P}$*-martingale. It is uniformly integrable if and only if* $\mathbf{Q} \ll \mathbf{P}$ *on* $\mathcal{F}_\infty$*. Furthermore, a process* $\mathcal{X}$ *is a* $\mathbf{Q}$*-martingale if and only if* $\mathcal{X}\mathcal{Z}$ *is a* $\mathbf{P}$*-martingale.*

Proof. 1. We define a projective family of probability measures on $E^{[0,\infty)}$ by means of the martingale $\mathcal{Z}$. For this purpose, let $J = \{t_1, ..., t_n\}$ with $t_1 < ... < t_n$. Then we define

$$\mathbf{Q}_J(A_1 \times ... \times A_n) := \mathbf{E_P}[Z_{t_n}, A_1 \times \cdots \times A_n].$$

This family is projective because for $H = \{t_1, ..., t_{n-1}$

$$\begin{aligned}(\pi^J_H)_*\mathbf{Q}_J(A_1 \times ... \times A_{n-1}) &:= \mathbf{E_P}[Z_{t_n}, A_1 \times \cdots \times A_{n-1}] = \mathbf{E_P}[Z_{t_{n-1}}, A_1 \times \cdots \times A_{n-1}] \\ &= \mathbf{Q}_H(A_1 \times ... \times A_{n-1})\end{aligned}$$

due to the martingale property of $\mathcal{Z}$. Thus, by Theorem 5.24 a probability measure $\mathbf{Q}$ exists with the required properties.

2. Let $\mathcal{X} = (X_t)_{t\geq 0}$. It is clear that X_t belongs to $L^1(\mathbf{Q})$ if and only if $X_t Z_t \in L^1(\mathbf{P})$. In this case, $\mathbf{E_Q}[X_s, A] = \mathbf{E_Q}[X_t, A]$ for $A \in \mathcal{F}_s$ (which means that $\mathcal{X}$ is a $\mathbf{Q}$-martingale) holds only if $\mathbf{E_P}[Z_s X_s, A] = \mathbf{E_P}[Z_t X_t, A]$. The latter is the case for all $s \leq t$ if and only if $\mathcal{X}\mathcal{Z}$ is a $\mathbf{P}$-martingale. Thus, we have shown the last assertion. If we choose $\mathcal{X} = 1$, it follows in particular that $\mathcal{Z}$ must be a $\mathbf{P}$-martingale.

It remains to be shown that $\mathbf{Q} \ll \mathbf{P}$ if and only if $\mathcal{Z}$ is equally integrable. First, assume that $\mathcal{Z}$ is equally integrable. Then, by Theorem **??** (and Theorem **??**) $\mathcal{Z}$ converges in L^1 to a random variable Z_∞. Therefore, $\mathbf{Q}(A) = \mathbf{E}[Z_\infty, A]$ for all $A \in \bigcup_{t\geq 0} \mathcal{F}_t$. This property can be extended as usual for all $A \in \mathcal{F}_\infty$, which means that $\mathbf{Q} = Z_\infty \cdot \mathbf{P}$. In particular, $\mathbf{Q} \ll \mathbf{P}$. If the reverse is true, $\mathbf{Q} \ll \mathbf{P}$, then there is a Z with $\mathbf{Q} = Z \cdot \mathbf{P}$ on $\mathcal{F}_\infty$. Thus, necessarily $Z_t = \mathbf{E_P}[Z|\mathcal{F}_t]$. Furthermore, by lemma 11.5, $\mathcal{Z}$ is equipollarly integrable. □

Lemma 17.13 (Measure transformation at stopping times)**.** *Let* $\mathbf{Q} = Z_t \cdot \mathbf{P}$ *on* $\mathcal{F}_t$*,* $t \geq 0$*. Then, for every stopping time* T*,*

$$\mathbf{Q} = Z_T \cdot \mathbf{P} \text{ on } \mathcal{F}_t \cap \{T < \infty\}.$$

Furthermore, a process $\mathcal{X}$ *is a local* $\mathbf{Q}$*-martingale if and only if* $\mathcal{X}\mathcal{Z}$ *is a local* $\mathbf{P}$*-martingale.*

Proof. Let $A \in \mathcal{F}_T$. Then, due to monotone convergence and the optional sampling theorem,

$$\begin{aligned}\mathbf{Q}(A \cap \{T < \infty\}) &= \lim_{t\geq 0} \mathbf{Q}(A \cap \{T \leq t\}) = \lim_{t\geq 0} \mathbf{E_P}[Z_t, A \cap \{T \leq t\}] \\ &= \lim_{t\geq 0} \mathbf{E_P}[Z_{T\wedge t}, A \cap \{T \leq t\}] = \mathbf{E_P}[Z_T, A \cap \{T < \infty\}].\end{aligned}$$

Therefore, $\mathbf{Q} = Z_T \cdot \mathbf{P}$ on $\mathcal{F}_t \cap \{T < \infty\}$. Furthermore, for a stopping time T, the process $\mathcal{X}^T$ is a $\mathbf{Q}$-martingale if and only if $\mathcal{X}^T \mathcal{Z}$ is a $\mathbf{P}$-martingale. If we write $\mathcal{Z} = \mathcal{Z}^T + (\mathcal{Z} - \mathcal{Z}^T)$, it is clear that $\mathcal{X}^T(\mathcal{Z} - \mathcal{Z}^T)$ is a $\mathbf{P}$-martingale. Thus, $\mathcal{X}^T \mathcal{Z}$ is a $\mathbf{P}$-martingale if and only if $\mathcal{X}^T \mathcal{Z}^T$ is a $\mathbf{P}$-martingale. Thus, all assertions are shown. □

Theorem 17.14 (Girsanov Transformation)**.** *1. Let $\mathbf{Q} = Z_t \cdot \mathbf{P}$ on $\mathcal{F}_t$, $t \geq 0$, where $\mathcal{Z} = (Z_t)_{t \geq 0}$ has continuous paths ($\mathbf{P}$-almost surely). If $\mathcal{M}$ is a continuous, local $\mathbf{P}$-martingale, then, $\mathbf{Q}$-almost surely, the process $\mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]$ is locally bounded and $\widetilde{\mathcal{M}} := \mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]$ is a local $\mathbf{Q}$-martingale.*

2. A continuous process $\mathcal{Z} > 0$ is a local $\mathbf{P}$-martingale if and only if

$$\mathcal{Z} = \exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}])$$

for a continuous local $\mathbf{P}$-martingale $\mathcal{N}$. In this case, $\mathcal{N}$ is almost surely unique, and for every continuous local $\mathbf{P}$-martingale $\mathcal{M}$, $[\mathcal{M}, \mathcal{N}] = \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]$. Thus, if $\mathbf{Q} = Z_t \cdot \mathbf{P}$ on $\mathcal{F}_t$, then $\widetilde{=}\mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}] = \mathcal{M} - [\mathcal{M}, \mathcal{N}]$ is a local Q-martingale.

Proof. 1. We first assume that $\mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}] < \infty$ is finite. Then $\widetilde{\mathcal{M}}$ is a continuous $\mathbf{P}$-semimartingale, and from the formula for partial integration, Theorem 16.48, it follows that

$$\begin{aligned} \widetilde{\mathcal{M}} \cdot \mathcal{Z} - \widetilde{M}_0 Z_0 &= \widetilde{\mathcal{M}} \cdot \mathcal{Z} + \mathcal{Z} \cdot \widetilde{\mathcal{M}} + [\widetilde{\mathcal{M}}, \mathcal{Z}] \\ &= \widetilde{\mathcal{M}} \cdot \mathcal{Z} + \mathcal{Z} \cdot \mathcal{M} - \mathcal{Z} \cdot \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}] + [\ widetilde{\mathcal{M}}, \mathcal{Z}] \\ &= \widetilde{\mathcal{M}} \cdot \mathcal{Z} + \mathcal{Z} \cdot \mathcal{M}. \end{aligned}$$

Therefore, $\widetilde{\mathcal{M}}\mathcal{Z}$ is a local P-martingale. By lemma 17.13, $\widetilde{\mathcal{M}}$ is a local

In the general case, we define $T_n := \inf\{t \geq 0 : Z_t \leq 1/n\}$. Then, just as above, $\widetilde{\mathcal{M}}^{T_n}$ is a local $\mathbf{Q}$-martingale. Furthermore, $T_n \uparrow \infty$ is measurable under $\mathbf{P}$, because

$$\begin{aligned} \lim_{n \to \infty} \mathbf{Q}(T_n \leq t) &= \lim_{n \to \infty} \mathbf{E}_{\mathbf{P}}[Z_t, T_n \leq t] = \lim_{n \to \infty} \mathbf{E}_{\mathbf{P}}[Z_{T_n \wedge t}, T_n \leq t] \\ &= \lim_{n \to \infty} \mathbf{E}_{\mathbf{P}}[Z_{T_n}, T_n \leq t] = \lim_{n \to \infty} \mathbf{E}_{\mathbf{P}}[1/n, T_n \leq t] = 0 \end{aligned}$$

for every $t \geq 0$. Thus, by Lemma 16.27, $\widetilde{\mathcal{M}}$ is a local $\mathbf{Q}$-martingale.

2. If $\mathcal{N}$ is a local $\mathbf{P}$-martingale, then by the Itô formula

$$\begin{aligned} \exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) - \exp(N_0) &= \exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) \cdot (\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) + \tfrac{1}{2} \exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) \cdot [\mathcal{N}] \\ &= \exp(\mathcal{N} - \tfrac{1}{2}[\mathcal{N}]) \cdot \mathcal{N}. \end{aligned}$$

In particular, $\exp(\mathcal{N} - \frac{1}{2}[\mathcal{N}])$ is a local $\mathbf{P}$-martingale. If, on the other hand, $\mathcal{Z} > 0$, then the Itô formula again yields

$$\log \mathcal{Z} - \log Z_0 = \mathcal{Z}^{-1} \cdot \mathcal{Z} - \tfrac{1}{2}\mathcal{Z}^{-2} \cdot [\mathcal{Z}] = \mathcal{Z}^{-1} \cdot \mathcal{Z} - \tfrac{1}{2}[\mathcal{Z}^{-1} \cdot \mathcal{Z}],$$

and the assertion can be read off for $\mathcal{N} = \log Z_0 + \mathcal{Z}^{-1} \cdot \mathcal{Z}$. The almost certain uniqueness of $\mathcal{N}$ follows from the uniqueness of $\log \mathcal{Z}$ and Theorem 16.26. Furthermore, for a local $\mathbf{P}$-martingale $\mathcal{M}$

$$[\mathcal{M}, \mathcal{N}] = [\mathcal{M}, \mathcal{Z}^{-1} \cdot \mathcal{Z}] = \mathcal{Z}^{-1} \cdot [\mathcal{M},\ ConZ].$$

□

The quadratic variation $[\mathcal{X}]$ of a continuous semimartingale $\mathcal{X} = \mathcal{A} + \mathcal{M}$ was defined as the almost uniquely process of locally bounded variation such that $\mathcal{M}^2 - [\mathcal{X}]$ is a local martingale. Since this latter property depends on the probability measure $\mathbf{P}$, $[\mathcal{X}]$ depends on $\mathbf{P}$. Therefore, we now write $[\mathcal{X}]^{\mathbf{P}}$ for the quadratic variation with respect to the probability measure $\mathbf{P}$. However, in the next result, it turns out that $[\mathcal{X}]$ is not changed by transformation of $\mathbf{P}$. Furthermore, we have defined $\mathcal{H} \cdot \mathcal{X}$ by a martingale property. Therefore, we write $(\mathcal{H} \cdot \mathcal{X})^{\mathbf{P}}$ for the stochastic integral with respect to $\mathbf{P}$. Again, the stochastic integral is not affected by a change of measure. (This is clear for $\mathcal{H} \in \mathbb{S}$.)

Proposition 17.15 (Quadratic variation does not change under change of measure)**.** *Let $\mathbf{Q} = Z_t \cdot \mathbf{P}$ on $\mathcal{F}_t$ and $\mathcal{Z} = (Z_t)_{t\geq 0}$ be continuous. Then every P-semimartingale $\mathcal{X}$ is also a and $[\mathcal{X}]^{\mathbf{P}} = [\mathcal{X}]^{\mathbf{Q}}$. Furthermore, it holds that $(\mathcal{H} \cdot \mathcal{X})^{\mathbf{P}} = (\mathcal{H} \cdot \mathcal{X})^{\mathbf{Q}}$ for $\mathbf{Q}$-almost sure $\mathcal{H} \in L_{\mathbf{P}}(\mathcal{X})$, hence in particular $L_{\mathbf{P}}(\mathcal{X}) \subseteq L_{\mathbf{Q}}(\mathcal{X})$. Furthermore, for every $\mathbf{P}$-martingale $\mathcal{M}$, we have*

$$\mathcal{H} \cdot \mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{H} \cdot \mathcal{M}, \mathcal{Z}] = \mathcal{H} \cdot (\mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]).$$

Proof. If $\mathcal{X} = \mathcal{A}+\mathcal{M}$ is a $\mathbf{P}$-semimartingale, we write $\mathcal{X} = (\mathcal{A}+\mathcal{Z}^{-1}\cdot[\mathcal{M},\mathcal{Z}])+(\mathcal{M}-\mathcal{Z}^{-1}\cdot[\mathcal{Z}]$ is the decomposition of $\mathcal{X}$ into the process of locally bounded variation $\mathcal{A}+\mathcal{Z}^{-1}\cdot[\mathcal{M},\mathcal{Z}]$ and the $\mathbf{Q}$-martingale $\mathcal{M} - \mathcal{Z}^{-1} \cdot [\mathcal{M}, \mathcal{Z}]$. Thus, $\mathcal{X}$ is also a $\mathbf{Q}$-semimartingale.

The assertion $[\mathcal{X}]^{\mathbf{P}} = [\mathcal{X}]^{\mathbf{Q}}$ follows from lemma 16.50. If $\mathcal{H} \in L_{\mathbf{P}}(\mathcal{X})$, then $\mathcal{H} \in L_{\mathbf{P}}(\mathcal{A})$, and thus also $\mathcal{H} \in L_{\mathbf{Q}}(\mathcal{A})$. Furthermore, is $\mathcal{H}^2 \in L_{\mathbf{P}}([\mathcal{X}])$ and it follows $\mathcal{H}^2 \in L_{\mathbf{Q}}([\mathcal{X}]) = L_{\mathbf{Q}}([\ \mathcal{M}]) = L_{\mathbf{Q}}([\ \mathcal{M}-\ \ Z^{-1}\cdot[\mathcal{H} \in L_{\mathbf{Q}}(\mathcal{Z}^{-1}\cdot[\mathcal{M},\mathcal{Z}])$. Since $\mathbf{Q}$-almost surely $\mathcal{Z} > 0$, it suffices to show that $\mathcal{H} \in L_{\mathbf{Q}}([\mathcal{M},\mathcal{Z}]) = L_{\mathbf{Q}}([\ \ Z^{-1}\cdot[\mathcal{M},\mathcal{Z}],\mathcal{Z}])$. Since $\mathcal{H}^2 \in L_{\mathbf{Q}}([\mathcal{M}-\mathcal{Z}^{-1}\cdot[\mathcal{M},\ \mathcal{Z}])$, the assertion follows from Proposition 16.33. □

Remark 17.16 (Completeness of the filtration)**.** *The usual conditions for a filtration to hold are right-continuity and completeness (with respect to a probability measure $\mathbf{P}$). We will abandon the latter in the next result, and for good reason: let $\mathcal{X}$ be a Brownian motion, then*

$$\frac{X_t}{t} \xrightarrow{t\to\infty} 0$$

$\mathbf{P}$-almost surely. In particular, for every $\mu \neq 0$, $\{X_t/t \xrightarrow{t\to\infty} \mu\}$ is a $\mathbf{P}$-null set. However, if we now transform as in Proposition 13.48 $\mathbf{Q} = Z_t \cdot \mathbf{P}$, so that $\mathcal{X}$ under $\mathbf{Q}$ becomes a Brownian motion with drift μ, then $\mathbf{Q}\{X_t/t \xrightarrow{t\to\infty} \mu\} = 1$. In particular, then hold $\mathbf{Q} \ll \mathbf{P}$ (on $\mathcal{F}_\infty$). Therefore, we do not require the completeness of the filtration $(\mathcal{F}_t)_{t\geq 0}$.

Theorem 17.17 (Deterministic Transformation of Brownian Motion)**.** *Let $\mathcal{X} = (\mathcal{X}^1, ..., \mathcal{X}^d)$ with $\mathcal{X}^i = (X_t^i)_{t\geq 0}$ be a Brownian motion in $\mathbb{R}^d$ with canonical filtration $(\mathcal{F}_t)_{t\geq 0}$, and $h : [0,\infty)$ to $\mathbb{R}^d$ with $h(0) = 0$. Further, let $\mathbf{P}_0$ be the distribution of $\mathcal{X}$ and $\mathbf{P}_h$ be the distribution of $\mathcal{X}+h$. Then $\mathbf{P}_0 \sim \mathbf{P}_h$ on every $\mathcal{F}_t$ if and only if $h_i = f_i \cdot \lambda$ for functions f_i with $(f_i^2 \cdot \lambda)_t < \infty$ for all $t \geq 0$. In this case holds*

$$\mathbf{P}_h = \exp\Big(\Big(\sum_{i=1}^d f_i \cdot \mathcal{X}^i - \tfrac{1}{2} f_i^2 \cdot \lambda\Big)_t\Big) \cdot \mathbf{P}_0$$

on $\mathcal{F}_t$.

Proof. First, let $\mathbf{P}_0 \sim \mathbf{P}_h$ on every $\mathcal{F}_t$. Then there is a $\mathbf{P}_0$-martingale $\mathcal{Z}$ such that $P_h = Z_t \cdot P_0$ on $\mathcal{F}_t$. According to Theorem 17.9, $\mathcal{Z}$ has continuous paths. To see that $\mathcal{Z} > 0$ holds, let T be the hitting time of 0 of $\mathcal{Z}$. Assuming $T < t$ with positive probability for a $t > 0$, then by lemma 17.13 there would be a set A with $\mathbf{P}_h(A) = 0$, $\mathbf{P}_0(A) > 0$, in contradiction to $\mathbf{P}_0 \sim \mathbf{P}_h$. By theorem 17.14.2 there is a local martingale $\mathcal{M}$ such that $\mathcal{Z} = \exp(\mathcal{M} - \frac{1}{2}[\ ConM])$. With Theorem 17.9 there are processes $\mathcal{H}^1, ..., \mathcal{H}^d$ with $((\mathcal{H}^i)^2 \cdot \lambda)_t < \infty$ for all $t \geq 0$, so that $\mathcal{M} = \sum_{i=1}^d \mathcal{H}^i \cdot \mathcal{X}^i$. Now, each of the processes $\mathcal{X}^i - [\mathcal{X}^i, \mathcal{M}] = \mathcal{X}^i - \mathcal{H}^i \cdot \lambda$ is, by Theorem 17.14, a continuous $\mathbf{P}_h$-martingale with quadratic variation process λ, and vanishing covariations. According to Theorem 17.3, it is therefore a $\mathbf{P}_h$-Brownian motion. Now

$$\mathcal{X}^i = (\mathcal{X}^i - \mathcal{H}^i \cdot \lambda) + \mathcal{H}^i \cdot \lambda = (\mathcal{X}^i - h_i) + h_i,$$

which is two decompositions of the $\mathbf{P}_h$-semimartingale $\mathcal{X}^i$. Since this decomposition is unique, it follows that $h_i = \mathcal{H}^i \cdot \lambda$. Thus, $\mathcal{H}^i$ is deterministic, i.e. $h_i = f^i \cdot \lambda$ for some suitable f^i. Therefore, $\mathcal{M} = \sum_{i=1}^d \mathcal{H}^i \cdot \mathcal{X}^i = \sum_{i=1}^d f^i \cdot \mathcal{X}^i$.

Conversely, let $h_i = f_i \cdot \lambda$ for suitable $f_1, ..., f_d$. Using the Itô formula, it is calculated that $\exp\left(\sum_{i=1}^d f_i \cdot \mathcal{X}^i - \frac{1}{2} f_i^2 \cdot \lambda\right)_{t \geq 0}$ is a $\mathbf{P}_0$-martingale. By lemma 17.12 there is a probability measure $\mathbf{Q}$ with $\mathbf{Q} = \exp\left(\sum_{i=1}^d f_i \cdot \mathcal{X}^i - \frac{1}{2} f_i^2 \cdot \lambda\right)_t \cdot \mathbf{P}_0$ on $\mathcal{F}_t$. Furthermore, Theorem 17.14 implies that $\mathcal{X}^i - [\mathcal{X}^i, \sum_{j=1}^d f_j \cdot \mathcal{X}^j] = \mathcal{X}^i - f_i \cdot \lambda = \ Con_i$ is a martingale with quadratic variation λ and vanishing covariation. Thus, $(\mathcal{X}^1 - h_1 \cdot \lambda, ..., \mathcal{X}^d - h_d \cdot \lambda)$ is a Q-Brownian motion, and therefore $\mathbf{Q} = \mathbf{P}_h$. In particular, $\mathbf{P}_0 \sim \mathbf{P}_h$ on $\mathcal{F}_t$. □

17.4 Local time

The Itô formula can only be applied to twice continuously differentiable functions f. However, it is sometimes desirable to describe other functions of continuous semimartingals. Let be a Brownian motion $\mathcal{X}$. Is then $|\mathcal{X}|$ a semimartingale? The answer to this question leads to local time and the representation

$$|X_t| - |X_0| = \int_0^t \operatorname{sgn}(X_s) dX_s + L_t \text{ or } |\mathcal{X}| - |X_0| = \operatorname{sgn}(\mathcal{X}) \cdot \mathcal{X} + \mathcal{L}, \tag{17.2}$$

where $\mathcal{L} = (L_t)_{t \geq 0}$ is called the local time of the Brownian motion at point 0 and

$$\operatorname{sgn}(x) := \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases} \tag{17.3}$$

where $\mathcal{L}$ is an almost surely non-decreasing process (and thus has locally finite variation). The right-hand side of (17.2) is best visualized as follows : if $X_t > 0$, then the change in $|\mathcal{X}|$ in the time interval $[t, t + dt]$ is identical to the change in $\mathcal{X}$. On the other hand, if $X_t < 0$, then this change is precisely the negative change of $\mathcal{X}$. Furthermore, if $X_t = 0$, then $|\mathcal{X}|$ in $[t, t + dt]$ increases. On the other hand, such times do not play a role for the integral with respect to $\mathcal{X}$, since these times are a Lebesgue null set. Therefore, $\mathcal{L}$ is is the non-decreasing process, which grows if and only if $|\mathcal{X}| = 0$, so that (17.2) is true. (One can show in particular that

$$L_t := \lim_{h \to 0} \frac{1}{h} \int_0^t 1_{\{|X_s| < h\}} ds$$

holds.

Since the construction of local time works not only for Brownian motion but for general continuous semimartingales, we formulate the existence of local time in this case.

Theorem 17.18 (Tanaka's Formula)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a continuous semimartingale. Then there exists a continuous non-decreasing adapted process $\mathcal{L} = (L_t)_{t\geq 0}$ with*

$$|X_t| - |X_0| = \int_0^t sgn(X_s)dX_s + L_t.$$

It is almost surely true that

$$\int_0^\infty 1_{X_s\neq 0}dL_s = 0,$$

i.e. $\mathcal{L}$ only increases if $X_s = 0$. Furthermore,

$$L_t = 0 \vee (-|X_0| - \inf_{s\leq t}(sgn(\mathcal{X})\cdot\mathcal{X})_s).$$

To prove the theorem, we need an elementary statement.

