Ruprecht-Karls-Universität Heidelberg
Institut für Angewandte Mathematik
Prof. Dr. Jan JOHANNES

Outline of the lecture course

PROBABILITY THEORY II

Summer semester 2016

Preliminary version: July 12, 2016

If you find **errors in the outline**, please send a short note
by email to johannes@math.uni-heidelberg.de

MΛTHEMΛTIKON, Im Neuenheimer Feld 205, 69120 Heidelberg
Telefon: +49 6221 54.14.190 – Fax: +49 6221 54.53.31
eMail: johannes@math.uni-heidelberg.de
Webseite zur Vorlesung: www.razbaer.eu/ag-johannes/vl/WT2-SS16/

Table of contents

1 Preliminaries **1**
1.1 Basic measure theory . . . 1
1.2 Random variables . . . 2
1.3 Independence . . . 4
1.4 Expectation . . . 5
1.5 Convergence of random variables . . . 6
1.6 Conditional expectation . . . 6

2 Stochastic processes **9**
2.1 Motivating examples . . . 9
2.1.1 The Poisson process . . . 9
2.1.2 Markov chains . . . 10
2.1.3 Brownian motion . . . 10
2.2 Definition of stochastic processes . . . 10
2.3 Probability measures on Polish spaces . . . 12
2.4 Adapted stochastic process and stopping times . . . 14

3 Martingale theory **17**
3.1 Positive {super}martingales . . . 17
3.2 Integrable {super/sub}martingales . . . 20
3.3 Regular integrable martingale . . . 21
3.4 Regular stopping times for an integrable martingale . . . 23
3.5 Regularity of integrable submartingales . . . 24
3.6 Doob decomposition and square variation . . . 25

4 Markov chains **29**
4.1 Time-homogeneous Markov chain . . . 29
4.2 Markov chains: recurrence and transience . . . 30
4.3 Invariant distributions . . . 31

5 Ergodic theory **33**
5.1 Stationary and ergodic processes . . . 33
5.2 Ergodic theorems . . . 34

6 Weak convergence **37**
6.1 Fundamental properties . . . 37
6.2 Prohorov's theorem . . . 39

7 Brownian motion **43**
7.1 Continuous versions . . . 43
7.2 Construction and path properties . . . 44
7.3 Markov properties . . . 46
7.4 Donsker's theorem . . . 48

Chapter 1

Preliminaries

This chapter presents elements of the lecture course PROBABILITY THEORY I along the lines of the textbook Klenke [2008], where far more details, examples and further discussions can be found.

1.1 Basic measure theory

In the following, let $\Omega \neq \emptyset$ be a nonempty set and let $\mathscr{A} \subset 2^{\Omega}$ (power set, set of all subsets of Ω) be a class of subsets of Ω. Later, Ω will be interpreted as the space of elementary events and $\mathscr{A}$ will be the system of observable events.

§1.1.1 **Definition.** (a) A pair $(\Omega, \mathscr{A})$ consisting of a nonempty set Ω and a σ-algebra $\mathscr{A}$ is called a *measurable space*. The sets $A \in \mathscr{A}$ are called *measurable sets*. If Ω is at most countably infinite and if $\mathscr{A} = 2^{\Omega}$, then the measurable space $(\Omega, 2^{\Omega})$ is called *discrete*.

(b) A triple $(\Omega, \mathscr{A}, \mu)$ is called a *measure space* if $(\Omega, \mathscr{A})$ is a measurable space and if μ is a measure on $\mathscr{A}$.

(c) A measure space $(\Omega, \mathscr{A}, \mathbb{P})$ is called a *probability space*, if in addition $\mathbb{P}(\Omega) = 1$. In this case, the sets $A \in \mathscr{A}$ are called *events*. □

§1.1.2 **Remark.** Let $\mathscr{A} \subset 2^{\Omega}$ and let $\mu : \mathscr{A} \to [0, \infty]$ be a set function. We say that μ is

(a) monotone, if $\mu(A) \leqslant \mu(B)$ for any two sets $A, B \in \mathscr{A}$ with $A \subset B$.

(b) *additive*, if $\mu\big(\biguplus_{i=1}^{n} A_i\big) = \sum_{i=1}^{n} \mu(A_i)$ for any choice of *finitely* many mutually disjoints sets $A_1, \dots, A_n \in \mathscr{A}$ with $\cup_{i=1}^{n} A_i \in \mathscr{A}$. The disjoint union of sets is denoted by the symbol $\biguplus$ which only stresses the fact that the sets involved are mutually disjoint.

(c) *σ-additive*, if $\mu\big(\biguplus_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} \mu(A_i)$ for any choice of *countably* many mutually disjoints sets $A_1, A_2, \dots \in \mathscr{A}$ with $\cup_{i=1}^{\infty} A_i \in \mathscr{A}$.

$\mathscr{A}$ is called an *algebra* if (i) $\Omega \in \mathscr{A}$, (ii) $\mathscr{A}$ is closed under complements, and (iii) $\mathscr{A}$ is closed under intersections. Note that, if $\mathscr{A}$ is closed under complements, then we have the equivalences between (i) $\mathscr{A}$ is closed under (countable) unions and (ii) $\mathscr{A}$ is closed under (countable) intersections. An algebra $\mathscr{A}$ is called *σ-algebra*, if it is closed under countable intersections. If $\mathscr{A}$ is an algebra and $\mu : \mathscr{A} \to [0, \infty]$ is a set function with $\mu(\emptyset) = 0$, then μ is called a

(d) *content*, if μ is additive,

(e) *premeasure*, if μ is σ-additive,

(f) *measure*, if μ is a premeasure and $\mathscr{A}$ is a σ-Algebra.

A content μ on an algebra $\mathscr{A}$ is called

(g) *finite*, if $\mu(A) < \infty$ for every $A \in \mathscr{A}$,

(h) *σ-finite*, if there is a sequence $\Omega_1, \Omega_2, \ldots \in \mathscr{A}$ such that $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$ and such that $\mu(\Omega_n) < \infty$ for all $n \in \mathbb{N}$. □

§1.1.3 **Examples.** (a) For any nonempty set Ω, the classes $\mathscr{A} = \{\emptyset, \Omega\}$ and $\mathscr{A} = 2^{\Omega}$ are the trivial examples of σ-algebras.

(b) Let $\mathcal{E} \subset 2^{\Omega}$. The smallest σ-algebra $\sigma(\mathcal{E}) = \bigcap\{\mathscr{A} : \mathscr{A} \text{ is } \sigma\text{-algebra and } \mathcal{E} \subset \mathscr{A}\}$ with $\mathcal{E} \subset \sigma(\mathcal{E})$ is called the σ-algebra *generated by* $\mathcal{E}$ and $\mathcal{E}$ is called a *generator* of $\sigma(\mathcal{E})$.

(c) Let (Ω, τ) be a topological space with class of open sets $\tau \subset 2^{\Omega}$. The σ-algebra $\mathscr{B}(\Omega)$ that is generated by the open sets is called the *Borel-σ-algebra* on Ω. The elements $B \in \mathscr{B}(\Omega)$ are called *Borel sets* or *Borel measurable sets*. We write $\mathscr{B} := \mathscr{B}(\mathbb{R})$, $\mathscr{B}^+ := \mathscr{B}(\mathbb{R}^+)$ and $\mathscr{B}^n := \mathscr{B}(\mathbb{R}^n)$ for the Borel-σ-algebra on $\mathbb{R}$, $\mathbb{R}^+ := [0, \infty)$ and $\mathbb{R}^n$, respectively, equipped with the usual Euclidean distance.

(d) Denote by $\mathbb{1}_A(x)$ the indicator function on a set A which takes the value one if $x \in A$ and zero otherwise. Let $\omega \in \Omega$ and $\delta_\omega(A) = \mathbb{1}_A(\omega)$. Then δ_ω is a probability measure on any σ-algebra $\mathscr{A} \subset 2^{\Omega}$. δ_ω is called the *Dirac measure* on the point ω.

(e) Let Ω be an (at most) countable nonempty set and let $\mathscr{A} = 2^{\Omega}$. Further let $(p_\omega)_{\omega \in \Omega}$ be non-negative numbers. Then $A \mapsto \mu(A) := \sum_{\omega \in \Omega} p_\omega \delta_\omega(A)$ defines a σ-finite measure. If $p_\omega = 1$ for every $\omega \in \Omega$, then μ is called *counting measure* on Ω. If Ω is finite, then so is μ. □

§1.1.4 **Theorem** (Carathéodory). *Let $\mathscr{A} \subset 2^{\Omega}$ be an algebra and let μ be a σ-finite premeasure on $\mathscr{A}$. There exists a unique measure $\tilde{\mu}$ on $\sigma(\mathscr{A})$ such that $\tilde{\mu}(A) = \mu(A)$ for all $A \in \mathscr{A}$. Furthermore, $\tilde{\mu}$ is σ-finite.*

Proof of Theorem §1.1.4. We refer to Klenke [2008], Theorem 1.41. □

§1.1.5 **Remark.** If μ is a finite content on an algebra $\mathscr{A}$, then *σ-continuity at $\emptyset$*, that is, $\mu(A_n) \to 0 = \mu(\emptyset)$ as $n \to \infty$ for any sequence $(A_n)_{n \in \mathbb{N}}$ in $\mathscr{A}$ with $\mu(A_n) < \infty$ for some (and then eventually all) $n \in \mathbb{N}$ and $A_n \downarrow \emptyset$ (i.e., $A_1 \supset A_2 \supset A_3 \supset \ldots$ and $\cap_{n=1}^{\infty} A_n = \emptyset$), implies σ-additivity. □

§1.1.6 **Example.** A probability measure $\mathbb{P}$ on the measurable space $(\mathbb{R}^n, \mathscr{B}^n)$ is uniquely determined by the values $\mathbb{P}((-\infty, b])$ (where $(-\infty, b] = \times_{i=1}^{n}(-\infty, b_i], b \in \mathbb{R}^n$). In particular, a probability measure $\mathbb{P}$ on $\mathbb{R}$ is uniquely determined by its *distribution function* $F : \mathbb{R} \to [0, 1]$, $x \mapsto \mathbb{P}((-\infty, x])$. □

1.2 Random variables

In this section $(\Omega, \mathscr{A})$, $(\mathcal{S}, \mathscr{S})$ and $(\mathcal{S}_i, \mathscr{S}_i)$, $i \in \mathcal{I}$, denote measurable spaces where $\mathcal{I}$ is an arbitrary index set.

§1.2.1 **Definition.** Let Ω be a nonempty set and let $X : \Omega \to \mathcal{S}$ be a map.

(a) X is called $\mathscr{A}$-$\mathscr{S}$-measurable (or, briefly, measurable) if $X^{-1}(\mathscr{S}) := \{X^{-1}(S) : S \in \mathscr{S}\} \subset \mathscr{A}$, that is, if $X^{-1}(S) \in \mathscr{A}$ for any $S \in \mathscr{S}$. A measurable map $X : (\Omega, \mathscr{A}) \to$

$(\mathcal{S}, \mathscr{S})$ is called a *random variable* (*r.v.*) with values in $(\mathcal{S}, \mathscr{S})$. If $(\mathcal{S}, \mathscr{S}) = (\mathbb{R}, \mathscr{B})$ or $(\mathcal{S}, \mathscr{S}) = (\mathbb{R}^+, \mathscr{B}^+)$, then X is called a *real* or *positive* random variable, respectively.

(b) The preimage $X^{-1}(\mathscr{S})$ is the smallest σ-algebra on Ω with respect to which X is measurable. We say that $\sigma(X) := X^{-1}(\mathscr{S})$ is the σ-algebra on Ω that is *generated by* X.

(c) For any, $i \in \mathcal{I}$, let $X_i : \Omega \to \mathcal{S}_i$ be an arbitrary map. Then $\sigma(X_i, i \in \mathcal{I}) := \bigvee_{i\in\mathcal{I}} \sigma(X_i) := \sigma\big(\cup_{i\in\mathcal{I}}\, \sigma(X_i)\big) = \sigma\big(\cup_{i\in\mathcal{I}}\, X_i^{-1}(\mathscr{S}_i)\big)$ is called the σ-algebra on Ω that is generated by $(X_i, i \in \mathcal{I})$. This is the the smallest σ-algebra with respect to which all X_i are measurable. $\square$

§1.2.2 **Properties**. *Let $\mathcal{I}$ be an arbitrary index set. Consider $S_i \in 2^{\mathcal{S}}$, $i \in \mathcal{I}$, and a map $X : \Omega \to \mathcal{S}$. Then*

(a) $X^{-1}(\cup_{i\in\mathcal{I}} S_i) = \cup_{i\in\mathcal{I}} X^{-1}(S_i)$, $X^{-1}(\cap_{i\in\mathcal{I}} S_i) = \cap_{i\in\mathcal{I}} X^{-1}(S_i)$,

(b) $X^{-1}(\mathscr{S})$ *is a σ-algebra on Ω and $\{S \in \mathcal{S} : X^{-1}(S) \in \mathscr{A}\}$ is a σ-algebra on $\mathcal{S}$.*

If $\mathcal{E}$ is a class of sets in $2^{\mathcal{S}}$, then $\sigma_\Omega(X^{-1}(\mathcal{E})) = X^{-1}(\sigma_{\mathcal{S}}(\mathcal{E}))$. $\square$

§1.2.3 **Examples**. (a) The *identity map* $\mathrm{Id} : \Omega \to \Omega$ is $\mathscr{A}$-$\mathscr{A}$-measurable.

(b) If $\mathscr{A} = 2^\Omega$ and $\mathscr{S} = \{\emptyset, \mathcal{S}\}$, then any map $X : \Omega \to \mathcal{S}$ is $\mathscr{A}$-$\mathscr{S}$-measurable.

(c) Let $A \subset \Omega$. The *indicator function* $\mathbb{1}_A : \Omega \to \{0, 1\}$ is $\mathscr{A}$-$2^{\{0,1\}}$-measurable, if and only if $A \in \mathscr{A}$. $\square$

For $x, y \in \mathbb{R}$ we agree on the following notations $\lfloor x\rfloor := \max\{k \in \mathbb{Z} : k \leqslant x\}$ (integer part), $x \vee y = \max(x, y)$ (maximum), $x \wedge y = \min(x, y)$ (minimum), $x^+ = \max(x, 0)$ (positive part), $x^- = \max(-x, 0)$ (negative part) and $|x| = x^- + x^+$ (modulus).$\mathbb{V}$ar

§1.2.4 **Properties**. (a) *If X, Y are real r.v.'s, then so are $X^+ := \max(X, 0)$, $X^- := \max(-X, 0)$, $|X| = X^+ + X^-$, $X + Y$, $X - Y$, $X \cdot Y$ and X/Y with $x/0 := 0$ for all $x \in \mathbb{R}$. In particular, X^+ and $\lfloor X\rfloor$ is $\mathscr{A}$-$\mathscr{B}^+$- and $\mathscr{A}$-$2^{\mathbb{Z}}$-measurable, respectively.*

(b) *If $X_1, X_2, \dots$ are real r.v.'s, then so are $\sup_{n\geqslant 1} X_n$, $\inf_{n\geqslant 1} X_n$, $\limsup_{n\to\infty} X_n := \inf_{k\geqslant 1} \sup_{n\geqslant k} X_n$ and $\liminf_{n\to\infty} X_n := \sup_{k\geqslant 1} \inf_{n\geqslant k} X_n$.*

(c) *Let $X_1, \dots, X_n : \Omega \to \mathbb{R}$ be maps and define $X := (X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$. Then X is a real r.v. (i.e., $\mathscr{A}$-$\mathscr{B}^n$-measurable), if and only if each X_i is a real r.v. (i.e., $\mathscr{A}$-$\mathscr{B}$-measurable).*

(d) *Let $\mathcal{E} = \{A_i \in 2^\Omega, i \in \mathcal{I}$, mutually disjoint and $\biguplus_{i\in\mathcal{I}} A_i = \Omega\}$ be a partition of Ω. A map $X : \Omega \to \mathbb{R}$ is $\sigma(\mathcal{E})$-$\mathscr{B}$-measurable, if there exist numbers $x_i \in \mathbb{R}$, $i \in \mathcal{I}$, such that $X = \sum_{i\in\mathcal{I}} x_i \mathbb{1}_{A_i}$.* $\square$

§1.2.5 **Definition**. (a) A real r.v. X is called *simple* if there is an $n \in \mathbb{N}$ and mutually disjoint measurable sets $A_i, \dots, A_n \in \mathscr{A}$ as well as numbers $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, such that $X = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$.