Lemma 17.19. *Let $f : [0,\infty) \to \mathbb{R}$ be continuous with $f(0) \geq 0$. Then there is exactly one non-decreasing continuous function g with $g(0) = 0$ such that $f+g \geq 0$ and $\int 1_{\{f+g>0\}}dg = 0$, namely*

$$g(t) := 0 \vee \sup_{s\leq t}(-f(s)).$$

Proof. OBdA is $f(0) = 0$. Otherwise, we assume the function $f(t+t_0)$, where t_0 is the first time with $f(t) = 0$. In this case, we claim that $g(t) := \sup_{s\leq t}(-f(s))$ is the only function with the given properties. For this function, we first have

$$f(t) + g(t) \geq f(t) - \inf_{s\leq t} f(s) \geq f(t) - f(t) = 0.$$

Furthermore, if $f + g > 0$ on the interval (t, t'), then $g(t) = -\inf_{s\leq t} f(s) = -\inf_{s\leq t'} f(s) = g(t')$, from which $\int 1_{\{f+g>0\}}dg = 0$ follows.

Now we check the uniqueness. Let g and g' be two functions with the given properties. If $g(t) < g'(t)$ for some $t \geq 0$, then we define $r := \sup\{s \leq t : g(s) = g'(s)\}$. Then $f + g' \geq f + g' - f - g = g' - g > 0$ on $(r, t]$. From $\int 1_{\{f+g'>0\}}dg' = 0$ follows that $g'(r) = g'(t)$. And thus $0 < g'(t) - g(t) \leq g'(r) - g(r) = 0$, which is a contradiction . □

Proof of Theorem 17.18. We will approximate the function $f(x) = |x|$ by $\mathcal{C}^\infty$ functions, and apply the Itô formula to these. For this purpose, let $\varphi \in \mathcal{C}^\infty(\mathbb{R})$ be a increasing function such that $\varphi(x) = -1$ for $x \leq 0$ and $\varphi(x) = 1$ for $x \geq 1$. Then we choose f_n such that $f_n'(x) = \varphi(nx)$ and $f_n(0) = 0$. Then $f_n \uparrow f$ uniformly for all $x \in \mathbb{R}$. The Itô formula yields

$$f_n(\mathcal{X}) - f_n(X_0) = f_n'(\ Con_X)\cdot\mathcal{X} + \mathcal{L}^n$$

with $\mathcal{L}^n = \frac{1}{2}f_n''(\ Con_X)\cdot[\ Con_X]$. Since $f_n''(x) = 0$ for $x \geq 1/n$, $1_{\{\mathcal{X}>1/n\}}\cdot\mathcal{L}^n = 1_{\{\mathcal{X}>1/n\}}f_n''(\ Con_n)\cdot[\mathcal{X}] = 0$. Furthermore, since $\sup_{t\geq 0}|f_n'(X_t) - \text{sgn}(X_t)| \xrightarrow{n\to\infty}_{fs} 0$, it follows from Proposition 16.47 that $f_n'(\mathcal{X})\cdot\mathcal{X} \xrightarrow{n\to\infty}_p \text{sgn}(\mathcal{X})\cdot\mathcal{X}$ uniformly on compact. Furthermore, $f_n(\mathcal{X}) \xrightarrow{n\to\infty} fs|\mathcal{X}|$ uniformly for all t, so $\mathcal{L}^n = f_n(\mathcal{X}) - f_n(X_0) - f_n'(\mathcal{X})\cdot\mathcal{X}$ converges in probability to $|\mathcal{X}| - |X_0| - \text{sgn}(\mathcal{X})\cdot\mathcal{X}$. We denote this limit by $\mathcal{L}$. Since $\mathcal{L}^n$ is non-decreasing, $\mathcal{L}$ is non-decreasing as well. If we regard $\mathcal{L}^n$ as the distribution function of a measure, then $\mathcal{L}^n$ converges weakly to $\mathcal{L}$. Thus, by Theorem 9.6(iii), since $\mathcal{X}$ is continuous ,

$$1_{\{\mathcal{X}>1/m\}} \cdot \mathcal{L} \leq \liminf_{n\to\infty} 1_{\{\mathcal{X}>1/m\}} \cdot \mathcal{L}.$$

With $m \to \infty$ the second assertion follows. The last assertion follows if we apply lemma 17.19 for the process $|X_0| + \int_0^t \operatorname{sgn}(X_s)dX_s$ instead of . Namely, it follows directly that the function g must be the local time $\mathcal{L}$ itself. □

Corollary 17.20. *The same situation as in Theorem 17.18 applies to*

$$X_t^+ - X_0^+ = \int_0^t 1_{\{X_s>0\}} dX_s + \tfrac{1}{2} L_t,$$
$$X_t^- - X_0^- = \int_0^t 1_{\{X_s<0\}} dX_s + \tfrac{1}{2} L_t.$$

Proof. We just note that $x^+ = \frac{1}{2}(x + |x|)$ and $x^- = \frac{1}{2}(x - |x|)$. □

As an application of local time, we recall Theorem 15.8. There we have for a Brownian motion $(X_t)_{t\geq 0}$ started at 0 and $M_t = \sup_{s\leq t} X_s$ that $M_t - X_t \stackrel{d}{=} |X_t|$. In remark 15.9, we claimed that this equality in distribution also holds along the entire path. We can now verify this.

Proposition 17.21 (Path-valued reflection principle)**.** *Let X be a Brownian motion started at 0 and L be the local time at 0 as in Theorem 17.18. Further, let $M = (M_t)_{t\geq 0}$ with $M_t = \sup_{s\leq t} X_s$. Then*

$$(\mathcal{L}, |\mathcal{X}|) \stackrel{d}{=} (\mathcal{M}, \mathcal{M} - \mathcal{X}).$$

Proof. We define $\mathcal{X}' = -\operatorname{sgn}\mathcal{X} \cdot \mathcal{X}$. Then $\mathcal{X}'$ is a continuous local martingale with $[\mathcal{X}] = (\operatorname{sgn}(\mathcal{X})^2 \cdot [\mathcal{X}] = \lambda$. Thus, $\mathcal{X}'$ is also a Brownian motion. Furthermore,

$$M_t' := \sup_{s\leq t} X_s' = -\inf_{s\leq t} (\operatorname{sgn}\mathcal{X} \cdot \mathcal{X})_s = L_t$$

and by the Tanaka formula

$$|\mathcal{X}| = -\mathcal{X}' + \mathcal{L} = \mathcal{M}' - \mathcal{X}'.$$

It follows that $(\mathcal{L}, |\mathcal{X}|) = (\mathcal{M}', \mathcal{M}' - \mathcal{X}') \stackrel{d}{=} (\mathcal{M}, \mathcal{M} - \ Con)$. □

18 Stochastic differential equations and diffusions

In many fields of application, such as biology, financial mathematics, or engineering, the processes described by stochastic differential equations, are particularly important. These are best thought of as the solution of ordinary differential equations, but with a noise term integrated in, so that the solution is a stochastic process.

18.1 Stochastic differential equations

In this section, we study equations of the form

$$dX^i = \mu^i(t, \underline{X}_t)dt + \sum_{j=1}^{n} \sigma^{ij}(t, \underline{X})dW_t^j, \tag{18.1}$$

with $\underline{X}_t = (X_t^1, ..., X_t^d)$, where $\mathcal{W} = (\mathcal{W}^j)_{j=1,...,n}$ with $\mathcal{W}^j = (W_t^j)_{t\geq 0}$ is an n-dimensional Brownian motion, and $t \mapsto \mu^i(t, \underline{X}_t), t \mapsto \sigma^{ij}(t, \underline{X}_t)$, $i = 1, ..., d, j = 1, ..., n$ are progressively measurable processes. Such equations are also called stochastic differential equations or SDEs (from *Stochastic Differential Equations*). Abbreviated we also write $\underline{\mu} = (\mu^i)_{i=1,...,d}, \underline{\underline{\sigma}} = (\sigma^{ij})_{i=1,...,d,j=1,...,n}$ and

$$d\underline{X} = \underline{\mu}(t, \underline{X})dt + \underline{\underline{\sigma}}(t, \underline{X})d\underline{W}.$$

or in integral notation

$$\underline{X}_t = \underline{X}_0 + \int_0^t \underline{\mu}(s, \underline{X}_s)ds + \int_0^t \underline{\underline{\sigma}}(s, \underline{X})d\underline{W}_s$$

or even shorter

$$\underline{\mathcal{X}} = \underline{\mu}(., \underline{\mathcal{X}}) \cdot \lambda + \underline{\underline{\sigma}}(., \underline{X}) \cdot \underline{\mathcal{W}}.$$

Example 18.1 (Some simple cases). *1. If $\sigma^{ij} = 0$ for all i, j, then* (18.1) *becomes*

$$dX^i = \mu^i(t, \underline{X})dt, \qquad \text{or} \qquad \dot{X}^i = \mu^i(t, \underline{X}).$$

Equations of this form are also referred to as (ordinary) differential equations.

2. *An especially simple differential equation is $\dot{X} = \mu X$, which is solved by $X_t = X_0 e^{\mu t}$. But what about*

$$dX = \mu X dt + \sigma X dW$$

? We can rewrite this as $d(lnX) = \mu dt + \sigma dW$. This formulation suggests applying the Itô formula to the (yet to be process $(\ln X_t)_{t\geq 0}$). This yields (because $[\mathcal{X}] = [\sigma\mathcal{X} \cdot \mathcal{W}] = \sigma^2 \mathcal{X}^2 \cdot \lambda$)

$$d \ln X = \frac{1}{X}dX - \frac{1}{X^2}d[\mathcal{X}] = \mu dt + \sigma dW - \tfrac{1}{2}\sigma^2 dt,$$

or also

$$\ln X_t = \ln X_0 + (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t.$$

Solving for X_t yields

$$X_t = X_0 \exp\Big((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\Big).$$

The process $\mathcal{X}$ is also called geometric Brownian motion and will be a model for the development of a share price in the section on financial mathematics .

It turns out that equations of the form (18.1) allow for different solution concepts.

Definition 18.2 (Strong and weak solutions of an SDE)**.** *We write* $\underline{\mathcal{X}} = (\mathcal{X}^i)_{i=1,\dots,d}$ *with* $\mathcal{X}^i = (X_t^i)_{t\geq 0}$.

1. *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space on which an n-dimensional Brownian motion* $\underline{\mathcal{W}}$ *is defined, and* $(\mathcal{F}_t)_{t\geq 0}$ *be the filtration generated by* $\underline{\mathcal{W}}$. *A* $(\mathcal{F}_t)_{t\geq 0}$*-adapted process* $\underline{\mathcal{X}}$, *for which* (18.1) *(with the given Brownian motion* $\underline{\mathcal{W}}$*) applies, is called a* strong solution of the SDE (18.1)*.*

2. *A probability space* $(\Omega, \mathcal{F}, \mathbf{P})$, *on which an n-dimensional Brownian motion* $\underline{\mathcal{W}}$ *with generated filtration* $(\mathcal{F}_t)_{t\geq 0}$, *and an adapted process* $\underline{\mathcal{X}}$ *are defined, for which* (18.1) *hold, is called a weak solution of the SDE* (18.1). *In short, we also say that* $(\underline{\mathcal{W}}, \underline{\mathcal{X}})$ *is a weak solution.*

The difference between a strong and a weak solution is that for strong solutions, the Brownian motion $\mathcal{W}$ is given in advance. For weak solutions, not only the process $\mathcal{X}$ that solves (18.1) is sought, but also the Brownian motion $\mathcal{W}$ that appears on the right-hand side in (18.1). It is clear that every strong solution also allows for a weak solution. The examples from Example (18.1) were all strong solutions.

Example 18.3 (An SDE that is weakly but not strongly solvable)**.** *We consider the SDE*

$$dX = sgn(X)dW, \tag{18.2}$$

where sgn is defined as in (17.3). *Since* $(sgn(X))^2 = 1$ *holds, no matter what the process* $\mathcal{X}$ *looks like, it follows that the solution of this SDE must be a Brownian motion. (After all, the solution is a local martingale with* $[\mathcal{X}]_t = t$*; see Theorem 17.3.)*

We now show that there is a weak solution to (18.2) *: For this, let* $\mathcal{X} = (X_t)_{t\geq 0}$ *be any Brownian motion and* $\mathcal{W} = sgn(\mathcal{X}) \cdot \mathcal{X}$. *Then, by Proposition* (16.46),

$$\mathcal{X} = sgn(\mathcal{X})^2 \cdot \mathcal{X} = sgn(\mathcal{X}) \cdot sgn(\mathcal{X}) \cdot \mathcal{X} = sgn(\mathcal{X}) \cdot \mathcal{W}.$$

Thus, $(\mathcal{W}, \mathcal{X})$ *is a weak solution of the SDE* (18.2).

On the other hand, we now show that there cannot be a strong solution of (18.2). *Suppose that* $\mathcal{X}$ *is a strong solution of* (18.2) *for a given Brownian motion* $\mathcal{W}$. *Then* $\mathcal{X}$ *is a Brownian motion for which*

$$\mathcal{W} = 1{\cdot}\mathcal{W} = (sgn(\mathcal{X}))^2{\cdot}\ Constance = sgn(\ Constance){\cdot}\ Constance = |\ Constance| - |X_0| - \mathcal{L}_{\mathcal{X}}$$

due to (17.2), *where* $\mathcal{L}_{\mathcal{X}}$ *is the local time of* $\mathcal{X}$ *at point 0. In particular, it follows that* $\sigma(X_s : s \leq t) \supsetneq \mathcal{F}_t$ *(after all, it is not possible to tell from the right-hand side of the last equation in which direction* $\mathcal{X}$ *has gone at points where* 0 *is met). This is in contradiction to the fact that* $\mathcal{X}$ *is adapted to* $(\mathcal{F}_t)_{t\geq 0}$. *Thus, there cannot be a strong solution of* (18.2).

We now come to the existence and uniqueness theorem for strong solutions of SDEs. In the following, we will use the notation $\lesssim$ from remark 10.13.

Theorem 18.4 (Strong Solutions of SDEs)**.** *Let* X_t^i *be a process on* $[0, \tau]$ *with the following SDE*

$$X_t^i = X_0^i + \int_0^t \mu^i(s, \underline{X}_s)ds + \sum_{j=1}^n \int_0^t \sigma^{ij}(s, \underline{X}_s)dW_s^j \tag{18.3}$$

on $[0,\tau]$ *with a Brownian motion* $\underline{\mathcal{W}} = (\mathcal{W}^j)_{j=1,\dots,n}$*, with* $\mu^i, i = 1, ..., d$ *and* $\sigma^{ij}, i = 1, ..., d, j = 1, ..., n$*, for which*

$$||\underline{\mu}(t,\underline{x})||_2 + ||\underline{\sigma}(t,\underline{x})||_2 \lesssim 1 + ||\underline{x}||_2, \tag{18.4}$$

$$||\underline{\mu}(t,\underline{x}) - \underline{\mu}(t,\underline{y})||_2 + ||\underline{\sigma}(t,\underline{x}) - \underline{\sigma}(t,\underline{y})||_2 \lesssim ||\underline{x} - \underline{y}||_2 \tag{18.5}$$

for $\underline{x}, \underline{y} \in \mathbb{R}^d, 0 \le t \le \tau$ *(where* $||\underline{\sigma}||_2^2 = \sum(\sigma^{ij})^2$*) holds. Furthermore, let* $\underline{Z}$ *be a random variable with values in* $\mathbb{R}^d$ *with* $\mathbf{E}[||\underline{Z}||_2^2] < \infty$*, which is independent of the sigma algebra generated by* $\underline{\mathcal{W}}$*. Then there is a unique, adapted to* $(\sigma((\underline{W}_s)_{s\le t}, Z))_{t\ge 0}$ *and t-continuous solution* $\underline{\mathcal{X}}^{\underline{Z}} = (X_t^{\underline{Z}})_{t\ge 0}$ *of* (18.3) *with* $\underline{X}_0^{\underline{Z}} = Z$ *and* $\mathbf{E}\big[\int_0^\tau ||\underline{X}_t||_2^2 dt\big] < \infty$*. Furthermore, the mapping* $\underline{z} \mapsto \underline{\mathcal{X}}^{\underline{z}}$ *(for deterministic* $\underline{z}$*) is continuous in* $\underline{z}$ *(where the space of paths of* $\underline{\mathcal{X}}$ *is equipped with the uniform convergence on compact).*

Before we prove the theorem, we need two lemmas.

Lemma 18.5. *We denote by* $(t, \underline{x}) \mapsto S^i(t, \underline{x})$ *the ith row of the right-hand side of the SDE* (18.3)*. Then under the same conditions as in Theorem 18.4 for two continuous adapted processes* $(\underline{X}_t)_{t\ge 0}$ *and* $(\underline{Y}_t)_{t\ge 0}$

$$\mathbf{E}[\sup_{0\le s\le t}(S^i(s,\underline{X}_s) - S^i(s,\underline{Y}_s))^2] \le c \cdot \mathbf{E}[||\underline{X}_0 - \underline{Y}_0||_2^2] + c(1+t)\cdot\int_0^t \mathbf{E}[||\underline{X}_s - \underline{Y}_s||_2^2]ds,$$

where $c > 0$ *is a constant that does not depend on the two processes .*

Proof. We use Doob's inequality and the Lipschitz continuity of the functions μ^i and σ^{ij} for

$$\begin{aligned}
\mathbf{E}[\sup_{0\le s\le t}(S^i(\underline{X}_s) - S^i(\underline{Y}_s))^2] &\lesssim \mathbf{E}[(\underline{X}_0^i - \underline{Y}_0^i)^2] + \mathbf{E}\Big[\sup_{0\le s\le t}\Big(\int_0^s \mu^i(r,\underline{X}_r) - \mu^i(r,\underline{Y}_r)dr\Big)^2\Big] \\
&\quad + \mathbf{E}\Big[\sup_{0\le s\le t}\Big(\int_0^s \sum_{j=1}^n \sigma^{ij}(r,\underline{X}_r) - \sigma^{ij}(r,\underline{Y}_r)dW_r^j\Big)^2\Big] \\
&\lesssim \mathbf{E}[||\underline{X}_0 - \underline{Y}_0||_2^2] + \mathbf{E}\Big[\Big(\int_0^t |\mu^i(s,\underline{X}_s) - \mu^i(s,\underline{Y}_s)|ds\Big)^2\Big] \\
&\quad + \mathbf{E}\Big[\Big(\int_0^t \sum_{j=1}^n \sigma^{ij}(s,\underline{X}_s) - \sigma^{ij}(s,\underline{Y}_s)dW_s^j\Big)^2\Big] \\
&= \mathbf{E}[||\underline{X}_0 - \underline{Y}_0||_2^2] + \mathbf{E}\Big[\Big(\int_0^t |\mu^i(s,\underline{X}_s) - \mu^i(s,\underline{Y}_s)|ds\Big)^2\Big] \\
&\quad + \mathbf{E}\left[\int_0^t \sum_{j=1}^n (\sigma^{ij}(s,\underline{X}_s) - \sigma^{ij}(s,\underline{Y}_s))^2 ds\right] \\
&\lesssim \mathbf{E}[||\underline{X}_0 - \underline{Y}_0||_2^2] + \mathbf{E}\left[\Big(\int_0^t ||\underline{X}_s - \underline{Y}_s||_1 ds\Big)^2\right] \\
&\quad + \mathbf{E}\Big[\int_0^t ||\underline{X}_s - \underline{Y}_s||_2^2 ds\Big] \\
&\lesssim \mathbf{E}[||\underline{X}_0 - \underline{Y}_0||_2^2] + (1+t)\int_0^t \mathbf{E}\big[||\underline{X}_s - \underline{Y}_s||_2^2\big]ds.
\end{aligned}$$

□

Lemma 18.6 (Gronwall's lemma)**.** *Let* $f : [0, \infty) \to \mathbb{R}$ *be a continuous function with*

$$f(t) \le a + b \int_0^t f(s)ds \qquad or \qquad \dot{f} \le bf$$

for $a, b \ge 0$*. Then*

$$f(t) \le ae^{bt}$$

for all $t \ge 0$*.*

Proof. From the assumption, it follows immediately

$$\frac{d}{dt}\Big(e^{-bt}\int_0^t f(s)ds\Big) = \Big(f(t) - b\int_0^t f(s)ds\Big)e^{-bt} \le ae^{-bt}.$$

Integration yields

$$\int_0^t f(s)ds \le e^{bt}a\int_0^t e^{-bs}ds = \frac{a}{b}(e^{bt} - 1).$$

Subtracting the result from the derivative at t yields the result. □

Proof of Theorem 18.4. First, we show uniqueness. Suppose there are two solutions $\underline{\mathcal{X}}^{\underline{Z}} = (X_t)_{t \ge 0}$ and $\underline{\mathcal{Y}}^{\underline{Z}} = (Y_t)_{t\ge 0}$ of (18.3) with the same initial value $\underline{Z}$. Then it follows from lemma 18.5 that

$$\begin{aligned}\mathbf{E}[\sup_{0\le s\le t} ||\underline{X}_s - \underline{Y}_s||_2^2] &\le \sum_{i=1}^d \mathbf{E}[\sup_{0\le s\le t}(S^i(s, \underline{X}_s) - S^i(s, \underline{Y}_s))^2] \\ &\lesssim (1+t)\int_0^t \mathbf{E}[\sup_{0\le s\le t} ||\underline{X}_s - \underline{Y}_s||_2^2].\end{aligned} \tag{18.6}$$

Therefore, the following applies to the function $v : t \mapsto \mathbf{E}[\sup_{0\le s\le t} ||\underline{X}_s - \underline{Y}_s||_2^2|||]$ (restricted to $t \in [0, \tau]$), that $\dot{v} \le cv$ for a constant c and $v(0) = 0$. Using the Gronwall lemma, it follows that $v = 0$. In particular, this means that $\underline{\mathcal{X}}^{\underline{Z}} = \underline{\mathcal{Y}}^{\underline{Z}}$.

We now come to the existence of the strong solution. For this, we define recursively processes $\underline{\mathcal{X}}^0 = (\underline{X}_t^0)_{t\ge 0}, \underline{\mathcal{X}}^1 = (\underline{X}_t^1)_{t\ge 0}, ...$ with $\underline{\mathcal{X}}^0 = \underline{Z}$, $\underline{X}_t^{n+1} = \underline{S}(t, \underline{X}_t^n)$ with $\underline{S} = (S^i)_{i=1,...,d}$ (and note that a solution of (18.3) exactly fulfills $\underline{X}_t = \underline{S}(t, \underline{X}_t)$). Then the following applies

$$\begin{aligned}\mathbf{E}[||\underline{X}_t^1 - \underline{X}_t^0||_2^2] = \mathbf{E}[||\underline{S}(t, \underline{Z}) - \underline{Z}||_2^2] &= \mathbf{E}\Big[\int_0^t \underline{\mu}(s, \underline{Z})ds + \int_0^t \underline{\underline{\sigma}}(s, \underline{Z})d\underline{W}_s\Big] \\ &\lesssim t(1 + \mathbf{E}[||\underline{Z}||_2^2]).\end{aligned}$$

because of (18.4), as well as

$$\begin{aligned}\mathbf{E}[||\underline{X}_t^{k+1} - \underline{X}_t^k||_2^2] &= \mathbf{E}[||\underline{S}(t, \underline{X}_t^k) - \underline{S}(t, \underline{X}_t^{k-1})||_2^2] \\ &\lesssim (1+t)\int_0^t \mathbf{E}[||\underline{X}_s^k - \underline{X}_s^{k-1}||_2^2]ds\end{aligned}$$

due to lemma 18.5. Now, by induction, the existence of a constant $c > 0$ follows, such that

$$\mathbf{E}[||\underline{X}_t^{k+1} - \underline{X}_t^k||_2^2] \le \frac{c^{k+1}t^k}{(k+1)!}$$

for $k = 0, 1, 2, ...$ and $0 \leq t \leq \tau$. Now, again using Lemma 18.5

$$\mathbf{P}(\sup_{0 \leq t \leq \tau} ||\underline{X}_t^{k+1} - \underline{X}_t^k||_2 > 2^{-k}) \leq 4^k \cdot \mathbf{E}[\sup_{0 \leq t \leq \tau} ||\underline{S}(t, \underline{X}_t^k) - \underline{S}(t, \underline{X}_t^{k-1})||_2^2]$$

$$\lesssim 4^k \int_0^\tau \mathbf{E}[||\underline{X}_t^k - \underline{X}_t^{k-1}||_2^2]dt \leq \frac{(4c\tau)^k}{k!}.$$

Since the right-hand side is summable, it follows from the Borel-Cantelli lemma that $\sup_{0 \leq t \leq \tau} ||\underline{X}_t^l - \underline{X}_t^k||_2 \xrightarrow{k,l \to \infty}_{fs} 0$. Therefore, $\underline{\mathcal{X}}^k$ converges uniformly on $[0, \tau]$ to a process $\underline{\mathcal{X}}$. With lemma 18.5 one also shows that the convergence also holds in L^2, so that $\mathcal{X}$ has the property $\mathbf{E}\big[\int_0^\tau ||\underline{X}_t||_2^2 dt\big] < \infty$. Since $\underline{X}^k, k = 1, 2, ...$ is adapted to $(\sigma((\underline{W}_s)_{s \leq t}, Z))_{t \geq 0}$ and continuous, the same is true for $\mathcal{X}$. It remains to be shown that $\mathcal{X}$ solves the SDE (18.3). However, this is however, since $\underline{X}_t = \lim_{k \to \infty} \underline{X}_t^k = \lim_{k \to \infty} \underline{X}_t^{k+1} = \lim_{k \to \infty} \underline{S}(t, \underline{X}_t^k) = \underline{S}(t, \underline{X}_t)$ due to the continuity of $\underline{S}$, which follows from lemma 18.5. With an analogous calculation to (18.6), it finally follows that continuity of $\underline{z} \mapsto \mathcal{X}^{\underline{z}}$ uniformly on compacts. □

Remark 18.7 (Itô-Diffusionen). *Given an SDE* (18.3)*, where* $\underline{\mu}$ *and* $\underline{\underline{\sigma}}$ *do not depend on* t*, so that* $\underline{\mu}(t, \underline{x}) = \underline{\mu}(\underline{x})$ *and* $\underline{\underline{\sigma}}(t, \underline{x}) = \underline{\underline{\sigma}}(\underline{x})$*. In this case, the solution of the SDE is called an Itô-diffusion. (More generally, a strong Markov process with continuous paths is called a diffusion. There are also other diffusions than Itô-diffusions, for example* $|\mathcal{X}|$ *is a diffusion if* $\mathcal{X}$ *is a Brownian motion.)*

In a sense, all solutions of SDEs are Itô-diffusions: namely, if we introduce a new variable $X_t^{d+1} := t$ *and supplement the SDE* (18.3) *by* $dX_t^{d+1} = dt$*, and if we write* $\underline{\mu}(t; x_1, ..., x_d) = \underline{\mu}(x_{d+1}; x_1, ..., x_d)$ *and* $\underline{\underline{\sigma}}(t; x_1, ..., x_d) = \underline{\underline{\sigma}}(x_{d+1}; x_1, ..., x_d)$*, we see that the right-hand sides of the last two equations are independent of* t*. For this reason, we will only deal with Itô-diffusions in the following, i.e. with SDEs of the form*

$$d\underline{X} = \underline{\mu}(\underline{X}_t)dt + \underline{\underline{\sigma}}(\underline{X}_t)d\underline{W}_t. \tag{18.7}$$

18.2 Martingale problems

Let X be a (weak) solution of the SDE (18.7), and let $G^{\underline{\mathcal{X}}} : \mathcal{C}_b^{(2)}(\mathbb{R}^d) \to \mathcal{C}(\mathbb{R}^d)$ be the operator (here, $\mathcal{C}_b^{(2)}(\mathbb{R}^d)$ is the space of real-valued, twice boundedly continuously differentiable functions on $\mathbb{R}^d$), given by

$$(G^{\underline{\mathcal{X}}} f)(\underline{x}) = \sum_{i=1}^d \mu^i(\underline{x}) \frac{\partial f}{\partial x_i}(\underline{x}) + \tfrac{1}{2} \sum_{i,j} (\sigma^{ij})^2(\underline{x}) \frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x}) \tag{18.8}$$

with $(\sigma^{ij})^2 = \sum_{k=1}^n \sigma^{ik}\sigma^{jk}$. Applying the Itô formula, we obtain

$$\begin{aligned} df(\underline{X}_t) &= \sum_{i=1}^d \frac{\partial f}{\partial X_t^i} dX^i + \tfrac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j} d[\mathcal{X}^i, \mathcal{X}^j]_t \\ &= \sum_{i=1}^d \mu^i(\underline{X}_t) \frac{\partial f}{\partial X_t^i} dt + \tfrac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 f}{\partial x_i \partial x_j} (\sigma^{ij})^2(\underline{X}_t)dt + dM_t^f \end{aligned}$$

for the local martingale $\mathcal{M}^f = (M_t^f)_{t\geq 0}$, where

$$M_t^f = \sum_{i=1}^d \sum_{j=1}^n \int_0^t \frac{\partial f}{\partial x_i}(\underline{X}_s)\sigma^{ij}(\underline{X}_s)dW_s^j.$$

In other words,

$$\mathcal{M}^f := \Big(f(\underline{X}_t) - f(\underline{X}_0) - \int_0^t (G^{\underline{\mathcal{X}}} f)(\underline{X}_s)ds\Big)_{t\geq 0} \tag{18.9}$$

is a local martingale for all $f \in \mathcal{C}_b^{(2)}(\mathbb{R}^d)$. The concept of (local) martingale problems highlights precisely this property. The question is therefore: given an operator $G^{\underline{\mathcal{X}}}$ (e.g. of the form (18.8)), there exists a probability space with a stochastic process $\underline{\mathcal{X}}$ such that $\mathcal{M}^f$ from (18.9) is a local martingale for all $f \in \mathcal{D}(G^{\mathcal{X}})$?

Definition 18.8 (Local martingale problem). *Let E be a Polish space, $\mathbf{P}_0$ a probability measure on E and G a bounded operator with domain $\mathcal{D}(G) \subseteq \mathcal{B}(E)$ (the set of bounded measurable functions). For a stochastic process $\mathcal{X} = (X_t)_{t\geq 0}$ with paths in $\mathcal{C}_E[0,\infty)$ and state space E, all derived processes $\mathcal{M}^f = (M_t^f)_{t\geq 0}$,*

$$M_t^f = f(X_t) - f(X_0) - \int_0^t (Gf)(X_s)ds \tag{18.10}$$

local martingales, and $X_0 \stackrel{d}{=} \mathbf{P}_0$, then $\mathcal{X}$ is called the solution of the local $(\mathbf{P}_0, G, \mathcal{D})$ martingale problem (on $\mathcal{C}_E[0,\infty)$).

If all $\mathcal{M}^f$ are even martingales, then it is said that $\mathcal{X}$ solves the $(\mathbf{P}_0, G, \mathcal{D})$ martingale problem . (If no distribution for X_0 is specified, one also speaks of the (local) $(G, \mathcal{D})$ martingale problem.) If there is exactly one solution $\mathcal{X}$ for the (local) $(\mathbf{P}_0, G, \mathcal{D})$ martingale problem, it is said to be well-posed.

Example 18.9 (A simple martingale problem). *Let $E = \mathbb{R}^d$, $\mathbf{P}_0 = \delta_{\underline{0}}$ and G the operator on $\mathcal{D}(G) = \{\underline{x} \mapsto x_i, \underline{x} \mapsto x_i x_j; i,j = 1,...,d\}$ with $Gx_i = 0, Gx_ix_j = \delta_{ij}$. Then only the d-dimensional Brownian motion solves the $(\mathbf{P}_0, G, \mathcal{D})$ martingale problem. In particular, this is well-posed.*

Because: According to Theorem 17.3, a process $\underline{\mathcal{X}} = (\underline{X}_t)_{t\geq 0}, \underline{X}_t = (X_t^i)_{i=1,\dots,d}$ is a d-dimensional Brownian movement if and only if both $(X_t^i)_{t\geq 0}$ and $(X_t^iX_t^j - \delta_{ij}t)_{t\geq 0}$ are local martingales. The latter is the case if and only if $\mathcal{X}$ solves the $(G, \mathcal{D})$ martingale problem.

Weak solutions of SDEs of the form (18.3) and solutions of local martingale problems for generators of the form (18.8) are closely related, as the following result shows.

Theorem 18.10 (Local martingale problems and weak solutions of SDEs). *Let $\underline{\mu} = (\mu^1, ..., \mu^d)$ and $\underline{\underline{\sigma}} = (\sigma^{ij})_{i=1,\dots,d,j=1,\dots,n}$ be continuous functions and $\mathbf{P}$ a probability measure on $\mathcal{C}_{\mathbb{R}^d}[0,\infty)$. Then the SDE (18.3) has a weak solution with distribution $\mathbf{P}$ if and only if a process distributed according to $\mathbf{P}$ solves the local martingale problem for $(G, \mathcal{C}_b^{(2)}(\mathbb{R}^d))$ with G as in (18.8).*

Proof. If $(\underline{\mathcal{W}}, \underline{\mathcal{X}})$ is a weak solution of (18.3), then we have already shown at the beginning of this section under (18.8) that the distribution of $\mathcal{X}$ solves the local $(G, \mathcal{C}_b^{(2)}(\mathbb{R}^d))$ martingale problem.

Conversely, let $\underline{\mathcal{X}}$ be a solution of the local $(G, \mathcal{C}_b^{(2)}(\mathbb{R}^d))$ martingale problem. Inserting $\underline{x} \mapsto x_i$ into G, we see that $\mathcal{M}^i = (M_t^i)_{t\geq 0}$ with

$$M_t^i = X_t^i - X_0^i - \int_0^t \mu^i(\underline{X}_s)ds$$

are local martingals. (Formally, we are only allowed to inserting bounded functions in G. From this, inserting bounded functions f with $f(\underline{x}) = x_i$ on $A_n := \{\underline{x} : ||\underline{x}||_2 \leq n\}$ initially yields that $(M_{t\wedge T}^i)_{t\geq 0}$ is a local martingale, where T is the hitting time of A_n. From Lemma 16.27, it follows that $\mathcal{M}^i$ is a local martingale.) Furthermore, if we substitute $\underline{x}_i x_j$ into G, we see analogously that $\mathcal{M}^{ij} = (M_t^{ij})_{t\geq 0}$ with

$$M_t^{ij} = X_t^i X_t^j - X_0^i X_0^j - \int_0^t X_s^i \mu^j(\underline{X}_s) + X_s^j \mu^i(\underline{X}_s) - (\sigma^{ij})^2(\underline{X}_s)ds$$

is a local martingale. Partial integration also yields

$$\begin{aligned}\mathcal{M}^{ij} &= \mathcal{X}^i \cdot \mathcal{X}^j + \mathcal{X}^j \cdot \mathcal{X}^i + [\mathcal{X}^i, \mathcal{X}^j] - (\ Con^i \mu^j + \mathcal{X}^j \mu^i + (\sigma^{ij})^2) \cdot \lambda \\ &= \mathcal{X}^i \cdot \mathcal{M}^j + \mathcal{X}^j \cdot \mathcal{M}^i + [\mathcal{M}^i, \mathcal{M}^j] - (\sigma^{ij})^2 \cdot \lambda.\end{aligned}$$

From this we read that

$$[\mathcal{M}^i, \mathcal{M}^j] = (\sigma^{ij})^2 \cdot \lambda = \sum_{k=1}^n \sigma^{ik}(\underline{\mathcal{X}})(\sigma^{ik})(\underline{\mathcal{X}}) \cdot \lambda.$$

From Theorem 17.10 it follows that there is an extension of the probability space and a Brownian motion $\underline{\mathcal{W}} = (\mathcal{W}^1, ..., \mathcal{W}^n)$ with

$$\mathcal{X}^i - \mu^i(\underline{\mathcal{X}}) \cdot \lambda = \mathcal{M}^i = \sum_{k=1}^n \sigma^{ik}(\underline{\mathcal{X}}) \cdot \mathcal{W}^k.$$

This means that $(\underline{\mathcal{W}}, \underline{\mathcal{X}})$ solves the SDE (18.3). □

Weak solutions only depend on the joint distribution of the underlying Brownian motion $\underline{\mathcal{W}}$ and the solution of the SDE $\underline{\mathcal{X}}$. Therefore, it is possible to create weak solutions of SDEs by measure transformations, i.e. by changing the distribution of the processes. This is done with the Girsanov transformation known from Theorem 17.14.

Theorem 18.11 (The Girsanov transformation of solutions of SDEs)**.** *On a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with filtration $(\mathcal{F}_t)_{t\geq 0}$, let $\underline{\mathcal{X}} = (\mathcal{X}^1, ..., \mathcal{X}^d)$ with $\mathcal{X}^i = (X_t^i)_{t\geq 0}$ Solution of the SDE*

$$d\underline{X}_t = \underline{\mu}(\underline{X}_t)dt + \underline{\underline{\sigma}}(\underline{X}_t)d\underline{W}_t. \tag{18.11}$$

for a $\mathbf{P}$-Brownian motion $\underline{\mathcal{W}} = (\mathcal{W}^1, ..., \mathcal{W}^d)$. Furthermore, let $\underline{\gamma} : \mathbb{R}^d \to \mathbb{R}^d$ be continuous and let there be a progressively measurable process $\underline{\mathcal{H}}$ in $\mathbb{R}^n$ with $\underline{\underline{\sigma}}(\underline{\mathcal{X}})\underline{\mathcal{H}} = \underline{\gamma}(\underline{\mathcal{X}})$ and such that

$$\mathcal{Z} = \exp\Big(\sum_{j=1}^n \mathcal{H}^j \cdot \mathcal{W}^j - \tfrac{1}{2}(\mathcal{H}^j)^2 \cdot \lambda\Big) \tag{18.12}$$

is a martingale. Then define the measure $\mathbf{Q}$ *by*

$$\mathbf{Q} = Z_t \cdot \mathbf{P} \text{ on } \mathcal{F}_t.$$

Then the process $\underline{\widetilde{\mathcal{W}}} := \underline{\mathcal{W}} - \underline{\mathcal{H}} \cdot \lambda$ *is a Brownian motion with respect to* $\mathbf{Q}$ *and*

$$d\underline{X}_t = (\underline{\mu}(\underline{X}_t) + \underline{\gamma}(\underline{X}_t))dt + \underline{\underline{\sigma}}(\underline{X}_t)d\underline{\widetilde{W}}_t. \tag{18.13}$$

In other words: the pair $(\underline{\widetilde{W}}, \underline{\mathcal{X}})$ *is a weak solution of* (18.13) *(under* $\mathbf{Q}$*).*

Proof. The proof is an application of Theorem 17.14. First of all, we note that according to this theorem, the process $\underline{\mathcal{W}} - [\underline{\mathcal{W}}, \sum \mathcal{H}^j \mathcal{W}^j] = \underline{\mathcal{W}} - \underline{\mathcal{H}} \cdot \lambda$ is a continuous local $\mathbf{Q}$-martingale. According to Proposition 17.15, this has the same covariation process as a Brownian motion and is therefore, according to Theorem 9.33, a Brownian motion under $\mathbf{Q}$. Furthermore,

$$(\underline{\mu}(\underline{\mathcal{X}}) + \underline{\gamma}(\underline{\mathcal{X}})) \cdot \lambda + \underline{\underline{\sigma}}(\underline{\mathcal{X}}) \cdot \underline{\widetilde{\mathcal{W}}} = (\underline{\mu}(\underline{\mathcal{X}}) + \underline{\gamma}(\underline{\mathcal{X}})) \cdot \lambda + \underline{\underline{\sigma}}(\underline{\mathcal{X}}) \cdot \underline{\mathcal{W}} - \underline{\underline{\sigma}}(\underline{\mathcal{X}})\underline{\mathcal{H}} \cdot \lambda = \underline{\mathcal{X}}.$$

Thus, all assertions are shown. □

Corollary 18.12. *Let* $\underline{\mathcal{X}}, \underline{\mathcal{W}}, \underline{\gamma}, \underline{\mathcal{H}}, \mathcal{Z}$ *be as in Theorem 18.11 and let* $\underline{\mathcal{Y}} = (\mathcal{Y}^1, ..., \mathcal{Y}^d)$ *with* $\mathcal{Y}^i = (Y_t^i)_{t \geq 0}$ *solution of the SDE*

$$d\underline{Y}_t = (\underline{\mu}(\underline{Y}_t) - \underline{\gamma}(\underline{Y}_t))dt + \underline{\underline{\sigma}}(\underline{Y}_t)d\underline{W}_t. \tag{18.14}$$

Then

$$d\underline{Y}_t = \underline{\mu}(\underline{Y}_t)dt + \underline{\underline{\sigma}}(\underline{Y}_t)d\underline{\widetilde{W}}_t. \tag{18.15}$$

In other words: Both $(\underline{\mathcal{W}}, \underline{\mathcal{X}})$*, and* $(\underline{\widetilde{\mathcal{W}}}, \underline{\mathcal{Y}})$ *solve the SDE* (18.11)*.* $(\underline{\mathcal{W}}, \underline{\mathcal{X}})$ *is a (weak) solution to* (18.11) *(under* $\mathbf{P}$*, then* $(\underline{\widetilde{\mathcal{W}}}, \underline{\mathcal{Y}})$ *solution to* (18.13)*) is also a weak solution to* (18.11) *(under* $\mathbf{Q}$*).*

Proof. It remains only to show that $(\widetilde{\mathcal{W}}, \underline{\mathcal{Y}})$ solves the SDE (18.15). To do this, we write

$$\underline{\mu}(\underline{\mathcal{Y}}) \cdot \lambda + \underline{\underline{\sigma}}(\underline{\mathcal{Y}}) \cdot \underline{\widetilde{\mathcal{W}}} = \underline{\mu}(\underline{\mathcal{Y}}) \cdot \lambda + \underline{\underline{\sigma}}(\underline{\mathcal{Y}}) \cdot \underline{\mathcal{W}} - \underline{\underline{\sigma}}(\underline{\mathcal{Y}})\underline{\mathcal{H}} \cdot \lambda = \underline{\mathcal{Y}}.$$

□

Remark 18.13 (When is $\mathcal{Z}$ a martingale?)**.** *According to Theorem 17.14.2,* $\mathcal{Z}$ *from* (18.12) *is always a local martingale. In applications, it is often important to show that the process defined by* $\mathcal{Z}$ *in* (18.12) *is a martingale, since only (true, i.e., nonlocal) martingales can be used to perform a change of measure. For this, it is sufficient, for example, (see remark 16.24.3), if* $\mathcal{Z}$ *is bounded. Another condition for the martingale property of* $\mathcal{Z}$ *is called the Novikov condition and is fulfilled if*

$$\mathbf{E}\Big[\exp\Big(\tfrac{1}{2}\Big(\sum_{j=1}^{n} \int_0^t (H_s^j)^2 ds\Big)\Big)\Big] < \infty$$

holds for all $t > 0$*. However, we will not show here that the martingale property of* $\mathcal{Z}$ *follows from this.*

Example 18.14 (Transformation of Geometric Brownian Motion)**.** *The following example plays an important role in financial mathematics. Let $\mathcal{W} = (W_t)_{t\geq 0}$ be a Brownian motion. In example 18.1.2, we saw that $\mathcal{X} = (X_t)_{t\geq 0}$ with*

$$X_t = \exp\big(\sigma W_t + (\mu - \tfrac{1}{2}\sigma^2)t\big)$$

is the solution of the SDE

$$dX = \mu X dt + \sigma X dW.$$

If we perform a measure transformation using $\mathcal{Z} = (Z_t)_{t\geq 0}$,

$$Z_t = \exp(-\tfrac{\mu}{\sigma} W_t - \tfrac{1}{2}\tfrac{\mu^2}{\sigma^2} dt)$$

to a measure $\mathbf{Q}$ (i.e. $\mathbf{Q} = Z_t \cdot \mathbf{P}$ on $\mathcal{F}_t$), then $\widetilde{W} = \sigma W + \frac{\mu}{\sigma} t$ is a Brownian motion under $\mathbf{Q}$ and $\mathcal{X}$ is a solution of

$$dX = \sigma X d\widetilde{W}.$$

If $\mathcal{X}$ models a stock price process, then there exists the measure $\mathbf{Q}$ with respect to which this price process is a martingale. Such a measure is also called a martingale measure (for the stock price process $\mathcal{X}$).

18.3 Markov property

The solution of an SDE of the form (18.1) is usually thought of in such a way that the μ^i indicate the direction of the process $\mathcal{X}^i$, and σ^{ij} indicate the fluctuations that depend only on and the infinitesimal increment of Brownian motion. In this description, it is clear that the future evolution of the process after t depends only on t and. In other words, we would expect to be a Markov process. We now show that this picture is correct.

Theorem 18.15 (Markov Property for Martingal Problems and SDEs)**.** *Let $\underline{\mu} = (\mu^1, ..., \mu^d)$ and $\underline{\underline{\sigma}} = (\sigma_{ij})_{i=1,...,d, j=1,...,n}$ be continuous functions such that the local $(\mathbf{P}_0, G, \mathcal{C}_b^{(2)}(\mathbb{R}^d))$ martingale problem with G as in (18.8) is well-posed for every probability measure $\mathcal{P}_0$ on $\mathbb{R}^d$ with solution $\underline{\mathcal{X}}$. Then $\underline{\mathcal{X}}$ is a Markov process. If furthermore, $\underline{x} \mapsto \mathbf{P}_{\delta_{\underline{x}}}$ is continuous, then $\mathcal{X}$ is strongly Markov. The generator $G^{\underline{\mathcal{X}}}$ of $\underline{\mathcal{X}}$ has domain $\mathcal{D}(G^{\underline{X}}) \supseteq \mathcal{C}_b^{(2)}(\mathbb{R}^d)$ and has the form (18.8) for $f \in \mathcal{C}_b^{(2)}(\mathbb{R}^d)$.*

Proof. Let $\mathcal{M}^f$ be as in (18.10) and $s, t \geq 0$ and θ_s the time-shift operator (i.e. $\theta_s((\omega_t)_{t\geq 0}) = (\omega_{t+s})_{t\geq 0}$. Then holds for $A \in \mathcal{F}_s$

$$\mathbf{E}[((M_t^f - M_0^f) \circ \theta_s, A] = \mathbf{E}[((M_{t+s}^f - M_s^f), A] = 0,$$

since $\mathcal{M}^f$ is a martingale. It follows immediately that

$$\mathbf{E}[((M_t^f - M_0^f) \circ \theta_s | \mathcal{F}_s] = 0,$$

from which it is also

$$\mathbf{E}[((M_t^f - M_0^f) \circ \theta_s | X_s] = 0,$$

immediately clear. Since the local martingale problem has a unique solution, the joint distribution of $M_t^f - M_0^f$ (for all f) uniquely determines the process $\mathcal{X}$ (and thus X_t). Thus, for $t \geq 0$,

$$\mathbf{P}(X_{t+s} \in .|\mathcal{F}_s) = \mathbf{P}(X_t \circ \theta_s \in .|\mathcal{F}_s) = \mathbf{P}(X_t \circ \theta_s \in .|X_s) = \mathbf{P}(X_{t+s} \in .|X_s).$$

However, this is precisely the Markov property of $\mathcal{X}$. The strong Markov property in the case of the continuous mapping $\underline{x} \mapsto \mathbf{P}_{\underline{x}}$ follows from Theorem 14.12. Since $\underline{\mathcal{X}}$ is now a Markov process, the Itô formula for $f \in \mathcal{C}_b^{(2)}(\mathbb{R}^d)$ applies

$$\begin{aligned}
(G^{\underline{\mathcal{X}}} f)(\underline{x}) &= \lim_{t\to 0} \frac{1}{t}\mathbf{E}[f(\underline{X}_t) - f(\underline{X}_0)|\underline{X}_0 = \underline{x}] \\
&= \lim_{t\to 0} \frac{1}{t}\mathbf{E}\Big[\int_0^t \sum_{i=1}^d \mu^i(\underline{X}_s)\frac{\partial f}{\partial x_i}(\underline{X}_s)ds + \tfrac{1}{2}\sum_{i,j=1}^d (\sigma^{ij})^2(\underline{X}_s)\frac{\partial^2 f}{\partial x_i \partial x_j}\Big|\underline{X}_0 = \underline{x}\Big] \\
&= \sum_{i=1}^d \mu^i(\underline{x})\frac{\partial f}{\partial x_i}(\underline{x}) + \tfrac{1}{2}\sum_{i,j=1}^d (\sigma^{ij})^2(\underline{x})\frac{\partial^2 f}{\partial x_i \partial x_j}(\underline{x})
\end{aligned}$$

because of the continuity of $\underline{\mu}$ and $\underline{\underline{\sigma}}^2$. □

Remark 18.16 (Infinitesimal Expected Value and (Co-)Variance)**.** *If we substitute* $f(\underline{x}) = x^i$ *into* $G^{\underline{\mathcal{X}}}$*, we obtain*

$$\lim_{t\to 0} \frac{1}{t}\mathbf{E}[X_t^i - X_0^i|\underline{X} = \underline{x}] = \mu^i(\underline{x}).$$

With $f(\underline{x}) = x^i x^j$*, we have*

$$\begin{aligned}
&\lim_{t\to 0} \frac{1}{t}\mathbf{E}[(X_t^i - X_0^i)(X_t^j - X_0^j)|\underline{X} = \underline{x}] \\
&\quad= \lim_{t\to 0} \frac{1}{t}\mathbf{E}[(X_t^i X_t^j - X_0^i X_0^j - X_0^i(X_t^j - X_0^j) - X_0^j(X_t^i - X_0^i)|\underline{X} = \underline{x}] \\
&\quad= x_i\mu^j(\underline{x}) + x_j\mu^i(\underline{x}) + \sigma_{ij}^2(\underline{x}) - x_j\mu^i(\underline{x}) - x_i\mu^j(\underline{x}) \\
&\quad= \sigma_{ij}^2(\underline{x}).
\end{aligned}$$

Therefore, $\underline{\mu}(\underline{X})$ *is called the vector of the* infinitesimal expectation value *of* $\mathcal{X}$ *and* $\sigma_{ij}^2(\mathcal{X})$ *the matrix of the* infinitesimal covariance *of* $\mathcal{X}$*. The diagonal elements* $(\sigma_{ii}^2(\underline{X}))_{i=1,\ldots,d}$ *are correspondingly called also called* infinitesimal variance*.*

Interestingly, Itô-diffusions can be helpful to solve partial differential equations given by an operator of the form (18.8) uniquely. The next theorem shows this connection.