(b) Assume that $X, X_1, X_2, \dots$ are maps $\Omega \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ such that $X_1(\omega) \leqslant X_2(\omega) \leqslant \dots$ and $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for any $\omega \in \Omega$. Then we write $X_n \uparrow X$ and say that $(X_n)_{n\in\mathbb{N}}$ *increases (point-wise)* to X. Analogously, we write $X_n \downarrow X$ if $(-X_n) \uparrow (-X)$. $\square$

§1.2.6 **Example**. Let us briefly consider the approximation of a positive r.v. by means of simple r.v.'s. Let $X : \Omega \to \mathbb{R}^+$ be a $\mathscr{A}$-$\mathscr{B}^+$-measurable. Define $X_n = (2^{-n}\lfloor 2^n X \rfloor) \wedge n$. Then X_n is a simple r.v. and clearly, $X_n \uparrow X$ uniformly on each interval $\{X \leqslant c\}$. □

§1.2.7 **Property**. *Let $X : (\Omega, \mathscr{A}) \to (\mathcal{S}, \mathscr{S})$ and $Y : (\Omega, \mathscr{A}) \to (\mathbb{R}, \mathscr{B})$ be r.v.'s. The real r.v. Y is $\sigma(X)$-$\mathscr{B}$-measurable if and only if there exists a $\mathscr{S}$-$\mathscr{B}$-measurable map $f : \mathcal{S} \to \mathbb{R}$ such that $Y = f(X)$.* □

§1.2.8 **Definition**. Let $X : (\Omega, \mathscr{A}) \to (\mathcal{S}, \mathscr{S})$ be a r.v..

(a) For $S \in \mathscr{S}$, we denote $\{X \in S\} := X^{-1}(S)$. In particular, we let $\{X \geqslant 0\} := X^{-1}([0, \infty))$ and define $\{X \leqslant b\}$ similarly and so on.

(b) Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathscr{A})$. The image probability measure $\mathbb{P}_X$ of $\mathbb{P}$ under the map X is the probability measure $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$ on $(\mathcal{S}, \mathscr{S})$ that is defined by $\mathbb{P}_X(S) := \mathbb{P}(X \in S) := \mathbb{P}(X^{-1}(S))$ for each $S \in \mathscr{S}$. $\mathbb{P}_X$ is called the *distribution* of X. We write $X \sim \mathbb{Q}$ if $\mathbb{Q} = \mathbb{P}_X$ and say X has distribution $\mathbb{Q}$.

(c) A family $(X_i)_{i \in \mathcal{I}}$ of r.v.'s is called *identically distributed* (*i.d.*) if $\mathbb{P}_{X_i} = \mathbb{P}_{X_j}$ for all $i, j \in \mathcal{I}$. We write $X \stackrel{d}{=} Y$ if $\mathbb{P}_X = \mathbb{P}_Y$ (d for distribution). □

1.3 Independence

In the sequel, $(\Omega, \mathscr{A}, \mathbb{P})$ is a probability space, the sets $A \in \mathscr{A}$ are the events and $\mathcal{I}$ is an arbitrary index set.

§1.3.1 **Definition**. (a) Let $(A_i)_{i \in \mathcal{I}}$ be an arbitrary family of events. The family $(A_i)_{i \in \mathcal{I}}$ is called *independent* if for any finite subset $\mathcal{J} \subset \mathcal{I}$ the product formula holds: $\mathbb{P}(\cap_{j \in \mathcal{J}} A_j) = \prod_{j \in \mathcal{J}} \mathbb{P}(A_j)$.

(b) Let $\mathcal{E}_i \subset \mathscr{A}$ for all $i \in \mathcal{I}$. The family $(\mathcal{E}_i)_{i \in \mathcal{I}}$ is called *independent* if, for any finite subset $\mathcal{J} \subset \mathcal{I}$ and any choice of $E_j \in \mathcal{E}_j$, $j \in \mathcal{J}$, the product formula holds: $\mathbb{P}(\cap_{j \in \mathcal{J}} E_j) = \prod_{j \in \mathcal{J}} \mathbb{P}(E_j)$. □

§1.3.2 **Lemma** (Borel-Cantelli). *Let $A_1, A_2, \dots$ be events and define $A^* := \limsup_{n \to \infty} A_n$.*

(a) *If $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(A^*) = 0$.*

(b) *If $(A_n)_{n \in \mathbb{N}}$ is independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(A^*) = 1$.*

Proof of Lemma §1.3.2. We refer to Klenke [2008], Theorem 2.7. □

§1.3.3 **Corollary** (Borel's 0-1 criterion). *Let $A_1, A_2, \dots$ be independent events and define $A^* := \limsup_{n \to \infty} A_n$, then*

(a) *$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ if and only if $\mathbb{P}(A^*) = 0$,*

(b) *$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ if and only if $\mathbb{P}(A^*) = 1$.* □

For each $i \in \mathcal{I}$, let $(\mathcal{S}_i, \mathscr{S}_i)$ be a measurable space and let $X_i : (\Omega, \mathscr{A}) \to (\mathcal{S}_i, \mathscr{S}_i)$ be a r.v. with generated σ-algebra $\sigma(X_i) = X_i^{-1}(\mathscr{S}_i)$.

§1.3.4 **Definition**. (a) The family $(X_i)_{i \in \mathcal{I}}$ of r.v.'s is called *independent* if the family $(\sigma(X_i))_{i \in \mathcal{I}}$ of σ-algebras is independent.

(b) Let $\mathcal{E}_i \subset \mathscr{A}$ for all $i \in \mathcal{I}$. The family $(\mathcal{E}_i)_{i\in\mathcal{I}}$ is called *independent* if, for any finite subset $\mathcal{J} \subset \mathcal{I}$ and any choice of $E_j \in \mathcal{E}_j$, $j \in \mathcal{J}$, the product formula holds: $\mathbb{P}(\cap_{j\in\mathcal{J}} E_j) = \prod_{j\in\mathcal{J}} \mathbb{P}(E_j)$.

§1.3.5 **Property**. *Let $\mathcal{K}$ be an arbitrary set and $\mathcal{I}_k$, $k \in \mathcal{K}$, arbitrary mutually disjoint index sets. Define $\mathcal{I} = \cup_{k\in\mathcal{K}}\mathcal{I}_k$. If the family $(X_i)_{i\in\mathcal{I}}$ of r.v.'s is independent, then the family of σ-algebras $(\sigma(X_j, j \in \mathcal{I}_k))_{k\in\mathcal{K}}$ is independent.* □

§1.3.6 **Definition**. Let $X_1, X_2, \dots$ be r.v.'s. The σ-algebra $\bigcap_{n\geqslant 1} \sigma(X_i, i \geqslant n)$ is called the *tail σ-algebra* and its elements are called *tail events*. □

§1.3.7 **Example**. $\{\omega : \sum_{n\geqslant 1} X_n(\omega)$ is convergent$\}$ is an tail event. □

§1.3.8 **Theorem** (Kolmogorov's 0-1 law). *The tail events of a sequence $(X_n)_{n\in\mathbb{N}}$ of independent r.v.'s have probability* 0 *or* 1.

Proof of Theorem §1.3.8. We refer to Klenke [2008], Theorem 2.37. □

1.4 Expectation

§1.4.1 **Definition**. We denote by $\mathcal{M} := \mathcal{M}(\Omega, \mathscr{A})$ the set of all real r.v.'s defined on the measurable space $(\Omega, \mathscr{A})$ and by $\mathcal{M}^+ := \mathcal{M}^+(\Omega, \mathscr{A}) \subset \mathcal{M}$ the subset of all positive r.v.'s. Given a probability measure $\mathbb{P}$ on $(\Omega, \mathscr{A})$ the *expectation* is the unique functional $\mathbb{E} : \mathcal{M}^+ \to [0, \infty]$ satisfying

(a) $\mathbb{E}(aX_1 + X_2) = a\mathbb{E}(X_1) + \mathbb{E}(X_2)$ for all $X_1, X_2 \in \mathcal{M}^+$ and $a \in \mathbb{R}^+$;

(b) Assume $X, X_1, X_2, \dots \in \mathcal{M}^+$ such that $X_n \uparrow X$ then $\mathbb{E}X_n \uparrow \mathbb{E}X$;

(c) $\mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$ for all $A \in \mathscr{A}$.

The *expectation* of $X \in \mathcal{M}$ is defined by $\mathbb{E}(X) := \mathbb{E}(X^+) - \mathbb{E}(X^-)$, if $\mathbb{E}(X^+) < \infty$ or $\mathbb{E}(X^-) < \infty$. Given $\|X\|_p := \big(\mathbb{E}(|X|^p)\big)^{1/p}$, $p \in [1, \infty)$, and $\|X\|_\infty := \inf\{c : \mathbb{P}(X > c) = 0\}$ for $p \in [1, \infty]$ set $\mathcal{L}_p(\Omega, \mathscr{A}, P) := \{X \in \mathcal{M}(\Omega, \mathscr{A}) : \|X\|_p < \infty\}$ and $L_p := L_p(\Omega, \mathscr{A}, P) := \{[X] : X \in \mathcal{L}_p(\Omega, \mathscr{A}, \mathbb{P})\}$ where $[X] := \{Y \in \mathcal{M}(\Omega, \mathscr{A}) : \mathbb{P}(X = Y) = 1\}$. □

§1.4.2 **Remark**. L_1 is the domain of definition of the expectation $\mathbb{E}$, that is, $\mathbb{E} : L_1 \to \mathbb{R}$. The vector space L_p equipped with the norm $\|\cdot\|_p$ is a Banach space and in case $p = 2$ it is a Hilbert space with norm $\|\cdot\|_2$ induced by the inner product $\langle X, Y\rangle_2 := \mathbb{E}(XY)$. □

§1.4.3 **Properties**. (a) *For r.v.'s $X, Y \in L_1$ we have the equivalences between* (i) $\mathbb{E}(X\mathbb{1}_A) \leqslant \mathbb{E}(Y\mathbb{1}_A)$ *for all $A \in \mathscr{A}$ and* (ii) $\mathbb{P}(X \leqslant Y) = 1$. *In particular, $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(Y\mathbb{1}_A)$ holds for all $A \in \mathscr{A}$ if and only if $\mathbb{P}(X = Y) = 1$.*

(b) *(Fatou's lemma) Assume $X_1, X_2, \dots \in \mathcal{M}^+$, then $\mathbb{E}(\liminf\limits_{n\to\infty} X_n) \leqslant \liminf\limits_{n\to\infty} \mathbb{E}(X_n)$.*

(c) *(Dominated convergence) Assume $X, X_1, X_2, \dots \in \mathcal{M}$ such that $\lim_{n\to\infty} |X_n(\omega) - X(\omega)| = 0$ for all $\omega \in \Omega$. If there exists $Y \in L_1$ with $\sup_{n\geqslant 1} |X_n| \leqslant Y$, then we have $\lim_{n\to\infty} \mathbb{E}|X_n - X| = 0$ which in turn implies $X \in L_1$ and $\lim_{n\to\infty} |\mathbb{E}X_n - \mathbb{E}X| = 0$.*

(d) *(Hölder's inequality) For $X, Y \in \mathcal{M}$ holds $\mathbb{E}|XY| \leqslant \|X\|_p \|Y\|_q$ with $p^{-1} + q^{-1} = 1$.*

(e) *(Cauchy-Schwarz inequality) For $X, Y \in \mathcal{M}$ holds $\mathbb{E}|XY| \leqslant \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$ and $|\mathbb{C}\mathrm{ov}(X, Y)| \leqslant \sqrt{\mathbb{V}\mathrm{ar}(X)}\sqrt{\mathbb{V}\mathrm{ar}(Y)}$.* □

1.5 Convergence of random variables

In the sequel we assume r.v.'s $X_1, X_2, \ldots \in \mathcal{M}(\Omega, \mathscr{A})$ and a probability measure $\mathbb{P}$ on $(\Omega, \mathscr{A})$.

§1.5.1 **Definition.** (a) Let $C := \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) \text{ exists and is finite}\}$. The sequence $(X_n)_{n\geqslant 1}$ *converges almost surely (a.s.)*, if $\mathbb{P}(C) = 1$. We write $X_n \stackrel{n\to\infty}{\longrightarrow} X$ $a.s.$, or briefly, $X_n \stackrel{a.s.}{\longrightarrow} X$.

(b) The sequence $(X_n)_{n\geqslant 1}$ *converges in probability*, if $\lim_{n\to\infty} P\big(|X_n - X| \geqslant \varepsilon\big) = 0$ for all $\varepsilon > 0$. We write $X_n \stackrel{n\to\infty}{\longrightarrow} X$ in $\mathbb{P}$, or briefly, $X_n \stackrel{\mathbb{P}}{\to} X$.

(c) The sequence $(X_n)_{n\in\mathbb{N}}$ *converges in distribution*, if $\mathbb{E}\big(f(X_n)\big) \stackrel{n\to\infty}{\longrightarrow} \mathbb{E}\big(f(X)\big)$ for any continuous and bounded function $f : \mathbb{R} \to \mathbb{R}$. We write $X_n \stackrel{n\to\infty}{\longrightarrow} X$ in $distribution$, or briefly, $X_n \stackrel{d}{\to} X$.