Theorem 18.17 (Backward Equation)**.** *Let* $\underline{\mathcal{X}}$ *be a Markov process with generator* G *of the form* (18.8) *and* $f \in \mathcal{C}^{(2)}(\mathbb{R})$*. For the boundary value problem*

$$\frac{\partial u}{\partial t} = Gu, \qquad t > 0, \underline{x} \in \mathbb{R}^d, \tag{18.16}$$

$$u(0, \underline{x}) = f(\underline{x}); \qquad \underline{x} \in \mathbb{R}^d. \tag{18.17}$$

(where Gu *describes the application of* G *to the* $\underline{x}$*-coordinates of* u*) is*

$$u(t, \underline{x}) := \mathbf{E}[f(\underline{X}_t)|\underline{X}_0 = \underline{x}]$$

the only solution.

Proof. It is to be shown that the given u solves the boundary value problem, as well as the uniqueness of the solution. Thus, let $u(t, \underline{x})$ be as given. Then $u(0, \underline{x}) = f(\underline{x})$ by definition and (18.16) holds by definition of the generator. Now we use Lemma 14.28, specifically (14.7). If $(T_t^{\underline{\mathcal{X}}})$ is the semigroup generated by $\underline{\mathcal{X}}$, then $u(t, \underline{x}) := (T_t^{\underline{\mathcal{X}}} f)(\underline{x})$. Thus, (18.16) follows directly from (14.7).

Now for the uniqueness: Let w be a solution of the boundary value problem, $t > 0$ and $\underline{x} \in \mathbb{R}^d$. It is to be shown that $w(t, \underline{x}) = \mathbf{E}[f(\underline{X}_t)|\underline{X}_0 = \underline{x}]$ holds. We define the process $\underline{\mathcal{Y}} = (\underline{Y}_s)_{0 \le s \le t}$ by means of $\underline{Y}_s = (t - s, \underline{X}_s)$, where $\underline{\mathcal{X}} = (\underline{X}_s)_{0 \le s \le t}$ in $\underline{X}_0 = \underline{x}$. Then the generator of $\mathcal{Y}$ is given by

$$(G^{\underline{Y}} f)(s, \underline{x}) = -\frac{\partial f}{\partial s} + (Gf)(s, \underline{x}),$$

where G acts on $\underline{x}$ in the last term. If we choose specifically $f = w$, then $(G^{\underline{Y}} w) = 0$ applies, since w solves the boundary value problem. Thus

$$w(t, \underline{x}) = w(\underline{Y}_0) = \mathbf{E}[w(\underline{Y}_t)] = \mathbf{E}[w(0, \underline{X}_t)|\underline{X}_0 = \underline{x}] = \mathbf{E}[f(\underline{X}_t)|\underline{X}_0 = \underline{x}] = u(t, \underline{x}).$$

□

18.4 One-dimensional diffusions

Diffusions are time-homogeneous, strong Markov processes with continuous paths. As became clear from Theorem 18.15 and Theorem 18.10, for example, unique solutions of SDEs that are continuous in the initial conditions are diffusions. We had processes Itô-diffusions. We turn to the special case of one-dimensional Itô-diffusions, i.e. solutions of equations of the form

$$dX = \mu(X)dt + \sigma(X)dW. \tag{18.18}$$

In this section, we will make the following assumptions throughout:

1. The equation (18.18) has continuous μ and σ and is uniquely solvable by a strong Markov process,

2. the solution does not leave a range $[l, r] \subseteq [-\infty, \infty]$ and $\sigma > 0$ on (l, r).

With regard to Theorem 18.15 or Theorem 18.4, the first condition is certainly fulfilled if μ and σ are locally Lipschitz continuous and do not grow faster than linearly. The second condition ($\sigma > 0$) means that the solution , i.e. $\mathcal{X}$ never stays in the interior of (l, r).

As explained in remark 18.16, μ is called the infinitesimal expectation and σ^2 the infinitesimal variance of $\mathcal{X}$. We have already seen that the solution of this SDE is a strong Markov process with generator

$$(G^{\mathcal{X}} f)(x) = \mu(x)f'(x) + \tfrac{1}{2}\sigma^2(x)f''(x),$$

where $f \in \mathcal{C}^{(2)}(\mathbb{R})$. We start with an example that will accompany us throughout this chapter .

Example 18.18 (Ornstein-Uhlenbeck Process)**.** *We consider the solution of the SDE* $\mathcal{X} = (X_t)_{t \ge 0}$ *with*

$$dX = -\mu X dt + dW$$

with generator

$$(Gf)(x) = -\mu x f'(x) + \tfrac{1}{2} f''(x).$$

From example 18.18 it follows that $\mathcal{X}$ *is a Gaussian process and* $X_t \xrightarrow{t\to\infty} N(0, 1/2\mu)$.

The backward equation states that in the case of one-dimensional SDEs, the boundary value problem

$$\begin{aligned} \frac{\partial u}{\partial t} &= \mu(x)\frac{\partial u}{\partial x} + \tfrac{1}{2}\sigma^2(x)\frac{\partial^2 u}{\partial x^2}, \\ u(0,x) &= f(x) \end{aligned}$$

has the solution $u(t,x) = \mathbf{E}[f(X_t)|X_0 = x]$, where $\mathcal{X} = (X_t)_{t\geq 0}$ is an Itô diffusion with infinitesimal expectation μ and infinitesimal variance σ^2. There is a second partial differential equation that plays an important role. To this end, we recall that every homogeneous Markov process with state space $\mathbb{R}$ defines a family of transition kernels $(\mu_t^{\mathcal{X}})_{t\geq 0}$.

Theorem 18.19 (Forward Equation)**.** *Let* $\mathcal{X} = (X_t)_{t\geq 0}$ *be an Itô diffusion with infinitesimal expectation* μ *and infinitesimal variance* σ^2*. If the transition kernels* $(\mu_t^{\mathcal{X}})_{t\geq 0}$ *have densities, i.e.*

$$\mu_t^{\mathcal{X}}(x, B) = \int_B p_t^{\mathcal{X}}(x,y)dy$$

for a family $(p_t^{\mathcal{X}})_{t\geq 0}$ *such that* $t \mapsto p_t^{\mathcal{X}}(x,y)$ *is twice continuously differentiable, then they satisfy the equation*

$$\frac{\partial p_t^{\mathcal{X}}(x,y)}{\partial t} = -\frac{\partial}{\partial y}\big(\mu(y)p_t^{\mathcal{X}}(x,y)\big) + \tfrac{1}{2}\frac{\partial^2}{\partial y^2}\big(\sigma^2(y)p_t^{\mathcal{X}}(x,y)\big).$$

Proof.

$$\begin{aligned} \frac{\partial p_t^{\mathcal{X}}(x,z)}{\partial t} &= \lim_{s\to 0}\frac{\partial}{\partial s}\int p_t^{\mathcal{X}}(x,y)p_s^{\mathcal{X}}(y,z)dy \\ &= \lim_{s\to 0}\int p_t^{\mathcal{X}}(x,y)\frac{\partial p_s^{\mathcal{X}}(y,z)}{\partial s}dy \\ &= \lim_{s\to 0}\int p_t^{\mathcal{X}}(x,y)\Big(\mu(y)\frac{\partial p_s^{\mathcal{X}}(y,z)}{\partial y} + \tfrac{1}{2}\sigma^2(y)\frac{\partial^2 p_s^{\mathcal{X}}(y,z)}{\partial y^2}\Big)dy \\ &= \lim_{s\to 0}\int -p_s^{\mathcal{X}}(y,z)\Big(\frac{\partial}{\partial y}\big(p_t^{\mathcal{X}}(x,y)\mu(y)\big) - \tfrac{1}{2}\frac{\partial^2}{\partial y^2}\big(p_t^{\mathcal{X}}(x,y)\sigma^2(y)\big)\Big) \\ &= -\frac{\partial}{\partial z}\big(\mu(z)p_t^{\mathcal{X}}(x,z)\big) + \tfrac{1}{2}\frac{\partial^2}{\partial z^2}\big(\sigma^2(z)p_t^{\mathcal{X}}(x,z)\big). \end{aligned}$$

□

Example 18.20 (Stationary distribution of the Ornstein-Uhlenbeck process)**.** *Let* $\mathcal{X} = (X_t)_{t\geq 0}$ *be an Ornstein-Uhlenbeck process with* $X_0 = x$*. As we have already seen in example***??***,* $X_t \overset{d}{=} N(e^{-\mu t}x, \frac{1}{2\mu}(1-e^{-2\mu t}))$*. In particular,* $N(0, \frac{1}{2\mu})$ *is the stationary distribution of the*

Ornstein-Uhlenbeck process. We will now verify this again using the forward equation, Theorem 18.19. We calculate with

$$\begin{aligned} p(y) := p_t(x,y) = \sqrt{\frac{\mu}{\pi}}\exp\big(-\mu y^2\big), \\ \frac{\partial p}{\partial y}(y) = -2\mu y p(y), \\ \frac{\partial p^2}{\partial y^2} = (4\mu^2y^2 - 2\mu)p(y), \end{aligned}$$

that

$$\begin{aligned} -\frac{\partial}{\partial y}\big(\mu(y)p(y)\big) + \tfrac{1}{2}\frac{\partial^2}{\partial y^2}\big(\sigma^2(y)p(y)\big) &= \frac{\partial}{\partial y}\mu y p(y) + \tfrac{1}{2}\frac{\partial^2}{\partial y^2}p(y) \\ &= (\mu - 2\mu^2y^2)p(y) + \tfrac{1}{2}(4\mu^2y^2 - 2\mu)p(y) = 0. \end{aligned}$$

Thus, if $X_0 \stackrel{d}{=} N(0, \frac{1}{2\mu})$, *then according to the forward equation*

$$\frac{\partial p_t(x,y)}{\partial t} = 0.$$

This means that the density of the distribution of X_t *in* t *does not change. Thus, we have (again) that* $N(0, \frac{1}{2\mu})$ *is a stationary distribution of* $\mathcal{X}$.

It turns out that one-dimensional Itô-diffusions allow for very explicit calculations. For this, we need the scale function of a one-dimensional Itô-diffusion.

Proposition 18.21 (Scale function)**.** *Let* $X = (X_t)_{t\geq 0}$ *be a solution of* (18.18) *with state space* $[l, r]$. *We recall the assumption* $\sigma > 0$ *on* (l, r) *and define*

$$S(x) := \int^x \exp\Big(-2\int^y \frac{\mu(z)}{\sigma^2(z)}dz\Big), \qquad x \in (l, r), \tag{18.19}$$

whereby the lower bounds of the integrals can be arbitrary in (l, r). *Then* $S(\mathcal{X}) = (S(X_t))_{t\geq 0}$ *is a local martingale.*

Proof. First, we note that S solves the ordinary differential equation

$$\frac{1}{2}\sigma^2 S'' + \mu S' = 0$$

Therefore, using the Itô formula, we obtain

$$\begin{aligned} S(\mathcal{X}) - S(X_0) &= S'(\mathcal{X})\cdot\mathcal{X} + \tfrac{1}{2}S''(\mathcal{X})\cdot[\mathcal{X}] \\ &= (S'(\mathcal{X})\mu(\mathcal{X}) + \tfrac{1}{2}\sigma^2(\mathcal{X})S''(\mathcal{X}))\cdot\lambda + S'(\mathcal{X})\cdot\mathcal{W} = S'(\mathcal{X})\cdot\mathcal{W}. \end{aligned}$$

□

Let τ_x be the hitting time at $x \in [l, r]$. We now recall the following calculation for the Brownian motion:

Example 18.22 (Hitting probabilities for the Brownian motion)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion with $X_0 = x$ and $a < x < b$. Then, due to the Optional Sampling and Optional Stopping Theorems,*

$$x = \mathbf{E}_x[X_{\tau_a \wedge \tau_b}] = a\mathbf{P}_x(\tau_a < \tau_b) + b(1 - \mathbf{P}_x(\tau_a > \tau_b)),$$

so

$$\mathbf{P}_x(\tau_a < \tau_b) = \frac{b-x}{b-a}.$$

We now generalize this calculation using the scale function.

Theorem 18.23 (Hit probabilities)**.** *Let X and S be as in Proposition 18.21, and $l < a < x < b < r$. Then*

$$\mathbf{P}_x(\tau_a < \tau_b) = \frac{S(b) - S(x)}{S(b) - S(a)}$$

as well as

$$\mathbf{P}_x(\tau_b < \tau_a) = \frac{S(x) - S(a)}{S(b) - S(a)}.$$

Proof. The function S is strictly increasing on (l, r). If τ_y^S denotes the first time $S(\mathcal{X})$ hits y, then

$$\mathbf{P}_x(\tau_a < \tau_b) = \mathbf{P}_x(\tau^S_{S(a)} < \tau^S_{S(b)}).$$

Since the process $S(\mathcal{X}^{\tau_a \wedge \tau_b})$ is a bounded martingale, it converges almost surely and it holds that

$$\begin{aligned} S(x) &= S(a)\mathbf{P}(\tau^S_{S(a)} < \tau^S_{S(b)} | S(X_0) = S(x)) + S(b)\mathbf{P}(\tau^S_{S(b)} < \tau^S_{S(a)} | S(X_0) = S(x)) \\ &= S(a)\mathbf{P}_x(\tau_a < \tau_b) + S(b)(1 - \mathbf{P}_x(\tau_a < \tau_b)). \end{aligned}$$

Solving for $\mathbf{P}_x(\tau_a < \tau_b)$ yields the claim. □

In addition to the hitting probabilities of the last theorem, we want to calculate expected hitting times. Again, we start with an example of Brownian motion.

Example 18.24 (Expected Hitting Times of Brownian Motion)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion, i.e. an Itô diffusion with $\mu = 0$ and $\sigma = 1$ and $\tau_a := \inf\{t \geq 0 : X_t = a\}$. Since $(X_t^2 - t)_{t\geq 0}$ is a martingale, we have*

$$\begin{aligned} \mathbf{E}_x[\tau_a \wedge \tau_b] &= \mathbf{E}_x[X^2_{\tau_a \wedge \tau_b}] - \mathbf{E}_x[X^2_{\tau_a \wedge \tau_b} - \tau_a \wedge \tau_b] = a^2\mathbf{P}(\tau_a < \tau_b) + b^2\mathbf{P}_x(\tau_b < \tau_a) - x^2 \\ &= (a^2 - x^2)\frac{b-x}{b-a} + (b^2 - x^2)\frac{x-a}{b-a} \\ &= \frac{(x-a)(b-x)}{b-a}(-a - x + b + x) = (x-a)(b-x). \end{aligned}$$

We now generalize this example.

Theorem 18.25 (Expected Meeting Times)**.** *Let $l < a < x < b < r$, and let $\mathcal{X}$ and S be as in Theorem 18.23. Let $\tau := \tau_a \wedge \tau_b$ and $g \in \mathcal{C}^{(2)}((l, r))$ such that $x \mapsto \mathbf{E}_x\Big[\int_0^\tau g(X_s)ds\Big]$ is twice continuously differentiable. Then holds*

$$\mathbf{E}_x\Big[\int_0^\tau g(X_s)ds\Big] = \int_a^b G(x, y)g(y)dy$$

with

$$G(x,y) = \begin{cases} 2\dfrac{S(x)-S(a)}{S(b)-S(a)} \cdot \big(S(b)-S(y)\big)m(y), & x \le y \le b, \\ 2\dfrac{S(b)-S(x)}{S(b)-S(a)} \cdot \big(S(y)-S(a)\big)m(y), & a \le y \le x, \end{cases}$$

where

$$m(x) := \frac{1}{\sigma^2(x)S'(x)}.$$

Here G is called the green function *of $\mathcal{X}$.*

Remark 18.26 (Interpretation)**.** *The value $G(x,y)dy$ should be interpreted as the time mean time spent by a path started at x in $[y, y+dy)$ before hitting a or b.*

Example 18.27 (Expected hitting times of the Brownian motion)**.** *For $\mu = 0, \sigma = 1$, we compute*

$$G(x,y) = \begin{cases} 2\dfrac{x-a}{b-a} \cdot (b-y), & x \le y \le b, \\ 2\dfrac{b-x}{b-a} \cdot (y-a), & a \le y \le x, \end{cases}$$

Thus,

$$\begin{aligned} \mathbf{E}_x[\tau_a \wedge \tau_b] &= 2\frac{b-x}{b-a}\int_a^x (y-a)dy + 2\frac{x-a}{b-a}\int_x^b (b-x)dy \\ &= \frac{(b-x)(x-a)^2}{b-a} + \frac{(x-a)(b-x)^2}{b-a} = \frac{(b-x)(x-a)}{b-a}(x-a+b-x) = (x-a)(b-x), \end{aligned}$$

just as in example 18.24.

Proof of Theorem 18.25. We set

$$w(x) = \mathbf{E}_x\Big[\int_0^T g(X_s)ds\Big].$$

Then, using the Markov property of $\mathcal{X}$, we have

$$w(x) = \mathbf{E}_x\Big[\int_0^h g(X_s)ds\Big] + \mathbf{E}_x\big[w(X_h)\big].$$

For small h, we now write, since w is twice continuously differentiable,

$$\begin{aligned} \mathbf{E}_x\Big[\int_0^h g(X_s)ds\Big] &= hg(x) + o(h^2), \\ \mathbf{E}_x[w(X_h)] &= \mathbf{E}_x[w(x) + (X_h - x)w'(x) + \tfrac{1}{2}(X_h - x)^2 w''(x) + o(h^2)] \\ &= w(x) + h\big(\mu(x)w'(x) + \tfrac{1}{2}\sigma^2(x)w''(x) + o(h)\big). \end{aligned}$$

From this we read that

$$\mu(x)w'(x) + \tfrac{1}{2}\sigma^2(x)w''(x) = -g(x), \qquad w(a) = w(b) = 0$$

or equivalently

$$\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big)\frac{2\mu(x)}{\sigma^2(x)}w'(x) + \exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big)w''(x) = -\frac{2g(x)}{\sigma^2(x)}\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big),$$

$$\frac{d}{dx}\Big(\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big)w'(x)\Big) = -\frac{2g(x)}{\sigma^2(x)}\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big).$$

It holds that

$$m(x) = \frac{1}{\sigma^2(x)S'(x)} = \frac{1}{\sigma^2(x)}\exp\Big(2\int_y^x \frac{\mu(z)}{\sigma^2(z)}dz\Big),$$

from which it follows that

$$\frac{d}{dx}\Big(\frac{w'(x)}{S'(x)}\Big) = -2m(x)g(x).$$

If both sides are integrated, the following is obtained

$$\begin{aligned}
\frac{w'(x)}{S'(x)} &= -2\int_a^x m(y)g(y)dy + \beta,\\
w(x) &= -2\int_a^x S'(\eta)\int_a^\eta m(y)g(y)dyd\eta + \beta\int_a^x S'(\eta)d\eta + \alpha\\
&= -2\int_a^x (S(x)-S(y))m(y)g(y)dy + \beta(S(x)-S(a)) + \alpha
\end{aligned}$$

for $\alpha, \beta \in \mathbb{R}$. Since $w(a) = 0$, it follows that $\alpha = 0$ and because $w(b) = 0$, it holds that

$$\beta = \frac{2}{S(b)-S(a)}\int_a^b \big(S(b)-S(y)\big)m(y)g(y)dy$$

From this we get

$$\begin{aligned}
w(x) = 2\int_a^x &\frac{(S(x)-S(a))(S(b)-S(y)) - (S(x)-S(y))(S(b)-S(a))}{S(b)-S(a)}m(y)g(y)dy\\
&+ 2\int_x^b \frac{(S(x)-S(a))(S(b)-S(y))}{S(b)-S(a)}m(y)g(y)dy.
\end{aligned}$$

Simplifying now shows the result. □

Example 18.28 (Geometric Brownian Motion and Feller's branching process)**.**

1. *As we know, the solution of the SDE* $dX = \sigma X dW$ *is the martingale* $(\exp(\sigma W_t - \frac{1}{\sigma}^2 t))_{t\geq 0}$. *This process, the geometric Brownian motion, does not hit* 0. *Using the last result, we can at least show that the expected hitting time is infinite. Namely,* $S(x) = x$ *and with* $X_0 = 1$

$$\begin{aligned}
\mathbf{E}[\tau_\varepsilon] &= \lim_{R\to\infty}\mathbf{E}[\tau_\varepsilon \wedge \tau_R]\\
&= \lim_{R\to\infty} 2\int_\varepsilon^1 \frac{(R-1)(y-\varepsilon)}{R-\varepsilon}\frac{1}{\sigma^2 y^2}dy + 2\int_1^R \frac{(1-\varepsilon)(R-y)}{R-\varepsilon}\frac{1}{\sigma^2 y^2}dy\\
&= 2\int_\varepsilon^1 \frac{y-\varepsilon}{\sigma^2 y^2}dy + \int_1^\infty \frac{1-\varepsilon}{\sigma^2 y^2}dy \xrightarrow{\varepsilon\to 0} \infty.
\end{aligned}$$

Since the first term of the Green's function considers such paths with $\tau_\varepsilon < \tau_R$, one shows analogously that

$$\mathbf{E}[\tau_0, \tau_0 < \tau_R] = \lim_{\varepsilon \to 0} 2 \int_\varepsilon^1 \frac{(R-1)(y-\varepsilon)}{R-\varepsilon} \frac{1}{\sigma^2 y^2} dy = 2 \int_0^1 \frac{(R-1)y}{R} \frac{1}{\sigma^2 y^2} dy = \infty.$$

2. *Geometric Brownian motion is a non-negative process that never reaches 0 in finite time. This does not apply to Feller's branching process, the solution of the SDE*

$$dX = \sqrt{X} dW.$$

(Incidentally, the existence and uniqueness of the solution here not from Theorem 18.4, but from other considerations.) Here, too, $S(x) = x$, but with $x = 1$

$$\mathbf{E}[\tau_0, \tau_0 < \tau_R] = \lim_{\varepsilon \to 0} \mathbf{E}[\tau_\varepsilon, \tau_\varepsilon < \tau_R] = 2 \int_0^1 \frac{(R-1)y}{R} \frac{1}{\sigma^2 y} dy < \infty.$$

This means that the 0 of the paths that do not meet R will almost surely be met in finite time.

We now want to deepen the outlook given in chapter **??** on applications of stochastic processes in financial mathematics. In contrast to chapter **??**, the models now dealt with will be continuous in time. Our acquired knowledge of stochastic integration will be used.

19 Introduction

19.1 Notation

To introduce models for financial markets in a mathematically sound way, we need some definitions for the following. For this, a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and an n-dimensional Brownian motion $\underline{\mathcal{W}} = (\mathcal{W}^1, ..., \mathcal{W}^n), \mathcal{W}^j = (W_t^j)_{0 \le t \le \tau}$ for some $\tau > 0$. Further, let $(\mathcal{F}_t)_{0 \le t \le \tau}$ be the filtration generated by $\underline{\mathcal{W}}$ with $\mathcal{F} = \mathcal{F}_\tau$. Thus, we model a stock market between times 0 and $\tau > 0$.

Definition 19.1 (Financial market). *A (financial) market is a $(\mathcal{F}_t)_{0 \le t \le \tau}$-adapted, $d+1$-dimensional process $\underline{\mathcal{X}} = (\mathcal{X}^0, ..., \mathcal{X}^d)$, $\mathcal{X}^i = (X_t^i)_{0 \le t \le \tau}$ of the form*

$$\begin{aligned} \mathcal{X}^0 &= X_0^0 + \mathcal{R}\mathcal{X}^0 \cdot \lambda, \\ \mathcal{X}^i &= X_0^i + \mu^i \mathcal{X}^i \cdot \lambda + \mathcal{X}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k, \qquad i = 1, ..., d \end{aligned} \tag{19.1}$$

for adapted processes $\mathcal{R} = (R_t)_{0 \le t \le \tau}, \mu^i = (\mu_t^i)_{0 \le t \le \tau}, \sigma^{ik} = (\sigma_t^{ik})_{0 \le t \le \tau}, i = 1, ..., d, k = 1, ..., n$. Here, $\mathcal{X}^0$ denotes a risk-free investment. The market is said to be normalized if $X_t^0 = 1$ for all $0 \le t \le \tau$. normalization of the market $\underline{\mathcal{X}}$ is the process $\underline{\widetilde{\mathcal{X}}} = (\widetilde{\mathcal{X}}^0, ..., \widetilde{\mathcal{X}}^d)$ with $\widetilde{X}_t^i = X_t^i / X_t^0$.

Remark 19.2 (Interpretation).

1. *In the above definition, it may come as a surprise that $\mathcal{X}^0$ is classified as risk-free, even though $\mathcal{R}$ can be a stochastic process. At least it follows from (19.1) that – in contrast to $\mathcal{X}^1, ..., \mathcal{X}^n$ – the quadratic variation of $\mathcal{X}^0$.*

2. *In applications, we will mostly assume that $\mathcal{X}^0, ..., \mathcal{X}^d$ is the solution of the SDE*

$$\begin{aligned} dX^0 &= r(t, \underline{X}_t) X_t^0 dt, \\ dX^i &= \mu^i(t, \underline{X}_t) X_t^i dt + X_t^i \sum_{k=1}^n \sigma^{ik}(t, \underline{X}_t) dW_t^k, \qquad i = 1, ..., n. \end{aligned} \tag{19.2}$$

   where $r, \mu^i, \sigma^{ik}, i = 1, ..., d, k = 1, ..., n$ are locally Lipschitz continuous functions. This market model is also called the generalized Black-Scholes model. Since the model (19.1) also allows processes μ^i, σ^{ik} that depend not only on $\underline{X}_t$ at time t but on the entire path $(\underline{X}_s)_{0 \le s \le t}$, this is a strong – but practical – restriction.

3. *The processes $\mathcal{X}^1, ..., \mathcal{X}^d$ are prices of risky investments, such as stocks. These prices are correlated, as can be seen from (19.1). It holds that*

$$\begin{aligned} [\mathcal{X}^i, \mathcal{X}^j] &= \Big[\mathcal{X}^i \sum_{k=1}^n \sigma^{ik} \cdot \ ConW^k, \mathcal{X}^j \sum_{l=1}^n \sigma^{jl} \cdot \mathcal{W}^l\Big] = \mathcal{X}^i \mathcal{X}^j \sum_{k,l=1}^n \sigma^{ik} \sigma^{jl} \cdot [\mathcal{W}^k, \ Convolve\mathcal{W}^l] \\ &= \sum_{k=1}^n \sigma^{ik} \sigma^{jk} \cdot \lambda. \end{aligned}$$

From this, for example,

$$\mathbf{COV}[X_t^i, X_t^j] = \mathbf{E}\Big[\sum_{k=1}^n \int_0^t X_s^i \sigma_s^{ik} X_s^j \sigma_s^{jk} ds\Big].$$

4. *The model* (19.1) *has the property that all prices are continuous semimartingales. In order to model the large price fluctuations that occur in practice in a short period of time, one can also switch to discontinuous semimartingales. However, since we have developed the theory of stochastic integration only for continuous integrators, we will not consider such models here.*

Definition 19.3 (Portfolio)**.** *In a market of the form* (19.1)*, a* portfolio *is an adapted stochastic process* $\underline{\Delta} = (\Delta^0, ..., \Delta^d)$*,* $\Delta^i = (\Delta_t^i)_{0\le t\le\tau}$*. Here,* Δ_t^i *is the number of securities of type* i *at time* t*. The value of the portfolio* $\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) = V_t(\underline{\Delta}, \underline{\mathcal{X}})$ *at time* t *is*

$$V_t(\underline{\Delta}, \underline{\mathcal{X}}) := \sum_{i=0}^d \Delta_t^i X_t^i =: \langle \underline{\Delta}_t, \underline{X}_t\rangle.$$

A portfolio is said to be self-financing *(with respect to* $\underline{\mathcal{X}}$*) if*

$$\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) = V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \sum_{i=0}^d \Delta^i \cdot \mathcal{X}^i, \tag{19.3}$$

d.h.

$$V_t(\underline{\Delta}, \underline{\mathcal{X}}) - V_0(\underline{\Delta}, \underline{\mathcal{X}}) := \sum_{i=0}^d \int_0^t \Delta_s^i dX_s^i$$

and

$$\begin{aligned}
\int_0^t |\Delta_s^0 R_s X_s^0| ds < \infty,& \\
\int_0^t |\Delta_s^i \mu_s^i X_s^i| ds < \infty,& \qquad i = 1, ..., d \\
\int_0^t (\Delta_s^i X_s^i \sigma_s^{ik})^2 ds < \infty,& \qquad i = 1, ..., d, k = 1, ..., n.
\end{aligned}$$

Remark 19.4 (Interpretation)**.** *The concept of a portfolio is self-explanatory. For the value of a (not necessarily self-financing) portfolio composed of semimartingals, the following applies because of the formula for partial integration, Theorem 16.48,*

$$\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) - V_0(\underline{\Delta}, \underline{\mathcal{X}}) = \sum_{i=0}^d \Delta^i \cdot \mathcal{X}^i + \mathcal{X}^i \cdot \Delta^i + [\ ConX^i, \Delta^i].$$

To understand self-financing portfolios, it is helpful to first discretize time. Let us divide the interval $[0, t]$ *into equally sized intervals of length* t/n *with large* n*, then a portfolio is self-financing if*

$$V_{(k+1)t/n}(\underline{\Delta}, \underline{\mathcal{X}}) - V_{kt/n}(\underline{\Delta}, \underline{\mathcal{X}}) = \sum_{i=0}^d \Delta_{kt/n}^i (X_{(k+1)t/n} - X_{kt/n}).$$

This means that changes in the value of the portfolio only come about because the prices of the securities change. After the price change has taken place, a change in the portfolio, i.e. a reinvestment in a different composition of securities, can follow.

It is often more practical to calculate in the normalized market $\widetilde{\mathcal{X}}$ instead of in the market model $\underline{\mathcal{X}}$. The following lemma shows how to do this and that self-financing portfolios remain self-financing under normalization .

Lemma 19.5 (Normalization of Markets and Self-Financing Strategies)**.** *Let $\underline{\mathcal{X}}$ and $\widetilde{X}$ be as in Definition 19.1, i.e., in particular, that $\widetilde{X_t}^i = X_t^i / X_t^0$. Then*

$$\widetilde{\mathcal{X}}^0 = 1,$$
$$\widetilde{\mathcal{X}}^i = X_0^i + (\mu^i - \mathcal{R})\widetilde{\mathcal{X}}^i \cdot \lambda + \widetilde{\mathcal{X}}^i \sum_{k=1}^{n} \sigma^{ik} \cdot \mathcal{W}^k, \qquad i = 1, ..., d.$$

Furthermore, let $\underline{\Delta}$ be a portfolio. Then

$$\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = \frac{1}{\mathcal{X}^0} \mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}})$$

and $\underline{\Delta}$ is self-financing with respect to $\underline{\mathcal{X}}$ if and only if $\underline{\Delta}$ is self-financing with respect to $\widetilde{\mathcal{X}}$.

Proof. The formula $\widetilde{\mathcal{X}}^0 = 1$ is clear from the definition. Furthermore,

$$X_t^0 = \exp\Big(\int_0^t R_s ds \Big) = 1 + \int_0^t R_s \exp\Big(\int_0^s R_r dr \Big) ds, \qquad \mathcal{X}^0 = 1 + \mathcal{R}\mathcal{X}^0 \cdot \lambda, \tag{19.4}$$

also

$$\widetilde{X}_t^i = X_t^i \exp\Big(- \int_0^t R_s ds \Big), \qquad \frac{1}{\mathcal{X}^0} = 1 - \frac{\mathcal{R}}{\mathcal{X}^0} \cdot \lambda.$$

Using the formula for partial integration, we write (since the quadratic variation of $\mathcal{X}^0$ disappears)

$$\begin{aligned} \widetilde{\mathcal{X}}^i &= X_0^i + \frac{1}{\mathcal{X}^0} \cdot \mathcal{X}^i + \mathcal{X}^i \cdot \frac{1}{\mathcal{X}^0} \\ &= X_0^i + \frac{1}{\mathcal{X}_0} \cdot \Big(\mu^i \mathcal{X}^i \cdot \lambda + \sum_{k=1}^{n} \sigma^{ik} \cdot \mathcal{W}^k \Big) - \Big(\mu^i - \frac{1}{\mathcal{X}^0} \cdot \lambda \Big) \cdot \Big(\frac{1}{\mathcal{X}^0} \cdot \lambda + \frac{1}{\mathcal{X}^0} \cdot \lambda \Big) \\ &= X_0^i + \widetilde{\mathcal{X}}^i (\mu^i - \frac{1}{\mathcal{R}}) \cdot \lambda + \widetilde{\mathcal{X}}^i \sum_{k=1}^{n} \sigma^{ik} \cdot \mathcal{W}^k. \end{aligned} \tag{19.5}$$

Let $\underline{\Delta}$ be a (with respect to $\underline{\mathcal{X}}$) self-financing portfolio, i.e. $\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) := \sum_i \Delta^i \mathcal{X}^i = \sum_i \Delta^i \cdot \mathcal{X}^i$.

Then, again with partial integration,

$$\begin{aligned}\mathcal{V}(\underline{\Delta}, \widetilde{\underline{\mathcal{X}}}) &= \sum_{i=0}^{d} \Delta^i \widetilde{\mathcal{X}}^i = \sum_{i=0}^{d} \frac{\Delta^i \mathcal{X}^i}{\mathcal{X}^0} = \frac{1}{\mathcal{X}^0} \sum_{i=0}^{d} \Delta^i \mathcal{X}^i = \sum_{i=0}^{d} \frac{1}{\mathcal{X}^0} (V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \Delta^i \cdot \mathcal{X}^i) \\ &= V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \sum_{i=0}^{d} \frac{1}{\mathcal{X}^0} \Big(\Delta^i \mu^i \mathcal{X}^i \cdot \lambda + \Delta^i \mathcal{X}^i \sum_{k=1}^{n} \sigma^{ik} \cdot \mathcal{W}^k \Big) \\ &= V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \sum_{i=0}^{d} \Delta^i \mu^i \widetilde{\mathcal{X}}^i \cdot \lambda + \Delta^i \widetilde{\mathcal{X}}^i \sum_{k=1}^{n} \sigma^{ik} \cdot \mathcal{W}^k - \Delta^i \mathcal{X}^i \frac{\mathcal{R}}{\mathcal{X}^0} \cdot \lambda \\ &= V_0(\underline{\Delta}, \underline{\mathcal{X}}) + \sum_{i=0}^{d} \Delta^i \cdot \widetilde{\mathcal{X}}^i. \end{aligned}$$

The reverse direction can be shown analogously. □

In the following, we will deal with financial derivatives and their fair prices, where the underlying stock prices follow certain stochastic processes. We will first define what financial derivatives are and what a fair price is. However, a fair price can only exist in arbitrage-free markets.

Definition 19.6 (Arbitrage)**.** *We use the notation from Definitions* **??** *and* **??**.

1. *A portfolio* $\underline{\Delta}$ *is* admissible *if the value* $\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}})$ *is almost surely bounded from below.*

2. *An admissible, self-financing portfolio* $\underline{\Delta}$ *is called* arbitrage *if almost surely*

$$V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0, \qquad V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0 \qquad \text{and} \qquad \mathbf{P}(V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) > 0) > 0.$$

Remark 19.7 (Interpretation)**.**

1. *Only portfolios in which the investor can take on a limited amount of debt are admissible in the above sense.*

2. *We have already seen the definition of arbitrage in remark* **??** *already. Roughly speaking, an arbitrage is a possibility to make a profit without risk and with limited financial resources. We note that the existence of arbitrage depends on the pricing process* $\underline{\mathcal{X}}$. *If there is no arbitrage, we also say that the market* $\underline{\mathcal{X}}$ *is arbitrage-free.*

Definition 19.8 (Derivatives, Achievable Strategies and the Fair Price)**.** *We use the notation from Definitions* **??** *and* **??**.

1. *A* (financial) derivative *is a* $\mathcal{F}_\tau$*-measurable, from below bounded, square-integrable random variable* S.

2. *The derivative* S *is called* achievable *if there exists a* $h \in \mathbb{R}$ *and a self-financing portfolio* $\underline{\Delta} = (\Delta^0, ...\Delta^d)$ *with* $\mathcal{V}_0(\underline{\Delta}, \underline{\mathcal{X}}) = h$, *so that*

$$S = \mathcal{V}_\tau(\underline{\Delta}, \underline{\mathcal{X}})$$

is almost surely. If the market is arbitrage-free, then $\mathcal{V}_t(\underline{\Delta}, \underline{\mathcal{X}})$ *is the fair price of the derivative* S *at time* t. *Furthermore,* $\underline{\Delta}$ *a* hedging portfolio *of* S.

Remark 19.9 (The fair price)**.** *The concept of the* fair price *of the derivative S is to be explained by the demand for the arbitrage-free market explain. If – say at time $t = 0$ – a derivative with a fair price h were to be sold at a price $\ell > h$, then one could sell this derivative at the price ℓ and, with $h < \ell$, set up a self-financing portfolio with seed capital that would almost certainly have the same value at all times as the sold derivative. At time τ, there is still $\ell - h$ of the starting capital left, without risk. For $\ell < h$, a similar procedure leads to an arbitrage opportunity. We will come back to the concept of arbitrage more often in the future.*

19.2 The Black-Scholes model

The Black-Scholes model is the simplest continuous-time model of a stock market of the form given in Definition 19.1. It consists of only two securities, where the process $\mathcal{R}$ and μ^1 and σ^{11} are constant. Thus, we have

$$\begin{aligned} dX^0 &= rX^0 dt, \\ dX^1 &= \mu X^1 dt + \sigma X^1 dW \end{aligned} \tag{19.6}$$

for $r, \mu \in \mathbb{R}, \sigma > 0$ and a Brownian motion $\mathcal{W} = (W_t)_{0 \leq t \leq \tau}$. This SDE can be solved explicitly, namely by

$$\begin{aligned} X_t^0 &= X_0^0 e^{rt}, \\ X_t^1 &= X_0^1 \exp\big((\mu - \tfrac{1}{2}\sigma^2)t + \sigma W_t\big), \end{aligned} \tag{19.7}$$

where X_0^1 is the (deterministic) starting price of the stock at time 0 and is assumed here to be 1. See examples 18.1.2 and 18.14. Using (19.7) we see that X_t^1 follows a logarithmic normal distribution for each t . Furthermore, we see that X_s^1 and $(X_t^1 - X_s^1)/X_s^1$ for $0 \leq s \leq t$ are independent. In addition, with $s \leq t \leq u$,

$$\frac{X_t^1 - X_s^1}{X_s^1} \stackrel{d}{=} \frac{X_{t-s}^1 - X_0^1}{X_0^1},$$

i.e. the relative increments are distributed independently and identically. Furthermore, we see from Lemma 19.5 that the normalization of (19.6), $\widetilde{X}^1$ by

$$d\widetilde{X}^1 = (\mu - r)\widetilde{X}^1 dt + \sigma \widetilde{X}^1 dW, \qquad \widetilde{X}_t^1 = X_0^1 \exp\big((\mu - r - \tfrac{1}{2}\sigma^2)t + \sigma W_t\big) \tag{19.8}$$

. We start with an important observation about the normalized process $\widetilde{X}^1$, which we actually already saw in Example (18.14). Throughout the section, we use the notation introduced above.

Proposition 19.10 (A measure under which $\widetilde{\mathcal{X}}^1$ is a martingale)**.** *We define a probability measure $\mathbf{Q}$ by*

$$\mathbf{Q} = Z_\tau \cdot \mathbf{P}, \qquad Z_\tau = \exp\Big(- \frac{\mu - r}{\sigma} W_\tau - \frac{1}{2}\frac{(\mu - r)^2}{\sigma^2}\tau\Big).$$

Then $\widetilde{\mathcal{X}}^1$ is a $\mathbf{Q}$-martingale, given by

$$\widetilde{\mathcal{X}}^1 = \sigma \widetilde{\mathcal{X}}^1 \cdot \widetilde{\mathcal{W}},$$

where $\widetilde{\mathcal{W}} = \widetilde{\mathcal{W}} + \frac{\mu - r}{\sigma} t$ is a Brownian motion with respect to $\mathbf{Q}$ and

$$\widetilde{X}_t^1 = X_0^1 \exp\Big(\sigma \widetilde{W}_t - \tfrac{1}{2}\sigma^2 t\Big). \tag{19.9}$$

Proof. See example 18.14. □

As we saw in Definition 19.8, the price of a derivative can only be meaningfully determined if the market is arbitrage-free. We will now show this for the Black-Scholes model .

Theorem 19.11 (Arbitrage-Freedom and Prices in the Black-Scholes Model)**.** *The Black-Scholes model is arbitrage-free. This means that for every bounded, self-financing portfolio $\underline{\Delta}$ with $V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0$ and $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$, it is almost certain that $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) = 0$. Furthermore, the fair price of a derivative S at time t is given by*

$$\mathbf{E_Q}[e^{-r(\tau-t)}S|\mathcal{F}_t], \tag{19.10}$$

where $\mathbf{Q}$ *is the probability measure from Proposition 19.10.*

Proof. First, the arbitrage-free property: Let $\underline{\Delta}$ be a portfolio with $V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0$ and $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ almost surely. Since $\underline{\Delta}$ is also self-financing with respect to $\widehat{\mathcal{X}}$, it holds that

$$\mathcal{V}(\underline{\Delta}, \widetilde{X}) = \sum_{i=0}^{d} \Delta^i \cdot \widetilde{X}^i.$$

Since by Proposition 19.10 the processes $\widetilde{X}^i, i = 0, 1$ are martingales with respect to $\mathbf{Q}$, $\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})$ is also a Q-martingale. By assumption, $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ is almost surely P-sure. Since P and Q are equivalent, $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ is also $\mathbf{Q}$-almost sure. Because of the martingale property of $\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})$ in addition, $\mathbf{E_Q}[V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})] = 0$. However, this is only possible if $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = 0$ is $\mathbf{Q}$-almost surely. Again, because of the equivalence of $\mathbf{P}$ and $\mathbf{Q}$, $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = 0$ follows from $\mathbf{P}$-almost surely, and arbitrage-freeness is shown.

We now show the general price formula (19.10). First, it is clear that $\mathcal{M} = (M_t)_{0\leq t\leq\tau}$, given by $M_t := \mathbf{E_Q}[e^{-r\tau}S|\mathcal{F}_t]$, is a $\mathbf{Q}$-martingale. According to Theorem 17.9, there is a process $\mathcal{H} = (H_t)_{0\leq t\leq\tau}$ with

$$M_t = \mathbf{E_Q}[e^{-r\tau}S] + \int_0^t H_s d\widetilde{X}_s.$$

Now we write $\Delta_t^1 := H_t/(\sigma\widetilde{X}_t^1), \Delta_t^0 = M_t - H_t/\sigma$, so that

$$V_t(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = \Delta_t^0 + \Delta_t^1\widetilde{X}_t^1 = M_t,$$

$$\mathbf{E_Q}[e^{-r\tau}S] + (\Delta^0 \cdot \widetilde{\mathcal{X}}^0 + \Delta^1 \cdot \widetilde{\mathcal{X}}^1)_t = \mathbf{E_Q}[e^{-r\tau}S] + \int_0^t H_s d\widetilde{W}_s = M_t.$$

Thus, $\underline{\Delta} = (\Delta^1, \Delta^2)$ is self-financing with respect to $\underline{\widetilde{\mathcal{X}}}$, according to lemma 19.5 also self-financing with respect to $\underline{\mathcal{X}}$. Because of the definition of $\mathcal{M}$, $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = M_\tau = e^{-r\tau}S$ is also almost surely true, so – again with lemma 19.5 –

$$V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) = S$$

almost surely. By definition, the fair price at time t is given by

$$V_t(\underline{\Delta}, \underline{\mathcal{X}}) = e^{rt}V_t(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = e^{rt}M_t = \mathbf{E_Q}[e^{-r(\tau-t)}S|\mathcal{F}_t].$$

With this, all assertions are shown. □

While the last theorem gives a general result for the fair price of a derivative, in practice you also need a hedging portfolio for the derivative. This problem is now solved for the Black-Scholes model.

Proposition 19.12 (Hedging Portfolio in the Black-Scholes Model)**.** *Let S be a derivative with price*

$$\mathbf{E_Q}[e^{-r(t-\tau)}S|\mathcal{F}_t] = F(X_t^1, t)$$

for a smooth function F. Then the portfolio $\underline{\Delta} = (\Delta_t^0, \Delta_t^1)_{0\leq t\leq\tau}$ with

$$\Delta_t^1 = \frac{\partial F(X_t^1, t)}{\partial x}, \qquad \Delta_t^0 = e^{-rt}F(X_t^1, t) - \Delta_t^1 \widetilde{X}_t^1$$

is a hedging portfolio for S.