(d) The sequence $(X_n)_{n\in\mathbb{N}}$ *converges in L_p*, if $\lim_{n\to\infty} \mathbb{E}|X_n - X|^p = 0$. We write $X_n \stackrel{n\to\infty}{\longrightarrow} X$ in L_p, or briefly, $X_n \stackrel{L_p}{\longrightarrow} X$. □

§1.5.2 **Remark.** In (a) the set $C = \bigcap_{k\geqslant 1} \bigcup_{n\geqslant 1} \bigcap_{i\geqslant 1} \{|X_{n+i}(\omega) - X_n(\omega)| < 1/k\}$ is measurable. Moreover, if $\mathbb{P}(C) = 1$ then there exists a r.v. $X \in \mathcal{M}$ such that $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$ where $X = \limsup\limits_{n\to\infty} X_n$ noting that $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ for $\omega \in C$. □

§1.5.3 **Properties.** (a) *We have $X_n \stackrel{a.s.}{\longrightarrow} X$ if and only if $\sup_{m>n} |X_m - X_n| \stackrel{n\to\infty}{\longrightarrow} 0$ in $\mathbb{P}$ if and only if $\sup_{j\geqslant n} |X_j - X| \stackrel{n\to\infty}{\longrightarrow} 0$ in $\mathbb{P}$ if and only if $\forall \varepsilon, \delta > 0$, $\exists N(\varepsilon, \delta) \in \mathbb{N}$, $\forall\, n \geqslant N(\varepsilon, \delta)$, $\mathbb{P}\big(\bigcap_{j\geqslant n} \{|X_j - X| \leqslant \varepsilon\}\big) \geqslant 1 - \delta$.*

(b) *If $X_n \stackrel{a.s.}{\longrightarrow} X$, then $X_n \stackrel{\mathbb{P}}{\to} X$.*

(c) *If $X_n \stackrel{a.s.}{\longrightarrow} X$, then $g(X_n) \stackrel{a.s.}{\longrightarrow} g(X)$ for any continuous function g.*

(d) *$X_n \stackrel{\mathbb{P}}{\to} X$ if and only if $\lim_{n\to\infty} \sup_{j\geqslant n} \mathbb{P}(|X_j - X_n| > \varepsilon) = 0$ for all $\varepsilon > 0$ if and only if any sub-sequence of $(X_n)_{n\in\mathbb{N}}$ contains a sub-sequence converging to X a.s..*

(e) *If $X_n \stackrel{\mathbb{P}}{\to} X$, then $g(X_n) \stackrel{\mathbb{P}}{\to} g(X)$ for any continuous function g.*

(f) *$X_n \stackrel{a.s.}{\longrightarrow} X \Rightarrow X_n \stackrel{\mathbb{P}}{\to} X \Leftarrow X_n \stackrel{L_p}{\longrightarrow} X$ and $X_n \stackrel{\mathbb{P}}{\to} X \Rightarrow X_n \stackrel{d}{\to} X$* □

1.6 Conditional expectation

In the sequel $(\Omega, \mathscr{A}, \mathbb{P})$ is a probability space and $\mathscr{F}$ is a sub-σ-algebra of $\mathscr{A}$.

§1.6.1 **Theorem.** *If $X \in \mathcal{M}^+(\Omega, \mathscr{A})$ or $X \in L_1(\Omega, \mathscr{A}, \mathbb{P})$ then there exists $Y \in \mathcal{M}^+(\Omega, \mathscr{F})$ or $Y \in L_1(\Omega, \mathscr{F}, \mathbb{P})$, respectively, such that $\mathbb{E}(X\mathbb{1}_F) = \mathbb{E}(Y\mathbb{1}_F)$ for all $F \in \mathscr{F}$, moreover Y is unique up to equality a.s..*

Proof of Theorem §1.6.1. We refer to Klenke [2008], Theorem 8.12. □

§1.6.2 **Definition.** For $X \in \mathcal{M}^+(\Omega, \mathscr{A})$ or $X \in L_1(\Omega, \mathscr{A}, \mathbb{P})$ each version Y as in Theorem §1.6.1 is called *conditional expectation* (bedingte Erwartung) of X given $\mathscr{F}$, symbolically

$\mathbb{E}(X|\mathscr{F}) := Y$. For $A \in \mathscr{A}$, $\mathbb{P}(A|\mathscr{F}) := \mathbb{E}(\mathbb{1}_A|\mathscr{F})$ is called a *conditional probability* of A given the σ-algebra $\mathscr{F}$. Given r.v.'s X_i, $i \in \mathcal{I}$, we set $\mathbb{E}(X|(X_i)_{i\in\mathcal{I}}) := \mathbb{E}(X|\sigma(X_i, i \in \mathcal{I}))$. □

§1.6.3 **Remark**. Employing Proposition §1.2.7 there exists a $\mathscr{B}$-$\mathscr{B}$-measurable function f such that $\mathbb{E}(Y|X) = f(X)$ a.s.. Therewith, we write $\mathbb{E}(Y|X = x) := f(x)$ (conditional expected value, bedingter Erwartungswert). Since conditional expectations are defined only up to equality a.s., all (in)equalities with conditional expectations are understood as (in)equalities a.s., even if we do not say so explicitly. □

§1.6.4 **Properties**. *Let $\mathscr{G} \subset \mathscr{F} \subset \mathscr{A}$ be σ-algebras and let $X, Y \in L_1(\Omega, \mathscr{A}, \mathbb{P})$. Then:*

(a) *(Linearity)* $\mathbb{E}(\lambda X + Y|\mathscr{F}) = \lambda\mathbb{E}(X|\mathscr{F}) + \mathbb{E}(Y|\mathscr{F})$.

(b) *(Monotonicity) If $X \geqslant Y$ a.s., then $\mathbb{E}(X|\mathscr{F}) \geqslant \mathbb{E}(Y|\mathscr{F})$.*

(c) *If $\mathbb{E}(|XY|) < \infty$ and Y is measurable with respect to $\mathscr{F}$, then $\mathbb{E}(XY|\mathscr{F}) = Y\mathbb{E}(X|\mathscr{F})$ and $\mathbb{E}(Y|\mathscr{F}) = \mathbb{E}(Y|Y) = Y$.*

(d) *(Tower property)* $\mathbb{E}(\mathbb{E}(X|\mathscr{F})|\mathscr{G}) = \mathbb{E}(\mathbb{E}(X|\mathscr{G})|\mathscr{F}) = \mathbb{E}(X|\mathscr{G})$.

(e) *(Triangle inequality)* $\mathbb{E}(|X|\,|\,\mathscr{F}) \geqslant |\mathbb{E}(X|\mathscr{F})|$.

(f) *(Independence) If $\sigma(X)$ and $\mathscr{F}$ are independent, then $\mathbb{E}(X|\mathscr{F}) = \mathbb{E}(X)$.*

(g) *If $\mathbb{P}(A) \in \{0, 1\}$ for any $A \in \mathscr{F}$, then $\mathbb{E}(X|\mathscr{F}) = \mathbb{E}(X)$.*

(h) *(Jensen's inequality) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be convex and let $\varphi(Y)$ be an element of $L_1(\Omega, \mathscr{A}, \mathbb{P})$. Then $\varphi(\mathbb{E}(Y|\mathscr{F})) \leqslant \mathbb{E}(\varphi(Y)|\mathscr{F})$.*

(i) *Assume $X, X_1, X_2, \ldots \in \mathcal{M}^+$ such that $X_n \uparrow X$ then $\sup_{n\in\mathbb{N}} \mathbb{E}[X_n|\mathscr{F}] = \mathbb{E}[X|\mathscr{F}]$.*

(j) *(Dominated convergence) Assume $Y \in L_1(\mathbb{P})$, $Y \geqslant 0$ and $(X_n)_{n\in\mathbb{N}}$ is a sequence of r.v.'s with $|X_n| \leqslant Y$ for $n \in \mathbb{N}$ and such that $X_n \xrightarrow{a.s.} X$. Then $\lim_{n\to\infty} \mathbb{E}(X_n|\mathscr{F}) = \mathbb{E}(X|\mathscr{F})$ a.s. and in $L_1(\mathbb{P})$.* □

§1.6.5 **Proposition**. *Let $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ be a Hilbert space equipped with induced norm $\|\cdot\|_{\mathbb{H}}$ and let $\mathcal{U}$ be a closed linear subspace of $\mathbb{H}$. For each $x \in \mathbb{H}$ there exists a unique element $u_x \in \mathcal{U}$ with $\|x - u_x\|_{\mathbb{H}} = \inf_{u\in\mathcal{U}} \|x - u\|_{\mathbb{H}}$.* □

§1.6.6 **Definition**. For a closed subspace $\mathcal{U}$ of the Hilbert space $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ the *orthogonal projection* $\Pi_{\mathcal{U}} : \mathbb{H} \to \mathcal{U}$ is defined by $\Pi_{\mathcal{U}}(x) = u_x$ with u_x as in Proposition §1.6.5. □

§1.6.7 **Properties**. *Let $\mathcal{U}^\perp$ be the orthogonal complement of $\mathcal{U}$ in $\mathbb{H}$. Then:*

(a) *(projection property)* $\Pi_{\mathcal{U}} \circ \Pi_{\mathcal{U}} = \Pi_{\mathcal{U}}$;

(b) *(orthogonality)* $x - \Pi_{\mathcal{U}}x \in \mathcal{U}^\perp$ *for each* $x \in \mathbb{H}$;

(c) *each $x \in \mathbb{H}$ can be decomposed uniquely as $x = \Pi_{\mathcal{U}}x + (x - \Pi_{\mathcal{U}}x)$ in the orthogonal sum of an element of $\mathcal{U}$ and an element of $\mathcal{U}^\perp$;*

(d) *$\Pi_{\mathcal{U}}$ is selfadjoint: $\langle\Pi_{\mathcal{U}}x, y\rangle_{\mathbb{H}} = \langle x, \Pi_{\mathcal{U}}y\rangle_{\mathbb{H}}$;*

(e) *$\Pi_{\mathcal{U}}$ is linear.* □

§1.6.8 **Lemma**. *Let $\mathscr{F}$ be a sub-σ-algebra of $\mathscr{A}$. Then $L_2(\Omega, \mathscr{F}, \mathbb{P})$ is embedded as closed linear subspace in the Hilbert space $L_2(\Omega, \mathscr{A}, \mathbb{P})$.* □

§1.6.9 **Corollary**. *Let $\mathscr{F} \subset \mathscr{A}$ be a sub-σ-algebra and let $X \in L_2(\Omega, \mathscr{A}, \mathbb{P})$ be a r.v.. Then $\mathbb{E}(X|\mathscr{F})$ is the orthogonal projection of X on $L_2(\Omega, \mathscr{F}, \mathbb{P})$. That is, for any $Y \in L_2(\Omega, \mathscr{F}, \mathbb{P})$, $\|X - Y\|_2^2 = \mathbb{E}[(X-Y)^2] \geqslant \mathbb{E}[(X - \mathbb{E}(X|\mathscr{F}))^2] = \|X - \mathbb{E}(X|\mathscr{F})\|_2^2$ with equality if and only if $Y = \mathbb{E}(X|\mathscr{F})$.* □

§1.6.10 **Example**. Let $X, Y \in L_1(\mathbb{P})$ be independent. Then $\mathbb{E}(X + Y|Y) = \mathbb{E}(X|Y) + \mathbb{E}(Y|Y) = \mathbb{E}(X) + Y$. □

§1.6.11 **Theorem**. *Let $p \in [1, \infty]$ and $\mathscr{F} \subset \mathscr{A}$ be a sub-σ-algebra. Then the linear map $L_p(\Omega, \mathscr{A}, \mathbb{P}) \to L_p(\Omega, \mathscr{F}, \mathbb{P})$, $X \mapsto \mathbb{E}(X|\mathscr{F})$, is a contraction (that is, $\|\mathbb{E}(X|\mathscr{F})\|_p \leqslant \|X\|_p$) and thus bounded and continuous. Hence, for $X, X_1, X_2, \ldots \in L_p(\Omega, \mathscr{A}, \mathbb{P})$ with $\|X_n - X\|_p \xrightarrow{n\to\infty} 0$ we have $\|\mathbb{E}(X_n|\mathscr{F}) - \mathbb{E}(X|\mathscr{F})\|_p \xrightarrow{n\to\infty} 0$.* □

§1.6.12 **Definition**. A family $(X_i)_{i\in\mathcal{I}}$ of r.v.'s in $L_1(\Omega, \mathscr{A}, \mathbb{P})$ with arbitrary index set $\mathcal{I}$ is called *uniformly integrable* if $\inf_{a\in[0,\infty)} \sup_{i\in\mathcal{I}} \mathbb{E}(\mathbb{1}_{\{|X_i|>a\}}|X_i|) = 0$ which is satisfied in case that $\sup_{i\in\mathcal{I}} \|X_i\|_1 \in L_1(\Omega, \mathscr{A}, \mathbb{P})$. □

§1.6.13 **Corollary**. *Let $(X_i)_{i\in\mathcal{I}}$ be uniformly integrable in $L_1(\Omega, \mathscr{A}, \mathbb{P})$ and let $(\mathscr{F}_j, j \in \mathcal{J})$ be a family of sub-σ-algebras of $\mathscr{A}$. Define $X_{i,j} := \mathbb{E}(X_i|\mathscr{F}_j)$. Then $(X_{i,j})_{i\in\mathcal{I}, j\in\mathcal{J}}$ is uniformly integrable in $L_1(\Omega, \mathscr{A}, \mathbb{P})$. In particular, for $X \in L_1(\Omega, \mathscr{A}, \mathbb{P})$ the family $\{\mathbb{E}(X|\mathscr{F}) : \mathscr{F}$ is sub-σ-algebra of $\mathscr{A}\}$ of r.v.'s in $L_1(\Omega, \mathscr{A}, \mathbb{P})$ is uniformly integrable.* □

§1.6.14 **Lemma**. *Every uniformly integrable sequence $(X_n)_{n\in\mathbb{N}}$ of real r.v.'s which converges a.s. also converges in L_1.*

Proof of Lemma §1.6.14 is given in the lecture. □

Chapter 2

Stochastic processes

2.1 Motivating examples

2.1.1 The Poisson process

§2.1.1 **Definition**. Let $(S_k)_{k\in\mathbb{N}}$ be positive r.v.'s on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with $0 \leqslant S_1(\omega) \leqslant S_2(\omega) \leqslant \dots$ for any $\omega \in \Omega$. The family $N = (N_t)_{t\geqslant 0}$ of $\mathbb{N}_o$-valued r.v.'s given by $N_t := \sum_{k=1}^{\infty} \mathbb{1}_{\{S_k \leqslant t\}}$, $t \geqslant 0$, is called *counting process* (Zählprozess) with *jump times* (Sprungzeiten) $(S_k)_{k\in\mathbb{N}}$. □

§2.1.2 **Definition**. A counting process $(N_t)_{t\geqslant 0}$ is called *Poisson process* of intensity $\lambda > 0$ if

(i) $\mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h)$ as $h \downarrow 0$;

(ii) $\mathbb{P}(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h)$ as $h \downarrow 0$;

(iii) *(independent increments)* $(N_{t_i} - N_{t_{i-1}})_{i=1}^n$ are independent for any numbers $0 = t_0 < t_1 < \dots < t_n$ in $\mathbb{R}^+$;

(iv) *(stationary increments)* $N_t - N_s \stackrel{d}{=} N_{t-s}$ for all numbers $t \geqslant s \geqslant 0$ in $\mathbb{R}^+$. □

§2.1.3 **Theorem**. *For a counting process $N = (N_t)_{t\geqslant 0}$ with jump times $(S_k)_{k\in\mathbb{N}}$ we have the equivalences between:*

(a) *N is a Poisson process;*

(b) *N satisfies the conditions (iii), (iv) in the Definition §2.1.2 of a Poisson ($\mathfrak{Poi}$) process and $N_t \sim \mathfrak{Poi}(\lambda t)$ holds for all $t > 0$;*

(c) *(waiting times) The r.v.'s $T_1 := S_1$ and $T_k := S_k - S_{k-1}$, $k = 2, 3, \dots$, are independent and identically $\mathfrak{Exp}(\lambda)$-distributed;*

(d) *$N_t \sim \mathfrak{Poi}(\lambda t)$ holds for all $t > 0$ and the conditional distribution of $(S_1, \dots, S_n)$ given $\{N_t = n\}$ has the density*

$$f(x_1, \dots, x_n) = \frac{n!}{t^n} \mathbb{1}_{\{0\leqslant x_1 \leqslant \dots \leqslant x_n \leqslant t\}}. \tag{2.1}$$