Proof. We immediately compute the value of the portfolio as

$$V_t(\underline{\Delta}, \underline{\mathcal{X}}) = \Delta_t^0 X_t^0 + \Delta_t^1 X_t^1 = F(X_t^1, t) - \Delta_t^1 \widetilde{X}_t^1 e^{rt} + \Delta_t^1 X_t^1 = F(X_t^1, t).$$

It remains to be shown that the portfolio is self-financing. For this, it is sufficient to show, due to Lemma 19.5, that the portfolio is self-financing for $\widetilde{X}_t^1$. We set

$$\widetilde{F}(x, t) := e^{-rt}F(xe^{rt}, t),$$

so that with the Itô formula

$$\begin{aligned}\widetilde{F}(\widetilde{X}_t^1, t) &= \widetilde{F}(\widetilde{X}_0^1, 0) + \int_0^t \frac{\partial \widetilde{F}(\widetilde{X}_s^1, s)}{\partial x} dX_s^1 + \int_0^t \frac{\partial \widetilde{F}(\widetilde{X}_s^1, s)}{\partial s} ds + \tfrac{1}{2}\int_0^t \frac{\partial^2 \widetilde{F}(\widetilde{X}_s^1, s)}{\partial x^2} d[\widetilde{\mathcal{X}}^1]_s \\ &= \widetilde{F}(\widetilde{X}_0^1, 0) + \int_0^t \frac{\partial \widetilde{F}(\widetilde{X}_s^1, s)}{\partial x} d\widetilde{X}_s^1 + \int_0^t K_s ds\end{aligned}$$

holds for a process $(K_t)_{0\leq t\leq\tau}$. Since $(\widetilde{F}(\widetilde{X}_t^1, t))_{0\leq t\leq\tau} = e^{-r\tau}(\mathbf{E_Q}[S|\mathcal{F}_t])_{0\leq t\leq\tau}$ is a Q-martingale, then by Theorem 16.26 $(K_t)_{0\leq t\leq\tau}$ must vanish. Thus, since $\frac{\partial \widetilde{F}(\widetilde{X}_t, t)}{\partial x} = \frac{\partial F(X_t, t)}{\partial x}$, we have that

$$\begin{aligned}V_t(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = \widetilde{F}(\widetilde{X}_t^1, t) &= F(X_0^1, 0) + \int_0^t \frac{\partial F(X_s^1, s)}{\partial x} d\widetilde{X}_s^1 \\ &= V_0(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) + \int_0^t \Delta_s^0 d\widetilde{X}_s^0 + \int_0^t \Delta_s^1 d\widetilde{X}_s^1,\end{aligned}$$

whereby $\underline{\Delta}$ is self-financing for $\widetilde{X}_s^1$ and □

19.3 Price and Hedging of European Options: The Black-Scholes Formula

In this section, we calculate the fair price of European call and put options in the Black-Scholes model. This leads to the well-known Black-Scholes formula for option pricing.

Definition 19.13 (European Call and Put Option)**.**

1. *A European call option (on a stock with price process $\mathcal{X}^1$) with maturity τ and strike price K is a derivative $S^c := (X_\tau^1 - K)^+$.*
2. *A European put option (on a stock with price process $\mathcal{X}^1$) with maturity τ and strike price K is a derivative $S^p := (K - X_\tau^1)^+$.*

Remark 19.14 (Interpretation). *1. The European option gives the owner, as in Section* **??***, the right (but not the obligation) to buy the share (with price process $\mathcal{X}^1$) at time τ at the price of K. The value at time τ is thus exactly the possible profit if the option is exercised. If the owner immediately sells the share bought for K at the price $X_\tau^1 > K$, he makes a profit of $X_\tau^1 - K$. If $K > X_\tau^1$, it is more favorable to let the option expire . Overall, the value of $S^c = (X_\tau^1 - K)^+$ at time τ*

2. The European put option gives the owner the right to sell *the stock at time τ at a price of K. Of course, this is only worthwhile if $X_\tau^1 < K$. In this case, the profit is then – similar to the call option – the profit is $K - X_\tau^1$. (Note that S^p is thus even bounded.)*

Since European call and put options complement each other in a sense, the following relationship applies to their prices.

Lemma 19.15 (Put/Call Parity for European Options). *Let S_t^c and S_t^p be the prices of European call and put options on a stock with price process $\mathcal{X}^1$ at the same maturity τ and strike K in the Black-Scholes model. Then*

$$S_t^c - S_t^p = X_t^1 - Ke^{-r(\tau-t)}.$$

Proof. It holds that $(X_t^1 - K)^+ - (K - X_t^1)^+ = (X_t^1 - K)^+ + (X_t^1 - K)^- = X_t^1 - K$. Thus, by applying Theorem 19.11 with the equivalent measure $\mathbf{Q}$ we get

$$\begin{aligned} S_t^c - S_t^p &= \mathbf{E_Q}[e^{-r(\tau-t)}(X_\tau^1 - K)|\mathcal{F}_t] = e^{rt}\mathbf{E_Q}[\widetilde{X}_\tau^1|\mathcal{F}_t] - Ke^{-r(\tau-t)} \\ &= e^{rt}\widetilde{X}_t^1 - K = X_t^1 - Ke^{-r(T-t)}. \end{aligned}$$

□

We now use Theorem 19.11 to calculate the price of European call and put options in the Black-Scholes model. This leads to the famous Black-Scholes formula for option pricing.

Theorem 19.16 (Black-Scholes Formula). *1. The fair price at time t of a European call option with maturity τ and strike price K in the Black-Scholes model is given by*

$$S_t^c = F(X_t^1, \tau - t)$$

with

$$F(x,t) = x\Phi(d_1(x,t)) - Ke^{-rt}\Phi(d_2(x,t)),$$

where Φ is the distribution function of the standard normal distribution and

$$d_1(x,t) = \frac{\log(x/K) + (r + \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}}, \qquad d_2(x,t) = d_1(x,t) - \sigma\sqrt{t}.$$

The portfolio $\underline{\Delta} = (\Delta_t^0, \Delta_t^1)_{0\le t\le\tau}$ with

$$\Delta_t^1 = \Phi(d_1(X_t^1, t)), \qquad \Delta_t^0 = e^{-rt}F(X_t^1, t) - \Delta_t^1\widetilde{X}_t^1$$

is a hedging portfolio for S^c.

2. Furthermore, the fair price at time t of a European put option with maturity τ and strike price K is given by

$$S_t^p = Ke^{-r(\tau-t)}\Phi(-d_2(X_t^1,\tau-t)) - X_t^1\Phi(-d_1(X_t^1,\tau-t)).$$

The portfolio $\underline{\Delta} = (\Delta_t^0,\Delta_t^1)_{0\le t\le\tau}$ with

$$\Delta_t^1 = -\Phi(-d_1(X_t^1,t)), \qquad \Delta_t^0 = e^{-rt}F(X_t^1,t) - \Delta_t^1\widetilde{X}_t^1$$

is a hedging portfolio for S^p.

Proof. Since $\Phi(-x)+\Phi(x) = 1$ for all $x\in\mathbb{R}$, it suffices to calculate the price of the European call option because of lemma 19.15. The proof of the hedging portfolio is also analogous.

To calculate the price of a call option, we have to calculate $S_t^c = \mathbf{E_Q}[e^{-r(\tau-t)}(X_\tau^1-K)^+|\mathcal{F}_t]$ for the equivalent measure $\mathbf{Q}$ with density Z_τ from Proposition 19.10. Due to (19.9) is

$$\begin{aligned} S_t^c &= \mathbf{E_Q}\Big[e^{-r(\tau-t)}\Big(X_t^1\frac{\widetilde{X}_\tau^1}{\widetilde{X}_t^1}e^{r(\tau-t)} - K\Big)^+\Big|\mathcal{F}_t\Big] \\ &= \mathbf{E_Q}[e^{-r(\tau-t)}(X_t^1e^{\sigma(\widetilde{W}_\tau^1-\widetilde{W}_t^1)+\left(r-\frac{1}{2}\sigma^2\right)(\tau-t)} - K)^+|\mathcal{F}_t] \\ &= \mathbf{E_Q}\big[\big(X_t^1e^{\sigma Z\sqrt{\tau-t}-\frac{1}{2}\sigma^2(\tau-t)} - Ke^{-r(\tau-t)}\big)^+|\mathcal{F}_t\big]. \end{aligned}$$

for a random variable $Z \stackrel{d}{=} N(0,1)$, which is independent of $\mathcal{F}_t$. Now

$$xe^{\sigma\sqrt{t}z-\frac{1}{2}\sigma^2t} > Ke^{-rt} \iff z > \frac{1}{\sigma\sqrt{t}}\Big(\tfrac{1}{2}\sigma^2t + \log\frac{Ke^{-rt}}{x}\Big) = \frac{-\log\frac{x}{K} - (r-\frac{1}{2}\sigma^2)t}{=} - d_2(x,t).$$

Therefore, $S_t^c = F(X_t^1,t)$ with

$$\begin{aligned} F(x,t) &= \frac{1}{\sqrt{2\pi}}\int_{-d_2(x,t-r)}^{\infty}(xe^{\sigma z\sqrt{\tau-t}-\frac{1}{2}\sigma^2(\tau-t)} - Ke^{-r(\tau-t)})e^{-\frac{z^2}{2}}dz \\ &\stackrel{y=z-\sigma\sqrt{\tau-t}}{=} \frac{1}{\sqrt{2\pi}}\Big(x\int_{-d_1(x,\tau-t)}^{\infty} e^{\sigma y\sqrt{\tau-t}+\frac{1}{2}\sigma^2(\tau-t)}e^{-(\frac{y^2}{2}+\sigma y\sqrt{\tau-t}+\frac{1}{2}\sigma^2(\tau-t))}dy \\ &\qquad - Ke^{-r(\tau-t)}\Big)\int_{-d_2(x,\tau-t)}^{\infty} e^{-\frac{z^2}{2}}dz\Big) \\ &= x\Phi(d_1(x,\tau-t)) - Ke^{-r(\tau-t)}\Phi(d_2(x,\tau-t)) \end{aligned}$$

and the option price formula is shown. For the hedging portfolio, we use Proposition 19.12 and calculate directly

$$\begin{aligned} \frac{\partial F(x,t)}{\partial x} &= \Phi(d_1(x,t)) + \frac{1}{\sqrt{2\pi}}\Big(x\exp\big(-\tfrac{1}{2}d_1^2(x,t)\big) - Ke^{-rt}\exp\big(-\tfrac{1}{2}(d_1(x,t)-\sigma\sqrt{t})^2\big)\frac{1}{x\sigma\sqrt{t}} \\ &= \Phi(d_1(x,t)) + \frac{1}{x\sigma\sqrt{2\pi t}}\exp\big(-\tfrac{1}{2}d_1^2(x,t)\big)\Big(x - Ke^{-rt}\exp\big(d_1(x,t)\sigma\sqrt{t}-\tfrac{1}{2}\sigma^2t\big)\frac{1}{x\sigma\sqrt{t}} \\ &= \Phi(d_1(x,t)) + \frac{1}{\sqrt{2\pi}}\exp\big(-d_1^2(x,t)\big)\Big(x - Ke^{-rt}\exp\big(\log(x/K)+rt\big)\Big)\frac{1}{x\sigma\sqrt{t}} \\ &= \Phi(d_1(x,t)). \end{aligned}$$

$\square$

Remark 19.17 (Volatility)**.** *It is noticeable that the price of the European call and put options no longer depends on μ, but only on r and σ, which is also volatility. Usually, r is known as the interest rate. To determine the volatility of a particular security, there are two different possibilities:*

1. *The* historical method *is based on the observations* $\log(X^1_t/X^1_0)$, $\log(X^1_{2t}/X^1_t)$, $\log(X^1_{3t}/X^1_{2t}), \ldots$ *are based. These random variables are independent in the Black-Scholes model and distributed according to* $N((\mu - \frac{1}{2}\sigma^2)t, \sigma^2 t)$*. From the empirical variance of these quantities, the volatility* σ *can be estimated.*

2. *The* implicit method *observes options traded on the market (European options) and their prices and uses the prices to calculate what volatility has been applied, so that the Black-Scholes formula is used to calculate the traded price.*

Remark 19.18 (The Greeks)**.** *In practice, one is interested in the dependence of the price of an option on the model parameters. The quantity*

$$\frac{\partial F(x,t)}{\partial x}$$

indicates the dependence of the option price on the current price of the stock and is called the option's delta*. (We just calculated in the proof that it is* $\Phi(d_1(x,t))$ *for a European call option.) To further determine the sensitivity with respect to the current price, calculate* $\frac{\partial^2 F(x,t)}{\partial x^2}$*, which is also referred to as the. Finally,* theta *is given as* $\frac{\partial F(x,t)}{\partial t}$.

19.4 Arbitrage-Freedom and Completeness

The goal of this section is to derive the two fundamental theorems of option pricing. The first one states that arbitrage-freedom is equivalent to the existence of a so-called equivalent martingale measure. (This is a measure with respect to which all securities are martingals.) The second one says that all derivatives can be hedged if and only if the equivalent martingale measure is unique. We will show these two theorems for the financial market from Definition **??**, i.e. the stock prices follow the price processes $\underline{\mathcal{X}} = (\mathcal{X}^0, \ldots, \mathcal{X}^d)$, $\mathcal{X}^i = (X^i_t)_{0 \le t \le \tau}$ of the form

$$\begin{aligned} \mathcal{X}^0 &= 1 + \mathcal{R}\mathcal{X}^0 \cdot \lambda, \\ \mathcal{X}^i &= X^i_0 + \mu^i \mathcal{X}^i \cdot \lambda + \mathcal{X}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k, \qquad i = 1, \ldots, d \end{aligned} \tag{19.11}$$

with an n-dimensional Brownian motion $\underline{\mathcal{W}}$ and a filtration generated by it $(\mathcal{F}_t)_{0 \le t \le \tau}$ such that $\mathcal{R} = (R_t)_{0 \le t \le \tau}, \mu^i = (\mu^i_t)_{0 \le t \le \tau}, \sigma^{ik} = (\sigma^{ik}_t)_{0 \le t \le \tau}, i = 1, \ldots, d, k = 1, \ldots, n$ are adapted. Further, we recall the normalized processes $\widetilde{X^i_t} = (X^i_0, \ldots, X^i_{d-1})/X^0 = (\widetilde{X^0}, \ldots, \widetilde{X^d})$ with $\widetilde{X^i} = \mathcal{X}^\rangle/\mathcal{X}'$.

We see from Lemma 19.5 that the normalized stock price processes satisfy the equations

$$\begin{aligned} \widetilde{\mathcal{X}}^0 &= 1, \\ \widetilde{\mathcal{X}}^i &= X^i_0 + (\mu^i - \mathcal{R})\widetilde{\mathcal{X}}^i \cdot \lambda + \widetilde{\mathcal{X}}^i \sum_{k=1}^n \sigma^{ik} \cdot \mathcal{W}^k, \qquad i = 1, \ldots, d \end{aligned} \tag{19.12}$$

.

Definition 19.19 (Equivalent martingale measures)**.** *An equivalent martingale measure (to* $\mathbf{P}$*) is a probability measure* $\mathbf{Q}$ *that (i) is equivalent to* $\mathbf{P}$ *(i.e., has the same null sets) and (ii) with respect to which all normalized processes* $\widetilde{X}^i$ *are martingales.*

Proposition 19.20. *Is there a process* $\underline{\mathcal{H}} = (\mathcal{H}^1, ..., \mathcal{H}^n)$ *such that*

$$\sum_{k=1}^{n} \sigma^{ik} \mathcal{H}^k = \mu^i - \mathcal{R}, \tag{19.13}$$

and

$$\mathcal{Z} = \exp\Big(- \sum_{k=1}^{n} \mathcal{H}^k \cdot \mathcal{W}^k - \tfrac{1}{2} (\mathcal{H}^k)^2 \cdot \lambda \Big)$$

is a martingale, then

$$\mathbf{Q} = Z_\tau \cdot \mathbf{P}$$

is an equivalent martingale measure with

$$\widetilde{\mathcal{X}}^i = X_0^i + \widetilde{\mathcal{X}}^i \sum_{k=1}^{n} \sigma^{ik} \cdot \widetilde{\mathcal{W}}^k,$$

where $\widetilde{\mathcal{W}}^k = \mathcal{W}^k + \mathcal{H}^k \cdot \lambda, k = 1, ..., n$ *are Brownian motions with respect to* $\mathbf{Q}$ *.*

Proof. All statements follow analogously to Theorem 18.11 with $\gamma = -\mu + \mathcal{R}$. □

The equations (19.13) are also called risk equations of the market. We now establish a connection between arbitrage and the existence of an equivalent martingale measure. For the definition of arbitrage see definition 19.6.

Theorem 19.21 (First Fundamental Theorem of Investment Valuation)**.** *If a financial market of the form* (19.11) *has an equivalent martingale measure, then it is arbitrage-free.*

Proof. The proof is almost analogous to the proof of the arbitrage-freeness of the Black-Scholes market from Theorem 19.11:

Let $\underline{\Delta}$ be a portfolio with $V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0$ and $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ almost surely. Since $\underline{\Delta}$ is also self-financing according to Lemma 19.5 with respect to $\widetilde{\mathcal{X}}$, it holds that

$$\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = V_0(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) + \sum_{i=0}^{d} \Delta^i \cdot \widetilde{\mathcal{X}}^i.$$

Since, according to the assumption, the processes $\widetilde{\mathcal{X}}^i, i = 0, ..., d$ are martingales with respect to $\mathbf{Q}$, $\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})$ is also a $\mathbf{Q}$-martingale. By assumption, $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ is almost surely P-sure. Since $\mathbf{P}$ and $\mathbf{Q}$ are equivalent, $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) \geq 0$ is also $\mathbf{Q}$-almost surely. Due to the martingale property of $\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})$, we also have $\mathbf{E_Q}[V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}})] = 0$. However, this is only possible if $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = 0$ $\mathbf{Q}$-almost surely. Again, due to the equivalence of $\mathbf{P}$ and $\mathbf{Q}$, $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = 0$ $\mathbf{P}$-almost surely, and the arbitrage-free property is established. □

Example 19.22 (A non-arbitrage-free market)**.** *We give a simple example in which no equivalent martingale measure exists, but arbitrage opportunities do. For this purpose, let* $i = 2$ *and* $k = 1$*, i.e. two stocks are driven by the same Brownian motion. We assume that* $\mathcal{R} =: r, \mu^i, \sigma^{i1} =: \sigma^i$ *are constant and* $X^1(0) = X^2(0) = 1$*. Then the risk equations of the market are given by*

$$\sigma^1 \mathcal{H} = \mu^1 - r,$$
$$\sigma^2 \mathcal{H} = \mu^2 - r.$$

This has a unique solution if and only if

$$\frac{\mu^1 - r}{\sigma^1} = \frac{\mu^2 - r}{\sigma^2}.$$

Let us assume that $a := \frac{\mu^1 - r}{\sigma^1} - \frac{\mu^2 - r}{\sigma^2} > 0$*. We set*

$$\Delta^1 = \frac{1}{\sigma^1 \mathcal{X}^1}, \qquad \Delta^2 = -\frac{1}{\sigma^2 \mathcal{X}^2}.$$

Furthermore, let Δ^0 *be chosen such that the portfolio is self-financing with* $V_0(\underline{\Delta}, \underline{\mathcal{X}}) = 0$*. Then we have*

$$\begin{aligned}
\Delta^0 &= \frac{1}{\sigma^2} - \frac{1}{\sigma^1}, \\
\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) &= \frac{1}{\sigma^1 \mathcal{X}^1} \cdot \mathcal{X}^1 - \frac{1}{\sigma^2 \mathcal{X}^2} \cdot \mathcal{X}^2 + r\Big(\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) - \frac{1}{\sigma^1} + \frac{1}{\sigma^2}\Big) \cdot \lambda \\
&= \frac{\mu^1 - r}{\sigma^1} \cdot \lambda - \frac{\mu^2 - r}{\sigma^2} \cdot \lambda + r\mathcal{V}(\underline{\Delta}, \underline{\mathcal{X}}) \cdot \lambda + \mathcal{W} - \mathcal{W}
\end{aligned}$$

and (with (19.5)*)*

$$\begin{aligned}
\mathcal{V}(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) &= V_0 + \frac{1}{\sigma^1 \mathcal{X}^1} \cdot \widetilde{\mathcal{X}}^1 - \frac{1}{\sigma^2 \mathcal{X}^2} \cdot \widetilde{\mathcal{X}}^2 \\
&= \Big(\frac{\mu^1 - r}{\sigma^1} - \frac{\mu^2 - r}{\sigma^2}\Big) \frac{1}{\mathcal{X}^0} \cdot \lambda = \frac{a}{\mathcal{X}^0} \cdot \lambda > 0
\end{aligned}$$

for $t > 0$*. This means that* $\underline{\Delta}$ *is a self-financing portfolio with* $V_0(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) = 0$ *and* $V_\tau(\underline{\Delta}, \underline{\widetilde{\mathcal{X}}}) > 0$ *almost surely, hence the market is not arbitrage-free.*

To price derivatives, it is important to know hedging portfolios for. We now come to conditions under which such portfolios exist.

Definition 19.23 (Complete Market)**.** *A market* (19.11) *is called* complete *if every derivative can be hedged, i.e. for every quadratically integrable random variable* S*, there is a self-financing strategy with a fixed starting value* $V_0(\underline{\Delta}, \underline{\mathcal{X}})$ *and* $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) = S$*.*

Theorem 19.24 (Second Fundamental Theorem of Investment Valuation)**.** *If a financial market of the form* (19.5) *has an equivalent martingale measure, then it is complete if and only if the martingale measure is unique.*

Before we begin with the proof, we introduce an elementary lemma from linear algebra.

Lemma 19.25. *Let* $\underline{\underline{A}}$ *be a* $d \times n$*-matrix,* $\underline{b}$ *a* d*-dimensional vector, such that*

$$\underline{\underline{A}}\,\underline{x}^\top = \underline{b}^\top$$

has a unique solution $\underline{x} \in \mathbb{R}^n$*. Then*

$$\underline{y}\,\underline{\underline{A}} = \underline{c}$$

has at least one solution $\underline{y} \in \mathbb{R}^d$ *for each* n*-dimensional vector* $\underline{c}$ *.*

Proof. We interpret $\underline{\underline{A}}$ as a map $\underline{\underline{A}} : \mathbb{R}^n \to \mathbb{R}^d$. Since the solution of $\underline{\underline{A}} \cdot \underline{x}^\top = \underline{b}^\top$ is unique, then $\ker(A) = \underline{0}$ (because with $\underline{x}$, $\underline{x} + \underline{x}_0$ is also a solution for $\underline{x}_0 \in \ker(A)$). Thus $n = \dim(\ker(A)) + \dim(\operatorname{im}(A)) = \dim(\operatorname{im}(A)) \leq d$, also $\operatorname{rg}(A) = \dim(\operatorname{im}(A)) = n$. We write the equation $\underline{y}\,\underline{\underline{A}} = \underline{c}$ as $\underline{\underline{A}}^\top\,\underline{y}^\top = \underline{c}^\top$. Since $\operatorname{rg}(A^\top) = \operatorname{rg}(A) = n$, every $\underline{c} \in \mathbb{R}^n$ lies in $\operatorname{im}(A^\top)$, in particular $\underline{\underline{A}}^\top\,\underline{y}^\top = \underline{c}^\top$ has at least one solution. □

Proof of Theorem 19.24. Assume that the market is complete and that $\mathbf{Q}$ and $\mathbf{R}$ are two equivalent martingale measures. The aim is to show that $\mathbf{Q} = \mathbf{R}$. For this, let $A \in \mathcal{F}_\tau = \mathcal{F}$ and $S = 1_A X^0_\tau$. Since the market is complete, S can be hedged, so there is a self-financing portfolio $\underline{\Delta}$ with $V_\tau(\underline{\Delta}, \underline{\mathcal{X}}) = S$, or also $V_\tau(\underline{\Delta}, \underline{\widetilde{X}^i} = 1_A$. Since $\widetilde{X}^i$ is both a $\mathbf{Q}$- and a $\mathbf{R}$-martingale, $i = 0, ..., d$, we have

$$\mathbf{E_Q}[1_A] = \mathbf{E_Q}[V_\tau(\underline{\Delta}, \underline{\widetilde{X}})] = \mathbf{E_Q}\Big[\sum_{i=0}^{d} \Delta^i_0 X^i_0 + (\Delta^i \cdot \widetilde{\mathcal{X}}^i)_\tau\Big] = \sum_{i=0}^{d} \Delta^i_0 X^i_0,$$

and analogously, $\mathbf{E_Q}[1_A] = \sum_{i=0}^d \Delta^i_0 X^i_0$. In particular, $\mathbf{Q}(A) = \mathbf{R}(A)$, and since A was arbitrary, it follows that $\mathbf{Q} = \mathbf{R}$.

Conversely, let $\mathbf{Q}$ be the only equivalent martingale. Then the risk equations of the market, (19.13), have exactly one solution $\mathcal{H}^1, ..., \mathcal{H}^n$. (Because: First, $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$ for a random variable $Z_\tau > 0$ according to the Radon-Nikodym theorem. $\mathcal{Z} = (Z_t)_{0 \leq t \leq \tau}$ with $Z_t := \mathbf{E_P}[Z_\tau | \mathcal{F}_t]$ is a positive $\mathbf{P}$-martingale, hence Theorem 17.14 implies that $\mathcal{Z} = \exp(\mathcal{N} - \frac{1}{2}[\mathcal{N}])$ for a (unique) local P-martingale $\mathcal{N}$, which, according to the martingale representation theorem 17.9, can be written as an integral of processes $\mathcal{H}^1, ..., \mathcal{H}^n$ with respect to the Brownian movements $\mathcal{W}^1, ..., \mathcal{W}^n$. Now, in order for $\widetilde{X}^i$ are $\mathbf{Q}$-martingales, the equations (19.13) must necessarily be fulfilled. Since this argumentation can be applied to any equivalent martingale measure, but this is unique according to the assumption, (19.13) has exactly one solution.) We set $\underline{\underline{\sigma}} := (\sigma^{ik})_{i=1,...,d;k=1,...,n}$, $\underline{\mu} := (\mu^i)_{i=1,...,d}$, $\underline{\mathcal{H}} := (\mathcal{H}^k)_{k=1,...,n}$. Then we can write these equations as a linear system of equations

$$\underline{\underline{\sigma}}\,\underline{\mathcal{H}}^\top = \underline{\mu}^\top - \mathcal{R}^\top \tag{19.14}$$

(with solution $\underline{\mathcal{H}}$). This equation has (for almost all ω and all t) thus exactly one solution.

We have to show that every derivative, given by a quadratically integrable random variable S, can be hedged. First, we note that

$$X^0_t = \exp\Big(\int_0^t R_s ds\Big)$$

must apply. We define the process $\mathcal{V} = (V_t)_{0\leq t\leq \tau}$ with

$$V_t := E_{\mathbf{Q}}[e^{-\int_t^\tau R_s ds} S|\mathcal{F}_t],$$

such that $\mathcal{V}/\mathcal{X}^0$ is a

$$V_t/X_t^0 = E_{\mathbf{Q}}[e^{-\int_0^\tau R_s ds} S|\mathcal{F}_t]$$

for all $0 \leq t \leq \tau$ is a $\mathbf{Q}$-martingale. With $\widetilde{W}^k = \mathcal{W}^k - \mathcal{H}^k \cdot \lambda$ and the martingale representation theorem 17.9, there are processes $\mathcal{L}^1, ..., \mathcal{L}^n$ with

$$\mathcal{V}/\mathcal{X}^0 = V_0 + \sum_{k=1}^n \mathcal{L}^k \cdot \widetilde{\mathcal{W}}^k.$$

To hedge S, it is sufficient to find a self-financing portfolio $\underline{\Delta}$ with $V_t(\underline{\Delta}, \underline{\mathcal{X}}) = V_t/X_t^0$. This implies that $\widetilde{\mathcal{X}}^i = X_0^i + \widetilde{\mathcal{X}}^i \sum_k \sigma^{ik} \cdot \widetilde{\mathcal{W}}^k$

$$V_0 + \sum_{k=1}^n \int_0^t L_s^k d\widetilde{W}_s^k = V_t/X_t^0 = V_0 + \sum_{k=1}^n \int_0^t \Delta_s^i \widetilde{X}_s^i \sigma_s^{ik} d\widetilde{W}_s^k.$$

So to hedge S, we need to find $\underline{\Delta}$ such that

$$\mathcal{L}^k = \Delta^i \widetilde{\mathcal{X}}^i \sigma^{ik},$$

or equivalently

$$\underline{\mathcal{L}} = \underline{\Delta\widetilde{\mathcal{X}}}\,\underline{\underline{\sigma}}$$

with $\underline{\Delta\widetilde{\mathcal{X}}} = (\Delta^i \widetilde{\mathcal{X}}^i)_{i=1,...d}$ and $\underline{\mathcal{L}} = (\mathcal{L}^k)_{k=1,...,n}$. Since (19.14) has exactly one solution by assumption, this system of equations has at least one solution according to Lemma 19.25 $\underline{\Delta}$. As just shown, this $\underline{\Delta}$ provides a hedging portfolio for S. □

20 Interest Rate Models

In the Black-Scholes model, the value of the risk-free security, $\mathcal{X}^0$, is deterministic. We will now abandon this condition, but instead consider no further securities. The market is thus determined by the equation

$$\mathcal{X}^0 = 1 + \mathcal{R} \cdot \mathcal{X}^0 \cdot \lambda, \tag{20.1}$$

whereby we always assume that the filtration to which $\mathcal{R} = (R_t)_{0\leq t\leq \tau}$ and thus $\mathcal{X}^0 = (X_t^0)_{0\leq t\leq \tau}$ are adapted, is generated by a Brownian motion $\mathcal{W}$ is generated. It is clear that

$$X_t^0 = \exp\Big(\int_0^t R_s ds\Big).$$

The aim of this section is to calculate the fair price of zero bonds. These are derivatives that are issued at a point in time $0 \leq t \leq \tau$ and whose owners receive a monetary unit at a point in time $t \leq u \leq \tau$ receive one monetary unit. The question is, therefore, what it is worth at time t to be sure of having one monetary unit at time u, given that the interest process is given by $\mathcal{R}$. We denote the price of the zero bond with maturity u at time t by S_t^u, which leads to a process $(S_t^u)_{0\leq t\leq u}$ with $S_u^u = 1$.

20.1 Basic

Again, we will use equivalent martingale measures to calculate the fair price of zero bonds; see Proposition 20.1. Subsequently, we can describe the price processes $(S_t^u)_{0\leq t\leq u}$ by means of a stochastic differential equation; see Proposition 20.4.

Proposition 20.1 (Price of a zero bond)**.** *If there is an equivalent martingale measure* $\mathbf{Q}$ *in the market* (20.1)*, (i.e., a measure with respect to which* $(S_t^u/X_t^0)_{0\leq t\leq u}$ *is a martingale for all* $0\leq u\leq\tau$*, then*

$$S_t^u = \mathbf{E_Q}\Big[\exp\Big(-\int_t^u R_s ds\Big)\Big|\mathcal{F}_t\Big]$$

is the fair price of the zero bond with maturity u *at time* t*.*

Proof. It is clear that $S_u^u/X_u^0 = \exp\Big(-\int_0^u R_s ds\Big)$. Since $(S_t^u/X_t^0)_{0\leq t\leq u}$ is a $\mathbf{Q}$-martingale, it follows that

$$S_t^u/X_t^0 = \mathbf{E_Q}[S_u^u/X_u^0|\mathcal{F}_t] = \mathbf{E_Q}\Big[\exp\Big(-\int_0^u R_s ds\Big)\Big|\mathcal{F}_t\Big],$$

hence the assertion follows after multiplication by X_t^0. □

Since the filtration $(\mathcal{F}_t)_{0\leq t\leq\tau}$ is generated by the Brownian motion, we can describe equivalent martingale measures in the market (20.1) well.

Proposition 20.2 (Characterization of equivalent measures)**.** *Let* $\mathbf{Q}$ *be equivalent to* $\mathbf{P}$*. Then there exists a random variable* $Z_\tau > 0$ *and a process* $\mathcal{H} = (H_t)_{0\leq t\leq\tau}$ *such that* $\mathbf{Q} = Z_\tau\cdot\mathbf{P}$ *and*

$$Z_\tau = \exp\Big(-\int_0^\tau H_s dW_s - \tfrac{1}{2}\int_0^\tau H_s^2 ds\Big). \tag{20.2}$$

Proof. According to the Radon-Nikodym theorem, there exists a density Z_τ of $\mathbf{Q}$ with respect to $\mathbf{P}$ such that $\mathbf{Q} = Z_\tau\cdot\mathbf{P}$. It is clear that $\mathcal{Z} = (Z_t)_{0\leq t\leq\tau}$ with $Z_t = \mathbf{E_P}[Z_\tau|\mathcal{F}_t]$ is a P-martingale. By Theorem 17.14.2, there exists a local $\mathbf{P}$-martingale $\mathcal{N}$ such that $\mathcal{Z} = \exp(\mathcal{N} - \frac{1}{2}[\ Con])$. Since the filtration $(\mathcal{F}_t)_{0\leq t\leq\tau}$ is generated by $\mathcal{W}$, theorem 17.9 implies that there is a process $\mathcal{H}$ with $\mathcal{N} = -\mathcal{H}\cdot\mathcal{W}$. Hence the claim follows. □

Corollary 20.3 (Preis eines Zero-Bonds)**.** *Is there an equivalent martingale measure in the market* (20.1) $\mathbf{Q}$ *with* $\mathbf{Q} = Z_\tau\cdot\mathbf{P}$ *for*

$$Z_\tau = \exp\Big(-\int_0^\tau H_s dW_s - \tfrac{1}{2}\int_0^\tau H_s^2 ds\Big).$$

and a process $\mathcal{H} = (H_t)_{0\leq t\leq\tau}$*, then*

$$S_t^u = \mathbf{E_P}\Big[\exp\Big(-\int_t^u R_s ds - \int_t^u H_s dW_s - \frac{1}{2}\int_t^u H_s^2 ds\Big)\Big|\mathcal{F}_t\Big]$$

is the fair price of the zero bond with maturity u *at time* t*.*

Proof. It suffices to show the statement for $u=\tau$. It follows immediately from the last two propositions, if one notes that for any non-negative random variable X for $0\leq t\leq\tau$ and $A\in\mathcal{F}_t$

$$\mathbf{E_P}[Z_t\mathbf{E_Q}[X|\mathcal{F}_t], A] = \mathbf{E_Q}[\mathbf{E_Q}[X|\mathcal{F}_t], A] = \mathbf{E_Q}[X, A] = \mathbf{E_P}[Z_\tau X, A] = \mathbf{E_P}[\mathbf{E_P}[Z_\tau X|\mathcal{F}_t], A],$$

which means that

$$\mathbf{E}_{\mathbf{Q}}[X|\mathcal{F}_t] = \frac{\mathbf{E}_{\mathbf{P}}[Z_\tau X|\mathcal{F}_t]}{Z_t}$$

holds. □

Proposition 20.4 (An SDE for the price process of zero bonds)**.** *Is there an equivalent martingale measure in the market* (20.1) $\mathbf{Q}$ *with* $\mathbf{Q} = Z_\tau \cdot \mathbf{P}$ *for*

$$Z_\tau = \exp\Big(-\int_0^\tau H_s dW_s - \tfrac{1}{2}\int_0^\tau H_s^2 ds\Big).$$

and a process $\mathcal{H} = (H_t)_{0\le t\le\tau}$*, then there is for each* $0 \le u \le \tau$ *an adapted stochastic process* $\mathcal{K}^u = (K_t^u)_{0\le t\le u}$ *such that*

$$dS_t^u = (R_t + K_t^u H_t)S_t^u dt + K_t^u S_t^u dW.$$

Remark 20.5 (Interpretation)**.** *We immediately notice that although the process* $\mathcal{R}$ *has finite variation, the processes* $(S_t^u)_{0\le t\le u}$ *do not. Furthermore, we observe that the expected interest rate, which can be calculated using a zero bond, differs from the process* $\mathcal{R}$ *differs. If* $\mathcal{H}$ *is positive (which is the normal case), then the expected interest rate is higher than for* $\mathcal{X}^0$*. However, this higher interest rate is associated with a higher risk due to a positive quadratic variation in the process* $(S_t^u)_{0\le t\le u}$*.*

Proof of Proposition 20.4. Since the process $(S_t^u/X_t^0)_{t\ge0}$ is a positive $\mathbf{Q}$-martingale, by lemma 17.12 $(Z_t S_t^u/X_t^0)_{t\ge0}$ is a positive $\mathbf{P}$-martingale. Thus, by Theorems 17.14 and Theorem 17.9, there is an adapted process $\mathcal{L}^u = (L_t^u)_{0\le t\le u}$ with

$$Z_t S_t^u/X_t^0 = S_0^u \exp\Big(\int_0^t L_s^u dW_s - \tfrac{1}{2}\int_0^t (L_s^u)^2 ds\Big),$$

or also

$$S_t^u = S_0^u \exp\Big(\int_0^t L_s^u + H_s dW_s + \int_0^t R_s - \tfrac{1}{2}((L_s^u)^2 - H_s^2) ds\Big).$$

Just as in example 18.1, we see that $(S_t^u)_{0\le t\le u}$ is the SDE

$$\begin{aligned} dS_t^u &= \big(R_t - \tfrac{1}{2}((L_t^u)^2 - (H_t)^2) + \tfrac{1}{2}(L_t^u + H_t)^2\big)S_t^u ds + (L_t^u + H_t)S_t^u dW \\ &= \big(R_t + L_t^u H_t + H_t^2\big)S_t^u ds + (L_t^u + H_t)S_t^u dW \end{aligned}$$

löst. Nun folgt die Aussage mit $K_t^u = L_t^u + H_t$. □

20.2 The Vasicek model

As we saw in Section 20.1, in particular in Proposition 20.1 and Corollary 20.3, we can calculate the price S_t^u of a zero bond if we know the price process $\mathcal{X}^0$ under the equivalent martingale measure $\mathbf{Q}$, or both $\mathcal{X}^0$ and $\mathcal{H}$ under $\mathbf{P}$. The (single-factor) Vasicek model is a simple model in which we describe $\mathcal{R}$ by an Ornstein-Uhlenbeck process. We will now briefly repeat this process .

Example 20.6 (Ornstein-Uhlenbeck Process). *We consider the solution of the SDE*

$$dX = -\mu X dt + \sigma dW$$

and claim that

$$X_t = e^{-\mu t}\Big(X_0 + \sigma \int_0^t e^{\mu s} dW_s\Big) \tag{20.3}$$

strong solution of this SDE. According to the formula for partial integration, the following applies to the process defined in this way and $\mathcal{A} = (A_t)_{t\geq 0}$ *with* $A_t = e^{-\mu t}$ *(and thus* $\mathcal{A} = 1 - \mu\mathcal{A}\cdot\lambda$*) and* $\mathcal{B} = (B_t)_{t\geq 0}$ *with* $B_t = X_0 + \sigma\int_0^t e^{\mu s} dW_s$*, i.e.* $\mathcal{B} = X_0 + \sigma\mathcal{A}^{-1}\cdot\mathcal{W}$

$$\mathcal{X} - X_0 = \mathcal{A}\cdot\mathcal{B} + \mathcal{B}\cdot\mathcal{A} = \sigma\cdot\mathcal{W} - \mu\mathcal{A}\mathcal{B}\cdot\lambda = -\mu\mathcal{X}\cdot\lambda + \sigma\cdot\mathcal{W}.$$

Furthermore, we assume that

$$Y_t = b + e^{-\mu t}\Big(Y_0 - b + \sigma\int_0^t e^{\mu s} dW_s\Big) \tag{20.4}$$

is a strong solution of the SDE

$$dY = \mu(b - Y)dt + \sigma dW \tag{20.5}$$

. *(Because if* $\mathcal{X}$ *solves the equation* (20.3)*, then* $\mathcal{Y} := \mathcal{X} + b$ *solves the equation* (20.4)*.)*

In the Vasicek model, the process $\mathcal{R}$ from (20.1) is modeled by the solution of (20.4). Furthermore, it is assumed that the equivalent martingale measure from Proposition (20.4) has the form $\mathbf{Q} = Z_\tau\cdot\mathbf{P}$ with Z_τ from (20.2) and $\mathcal{H} = \rho$ for some $\rho\in\mathbb{R}$. In this fairway, we state some elementary results.

Lemma 20.7 (Modeling in the Vasicek model). *Let* $\mathcal{Y} = Y_{0\leq t\leq\tau}$ *be the solution of* (20.5) *and* $\mathbf{Q} = Z_\tau\cdot\mathbf{P}$ *with* $Z_\tau = \exp(-\rho W_\tau - \frac{1}{2}\rho^2 t)$*. Then* $\widetilde{W} = W + \rho t$ *is a Brownian motion with respect to* $\mathbf{Q}$ *and* $\mathcal{Y}$ *is the only strong solution of*

$$dY = \mu(b^* - Y_t)dt + \sigma\widetilde{W}_t$$

with $b^* = b - \rho\sigma/\mu$.

Proof. This follows directly from the Girsanov transformation of Theorem 18.11. □

Theorem 20.8 (Prices of Zero-Bonds in the Vasicek Model). *We use the notation from Proposition 20.1. Let* $\mathcal{R} = R_{0\leq t\leq\tau}$ *the solution of* (20.5) *and* $\mathbf{Q} = Z_\tau\cdot\mathbf{P}$ *with* $Z_\tau = \exp(-\rho W_\tau - \frac{1}{2}\rho^2 t)$*. Then*

$$S_t^u = F(R_t, u - t)$$

with $b^* = b - \rho\sigma/\mu$ *and*

$$F(x,t) = \exp\Big(-\Big(b^* - \frac{\sigma^2}{2\mu^2}\Big)t + \frac{1}{\mu}\Big(b^* - \frac{\sigma^2}{2\mu^2} - x\Big)(1 - e^{-\mu t}) - \frac{\sigma^2}{4\mu^3}(1 - e^{-\mu t})^2\Big)$$

the fair price of the zero bond with maturity u *at time* t.

Remark 20.9 (The long-term achievable interest rate)**.** *For large t, approximately*

$$F(x,t) \approx C \exp\Big(-\Big(b^* - \frac{\sigma^2}{2\mu^2}\Big)t\Big)$$

for a constant C. This means that $b^* - \sigma^2/(2\mu^2)$ is the long-term interest rate in the Vasicek model. In practice, however, it is assumed that the long-term interest rate depends on the current interest rate, R_t.

Proof of Theorem 20.8. Due to Proposition 20.1, only $\mathbf{E_Q}[\exp(-\int_t^u R_s ds)|\mathcal{F}_t]$ needs to be calculated. Since $\mathcal{R}$ solves the SDE

$$dR = \mu(b^* - R_t)dt + \sigma d\widetilde{W} \tag{20.6}$$

with the $\mathbf{Q}$-Brownian motion $\widetilde{W}$ and $b^* = b - \rho\sigma/\mu$, and is therefore a Markov process we write

$$\mathbf{E_Q}\Big[\exp\Big(-\int_t^u R_s ds\Big)\Big|\mathcal{F}_t\Big] = F(X_t, u-t)$$

with

$$F(x,t) = \mathbf{E}\Big[\exp\Big(-\int_0^t R_s ds\Big)\Big|R_0 = x\Big], \tag{20.7}$$

where $\mathcal{R} = (R_t)_{0\le t\le\tau}$ solves the SDE (20.6). We can use the fact that

$$R_t = b^* + e^{-\mu t}\Big(x - b^* + \sigma\int_0^t e^{\mu s} d\widetilde{W}_s\Big).$$

Since R_s is normally distributed in (20.7) for each s, $\int_0^t R_s ds$ is also normally distributed. We calculate the expectation and variance as

$$\begin{aligned}
\widehat{\mu} := \mathbf{E}\Big[\int_0^t R_s ds\Big] &= \int_0^t b^* + e^{-\mu s}(x - b^*)ds = b^*t + \frac{x-b^*}{\mu}(1 - e^{-\mu t}),\\
\widehat{\sigma}^2 := \mathbf{V}\Big[\int_0^t R_s ds\Big] &= 2\int_0^t\int_0^s \mathbf{COV}[R_s, R_r]drds\\
&= 2\sigma^2\int_0^t\int_0^s e^{-\mu r}e^{-\mu s}\mathbf{E}\Big[\int_0^s e^{\mu s}d\widetilde{W}_s\int_0^r e^{\mu r}d\widetilde{W}_r\Big]drds\\
&= 2\sigma^2\int_0^t\int_0^s\int_0^r e^{-\mu r}e^{-\mu s}e^{2\mu v}dvdsdr\\
&= \frac{\sigma^2}{\mu}\int_0^t e^{-\mu s}\int_0^s e^{\mu r} - e^{-\mu r}drds\\
&= \frac{\sigma^2}{\mu^2}\int_0^t 1 - 2e^{-\mu s} + e^{-2\mu s}ds\\
&= \frac{\sigma^2 t}{\mu^2} - \frac{2\sigma^2}{\mu^3}(1 - e^{-\mu t}) + \frac{\sigma^2}{2\mu^3}(1 - e^{-2\mu t})\\
&= \frac{\sigma^2 t}{\mu^2} - \frac{\sigma^2}{\mu^3}(1 - e^{-\mu t}) - \frac{\sigma^2}{2\mu^3}(1 - e^{-\mu t})^2.
\end{aligned}$$

Since by Example 6.13.3 for a $N(\widehat{\mu},\ building\sigma^2)$-distributed random variable Z it holds that $\mathbf{E}[e^{-Z}] = e^{-\ building\mu + \ building\sigma^2/2}$, follows

$$\begin{aligned} F(x,t) &= e^{-\hat{\mu}+\hat{\sigma}^2/2} \\ &= \exp\Big(- \Big(b^* - \frac{\sigma^2}{2\mu^2}\Big)t + \frac{1}{\mu}\Big(b^* - \frac{\sigma^2}{2\mu^2} - x\Big)(1-e^{-\mu t}) - \frac{\sigma^2}{4\mu^3}(1-e^{-\mu t})^2\Big) \end{aligned}$$

and thus the assertion. □

Part V

Appendix

A Some Topology

A topology is used in mathematics whenever a notion of convergence is introduced. Even if topologies have only been treated as a sideline in most of your lectures so far, some concepts of convergence are well known. There are also many connections between measure theory and topology; see for example the notion of a Borel σ-algebra in Definition 1.7. Therefore, we repeat basic notions of topology here.

A.1 Basics

By a *topology* we understand a family of open subsets of a space Ω.[22] In metric spaces one calls a set A *open* if for every $\omega \in A$ there is an open ball[23] $B_\varepsilon(\omega) \subseteq A$ for some $\varepsilon > 0$. This case of metric spaces is in practice the most important.

In measure theory, the case of separable topologies, which are generated by complete metrics, is of particular importance. Such spaces are called *Polish*.

Definition A.1 (Metric space, topological space)**.** *Let Ω be some set.*

1. *A function $r : \Omega \times \Omega \to \mathbb{R}_+$ is called a* metric *if (i) $r(\omega,\omega') \neq 0$ for $\omega \neq \omega'$, (ii) $r(\omega,\omega') = r(\omega',\omega)$ for all $\omega,\omega' \in \Omega$, and (iii) $r(\omega,\omega'') \leq r(\omega,\omega') + r(\omega',\omega'')$ for all $\omega,\omega',\omega'' \in \Omega$. The pair (Ω, r) is a* metric space.