(e) *N satisfies the condition (iii) in the Definition §2.1.2 of a Poisson process, $\mathbb{E}(N_1) = \lambda$ and (2.1) is the conditional density of $(S_1, \dots, S_n)$ given $\{N_t = n\}$.*

Proof of Theorem §2.1.3 is given in the lecture. □

§2.1.4 **Remark**. Let $(U_i)_{i=1}^n$ be independent and identically $\mathfrak{U}([0,t])$-distributed r.v.'s and let $(U_{(i)})_{i=1}^n$ be their order statistics where $U_{(1)} = \min\{U_i\}_{i=1}^n$ and $U_{(k+1)} = \min\{U_i\}_{i=1}^n \setminus \{U_{(i)}\}_{i=1}^k$, $k = 2, \dots, n$. Then the joint density of $(U_{(i)})_{i=1}^n$ is given exactly by (2.1). The characterisations give rise to three simple methods to simulate a Poisson process: the definition §2.1.2 gives an approximation for small h (forgetting the $o(h)$-term), part (iii) in §2.1.3 just uses exponentially distributed inter-arrival times T_k and part (iv) uses the value at a specified right-end point and then uses the uniform order statistics as jump times in-between (write down the details!). □

2.1.2 Markov chains

§2.1.5 **Definition.** Let $\mathbb{T} = \mathbb{N}_0$ (discrete time) or $\mathbb{T} = [0,\infty)$ (continuous time), let $\mathcal{S}$ be a (at most) countable nonempty set (state space) and let $\mathscr{S} = 2^{\mathcal{S}}$. A family $(X_t)_{t\in\mathbb{T}}$ of $\mathcal{S}$-valued r.v.'s forms a *Markov chain* if for all $n \in \mathbb{N}$, all $t_1 < t_2 < \ldots < t_n < t$ in $\mathbb{T}$ and all $s_1,\ldots,s_n,s$ in $\mathcal{S}$ with $\mathbb{P}(X_{t_1} = s_1,\ldots,X_{t_n} = s_n) > 0$ the *Markov property* is satisfied: $\mathbb{P}(X_t = s|X_{t_1} = s_1,\ldots,X_{t_n} = s_n) = \mathbb{P}(X_t = s|X_{t_n} = s_n)$. For a Markov chain $(X_t)_{t\in\mathbb{T}}$ and $t_1 \leqslant t_2$ in $\mathbb{T}$, $i,j \in \mathcal{S}$ the *transition probability* to reach state j at time t_2 from state i at time t_1 is defined by $p_{ij}(t_1,t_2) := \mathbb{P}(X_{t_2} = j|X_{t_1} = i)$ (or arbitrary if not well-defined). The *transition matrix* is given by $P(t_1,t_2) := \big(p_{ij}(t_1,t_2)\big)_{i,j\in\mathcal{S}}$. The transition matrix and the Markov chain are called *time-homogeneous* if$P(t_1,t_2) = P(0,t_2-t_1) =: P(t_2-t_1)$ holds for all $t_1 \leqslant t_2$. □

§2.1.6 **Proposition.** *The transition matrices satisfy the Chapman-Kolmogorov equation, that is, for any $t_1 \leqslant t_2 \leqslant t_3$ in $\mathbb{T}$, $P(t_1,t_3) = P(t_1,t_2)P(t_2,t_3)$ (matrix multiplication). In the time-homogeneous case this gives the semigroup property $P(t_1+t_2) = P(t_1)P(t_2)$ for all $t_1,t_2 \in \mathbb{T}$, and in particular $P(n) = P(1)^n$ for $n \in \mathbb{N}$.*

Proof of Proposition §2.1.6 is given in the lecture. □

2.1.3 Brownian motion

§2.1.7 **Definition.** A family $(W_t)_{t\geqslant 0}$ of real r.v.'s is called a *Brownian motion* if

(a) $W_0 = 0$ a.s.;

(b) *(independent increments)* $(W_{t_i} - W_{t_{i-1}})_{i=1}^n$ are independent for any numbers $0 = t_0 < t_1 < \ldots < t_n$ in $\mathbb{R}^+$;

(c) *(stationary increments)* $W_t - W_s \overset{d}{=} W_{t-s} \sim \mathfrak{N}(0,t-s)$ for all numbers $0 \leqslant s < t$ in $\mathbb{R}^+$;

(d) $t \mapsto W_t$ is continuous a.s.. □

§2.1.8 **Remark.** Questions:

(i) Existence?

(ii) $W := (W_t)_{t\geqslant 0}$ r.v. on which space?

(iii) For which functions f is $f(W)$ a r.v.? (e.g. $f(W) = \sup_{0\leqslant t\leqslant 1} W_t$)

Importance of the Brownian motion:

- If $X_1, X_2, \ldots$ are i.i.d. with $\mathbb{E}(X_i) = 0$ and $\mathbb{V}\mathrm{ar}(X_i) = \sigma^2 < \infty$ then W is a "limit" of $S_t^n = \frac{1}{\sigma\sqrt{n}}\sum_{1\leqslant i\leqslant nt} X_i$ (Donsker's theorem).
- W is a central element in stochastic differential equations $X_t = \int_0^t \sigma(X_s)dW_s + \int_0^t b(X_s)ds$. How to define the first integral? ("Ito integral") □

2.2 Definition of stochastic processes

§2.2.1 **Definition.** A family $X = (X_t)_{t\in\mathbb{T}}$ of r.v.'s on a common probability space $(\Omega,\mathscr{A},\mathbb{P})$ is called *stochastic process*. We call X *time-discrete* if $\mathbb{T} \subset \mathbb{Z}$ and *time-continuous* if $(a,b) \subset$

$\mathbb{T} \subset \mathbb{R}$ for some real numbers $a < b$. If all X_t take values in $(\mathcal{S}, \mathscr{S})$, then $(\mathcal{S}, \mathscr{S})$ is called the *state space* (Zustandsraum) of X. For each fixed $\omega \in \Omega$ the map $t \mapsto X_t(\omega)$ is called *sample path* (Pfad), *trajectory* (Trajektorie) or *realisation* (Realisierung) of X. If $\mathbb{T} = \mathbb{N}_0$ or $\mathbb{T} = \mathbb{R}^+$ the law of X_0 is called *initial distribution*. □

§2.2.2 **Remark.** We are particularly interested in the "random functions" $t \mapsto X_t$ rather than in a single r.v. X_t. For this reason, we identify $X = (X_t)_{t \in \mathbb{T}}$ as a r.v. with values in $\mathcal{S}^{\mathbb{T}}$ which forces us to specify a σ-algebra on $\mathcal{S}^{\mathbb{T}}$. □

§2.2.3 **Definition.** Let $(\mathcal{S}_i, \mathscr{S}_i)$, $i \in \mathcal{I}$, be an arbitrary family of measurable spaces.

(a) The set $\times_{i \in \mathcal{I}} \mathcal{S}_i$ of maps $(s_i)_{i \in \mathcal{I}} : \mathcal{I} \to \cup_{i \in \mathcal{I}} \mathcal{S}_i$ such that $s_i \in \mathcal{S}_i$ for all $i \in \mathcal{I}$ is called *product space*. For $\mathcal{J} \subset \mathcal{I}$, let $\mathcal{S}_{\mathcal{J}} := \times_{j \in \mathcal{J}} \mathcal{S}_j$. If, in particular, all the $\mathcal{S}_i$ are equal, say $\mathcal{S}_i = \mathcal{S}$, then we write $\times_{i \in \mathcal{I}} \mathcal{S}_i = \mathcal{S}^{\mathcal{I}}$.

(b) If $j \in \mathcal{I}$, then $\Pi_j : \mathcal{S}_{\mathcal{I}} \to \mathcal{S}_j$, $(s_i)_{i \in \mathcal{I}} \mapsto s_j$ denotes the jth *coordinate map*. More generally, for $\mathcal{J} \subset \mathcal{K} \subset \mathcal{I}$, the restricted map $\Pi^{\mathcal{K}}_{\mathcal{J}} : \mathcal{S}_{\mathcal{K}} \to \mathcal{S}_{\mathcal{J}}$, $(s_k)_{k \in \mathcal{K}} \mapsto (s_j)_{j \in \mathcal{J}}$ are called *canonical projection*. In particular, we write $\Pi_{\mathcal{J}} := \Pi^{\mathcal{I}}_{\mathcal{J}}$.

(c) The product-σ-algebra $\mathscr{S}_{\mathcal{I}} := \bigotimes_{i \in \mathcal{I}} \mathscr{S}_i$ is the smallest σ-algebra on the product space $\mathcal{S}_{\mathcal{I}}$ such that for every $j \in \mathcal{I}$ the coordinate map $\Pi_j : \mathcal{S}_{\mathcal{I}} \to \mathcal{S}_j$ is measurable with respect to $\mathscr{S}_{\mathcal{I}}$-$\mathscr{S}_j$, that is, $\mathscr{S}_{\mathcal{I}} = \bigotimes_{i \in \mathcal{I}} \mathscr{S}_i = \sigma(\Pi_i, i \in \mathcal{I}) := \bigvee_{i \in \mathcal{I}} \Pi_i^{-1}(\mathscr{S}_i)$. For $\mathcal{J} \subset \mathcal{I}$, let $\mathscr{S}_{\mathcal{J}} = \bigotimes_{j \in \mathcal{J}} \mathscr{S}_j$. If $(\mathcal{S}_i, \mathscr{S}_i) = (\mathcal{S}, \mathscr{S})$ for all $i \in \mathcal{I}$, then we also write $\bigotimes_{i \in \mathcal{I}} \mathscr{S}_i = \mathscr{S}^{\otimes \mathcal{I}}$. □

§2.2.4 **Lemma.** *For a stochastic process $X = (X_t)_{t \in \mathbb{T}}$ with state space $(\mathcal{S}, \mathscr{S})$ the mapping $X : \Omega \to \mathcal{S}^{\mathbb{T}}$, $\omega \mapsto (X_t(\omega))_{t \in \mathbb{T}}$ is a $(\mathcal{S}^{\mathbb{T}}, \mathscr{S}^{\otimes \mathbb{T}})$-valued r.v.*

Proof of Lemma §2.2.4 is given in the lecture. □

§2.2.5 **Remark.** Later on, we shall also consider smaller function spaces than $\mathcal{S}^{\mathbb{T}}$, e.g. $C(\mathbb{R}^+)$ instead of $\mathbb{R}^{\mathbb{R}^+}$. □

§2.2.6 **Definition.** The distribution $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ of a stochastic process $X = (X_t)_{t \in \mathbb{T}}$ defined on $(\Omega, \mathscr{A}, \mathbb{P})$ with values in $(\mathcal{S}^{\mathbb{T}}, \mathscr{S}^{\otimes \mathbb{T}})$ is the image probability measure of $\mathbb{P}$ under the map X. □

§2.2.7 **Remark.** The distribution of a stochastic process is often complicate and generally there does not exists an explicit formula. Therefore, we are interested in a characterisation exploiting the distributions of the r.v.'s X_t. □

§2.2.8 **Definition.** Let $X = (X_t)_{t \in \mathbb{T}}$ be a stochastic process with distribution $\mathbb{P}_X$. For any finite $\mathcal{T} \subset \mathbb{T}$ let $\mathbb{P}^{\mathcal{T}}_X := \mathbb{P}_{\Pi_{\mathcal{T}} \circ X}$ be the distribution of the r.v. $(X_t)_{t \in \mathcal{T}} = \Pi_{\mathcal{T}} \circ X$. The family $\{\mathbb{P}^{\mathcal{T}}_X, \mathcal{T} \subset \mathbb{T} \text{ finite}\}$ is called family of the *finite-dimensional distributions* of X or $\mathbb{P}_X$. □

§2.2.9 **Definition.** A family $\{\mathbb{P}_{\mathcal{J}}, \mathcal{J} \subset \mathcal{I} \text{ finite}\}$ of probability measures is called *consistent* on $(\mathcal{S}_{\mathcal{I}}, \mathscr{S}_{\mathcal{I}})$ if for any finite $\mathcal{J} \subset \mathcal{K} \subset \mathcal{I}$ the canonical projection $\Pi^{\mathcal{K}}_{\mathcal{J}}$ as in §2.2.3 (c) and the probability measure $\mathbb{P}_{\mathcal{J}}$ and $\mathbb{P}_{\mathcal{K}}$ on $(\mathcal{S}_{\mathcal{J}}, \mathscr{S}_{\mathcal{J}})$ and $(\mathcal{S}_{\mathcal{K}}, \mathscr{S}_{\mathcal{K}})$, respectively, satisfy $\mathbb{P}_{\mathcal{J}} = \mathbb{P}_{\mathcal{K}} \circ (\Pi^{\mathcal{K}}_{\mathcal{J}})^{-1}$. □

§2.2.10 **Remark.** Let $\mathbb{P}_X$ be the distribution of a stochastic process X on $(\mathcal{S}^{\mathbb{T}}, \mathscr{S}^{\otimes \mathbb{T}})$ then its family $\{\mathbb{P}^{\mathcal{T}}_X, \mathcal{T} \subset \mathbb{T} \text{ finite}\}$ of finite-dimensional distributions is consistent. Indeed, for $\mathcal{J} \subset$

$\mathcal{K} \subset \mathcal{I}$ finite, $\mathbb{P}_X^{\mathcal{J}} = \mathbb{P}_X \circ \Pi_{\mathcal{J}}^{-1} = \mathbb{P}_X \circ (\Pi_{\mathcal{J}}^{\mathcal{K}} \circ \Pi_{\mathcal{K}})^{-1} = \mathbb{P}_X \circ (\Pi_{\mathcal{K}})^{-1} \circ (\Pi_{\mathcal{J}}^{\mathcal{K}})^{-1} = \mathbb{P}_X^{\mathcal{K}} \circ (\Pi_{\mathcal{J}}^{\mathcal{K}})^{-1}$. □

§2.2.11 **Definition.** Two processes $(X_t)_{t\in\mathbb{T}}$, $(Y_t)_{t\in\mathbb{T}}$ on $(\Omega, \mathscr{A}, \mathbb{P})$ are called

(a) *indistinguishable* (ununterscheidbar) if $\mathbb{P}(\forall\, t \in \mathbb{T} : X_t = Y_t) = 1$;

(b) *versions* or *modifications* (Versionen, Modifikationen) of each other, if $\mathbb{P}(X_t = Y_t) = 1$ for all $t \in \mathbb{T}$. □

§2.2.12 **Remark.** (a) Obviously, indistinguishable processes are versions of each other. The converse is in general false.

(b) If X is a version of Y, then X and Y share the same finite-dimensional distributions. Processes with the same finite-dimensional distributions need not even be defined on the same probability space and will in general not be versions of each other.

(c) Suppose $(X_t)_{t\in\mathbb{R}^+}$ and $(Y_t)_{t\in\mathbb{R}^+}$ are real-valued stochastic processes with right-continuous sample paths. Then they are indistinguishable already if they are versions of each other. □

§2.2.13 **Definition.** A stochastic processes $(X_t)_{t\in\mathbb{R}^+}$ is called *continuous* if all sample paths are continuous. It is called *stochastically continuous*, if $t_n \xrightarrow{n\to\infty} t$ always implies $X_{t_n} \xrightarrow{\mathbb{P}} X_t$ (convergence in probability). □

§2.2.14 **Remark.** Every continuous stochastic process is stochastically continuous since a.s. convergence implies convergence in probability. On the other hand, the Poisson process is obviously not continuous but stochastically continuous, since $\lim_{t_n\to t} \mathbb{P}(|N_t - N_{t_n}| > \varepsilon) = \lim_{t_n\to t}(1 - e^{-\lambda|t-t_n|}) = 0$ for all $\varepsilon \in (0, 1)$. □

2.3 Probability measures on Polish spaces

§2.3.1 **Definition.** A metric space $(\mathcal{S}, d)$ is called *Polish space* if it is *separable* and *complete*. More generally, a separable completely metrisable topological space is called *Polish*. Canonically, it is equipped with its Borel-σ-algebra $\mathscr{B}(\mathcal{S})$ generated by the open sets. □

§2.3.2 **Remark.** Let (Ω, τ) be a topological space. For $A \subset \Omega$ we denote by $\overline{A}$ the closure of A, by A° the interior and by ∂A the boundary of A. A set $A \subset \Omega$ is called *dense* if $\overline{A} = \Omega$. A set $A \subset \Omega$ is called *compact* if each open cover $\mathcal{U} \subset \tau$ of A (that is, $A \subset \cup\{U; U \in \mathcal{U}\}$) has a finite subcover; that is, a finite $\mathcal{U}' \subset \mathcal{U}$ with $A \subset \cup\{U; U \in \mathcal{U}'\}$. Compact sets are closed. $A \subset \Omega$ is called *relatively compact* if $\overline{A}$ is compact. On the other hand, A is called *sequentially compact* (respectively *relatively sequentially compact*) if any sequence $(\omega_n)_{n\in\mathbb{N}}$ with values in A has a subsequence $(\omega_{n_k})_{k\in\mathbb{N}}$ that converges to some $\omega \in A$ (respectively $\omega \in \overline{A}$).

(Ω, τ) is called *metrisable* if there exists a metric d on Ω such that τ is induced by the open balls $B_\varepsilon(x) = \{\omega \in \Omega : d(x, \omega) < \varepsilon\}$. In metrisable spaces, the notions compact and sequentially compact coincide. A metric d on Ω is called *complete* if any Cauchy sequence with respect to d converges in Ω. (Ω, τ) is called *completely metrisable* if there exists a complete metric on Ω that induces τ. A metrisable space (Ω, τ) is called *separable* if there exists a countable dense subset of Ω. Separability in metrisable spaces is equivalent to the existence of a countable base of the topology; that is, a countable set $\mathcal{U} \subset \tau$ with $A = \bigcup\{U; U \subset A, U \in \mathcal{U}\}$ for all $A \in \tau$. A compact metric space is always separable (simply choose the union of finite covers comprising balls of radius $1/n$).

Two measurable spaces $(\Omega_1, \mathscr{B}_1)$, $(\Omega_2, \mathscr{B}_2)$ with Borel-σ-algebra $\mathscr{B}_1$, $\mathscr{B}_2$, respectively, are called *Borel-isomorphic*, if there exists a bijective map $g : \Omega_1 \to \Omega_2$, such that g and g^{-1} are measurable. In particular, each Polish space is Borel-isomorphic to a Borel subset of $[0, 1]$.

Two topological spaces (Ω_1, τ_1) (Ω_2, τ_2) are called *homeomorphic* if there exists a bijective map $g : \Omega_1 \to \Omega_2$ such that g and g^{-1} are continuous. Therewith, each Polish space is homeomorphic to a subset of $[0, 1]^{\mathbb{N}}$, equipped with its product topology. □

§2.3.3 **Examples**. $\mathbb{R}$, $\mathbb{R}^n$, $\ell_p \subset \mathbb{R}^{\mathbb{N}}$ and $L_p([0, 1])$ equipped with their usual distance are Polish spaces. □

§2.3.4 **Definition**. Let $(\mathcal{S}_i, d_i)$, $i \in \mathcal{I} \subset \mathbb{N}$, be a finite or countable family of metric spaces. The *product space* $\bigtimes_{i\in\mathcal{I}} \mathcal{S}_i$ is canonically equipped with the *product metric* $d((s_i)_{i\in\mathcal{I}}, (s'_i)_{i\in\mathcal{I}}) := \sum_{i\in\mathcal{I}} 2^{-i}(d_i(s_i, s'_i) \wedge 1)$ generating the product topology on $\bigtimes_{i\in\mathcal{I}} \mathcal{S}_i$ in which a vector/sequence converges if and only if all coordinates converge, that is, $d(s^{(n)}, s) \xrightarrow{n\to\infty} 0 \Leftrightarrow d_i(s_i^{(n)}, s_i) \xrightarrow{n\to\infty} 0$ for all $i \in \mathcal{I}$. □

§2.3.5 **Lemma**. *Let $(\mathcal{S}_n, d_n)$, $n \in \mathbb{N}$, be a family of Polish spaces, then the Borel-σ-Algebra $\mathscr{B}(\bigtimes_{n\in\mathbb{N}} \mathcal{S}_n)$ on the product space $\bigtimes_{n\in\mathbb{N}} \mathcal{S}_n$ equals the product Borel-σ-algebra $\bigotimes_{n\in\mathbb{N}} \mathscr{B}(\mathcal{S}_n)$.*
Proof of Lemma §2.3.5 is given in the lecture. □

§2.3.6 **Remark**. The $\supseteq$-relation holds for all topological spaces and products of any cardinality with the same proof. The $\subseteq$-property can already fail for the product of two topological (non-Polish) spaces. □

§2.3.7 **Definition**. Let $(\mathcal{S}, d)$ be a metric space equipped with its Borel-σ-algebra $\mathscr{B}(\mathcal{S})$. A probability measure $\mathbb{P}$ on $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$ is called

(a) *tight* (straff) if for all $\varepsilon > 0$ there is a compact set K such that $\mathbb{P}(K) \geqslant 1 - \varepsilon$,

(b) *regular* (regulär) if $B \in \mathscr{B}(\mathcal{S})$ and $\varepsilon > 0$ then there exist a compact set K and an open set O such that $K \subset B \subset O$ and $\mathbb{P}(O \backslash K) \leqslant \varepsilon$.

A family $\mathcal{P}$ of probability measures on $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$ is called (uniformly) tight, if for all $\varepsilon > 0$ there is a compact set K such that $\mathbb{P}(K) \geqslant 1 - \varepsilon$ for all $\mathbb{P} \in \mathcal{P}$. □

§2.3.8 **Remark**. Considering a probability measure $\mathbb{P}$ on a metric space $\mathcal{S}$ we have the equivalences between (i) $\mathbb{P}$ is tight and (ii) $\mathbb{P}(B) = \sup\{\mathbb{P}(K) : K \subseteq B \text{ compact}\}$ for all $B \in \mathscr{B}(\mathcal{S})$, and on the other hand between (i) $\mathbb{P}$ is regular and (ii) $\sup\{\mathbb{P}(K) : K \subseteq B \text{ compact}\} = \mathbb{P}(B) = \inf\{\mathbb{P}(O) : O \supseteq B \text{ open}\}$ for all $B \in \mathscr{B}(\mathcal{S})$. □

§2.3.9 **Proposition** (Ulam (1939)). *Every probability measure on a Polish space is tight.*
Proof of Proposition §2.3.9 is given in the lecture. □

§2.3.10 **Theorem**. *Every probability measure on a Polish space is regular.*
Proof of Theorem §2.3.10 is given in the lecture. □

§2.3.11 **Theorem** (Kolmogorov's consistency theorem). *Let $\mathcal{I}$ be an arbitrary index set and let $(\mathcal{S}_i, \mathscr{B}_i)$ be Polish spaces, $i \in \mathcal{I}$. Let $\{\mathbb{P}_{\mathcal{J}}, \mathcal{J} \subset \mathcal{I} \text{ finite}\}$ be a consistent family of probability measures on the product space $(\mathcal{S}_{\mathcal{I}}, \mathscr{B}_{\mathcal{I}})$ as in §2.2.9. Then there exists a unique probability measure $\mathbb{P}$ on $(\mathcal{S}_{\mathcal{I}}, \mathscr{B}_{\mathcal{I}})$ having $\{\mathbb{P}_{\mathcal{J}}, \mathcal{J} \subset \mathcal{I} \text{ finite}\}$ as family of finite dimensional distributions, that is, $\mathbb{P}_{\mathcal{J}} = \mathbb{P} \circ \Pi_{\mathcal{J}}^{-1}$ for any $\mathcal{J} \subset \mathcal{I}$ finite.*

Proof of Theorem §2.3.11 is given in the lecture. □

§2.3.12 **Corollary**. *Let $\mathcal{I}$ be an arbitrary index set and let $(\mathcal{S}, \mathscr{B})$ be Polish space. Let $\{\mathbb{P}_{\mathcal{J}}, \mathcal{J} \subset \mathcal{I}$ finite$\}$ be a consistent family of probability measures on the product space $(\mathcal{S}^{\mathcal{I}}, \mathscr{B}^{\otimes \mathcal{I}})$ as in §2.2.9. Then there exists a stochastic process $(X_t)_{t \in \mathcal{I}}$ whose family of finite dimensional distributions is given by $\{\mathbb{P}_{\mathcal{J}}, \mathcal{J} \subset \mathcal{I}$ finite$\}$, that is, $(X_t)_{t \in \mathcal{J}} \sim \mathbb{P}_{\mathcal{J}}$ for any $\mathcal{J} \subset \mathcal{I}$ finite.*

Proof of Corollary §2.3.12 is given in the lecture. □

§2.3.13 **Corollary**. *Let $\mathcal{I}$ be an arbitrary index set and let $(\mathcal{S}, \mathscr{B})$ be Polish space. Let $(\mathbb{P}_i)_{i \in \mathcal{I}}$ be a family of probability measures on $(\mathcal{S}, \mathscr{B})$. Then there exists the product measure $\bigotimes_{i \in \mathcal{I}} \mathbb{P}_i$ on the product space $(\mathcal{S}^{\mathcal{I}}, \mathscr{B}^{\otimes \mathcal{I}})$. In particular, there exists a family $X = (X_i)_{i \in \mathcal{I}}$ of independent r.v.'s admitting the image probability measure $\mathbb{P}_X = \bigotimes_{i \in \mathcal{I}} \mathbb{P}_i$.*

Proof of Corollary §2.3.13 is given in the lecture. □

§2.3.14 **Remark**. Kolmogorov's consistency theorem does not hold for general measure spaces $(\mathcal{S}, \mathscr{S})$. The Ionescu-Tulcea Theorem, however, shows the existence of the probability measure on general measure spaces under a Markovian dependence structure, see e.g. Klenke [2008], Theorem 14.32. □

2.4 Adapted stochastic process and stopping times

In the sequel, the index set $\mathbb{T}$ is a subset of $\mathbb{R}$, $X = (X_t)_{t \in \mathbb{T}}$ is a stochastic process on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with state space $(\mathcal{S}, \mathscr{S})$ and image probability measure $\mathbb{P}_X$ on $(\mathcal{S}^{\mathbb{T}}, \mathscr{S}^{\otimes \mathbb{T}})$.

§2.4.1 **Definition**. A family $\mathscr{F} = (\mathscr{F}_t)_{t \in \mathbb{T}}$ of σ-algebras with $\mathscr{F}_t \subset \mathscr{A}$, $t \in \mathbb{T}$, is called a *filtration* if $\mathscr{F}_s \subset \mathscr{F}_t$ for all $s, t \in \mathbb{T}$ with $s \leqslant t$. $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F})$ is called *filtered probability space*. □

§2.4.2 **Definition**. A stochastic process $X = (X_t)_{t \in \mathbb{T}}$ is called *adapted* to the filtration $\mathscr{F} = (\mathscr{F}_t)_{t \in \mathbb{T}}$ if X_t is $\mathscr{F}_t$-measurable for all $t \in \mathbb{T}$. If $\mathscr{F}_t = \sigma(X_s, s \leqslant t)$ for all $t \in \mathbb{T}$, then we denote by $\mathcal{F}^X = \sigma(X)$ the *natural filtration* generated by X. □

§2.4.3 **Remark**. Clearly, a stochastic process is always adapted to the natural filtration it generates. The natural filtration is the smallest filtration to which the process is adapted. Moreover, $\mathscr{F}_\infty = \bigvee_{t \in \mathbb{T}} \mathscr{F}_t$. □

§2.4.4 **Definition**. A stochastic process $X = (X_n)_{n \in \mathbb{N}_0}$ is called *predictable* (or *previsible*) with respect to a filtration $\mathscr{F} = (\mathscr{F}_n)_{n \in \mathbb{N}_0}$ if X_0 is constant (i.e. $\mathscr{F}_0$-measurable) and if, for every $n \in \mathbb{N}$, X_n is $\mathscr{F}_{n-1}$-measurable. X is called an *increasing* process if it is a predictable process of finite r.v.'s such that $0 = X_0 \leqslant X_1 \leqslant X_2 \leqslant \dots$ a.s. on Ω. □

§2.4.5 **Remark**. It is important to note that for a predictable process and in particular for an increasing process, not only, $(X_n)_{n \in \mathbb{N}_0}$ but also the sequence $(X_{n+1})_{n \in \mathbb{N}_0}$ is adapted to the filtration $(\mathscr{F}_n)_{n \in \mathbb{N}_0}$. □

§2.4.6 **Definition**. A r.v. τ with values in $\mathbb{T} \cup \{\sup\{\mathbb{T}\}\}$ is called a *stopping time* (with respect to the filtration $\mathscr{F}$) if for any $t \in \mathbb{T}$, $\{\tau \leqslant t\} \in \mathscr{F}_t$, that is, if the process $X_t := \mathbb{1}_{\{\tau \leqslant t\}}$ is adapted. □

§2.4.7 **Proposition**. *Let $\mathbb{T}$ be countable, τ is a stopping time if and only if $\{\tau = t\} \in \mathscr{F}_t$ for all $t \in \mathbb{T}$.*

Proof of Proposition §2.4.7 is left as an exercise. □

§2.4.8 **Examples**. (a) Let $t_o \in \mathbb{T}$, then $\tau \equiv t_o$ (constant) is a stopping time where $\sigma(\tau) = \{\emptyset, \Omega\}$.