   For $\omega \in \Omega$ and $\varepsilon > 0$, we denote by $B_\varepsilon(\omega) := \{\omega' \in \Omega : r(\omega,\omega') < \varepsilon\}$ the open ball *around ω with radius ε.*

2. *A metric r on Ω is called* complete *if every Cauchy sequence converges. That is, if $\omega_1, \omega_2, \ldots \in \Omega$ with*

   $$\forall \varepsilon > 0\ \exists N \in \mathbb{N}\ \forall n,m \geq N : \ \ r(\omega_n,\omega_m) < \varepsilon,$$

   then there is $\omega \in \Omega$ with $r(\omega_n,\omega) \xrightarrow{n\to\infty} 0$.

3. *A set system $\mathcal{O} \subseteq 2^\Omega$ is called* topology *if (i) $\emptyset, \Omega \in \mathcal{O}$; (ii) if $A, B \in \mathcal{O}$, then $A \cap B \in \mathcal{O}$; (iii) if I is arbitrary and if $A_i \in \mathcal{O}, i \in I$, then $\bigcup_{i\in I} A_i \in \mathcal{O}$. The pair $(\Omega, \mathcal{O})$ is called* topological space. *Its members, i.e. every $A \in \mathcal{O}$, is called* open*; any set $A \subseteq \Omega$ with $A^c \in \mathcal{O}$ is called* closed.

[22] We will write 2^Ω for the set of all subsets of Ω.

[23] We define $B_\varepsilon(\omega) := \omega' : r(\omega,\omega') < \varepsilon\}$.

4. *Let* $(\Omega, \mathcal{O})$ *be a topological space and* $A \subseteq \Omega$. *Then*

$$A^\circ := \bigcup\{O \subseteq A : O \in \mathcal{O}\}$$

is called the interior of A *and*

$$\overline{A} := \bigcap\{F \supseteq A : F^c \in \mathcal{O}\}$$

is called closure of A.

5. *A topological space* $(\Omega, \mathcal{O})$ *is called* separable *if there is a countable set* $\Omega' \subseteq \Omega$ *with* $\overline{\Omega}' = \Omega$.

6. *Let* $(\Omega, \mathcal{O})$ *be a topological space and* $\mathcal{B} \subseteq \mathcal{O}$. *Then* $\mathcal{B}$ *is called a* base *of* $\mathcal{O}$ *if*

$$\forall A \in \mathcal{O} \ \ \forall \omega \in A \ \ \exists B \in \mathcal{B} : \omega \in B \subseteq A.$$

This is exactly the case if

$$\mathcal{O} = \{A \subseteq \Omega : \ \ \forall \omega \in A \, \exists B \in \mathcal{B} : \omega \in B \subseteq A\}. \tag{A.1}$$

or (equivalently)

$$\mathcal{O} = \Big\{ \bigcup_{B \in \mathcal{C}} B : \mathcal{C} \subseteq \mathcal{B} \Big\}. \tag{A.2}$$

7. *Let* $\mathcal{B} \subseteq 2^\Omega$. *Then, the right hand sides of* (A.1) *and* (A.2) *define the topology* generated by $\mathcal{B}$, *which we denote by* $\mathcal{O}(\mathcal{B})$.

8. *Let* (Ω, r) *be a metric space and*

$$\mathcal{B} := \{B_\varepsilon(\omega) : \varepsilon > 0, \omega \in \Omega\}. \tag{A.3}$$

Then $\mathcal{O}(\mathcal{B})$ *is the* topology generated by r. *If specifically* $\Omega \subseteq \mathbb{R}^d$ *and* r *is the Euclidean distance, then the topology generated in* (A.1) *or* (A.2) *is called the* euclidean topology.

9. *The space* $(\Omega, \mathcal{O})$ *is called (completely) metrizable if there exists a (complete) metric* r *on* Ω *such that* (A.1) *holds with* $\mathcal{B}$ *from* (A.3). *The space* $(\Omega, \mathcal{O})$ *is called* Polish *if it is separable and completely metrizable.*

10. *Let* $(\Omega, \mathcal{O})$ *and* $(\Omega', \mathcal{O}')$ *be topological spaces. Then a mapping* $f : \Omega \to \Omega'$ *is called* continuous *if* $f^{-1}(A') \in \mathcal{O}$ *for all* $A' \in \mathcal{O}'$.

Example A.2 (The space $\overline{\mathbb{R}}$)**.** *We will often use functions with values in*[24]

$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\} \quad \text{or} \quad \overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$$

[24] The notation $\overline{\mathbb{R}}$ suggests that the termination of $\mathbb{R}$ is meant here. This is not true, since the added elements $-\infty, \infty$ do not lie in $\mathbb{R}$, but closures of sets always contain at most the elements of the basic space can contain. Topologically, $\overline{\mathbb{R}}$ is the two-point compactification of $\mathbb{R}$

In order to be able to consider these spaces as topological spaces, we set

$$\varphi : \begin{cases} \overline{\mathbb{R}} & \to [-1,1], \\ x & \mapsto \begin{cases} \frac{2}{\pi}\arctan(x), & x \in \mathbb{R}, \\ 1, & x = \infty, \\ -1, & x = -\infty \end{cases} \end{cases}$$

and define the metric

$$r_{\overline{\mathbb{R}}}(x,y) := |\varphi(x) - \varphi(y)|, \qquad x, y \in \overline{\mathbb{R}}.$$

The topological space defined by $r_{\overline{\mathbb{R}}}(\overline{\mathbb{R}}, \overline{\mathcal{O}})$ *extends the Euclidean topology* $(\mathbb{R}, \mathcal{O})$ *to* $\mathbb{R}$ *in the sense that* $\{A \cap \mathbb{R} : A \in \overline{\mathcal{O}}\} = \mathcal{O}$*. This is true because* φ *is continuous on* $\mathbb{R}$ *with a continuous inverse function. It further holds that* $(\overline{\mathbb{R}}, \overline{\mathcal{O}})$ *is separable and* $r_{\overline{\mathbb{R}}}$ *is a complete metric.*

On $\overline{\mathbb{R}}$ *one can calculate as usual in calculus. For example,* $a \cdot \infty = \infty$ *for* $a > 0$*. However, expressions like* $\infty - \infty$ *and* ∞/∞ *are not defined.*

Remark A.3 (Metric and topological spaces)**.** *Let* $(\Omega, \mathcal{O})$ *be a topological space and* $\omega, \omega_1, \omega_2, \dots \in \Omega$*. We define*

$$\omega_n \xrightarrow{n\to\infty} \omega :\iff (\forall O \in \mathcal{O} : \omega \in O \Rightarrow \omega_n \in O \text{ for almost all } n \in \mathbb{N}). \tag{A.4}$$

In particular, this gives any topology on Ω *a notion of convergence for sequences in* Ω*.*

This notion of convergence agrees with the well-known notion on metric spaces: namely, if r *is a metric on* Ω*, which generates* $\mathcal{O}$*, then the right-hand side of* (A.4) *holds if and only if for all* $\varepsilon > 0$*, we have* $r(\omega_n, \omega) < \varepsilon$ *for almost all* $n \in \mathbb{N}$*.*

Using the notion of convergence from (A.4), we state the following well-known property:

Lemma A.4 (Closure of a metric space)**.** *Let* (Ω, r) *be a metric space and* $\mathcal{O}$ *be the topology generated by* r*. For* $F \subseteq \Omega$ *the following are equivalent:*

1. *F is closed.*
2. *If* $\omega_1, \omega_2, \dots \in F$ *and* $\omega \in \Omega$ *are such that* $\omega_n \xrightarrow{n\to\infty} \omega$*, then* $\omega \in F$*.*

In particular, for every $A \subseteq \Omega$ *there exists the closure* $\overline{A}$ *consists exactly of the cluster points*[25] *of* A*.*

Proof. '1.⇒2.' Assume there is a sequence $\omega_1, \omega_2, \dots \in F$ with $\omega_n \xrightarrow{n\to\infty} \omega \in F^c$. Then, since $F^c \in \mathcal{O}$, we find $\omega_n \in F^c$ for almost all n. This is in contradiction with the assumption.
'2.⇒1.': Suppose F was not closed, i.e. F^c, is not open. Then there is $\omega \in F^c$ such that for all $\varepsilon > 0$ it holds that $B_\varepsilon(\omega) \not\subseteq F^c$. Choose $\varepsilon_1, \varepsilon_2, \dots > 0$ with[26] $\varepsilon_n \downarrow 0$ and $\omega_n \in B_{\varepsilon_n}(\omega) \cap F$. Then $\omega_1, \omega_2, \dots \in F$ with $\omega_n \xrightarrow{n\to\infty} \omega$, but $\omega \in F^c$. □

Lemma A.5 (Countable base and separable spaces)**.** *Let* (Ω, r) *be a separable metric space,* $\mathcal{O}$ *be the topology generated by* r*,* Ω' *countable with* $\overline{\Omega'} = \Omega$ *and*

$$\widetilde{\mathcal{B}} := \{B_\varepsilon(\omega) : \varepsilon \in \mathbb{Q}_+, \omega \in \Omega'\}.$$

Then $\widetilde{\mathcal{B}}$ *is countable and* $\mathcal{O}(\widetilde{\mathcal{B}}) = \mathcal{O}$*.*

[25] A cluster point of A is any limit of a convergent sequence $\omega_1, \omega_2, \dots \in A$.

[26] We write $\varepsilon_n \downarrow 0$ if $\varepsilon_1 \geq \varepsilon_2 \geq \dots$ and $\varepsilon_n \xrightarrow{n\to\infty} 0$

Proof. Clearly, $\widetilde{\mathcal{B}}$ is countable and $\mathcal{O}(\widetilde{\mathcal{B}}) \subseteq \mathcal{O}$. Let $\mathcal{B}$ as in (A.3). Then for $B_\varepsilon(\omega) \in \mathcal{B}$

$$B_\varepsilon(\omega) = \bigcup_{\widetilde{\mathcal{B}} \ni B \subseteq B_\varepsilon(\omega)} B,$$

thus $\mathcal{B} \subseteq \mathcal{O}(\widetilde{\mathcal{B}})$ and thus $\mathcal{O} = \mathcal{O}(\mathcal{B}) \subseteq \mathcal{O}(\widetilde{\mathcal{B}})$. □

Example A.6 (Two Polish spaces)**.** *1. Let $\mathcal{O}$ be the Euclidean topology on $\mathbb{R}^d$, as given in Definition A.1.9 by the Euclidean metric. From your lecture* Analysis I*, it is known that this metric is complete. Further, $\mathbb{Q}^d$ is countable and every $\omega \in \mathbb{R}^d$ is a cluster point of a sequence in $\mathbb{Q}^d$. Thus, in particular, $\overline{\mathbb{Q}^d} = \mathbb{R}^d$ by Lemma A.4, so $\mathbb{R}^d$ is separable. So overall, $(\mathbb{R}^d, \mathcal{O})$ is Polish.*

2. *Let $K \subseteq \mathbb{R}$ be compact (i.e. closed and bounded) and $\Omega = \mathcal{C}_\mathbb{R}(K)$ be the set of continuous functions $\omega : K \to \mathbb{R}$. On Ω let*

$$r(\omega_1, \omega_2) := \sup_{x \in K} |\omega_1(x) - \omega_2(x)|$$

be the supremum distance. It is known from Analysis II *that r is complete is complete. Furthermore, every $\omega \in \Omega$ can be calculated according to the Weierstrass' approximation theorem can be uniformly approximated by polynomials by polynomials. Let Ω' be the countable set of polynomials with rational coefficients. Then it also holds that $\overline{\Omega'} = \Omega$. Thus $(\Omega, \mathcal{O})$ is separable, i.e. Polish.*

A.2 Compact sets

Topological spaces can be very large. Just think of the space $\mathbb{R}$, in which there are sequences that diverge. Now *compact set* are considered as smaller subsets of a topological space. In such compact sets there are always convergent subsequences.

Definition A.7 (Relatively compact, compact, relatively sequenctially compact, totally restricted)**.** *Let $(\Omega, \mathcal{O})$ be a topological space and $K \subseteq \Omega$.*

1. *The set K is called* compact *if every open cover has a finite partial cover. That is: If $O_i \in \mathcal{O}, i \in I$ and $K \subseteq \bigcup_{i \in I} O_i$, then there is*[27]*. $J \subset I$ with $K \subseteq \bigcup_{i \in J} O_i$.*
2. *The set K is called* relatively compact *if $\overline{K}$ is compact.*
3. *The set K is called* relatively sequentially compact *if for every sequence $\omega_1, \omega_2, \ldots \in K$ there is a convergent subsequence, i.e. there is an increasing sequence $k_1, k_2, \ldots \uparrow \infty$ and $\omega \in \Omega$ with $\omega_{k_n} \xrightarrow{n \to \infty} \omega$ as in* (A.4)*.*
4. *Let r be a metric that generates $\mathcal{O}$. Then we call $K \subseteq \Omega$* totally bounded *if for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ and $\omega_1, \ldots, \omega_N \in K$ such that $K \subseteq \bigcup_{n=1}^N B_\varepsilon(\omega_n)$. In other words, for every radius $\varepsilon > 0$, there is a finite number of balls with this radius covering K.*

Lemma A.8 (Compact sets are closed)**.** *. Let (Ω, r) be a metric space and $\mathcal{O}$ the topology generated by r. If $K \subseteq \Omega$ is compact, then K is also closed.*

[27] We write $J \subseteq_f I$ if $J \subseteq I$ and J is finite

Proof. We show that K^c is open. For this, let $\omega \in K^c$. For all $\omega' \in K$ we choose $\delta_{\omega'}$ and $\varepsilon_{\omega'}$ such that $B_{\delta_{\omega'}}(\omega) \cap B_{\varepsilon_{\omega'}}(\omega') = \emptyset$. Then obviously $\bigcup_{\omega' \in K} B_{\varepsilon_{\omega'}}(\omega') \supseteq K$, so there is $J \subseteq_f K$ with $K \subseteq \bigcup_{\omega' \in J} B_{\varepsilon_{\omega'}}(\omega')$. Set $\delta := \min_{\omega' \in J} \delta_{\omega'} > 0$. Then $B_\delta(\omega) is \cap K \subseteq B_\delta(\omega) \cap \bigcup_{\omega' \in J} B_{\varepsilon_{\omega'}}(\omega') = \emptyset$, i.e. $B_\delta(\omega) \subseteq K^c$. Since $\omega \in K^c$ was arbitrary, K^c is open, so K is closed. □

The following theorem about compact sets gives a complete characterization of compact sets in Polish spaces.

Proposition A.9 (Characterising relatively compact sets)**.** . *Let (Ω, r) be a metric space, $\mathcal{O}$ be the topology generated by r and $K \subseteq \Omega$. Consider the following statements:*

1. *K is relatively compact.*
2. *If $F_i \subseteq \overline{K}$ is closed, $i \in I$, and $\bigcap_{i \in I} F_i = \emptyset$, then there is $J \subseteq_f I$ with $\bigcap_{i \in J} F_i = \emptyset$.*
3. *K is relatively sequentially compact.*
4. *K is totally bounded.*

Then

$$4. \Longleftarrow 1. \iff 2. \Longrightarrow 3.$$

Furthermore, $3. \Longrightarrow 2.$ *also holds if $(\Omega, \mathcal{O})$ is separable and* $4. \Longrightarrow 3.$ *if (Ω, r) is complete. In particular, all four statements are equivalent, if $(\Omega, \mathcal{O})$ is Polish.*

Corollary A.10. *Let (Ω, r) be a metric space, $\mathcal{O}$ the topology generated by r. Then closed subsets of compact sets are compact.*

Proof. Let $K \subseteq \Omega$ be compact and $A \subseteq K$ closed. A closed set is compact if and only if it is relatively compact. From Proposition A.9.2 one reads, because of the relative compactness of K, that for F_i closed, $i \in I$, with $F_i \subseteq A \subseteq K$ and $\bigcap_{i \in I} F_i = \emptyset$ a $J \subset I$ exists with $\bigcap_{i \in J} F_i = \emptyset$. Again with Proposition A.9.2 it follows that A is relatively compact, i.e. compact. □

Proof of Proposition A.9. '1.⇒4.' Let $\overline{K}$ be compact and $\varepsilon > 0$. Obviously, $\bigcup_{\omega \in K} B_\varepsilon(\omega) \supseteq \overline{K}$ is an open covering. Thus, since $\overline{K}$ is compact, there is a finite subcover, i.e. there is $\omega_1, \dots, \omega_N$ with $\overline{K} \subseteq \bigcup_{n=1}^N B_\varepsilon(\omega_n)$. Since $\varepsilon > 0$ was arbitrary, the assertion follows.

'1.⇒2.' Now let $F_i, i \in I$ be as stated. Then $\bigcup_{i \in I} F_i^c = \left(\bigcap_{i \in I} F_i\right)^c = \Omega \supseteq \overline{K}$. Since $\overline{K}$ is compact, there is $J \subseteq_f I$ with $\overline{K} \subseteq \bigcup_{i \in J} F_i^c$. Thus $\bigcap_{i \in J} F_i = \left(\bigcup_{i \in J} F_i^c\right) \subseteq \overline{K}^c$. But since $F_i \subseteq \overline{K}$ was assumed, $\bigcap_{i \in J} F_i = \emptyset$.

'2.⇒1.' Let $O_i \in \mathcal{O}, i \in I$ be a covering of $\overline{K}$, i.e. $\overline{K} \subseteq \bigcup_{i \in I} O_i$. Set $F_i = O_i^c \cap \overline{K}$, then $F_i^c \in \mathcal{O}$ and $\bigcap_{i \in I} F_i = \overline{K} \cap \left(\bigcup_{i \in I} O_i\right)^c = \emptyset$. So there is $J \subseteq_f I$ with $\bigcap_{i \in J} F_i = \emptyset$. Therefore, $\overline{K}^c \cup \bigcup_{i \in J} O_i = \bigcup_{i \in J} F_i^c = \Omega$, so $\bigcup_{i \in J} O_i \supseteq \overline{K}$. So we found a finite subcovering. In other words, $\overline{K}$ is compact.

'2,⇒3.' Let $\omega_1, \omega_2, \dots \in K$. We set $F_n = \overline{\{\omega_n, \omega_{n+1}, \dots\}} \subseteq \overline{K}$. Suppose there is no convergent subsequence of $\omega_1, \omega_2, \dots$ Then $\bigcap_{n=1}^\infty F_n = \emptyset$. From 2. it then follows that there is a $N \in \mathbb{N}$

with $\emptyset = .\bigcap_{n=1}^N F_n = F_N$. This is a contradiction, since F_N is not empty by construction; therefore there is a convergent subsequence.

'3.⇒1.' if $(\Omega, \mathcal{O})$ is separable. Let Ω' be countable with $\overline{\Omega'} = \Omega$ and $\mathcal{B} := \{B_{1/n}(\omega) : \omega \in \Omega', n \in \mathbb{N}\}$. Then, $\mathcal{B}$ is a countable basis of $\mathcal{O}$. We write $\mathcal{B} = \{B_1, B_2, \ldots\}$.

Suppose $\overline{K}$ is not compact. That is, there is a cover $A_i \in \mathcal{O}, i \in I$ (for some infinite I) with $\overline{K} \subseteq \bigcup_{i\in I} A_i$ and there is no finite subcover. We set for $i \in I$

$$J_i = \{j \in \mathbb{N} : B_j \subseteq A_i\} \subseteq \mathbb{N}$$

and $J := \bigcup_{i\in I} J_i \subseteq \mathbb{N}$. Thus $A_i = \bigcup_{j\in J_i} B_j$, and

$$\overline{K} \subseteq \bigcup_{i\in I} A_i = \bigcup_{i\in I}\bigcup_{j\in J_i} B_j = \bigcup_{j\in J} B_j.$$

This shows that $B_j \in \mathcal{O}, j \in J$ is a countable cover of $\overline{K}$. Since there is no finite subcover for $A_i, i \in I$, there can also be no finite finite subcover for $B_j, j \in J$. (If there would be a finite subcover $B_j, j \in J$, we could take $A_j \supseteq B_j, j \in J$ and find a finite subcover $A_j, j \in J$, contradiction.) We write $J = \{j_1, j_2, ...\}$. For $n \in \mathbb{N}$ we set $\omega_n \in \overline{K} \setminus \bigcup_{i=1}^n B_{j_i}$. (Note that this set is non-empty, otherwise a finite subcover would exist.) By assumption, the sequence $\omega_1, \omega_2, \ldots \in \overline{K}$ has a cluster point $\omega \in \overline{K}$. Since $\overline{K} \subseteq \bigcup_{j\in J} B_j$, there is $k \in J \subseteq \mathbb{N}$ with $\omega \in B_k$. So, on the one hand (since B_k is open) there are infinitely many of the ω_n in B_k, on the other hand, $\omega_i \notin B_k$ for all $i \geq k$ by construction. This is a contradiction, so $\overline{K}$ is compact.

'4.⇐'3. If (Ω, r) is complete: Let $\omega_1, \omega_2, \ldots \in K$. We are going to construct a Cauchy subsequence. This converges since (Ω, r) is complete and K is found to be relatively sequentially compact. In order to construct the subsequence, choose a sequence $\varepsilon_1, \varepsilon_2, \ldots > 0$ with $\varepsilon_n \downarrow 0$. Since K is totally bounded, there are finitely many ε_1-balls covering K. At least one of these balls must contain infinitely many of the ω_n. These have each at most distance $2\varepsilon_1$. Choose ω_{k_1} as one of these infinitely many points. Since this ε_1-ball is covered by finitely many ε_2-balls, there is one of these ε_2-balls, which contains infinitely many of the ω_n. These each have at most distance $2\varepsilon_2$. Choose $\omega_{k_2} \neq \omega_{k_1}$ as one of these infinitely many points. By proceeding further we obtain a sequence $\omega_{k_1}, \omega_{k_2}, \ldots \in K$ such that $r(\omega_{k_n}, \omega_{k_m}) \leq 2\varepsilon_{m\wedge n}$. With other words, as announced, we have found a Cauchy subsequence in K. □

Lemma A.11 (Compact metric spaces are Polish)**.** *. Let (Ω, r) be a metric space and $\mathcal{O}$ be the topology generated by r. If Ω is compact, then $(\Omega, \mathcal{O})$ is Polish.*

Proof. For the proof, we need to show both, completeness of (Ω, r) and separability of $(\Omega, \mathcal{O})$. For completeness, let $\omega_1, \omega_2, \ldots \in \Omega$ be a Cauchy sequence. Since K is relatively sequentially compact according to Proposition A.9, there is $\omega \in \Omega$ and a subsequence $\omega_{k_1}, \omega_{k_2}, \ldots$ converging to ω. Let $\varepsilon > 0$ and $N \in \mathbb{N}$ be such that $r(\omega_m, \omega_n) < \varepsilon/2$ for $m, n > N$ and $r(\omega_{k_n}, \omega) < \varepsilon/2$ for $k_n > N$. Then for $m > N$ it holds that $r(\omega_m, \omega) \leq r(\omega_m, \omega_{k_n}) + r(\omega_{k_n}, \omega) \leq \varepsilon$. It follows that $\omega_n \xrightarrow{n\to\infty} \omega$. For separability of $(\Omega, \mathcal{O})$, let $\varepsilon_1, \varepsilon_2, \ldots > 0$ with $\varepsilon_n \downarrow 0$. Since K is totally bounded, for all $n \in \mathbb{N}$ there is a k_n and $\omega_{n1}, \ldots, \omega_{nk_n}$ with $K \subseteq \bigcup_{k=1}^{k_n} B_{\varepsilon_n}(\omega_{nk_n})$. Let $\Omega' = \omega_{nk} : n \in \mathbb{N}, k = 1, \ldots, k_n\}$. Then Ω' is countable and for each $\omega \in \Omega$ and each $n \in \mathbb{N}$ there is a $k(\omega, n) \in \{1, \ldots, k_n\}$ with $r(\omega_{k(\omega,n)}, \omega) < \varepsilon_n$. Thus, $(\omega_{k(\omega,n)}, \omega) \xrightarrow{n\to\infty} \omega$. So $\overline{\Omega'} = \Omega$. □