(b) Let $X = (X_n)_{n \in \mathbb{N}_0}$ be a stochastic process adapted to a filtration $\mathscr{F} = (\mathscr{F}_n)_{n \in \mathbb{N}_0}$. For $S \in \mathscr{S}$ we call *hitting time* the first time that X is in S, that is,

$$\tau_S(\omega) := \begin{cases} \inf\{n \in \mathbb{N}_0 : X_n(\omega) \in S\}, & \text{if } \omega \in \bigcup_{n \in \mathbb{N}_0} X_n^{-1}(S), \\ \infty, & \text{otherwise} \end{cases}$$

Then τ_S is a stopping time with respect to $\mathscr{F}$. Note that $\tau_\emptyset \equiv \infty$ and $\tau_{\mathcal{S}} \equiv 0$. □

§2.4.9 **Lemma**. *Let τ and σ be stopping times. Then*

(a) *$\tau \vee \sigma$ and $\tau \wedge \sigma$ are stopping times.*

(b) *If $\tau, \sigma \geqslant 0$, then $\tau + \sigma$ is also a stopping time.*

(c) *If $s \in \mathbb{R}^+$, then $\tau + s$ is a stopping time. However, in general, $\tau - s$ is not.*

Proof of Lemma §2.4.9 is left as an exercise. □

§2.4.10 **Remark**. We note that (a) and (c) are properties we would expect of stopping times. With (a), the interpretation is clear. For (c), note that $\tau - s$ peeks into the future by s time units (in fact, $\{\tau - s \leqslant t\} \in \mathscr{F}_{t+s}$), while $\tau + s$ looks back s time units. For stopping times, however, only retrospection is allowed. □

§2.4.11 **Example**. Let $X = (X_n)_{n \in \mathbb{N}_0}$ be a stochastic process adapted to a filtration $(\mathscr{F}_n)_{n \in \mathbb{N}_0}$. For $S_1, S_2 \in \mathscr{S}$ let τ_{S_1} and τ_{S_2} be hitting times as in §2.4.8 (b), then $\tau_{S_1} \geqslant \tau_{S_2}$ whenever $S_1 \subset S_2$. In particular, it follows that $\tau_{S_1} \wedge \tau_{S_2} \geqslant \tau_{S_1 \cup S_2}$ and $\tau_{S_1 \cap S_2} \geqslant \tau_{S_1} \vee \tau_{S_2}$. □

§2.4.12 **Definition**. Let τ be a stopping time. Then

$$\mathscr{F}_\tau := \{A \in \mathscr{A} : A \cap \{\tau \leqslant t\} \in \mathscr{F}_t \text{ for any } t \in \mathbb{T}\}$$

is called the *σ-algebra of τ-past*. □

§2.4.13 **Example**. If $\tau \equiv t_o$ is a constant stopping time at $t_0 \in \mathbb{T}$, then $\mathscr{F}_\tau = \mathscr{F}_{t_0}$. □

§2.4.14 **Lemma**. *If τ and σ are stopping times then* (i) $\mathscr{F}_\sigma \cap \{\sigma \leqslant \tau\} \subset \mathscr{F}_{\tau \wedge \sigma} = \mathscr{F}_\tau \cap \mathscr{F}_\sigma$, (ii) $\mathscr{F}_\tau = \mathscr{F}_t$ *on* $\{\tau = t\}$ *for all* $t \in \mathbb{T}$ *and* (iii) $\mathscr{F}_{\tau \vee \sigma} = \mathscr{F}_\tau \vee \mathscr{F}_\sigma$. *In particular, we see from* (i) *that* $\{\sigma \leqslant \tau\} \in \mathscr{F}_\sigma \cap \mathscr{F}_\tau$, *that* $\mathscr{F}_\sigma = \mathscr{F}_\tau$ *on* $\{\sigma = \tau\}$, *and that* $\mathscr{F}_\tau \subset \mathscr{F}_\sigma$ *whenever* $\tau \leqslant \sigma$.

Proof of Lemma §2.4.14 is given in the lecture. □

§2.4.15 **Definition**. For a stopping time τ define $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$ for all $\omega \in \{\tau < \infty\}$ or equivalently $X_\tau := X_t$ on $\{\tau = t\}$ for all $t \in \mathbb{T}$. □

§2.4.16 **Lemma**. *Let $\mathbb{T}$ be countable, let X be adapted and let τ be a stopping time. Then X_τ is measurable with respect to $\mathscr{F}_\tau$. In particular, τ is $\mathscr{F}_\tau$-measurable.*

Proof of Lemma §2.4.16 is given in the lecture. □

§2.4.17 **Remark**. For uncountable $\mathbb{T}$ and for fixed ω, in general, the map $\mathbb{T} \to \mathcal{S}$, $t \mapsto X_t(\omega)$ is not measurable; hence neither is the composition X_τ always measurable. Here one needs assumptions on the regularity of the paths $t \mapsto X_t(\omega)$; for example, right continuity (cf. Kallenberg [2002], Lemma 7.5, p.122). □

§2.4.18 **Corollary**. *Let $\mathbb{T}$ be countable, let X be adapted and let $(\tau_t)_{t\in\mathbb{T}}$ be a family of stopping times with $\tau_t \leqslant \tau_s < \infty$, $s, t \in \mathbb{T}$, $t \leqslant s$. Then the process $(X_{\tau_t})_{t\in\mathbb{T}}$ is adapted to the filtration $(\mathscr{F}_{\tau_t})_{t\in\mathbb{T}}$. In particular, $(X_{\tau\wedge t})_{t\in\mathbb{T}}$ is adapted to both filtration $(\mathscr{F}_{\tau\wedge t})_{t\in\mathbb{T}}$ and $(\mathscr{F}_t)_{t\in\mathbb{T}}$.*

Proof of Corollary §2.4.18 is given in the lecture. □

§2.4.19 **Definition**. Let $\mathbb{T}$ be countable, let $(X_t)_{t\in\mathbb{T}}$ be adapted and let τ be a stopping time. We define the *stopped process* $X^\tau = (X^\tau_t)_{t\in\mathbb{T}}$ by $X^\tau_t = X_{\tau\wedge t}$ for any $t \in \mathbb{T}$ which is adapted to both filtration $\mathscr{F}^\tau = (\mathscr{F}^\tau_t)_{t\in\mathbb{T}} = (\mathscr{F}_{\tau\wedge t})_{t\in\mathbb{T}}$ and $\mathscr{F} = (\mathscr{F}_t)_{t\in\mathbb{T}}$. □

Chapter 3

Martingale theory

3.1 Positive {super}martingales

In the following, let $\mathbb{T} \subset \mathbb{R}$ be an index set, let $\mathscr{F} = (\mathscr{F}_t)_{t\in\mathbb{T}}$ be a filtration and let $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F})$ be a filtered probability space. For $a, b \in \mathbb{R}$, $a < b$, we denote by $[\![a, b]\!] := [a, b] \cap \mathbb{Z}$ the set of all integers contained in the closed interval $[a, b]$.

§3.1.1 **Definition**. Let $X = (X_t)_{t\in\mathbb{T}}$ be a positive adapted stochastic process on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F})$. X is called (with respect to $\mathscr{F}$) a

positive supermartingale if $X_s \geqslant \mathbb{E}(X_t|\mathscr{F}_s)$ for all $s, t \in \mathbb{T}$ with $t > s$,

positive martingale if $X_s = \mathbb{E}(X_t|\mathscr{F}_s)$ for all $s, t \in \mathbb{T}$ with $t > s$.

A $\mathbb{R}^d$-valued adapted stochastic process $X = ((X_t^1, \dots, X_t^d))_{t\in\mathbb{T}}$ on $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F})$ is called a *positive {super}martingale* if each coordinate process $X^k = (X_t^k)_{t\in\mathbb{T}}$ is a *positive {super}martingale*. □

§3.1.2 **Remark**. (a) Clearly, for a supermartingale, we have $\mathbb{E}(X_r|\mathscr{F}_s) \geqslant \mathbb{E}(X_t|\mathscr{F}_s)$ for all $s < r \leqslant t$, i.e., $(\mathbb{E}(X_t|\mathscr{F}_s))_{t>s}$ decreases (point-wise), the map $t \mapsto \mathbb{E}[X_t]$ is monotone decreasing and for martingales it is constant.

(b) If $\mathbb{T} = \mathbb{N}$, $\mathbb{T} = \mathbb{N}_0$ or $\mathbb{T} = \mathbb{Z}$, then it is enough to consider at each instant s only $t = s+1$. In fact, by the tower property of the conditional expectation, we get $\mathbb{E}(X_{s+2}|\mathscr{F}_s) \geqslant \mathbb{E}(\mathbb{E}(X_{s+1}|\mathscr{F}_{s+1})|\mathscr{F}_s) = \mathbb{E}(X_{s+1}|\mathscr{F}_s)$. Thus, if the defining inequality (or equality) holds for any time step of size one, by induction it holds for all times.

(c) If we do not explicitly mention the filtration $\mathscr{F}$, we tacitly assume that $\mathscr{F} = \sigma(X)$ is the natural filtration generated by X.

(d) Let $\mathscr{F}$ and $\mathscr{F}^\circ$ be filtrations with $\mathscr{F}_t \subset \mathscr{F}_t^\circ$ for all t, and let X be a positive $\mathscr{F}^\circ$-{super}martingale that is adapted to $\mathscr{F}$. Then X is also a positive {super}martingale with respect to the smaller filtration $\mathscr{F}$. Indeed, for $s < t$ and for the case of a supermartingale, $\mathbb{E}(X_t|\mathscr{F}_s) = \mathbb{E}(\mathbb{E}(X_t|\mathscr{F}_s^\circ)|\mathscr{F}_s) \leqslant \mathbb{E}(X_s|\mathscr{F}_s) = X_s$. In particular, a positive $\mathscr{F}$-{super}martingale X is always a {super}martingale with respect to its own natural filtration $\sigma(X)$. □

§3.1.3 **Theorem**. (a) *Let X and Y be positive {super}martingales and $a, b \geqslant 0$. Then $(aX + bY)$ is a positive {super}martingale.*

(b) *Let X and Y be positive supermartingales. Then $Z := X \wedge Y = (\min(X_t, Y_t))_{t\in\mathbb{T}}$ is a positive supermartingale.*

(c) *If $(X_n)_{n\in\mathbb{N}}$ is a positive supermartingale, $\mathbb{E}(X_k) \geqslant \mathbb{E}(X_1)$ for some $k \in \mathbb{N}$, then $(X_n)_{n\in[\![1,k]\!]}$ is a positive martingale. If there exists a sequence $k_n \uparrow \infty$ with $\mathbb{E}(X_{k_n}) \geqslant \mathbb{E}(X_1)$, $n \in \mathbb{N}$, then X is a positive martingale.*

(d) *Let $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ be positive supermartingales and let τ be a stopping time such that $X_\tau(\omega) \geqslant Y_\tau(\omega)$ for all $\omega \in \{\tau < \infty\}$. Then $Z := (X_n \mathbb{1}_{\{n<\tau\}} + Y_n \mathbb{1}_{\{\tau\leqslant n\}})_{n\in\mathbb{N}_0}$ is a positive supermartingale.*

Proof of Theorem §3.1.3 is given in the lecture. □

§3.1.4 **Proposition** (Maximal inequality). *Let $(X_n)_{n\in\mathbb{N}}$ be a positive supermartingale. Then $\sup_{n\in\mathbb{N}} X_n$ is a.s. finite on the set $\{X_1 < \infty\}$ and satisfies for any number $a > 0$:*

$$\mathbb{P}(\sup_{n\in\mathbb{N}} X_n \geqslant a|\mathscr{F}_1) := \mathbb{E}\big[\mathbb{1}_{\{\sup_{n\in\mathbb{N}} X_n \geqslant a\}}|\mathscr{F}_1\big] \leqslant \min(X_1/a, 1).$$

Proof of Proposition §3.1.4 is given in the lecture. □

§3.1.5 **Remark**. The last results still holds true when replacing the constant a by a positive, $\mathscr{F}_1$-measurable r.v. A, that is, $\mathbb{P}\big(\sup_{n\in\mathbb{N}} X_n \geqslant A|\mathscr{F}_1\big) \leqslant \min\big(\frac{X_1}{A}, 1\big)$ on the set $\{A > 0\}$. Consequently:

(a) For any positive supermartingale $(X_n)_{n\in\mathbb{N}}$, any positive $\mathscr{F}_1$-measurable r.v. A such that $A \leqslant \sup_{n\in\mathbb{N}} X_n$ it follows that $1 = \mathbb{P}\big(\sup_{n\in\mathbb{N}} X_n \geqslant A|\mathscr{F}_1\big) \leqslant \min\big(\frac{X_1}{A}, 1\big)$ and, hence $A \leqslant X_1$. In other words, X_1 is the largest $\mathscr{F}_1$-measurable lower bound of $\sup_{n\in\mathbb{N}} X_n$.

(b) More generally: $\sup_{n\in[\![1,k]\!]} X_n$, $k \in \mathbb{N}$, is the largest $\mathscr{F}_k$-measurable lower bound of $\sup_{n\in\mathbb{N}} X_n$. Indeed, $(\sup_{n\in[\![1,k]\!]} X_n, X_{k+1}, X_{k+2}, \dots)$ is a supermartingale adapted to the filtration $(\mathscr{F}_k, \mathscr{F}_{k+1}, \dots)$ and, hence by employing Proposition §3.1.4 any positive $\mathscr{F}_k$-measurable r.v. A such that $A \leqslant \sup_{n\in\mathbb{N}} X_n$ satisfies $A \leqslant \sup_{n\in[\![1,k]\!]} X_n$. □

§3.1.6 **Definition**. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$. For $a, b \in \mathbb{R}$ with $a < b$ defining inductively the integers $\tau_0 := 1$, $\sigma_k := \inf\{n \geqslant \tau_k : x_n \leqslant a\}$ and $\tau_{k+1} := \inf\{n \geqslant \sigma_k : x_n \geqslant b\}$, $k = 0, 1, 2, \dots$, the number of upcrossing (aufsteigende Überquerungen) of the interval $[a, b]$ by the sequence $(x_n)_{n\in\mathbb{N}}$ is denoted by $\beta_{a,b} := \sup\{k \geqslant 1 : \tau_k < \infty\}$. □

§3.1.7 **Remark**. Clearly, if $\liminf_{n\to\infty} x_n < a < b < \limsup_{n\to\infty} x_n$ then $\beta_{a,b} = \infty$ which in turn implies $\liminf_{n\to\infty} x_n \leqslant a < b \leqslant \limsup_{n\to\infty} x_n$. In other words, the sequence $(x_n)_{n\in\mathbb{N}}$ in $\overline{\mathbb{R}}$ is convergent if and only if $\beta_{a,b} < \infty$ for all $a < b$ in $\mathbb{R}$ (or in $\mathbb{Q}$). □

§3.1.8 **Lemma**. *For any sequence of real r.v.'s $(X_n)_{n\in\mathbb{N}}$ and any $a < b$ in $\mathbb{R}$ (or $\mathbb{Q}$) the upcrossing numbers $\beta_{a,b}(\omega)$ associated with each sequence $(X_n(\omega))_{n\in\mathbb{N}}$ define a r.v..*

Proof of Lemma §3.1.8 is left as an exercise. □

§3.1.9 **Remark**. Note that τ_k (and σ_k) as in §3.1.6 defines for each $k = 0, 1, \dots$ a stopping time since $\{\tau_k = n\}$ (and $\{\sigma_k = n\}$) depends only on $\{X_m, m \leqslant n\}$ and, hence belongs to $\mathscr{F}_n$. In addition, $\tau_k \leqslant \tau_{k+1}$, $k \in \mathbb{N}$. □

§3.1.10 **Lemma**. *A sequence of real r.v.'s $(X_n)_{n\in\mathbb{N}}$ converges a.s. if and only if the upcrossing numbers $\beta_{a,b}$ are finite a.s. for any $a < b$ in $\mathbb{R}$ (or $\mathbb{Q}$).*

Proof of Lemma §3.1.10 is left as an exercise. □

§3.1.11 **Lemma** (Dubin's inequality). *Let $(X_n)_{n\in\mathbb{N}}$ be a positive supermartingale. For any $k \in \mathbb{N}$ and any numbers $0 < a < b < \infty$ the associated upcrossing numbers $\beta_{a,b}$ satisfy the inequality*

$$P\big(\beta_{a,b} \geqslant k|\mathscr{F}_1\big) \leqslant \big(a/b\big)^k \min\left(\frac{X_1}{a}, 1\right)$$

The r.v.'s $\beta_{a,b}$ are hence a.s. finite.

Proof of Lemma §3.1.11 is given in the lecture. □

§3.1.12 **Remark**. Note that, if $(X_z)_{z\in\mathbb{Z}}$ is a positive supermartingale, then $P\big(\beta_{a,b} \geqslant k|\mathscr{F}_1\big) \leqslant \big(a/b\big)^k \min\Big(\frac{\sup_{z\leqslant 1} X_z}{a}, 1\Big)$. □

§3.1.13 **Theorem**. *Every positive supermartingale $(X_n)_{n\in\mathbb{N}}$ converges a.s., i.e., $X_n \xrightarrow{a.s.} X_\infty$. Furthermore, the a.s. limit X_∞ satisfies $\mathbb{E}[X_\infty|\mathscr{F}_n] \leqslant X_n$ for all $n \in \mathbb{N}$.*

Proof of Theorem §3.1.13 is given in the lecture. □

§3.1.14 **Remark**. (a) Since $\mathbb{E}[X_\infty|\mathscr{F}_n] \leqslant X_n$ holds for all $n \in \mathbb{N}$ it follows that $X_\infty < \infty$ a.s. on the complement of the event $\cap_{n\in\mathbb{N}}\{X_n = \infty\}$. Indeed, for all n, X_∞ is integrable on each event $\{\mathbb{E}[X_\infty|\mathscr{F}_n] \leqslant a\}$, $a \in \mathbb{R}_+$ and hence finite on the event $\{\mathbb{E}[X_\infty|\mathscr{F}_n] < \infty\}$.

(b) If $(X_n)_{n\in\mathbb{N}}$ is an integrable positive supermartingale, that is, $X_n \in L_1$ for all $n \in \mathbb{N}$, then $\mathbb{E}[X_\infty|\mathscr{F}_n] \leqslant X_n$ implies $X_\infty \in L_1$. However, in general, an integrable positive supermartingale does not converge to X_∞ in L_1.

(c) If $(X_n)_{n\in\mathbb{N}}$ is a positive martingale, that is, $X_n = \mathbb{E}[X_{n+1}|\mathscr{F}_n]$ a.s. for all $n \in \mathbb{N}$, then by Theorem §3.1.13 $X_n \xrightarrow{a.s.} X_\infty$ and $\mathbb{E}[X_\infty|\mathscr{F}_n] \leqslant X_n$ for all $n \in \mathbb{N}$, where the inequality does generally not become an equality. The next proposition provides a situation in which this phenomena not arrives. □

§3.1.15 **Proposition**. *Let $p \in [1,\infty)$. For all $Z \in L_p^+ := L_p \cap \mathcal{M}^+$ the stochastic process $(Z_n)_{n\in\mathbb{N}}$ given by $Z_n := \mathbb{E}[Z|\mathscr{F}_n]$, $n \in \mathbb{N}$, is a positive martingale which converges a.s. and in L_p to $Z_\infty := \mathbb{E}[Z|\mathscr{F}_\infty]$ with $\mathscr{F}_\infty := \bigvee_{n\in\mathbb{N}} \mathscr{F}_n$.*

Proof of Proposition §3.1.15 is given in the lecture. □

§3.1.16 **Remark**. (a) A positive martingale $(Z_n)_{n\in\mathbb{N}}$ as in §3.1.15 and its a.s.-limit Z_∞ verify the equality $Z_n = \mathbb{E}[Z_\infty|\mathscr{F}_n]$ a.s. for all $n \in \mathbb{N}$ by employing that $\mathbb{E}[Z_\infty|\mathscr{F}_n] = \mathbb{E}[\mathbb{E}[Z|\mathscr{F}_\infty]|\mathscr{F}_n] = \mathbb{E}[Z|\mathscr{F}_n] = Z_n$.

(b) Let $(X_n)_{n\in\mathbb{N}}$ be a positive martingale which converges in L_p, i.e., $X_n \xrightarrow{L_p} X_\infty$. Then, the equality $X_n = \mathbb{E}[X_m|\mathscr{F}_n]$ a.s. for all $m \geqslant n$ and the continuity of the conditional expectation on L_p imply together that $X_n = \mathbb{E}[X_\infty|\mathscr{F}_n]$ a.s. for all $n \in \mathbb{N}$. Thereby, Proposition §3.1.15 implies that the martingales of the form $(\mathbb{E}[Z|\mathscr{F}_n])_{n\in\mathbb{N}}$ with $Z \in L_p^+$ are exactly the positive martingales in L_p which converge in L_p as $n \to \infty$. A positive martingale $(X_n)_{n\in\mathbb{N}}$ is called *closable* (abschließbar) in L_p, if there exists an $X \in L_p^+$ with $X_n = \mathbb{E}[X|\mathscr{F}_n]$, for all $n \in \mathbb{N}$.

(c) Considering $Z = Z^+ - Z^-$ allows to extend immediately the last proposition to a r.v. $Z \in L_p$. □

§3.1.17 **Corollary**. *For any positive r.v. Z we have $\mathbb{E}[Z|\mathscr{F}_n] \xrightarrow{a.s.} \mathbb{E}[Z|\mathscr{F}_\infty]$ on the complement of the event $\cap_{n\in\mathbb{N}}\{\mathbb{E}[Z|\mathscr{F}_n] = \infty\}$.*

Proof of Corollary §3.1.17 is left as an exercise. □

§3.1.18 **Remark**. Note that in the preceding corollary integrability is not assumed. However, the result cannot be improved. In Neveu [1975], p.31, for example, a r.v. Z is constructed which is $\mathscr{F}_\infty$-measurable and a.s. finite such that $\mathbb{E}[Z|\mathscr{F}_n] = \infty$ a.s. for all $n \in \mathbb{N}$. In this case, $\mathbb{E}[Z|\mathscr{F}_n] \xrightarrow{L_p} \mathbb{E}[Z|\mathscr{F}_\infty] = Z$ holds only on a negligible set. □

§3.1.19 **Lemma**. *For any positive {super}martingale $(X_n)_{n\in\mathbb{N}}$ and for any stopping time τ, the stopped process $X^\tau = (X_{\tau\wedge n})_{n\in\mathbb{N}}$ is a positive {super}martingale.*
Proof of Lemma §3.1.19 is left as an exercise. □

§3.1.20 **Theorem** (Optional stopping). *Let $(X_n)_{n\in\mathbb{N}}$ be a positive supermartingale and X_∞ its a.s.-limit. Then, for any stopping times τ and σ we have*

$$X_\tau \geqslant \mathbb{E}[X_\sigma|\mathscr{F}_\tau] \ a.s. \quad \text{on the event} \quad \{\tau \leqslant \sigma\}.$$

Proof of Theorem §3.1.20 is given in the lecture. □

§3.1.21 **Remark**. If $(X_n)_{n\in\mathbb{N}}$ is a positive martingale, then the inequality $X_\tau \geqslant \mathbb{E}[X_\sigma|\mathscr{F}_\tau]$ does generally not become an equality. □

3.2 Integrable {super/sub}martingales

§3.2.1 **Definition**. Let $X = (X_t)_{t\in\mathbb{T}}$ be an adapted stochastic process on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F})$ with $X_t \in L_1(\Omega, \mathscr{A}, \mathbb{P})$ for all $t \in \mathbb{T}$. X is called (with respect to $\mathscr{F}$) a

(integrable) **supermartingale** if $X_s \geqslant \mathbb{E}(X_t|\mathscr{F}_s)$ for all $s, t \in \mathbb{T}$ with $t > s$,

(integrable) **submartingale** if $X_s \leqslant \mathbb{E}(X_t|\mathscr{F}_s)$ for all $s, t \in \mathbb{T}$ with $t > s$,

(integrable) **martingale** if $X_s = \mathbb{E}(X_t|\mathscr{F}_s)$ for all $s, t \in \mathbb{T}$ with $t > s$.

An $\mathbb{R}^d$-valued adapted stochastic process $X = ((X_t^1, \dots, X_t^d))_{t\in\mathbb{T}}$ is called an *(integrable)* **{super/sub}martingale** if each coordinate process $X^k = (X_t^k)_{t\in\mathbb{T}}$ is an (integrable) {super/sub} martingale. □

§3.2.2 **Remark**.
(a) The integrability assumption is often replaced by the weaker assumption $\mathbb{E}(X_t^+) < \infty$ for all $t \in \mathbb{T}$. This generalisation is only helpful in case of a negative submartingale (by changing the sign a positive supermartingale).
(b) The a.s. convergence of an integrable submartingale is essentially a corollary of Theorem §3.1.13 which establishes the convergence for positive supermartingales with the only difference, that any positive supermartingale converges a.s. but not every integrable submartingale converges a.s.. □

§3.2.3 **Lemma**. *Let M be a $\mathbb{R}^d$-valued integrable martingale and consider a convex function $f : \mathbb{R}^d \to \mathbb{R}$ such that $X = f(M)$ is integrable. Then X is a submartingale. The statement remains true for any real-valued integrable submartingale M, provided that f is also non-decreasing.*
Proof of Lemma §3.2.3 is left as an exercise. □

§3.2.4 **Remark**. The last result is often applied with $f(x) = \|x\|_p^p$, for some $p \geqslant 1$ or, for $d = 1$, with $f(x) = x^+$. □

§3.2.5 **Theorem**. *Every integrable submartingale $(X_n)_{n\in\mathbb{N}}$ satisfying $\sup_{n\in\mathbb{N}} \mathbb{E}(X_n^+) < \infty$ converges a.s., i.e., $X_n \xrightarrow{a.s.} X_\infty$. Furthermore, the a.s. limit X_∞ is integrable. In case of an integrable martingale the condition $\sup_{n\in\mathbb{N}} \mathbb{E}(X_n^+) < \infty$ is equivalent to $\sup_{n\in\mathbb{N}} \|X_n\|_1 < \infty$.*

Proof of Theorem §3.2.5 is given in the lecture. □

§3.2.6 **Remark**. The decomposition $X_n = M_n - A_n$, $n \in \mathbb{N}$, into a positive integrable martingale $(M_n)_{n\in\mathbb{N}}$ and a positive integrable supermartingale $(A_n)_{n\in\mathbb{N}}$ obtained in the proof of Theorem §3.2.5 is called Krickeberg decomposition. □

§3.2.7 **Lemma**. *Let $(X_n)_{n\in\mathbb{N}}$ be an integrable martingale and let τ be a bounded stopping time, that is, $\tau \leqslant K$ for some $K \in \mathbb{N}$. Then $X_\tau = \mathbb{E}[X_K|\mathscr{F}_\tau]$ and in, particular $\mathbb{E}(X_\tau) = \mathbb{E}(X_1)$. Assume that, more generally, $(X_n)_{n\in\mathbb{N}}$ is only adapted and integrable. Then $(X_n)_{n\in\mathbb{N}}$ is an integrable martingale if and only if $\mathbb{E}(X_\tau) = \mathbb{E}(X_1)$ for any bounded stopping time τ.*

Proof of Lemma §3.2.7 is given in the lecture. □

§3.2.8 **Definition**. Let $(X_n)_{n\in\mathbb{N}_0}$ be an adapted real-valued process and let $(H_n)_{n\in\mathbb{N}}$ be a real-valued predictable process as defined in §2.4.4. The *discrete stochastic integral* of H with respect to X is the adapted stochastic process $H\bullet X = ((H\bullet X)_n)_{n\in\mathbb{N}_0}$ defined by $(H\bullet X)_0 := 0$ and $(H \bullet X)_n := \sum_{k=1}^{n} H_k(X_k - X_{k-1})$ for $n \in \mathbb{N}$. If X is a martingale, then $H \bullet X$ is also called the *martingale transform* of X. □

§3.2.9 **Example**. Let X be a (possibly unfair) game where $X_n - X_{n-1}$ is the gain per euro in the nth round. We interpret H_n as the number of euros we bet in the nth game. H is then a gambling strategy. Clearly, the value of H_n has to be decided at time $n-1$; that is, before the result of X_n is known. In other words, H must be predictable. Now assume that X is a fair game (that is, a martingale) and H is locally bounded (that is, each H_n is bounded). From $\mathbb{E}[X_{n+1}-X_n|\mathscr{F}_n] = 0$ follows that $\mathbb{E}[(H\bullet X)_{n+1}|\mathscr{F}_n] = \mathbb{E}[(H\bullet X)_n+H_{n+1}(X_{n+1}-X_n)|\mathscr{F}_n] = (H \bullet X)_n + H_{n+1}\mathbb{E}[X_{n+1} - X_n|\mathscr{F}_n] = (H \bullet X)_n$. Thus $H \bullet X$ is a martingale. The next result says that the converse also holds; that is, X is a martingale if, for sufficiently many predictable processes, the stochastic integral is a martingale. □

§3.2.10 **Proposition**. *Let $(X_n)_{n\in\mathbb{N}_0}$ be an adapted, real-valued process with $X_0 \in L_1$.*

(a) *X is an integrable martingale if and only if, for any locally bounded predictable process H, the stochastic integral $H \bullet X$ is an integrable martingale.*

(b) *X is an integrable submartingale (supermartingale) if and only if $H \bullet X$ is an integrable submartingale (supermartingale) for any locally bounded positive predictable process H.*

Proof of Proposition §3.2.10 is given in the lecture. □

§3.2.11 **Remark**. The preceding proposition says, in particular, that we cannot find any locally bounded gambling strategy that transforms a martingale (or, if we are bound to non-negative gambling strategies, as we are in real life, a supermartingale) into a submartingale. Quite the contrary is suggested by the many invitations to play all kinds of "sure winning systems" in lotteries. □

3.3 Regular integrable martingale

§3.3.1 **Proposition**. *For every integrable martingale $(X_n)_{n\in\mathbb{N}}$ on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F})$ the following conditions are equivalent*

(i) *The sequence $(X_n)_{n\in\mathbb{N}}$ converges in L_1 as $n \to \infty$;*

(ii) $\sup_{n\in\mathbb{N}} \|X_n\|_1 < \infty$ *and the a.s. limit* $X_\infty = \lim_{n\to\infty} X_n$ *of the martingale which exists in* L_1 *due to Theorem §3.2.5 satisfies the equalities* $X_n = \mathbb{E}[X_\infty|\mathscr{F}_n]$ *for all* $n \in \mathbb{N}$*;*

(iii) *The martingale is closable, that is, there exists a r.v.* $X \in L_1(\Omega, \mathscr{A}, \mathbb{P})$ *such that* $X_n = \mathbb{E}[X|\mathscr{F}_n]$ *for all* $n \in \mathbb{N}$*;*

(iv) *The sequence* $(X_n)_{n\in\mathbb{N}}$ *is uniformly integrable in* $L_1(\Omega, \mathscr{A}, \mathbb{P})$*, that is,* $\lim_{a\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}\big(\mathbb{1}_{\{|X_n|\geqslant a\}}|X_n|\big) = 0$ *which is satisfied whenever* $\sup_{n\in\mathbb{N}} \|X_n\|_1 \in L_1$*.*

The integrable martingale $(X_n)_{n\in\mathbb{N}}$ *will be called* ***regular*** *if it satisfies one of these equivalent conditions.*

Proof of Proposition §3.3.1 is given in the lecture. □

§3.3.2 **Corollary**. *Let* $(X_n)_{n\in\mathbb{N}}$ *be a regular integrable martingale.* (i) *For every stopping time* τ*, the r.v.* X_τ *is integrable.* (ii) *The family* $\{X_\tau; \tau \text{ is a finite stopping time}\}$ *is uniformly integrable.* (iii) *For every pair of stopping times* τ, σ *such that* $\tau \leqslant \sigma$ *a.s., the "martingale equality"* $X_\tau = \mathbb{E}[X_\sigma|\mathscr{F}_\tau]$ *is also satisfied.*

Proof of Corollary §3.3.2 is given in the lecture. □

§3.3.3 **Remark**. For a regular integrable martingale the limit $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. and the r.v. X_τ (resp. X_σ) by definition equals X_∞ on $\{\tau = \infty\}$ (resp. $\{\sigma = \infty\}$). Since $\tau \wedge \sigma \leqslant \sigma$ a.s. the corollary implies $X_{\tau\wedge\sigma} = \mathbb{E}[X_\sigma|\mathscr{F}_{\tau\wedge\sigma}]$. Furthermore $\mathbb{E}[X_\sigma|\mathscr{F}_\tau] = \mathbb{E}[X_\sigma|\mathscr{F}_{\tau\wedge\sigma}]$, and hence, for any stopping time τ, σ we have $X_{\tau\wedge\sigma} = \mathbb{E}[X_\sigma|\mathscr{F}_\tau]$. Indeed, for all $A \in \mathscr{F}_\tau$ we have

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X_\sigma|\mathscr{F}_\tau]\mathbb{1}_A] = \mathbb{E}[X_\sigma\mathbb{1}_A] &= \mathbb{E}[X_\sigma\mathbb{1}_{\underbrace{A\cap\{\tau\leqslant\sigma\}}_{\in\mathscr{F}_{\tau\wedge\sigma}}}] + \mathbb{E}[X_\sigma\mathbb{1}_{\underbrace{A\cap\{\tau>\sigma\}}_{\in\mathscr{F}_{\tau\wedge\sigma}}}]\\
&= \mathbb{E}[\mathbb{E}[X_\sigma|\mathscr{F}_{\tau\wedge\sigma}]\mathbb{1}_{A\cap\{\tau\leqslant\sigma\}}] + \mathbb{E}[X_{\tau\wedge\sigma}\mathbb{1}_{A\cap\{\tau>\sigma\}}]\\
&= \mathbb{E}\Big[\big\{\mathbb{E}[X_\sigma|\mathscr{F}_{\tau\wedge\sigma}]\mathbb{1}_{\{\tau\leqslant\sigma\}} + X_{\tau\wedge\sigma}\mathbb{1}_{\{\tau>\sigma\}}\big\}\mathbb{1}_A\Big]
\end{aligned}$$

Thereby, $\mathbb{E}[X_\sigma|\mathscr{F}_\tau] = \mathbb{E}[X_\sigma|\mathscr{F}_{\tau\wedge\sigma}]\mathbb{1}_{\{\tau\leqslant\sigma\}} + X_{\tau\wedge\sigma}\mathbb{1}_{\{\tau>\sigma\}}$ is $\mathscr{F}_{\tau\wedge\sigma}$-measurable, which in turn implies, $\mathbb{E}[X_\sigma|\mathscr{F}_\tau] = \mathbb{E}\big[\mathbb{E}[X_\sigma|\mathscr{F}_\tau]|\mathscr{F}_{\tau\wedge\sigma}\big] = \mathbb{E}[X_\sigma|\mathscr{F}_{\tau\wedge\sigma}]$ by employing that $\mathscr{F}_{\tau\wedge\sigma} \subset \mathscr{F}_\tau$. □

§3.3.4 **Proposition**. *Every martingale* $(X_n)_{n\in\mathbb{N}}$ *which is bounded in* L_p *for some* $p > 1$ *in the sense that* $\sup_{n\in\mathbb{N}} \|X_n\|_p < \infty$*, is regular. Furthermore, the martingale converges in* L_p *to an a.s. limit* X_∞*.*

Proof of Proposition §3.3.4 is given in the lecture. □

§3.3.5 **Remark**. The last proposition is false for $p = 1$. □

§3.3.6 **Lemma**. *Every positive and integrable submartingale* $(X_n)_{n\in\mathbb{N}}$ *satisfies the inequalities* $a\mathbb{P}(\sup_{m\in[\![1,n]\!]} X_m > a) \leqslant \mathbb{E}\big(\mathbb{1}_{\{\sup_{m\in[\![1,n]\!]} X_m > a\}}X_n\big)$ *for all* $n \in \mathbb{N}$ *and all* $a > 0$*.*

Proof of Lemma §3.3.6 is given in the lecture. □

§3.3.7 **Proposition**. *For every martingale* $(X_n)_{n\in\mathbb{N}}$ *which is bounded in* L_p *for some* $p > 1$ *the r.v.* $\sup_{n\in\mathbb{N}}|X_n|$ *belongs to* L_p *and satisfies* $\|\sup_{n\in\mathbb{N}}|X_n|\|_p \leqslant \frac{p}{p-1} \sup_{n\in\mathbb{N}} \|X_n\|_p$*.*

Proof of Proposition §3.3.7 is given in the lecture. □

§3.3.8 **Remark.** The last proposition is false for $p = 1$. However, for every martingale $(X_n)_{n\in\mathbb{N}}$ satisfying the condition $\sup_{n\in\mathbb{N}} \mathbb{E}\big[|X_n|\big(\log|X_n|\big)_+\big] < \infty$, the r.v. $\sup_{n\in\mathbb{N}}|X_n|$ is integrable and the martingale $(X_n)_{n\in\mathbb{N}}$ is therefore regular (c.f. Neveu [1975], Proposition IV-2-10, p.70). □

The concepts of filtration and martingale do not require the index set $\mathbb{T}$ (interpreted as time) to be a subset of $[0, \infty)$. Hence we can consider the case $\mathbb{T} = -\mathbb{N}_0$.

§3.3.9 **Definition.** Let $(\mathscr{F}_n)_{n\in-\mathbb{N}_0}$ be a filtration where $\mathscr{F}_{-n-1} \subset \mathscr{F}_{-n}$, $n \in \mathbb{N}_0$ and let $(X_n)_{n\in-\mathbb{N}_0}$ be an integrable martingale with respect to $(\mathscr{F}_n)_{n\in-\mathbb{N}_0}$, that is, $X_{-n} \in L_1$, X_{-n} is $\mathscr{F}_{-n}$-measurable and $\mathbb{E}[X_{-n}|\mathscr{F}_{-n-1}] = X_{-n-1}$ hold for all $n \in \mathbb{N}_0$. Then $X = (X_{-n})_{n\in\mathbb{N}_0}$ is called an *(integrable) backwards martingale*. □

§3.3.10 **Remark.** A backwards martingale is always uniformly integrable and hence regular. This follows from Corollary §1.6.13 and the fact that $X_{-n} = E[X_0|\mathscr{F}_{-n}]$ for any $n \in \mathbb{N}_0$. □

§3.3.11 **Proposition.** *Let $(X_{-n})_{n\in\mathbb{N}_0}$ be a backward martingale with respect to $(\mathscr{F}_{-n})_{n\in\mathbb{N}_0}$. Then there exists $X_{-\infty} = \lim_{n\to\infty} X_{-n}$ a.s. and in L_1. Furthermore, $X_{-\infty} = \mathbb{E}[X_0|\mathscr{F}_{-\infty}]$ where $\mathscr{F}_{-\infty} = \cap_{n=1}^{\infty}\mathscr{F}_{-n}$ is called terminal or tail σ-algebra.*

Proof of Proposition §3.3.11 is given in the lecture. □

§3.3.12 **Example** *(Kolmogorov's strong law of large numbers).* Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. real-valued r.v.'s in L_1, then $n^{-1}\sum_{k=1}^n X_k \overset{n\to\infty}{\longrightarrow} \mathbb{E}(X_1)$ a.s. and in L_1. □

3.4 Regular stopping times for an integrable martingale

§3.4.1 **Lemma.** *Let $(X_n)_{n\in\mathbb{N}}$ be an integrable {super/sub}martingale. For every stopping time τ, the stopped process $X^\tau = (X_n^\tau)_{n\in\mathbb{N}}$ with $X_n^\tau := X_{\tau\wedge n}$ for any $n \in \mathbb{N}$ is again an integrable {super/sub}martingale.*

Proof of Lemma §3.4.1 is left as an exercise. □

§3.4.2 **Definition.** A stopping time τ is called **regular** for an integrable martingale $(X_n)_{n\in\mathbb{N}}$ if the stopped process $X^\tau = (X_n^\tau)_{n\in\mathbb{N}}$ is regular. □

§3.4.3 **Proposition.** *For every integrable martingale $(X_n)_{n\in\mathbb{N}}$ on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F})$ and for every stopping time τ the following conditions are equivalent*

(a) *the stopping time is regular;*

(b) *the stopping time satisfies the following conditions:* (i) *the limit $X_\infty = \lim_{n\to\infty} X_n$ exists a.s. on $\{\tau = \infty\}$;* (ii) *the r.v. X_τ which is defined a.s., is integrable and* (iii) *$X_{\tau\wedge n} = \mathbb{E}[X_\tau|\mathscr{F}_n]$ a.s. for all $n \in \mathbb{N}$.*

(c) *the stopping time satisfies the following conditions:* (i) *$(X_n\mathbb{1}_{\{\tau>n\}})_{n\in\mathbb{N}}$ is a uniformly integrable sequence and* (ii) *$\mathbb{E}(\mathbb{1}_{\{\tau<\infty\}}|X_\tau|) < \infty$.*

Proof of Proposition §3.4.3 is given in the lecture. □

§3.4.4 **Remark.** Condition (c) (ii) is automatically satisfied by every martingale $(X_n)_{n\in\mathbb{N}}$ such that $\sup_{n\in\mathbb{N}} \mathbb{E}|X_n| < \infty$, in particular by every positive integrable martingale ($\mathbb{E}|X_n| = \mathbb{E}X_n = \mathbb{E}X_1$). □

§3.4.5 **Proposition**. *Let τ be a regular stopping time. For every pair σ_1, σ_2 of stopping times such that $\sigma_1 \leqslant \sigma_2 \leqslant \tau$, for such such a pair the r.v.'s X_{σ_1} and X_{σ_2} both exist, are integrable, and satisfy the "martingale identity" $X_{\sigma_1} = \mathbb{E}[X_{\sigma_2}|\mathscr{F}_{\sigma_1}]$ a.s..*

Proof of Proposition §3.4.5 is given in the lecture. □

§3.4.6 **Corollary**. *Let τ and σ be two stopping times such that $\tau \leqslant \sigma$ a.s.. For a given martingale $(X_n)_{n\in\mathbb{N}}$ the stopping time τ is regular whenever the stopping time σ is regular.*

Proof of Corollary §3.4.6 is given in the lecture. □

§3.4.7 **Remark**. Corollary §3.4.6 shows in particular that for a regular martingale, every stopping time is regular (take $\sigma = +\infty$). On the other hand, for an integrable martingale every constant stopping time is regular, and hence, by Corollary §3.4.6 every bounded stopping time is regular too. □

§3.4.8 **Corollary**. *For every martingale $(X_n)_{n\in\mathbb{N}}$ such that $\sup_{n\in\mathbb{N}} \mathbb{E}|X_n| < \infty$, in particular for every positive and integrable martingale, the hitting time τ_a defined by $\tau_a := \inf\{n \in \mathbb{N} : |X_n| > a\}$ is regular for all $a > 0$.*

Proof of Corollary §3.4.8 is given in the lecture. □

§3.4.9 **Proposition**. *Let $(X_n)_{n\in\mathbb{N}}$ be an integrable martingale. In order that the stopping time τ be regular for this martingale and that also $\lim_{n\to\infty} X_n = 0$ a.s. on $\{\tau = \infty\}$, it is necessary and sufficient that the following two conditions be satisfied:* (i) *$\mathbb{E}\mathbb{1}_{\{\tau<\infty\}}|X_\tau| < \infty$ and* (ii) *$\lim_{n\to\infty} \mathbb{E}\mathbb{1}_{\{\tau>n\}}|X_n| = 0$.*

Proof of Proposition §3.4.9 is given in the lecture. □

§3.4.10 **Example** *(Wald identity)*. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of i.i.d. real-valued r.v.'s defined on a filtered probability space $(\Omega, \mathscr{A}, \mathbb{P}, \mathscr{F}^X)$ with natural filtration $\mathscr{F}^X$. Assuming further that $X_1 \in L_2$ the processes $(S_n - n\mathbb{E}X_1)_{n\in\mathbb{N}}$ with $S_n := \sum_{i=1}^n X_i$, $n \in \mathbb{N}$, and $((S_n - n\mathbb{E}X_1)^2 - n\,\mathbb{V}\mathrm{ar}\, X_1)_{n\in\mathbb{N}}$ are integrable martingales which are not regular since they diverge a.s. when $n \to \infty$. However, every stopping time τ such that $\mathbb{E}(\tau) < \infty$ is regular for each of the two martingales $(S_n - n\mathbb{E}X_1)_{n\in\mathbb{N}}$ and $((S_n - n\mathbb{E}X_1)^2 - n\,\mathbb{V}\mathrm{ar}\, X_1)_{n\in\mathbb{N}}$. Such a stopping time satisfies the *Wald identities* (i) $\mathbb{E}(S_\tau) = \mathbb{E}(\tau)\mathbb{E}(X_1)$ and (ii) $\mathbb{E}[S_\tau - \tau\mathbb{E}(X_1)]^2 = \mathbb{E}(\tau)\,\mathbb{V}\mathrm{ar}(X_1)$. Moreover, if in addition $\mathbb{E}(\tau^2) < \infty$ then $\mathbb{V}\mathrm{ar}(S_\tau) = \mathbb{V}\mathrm{ar}(\tau)(\mathbb{E}X_1)^2 + \mathbb{E}(\tau)\,\mathbb{V}\mathrm{ar}(X_1)$. □

3.5 Regularity of integrable submartingales

The study of integrable martingales can be very easily extended to integrable submartingales by using the Krickeberg decomposition of such submartingales.

§3.5.1 **Proposition**. *For every integrable submartingale $(X_n)_{n\in\mathbb{N}}$, the following conditions are equivalent:*

(a) *The sequence $(X_n^+)_{n\in\mathbb{N}}$ converges in L_1;*

(b) *$\sup_{n\in\mathbb{N}} \mathbb{E}X_n^+ < \infty$ and the a.s. limit $X_\infty = \lim_{n\to\infty} X_n$ of the submartingale $(X_n)_{n\in\mathbb{N}}$ which exists and is integrable by Theorem §3.2.5, satisfies the inequalities $X_n \leqslant \mathbb{E}[X_\infty|\mathscr{F}_n]$ a.s. for all $n \in \mathbb{N}$;*

(c) *There exists an integrable r.v. Y such that $X_n \leqslant \mathbb{E}[Y|\mathscr{F}_n]$ for all $n \in \mathbb{N}$;*

(d) *The sequence $(X_n^+)_{n\in\mathbb{N}}$ satisfies the uniform integrability condition*

$$\lim_{a\to\infty}\sup_{n\in\mathbb{N}}\mathbb{E}\mathbb{1}_{\{X_n^+>a\}}X_n^+=0$$

which holds particularly if $\mathbb{E}\sup_{n\in\mathbb{N}}X_n^+<\infty$.)

The integrable submartingale $(X_n)_{n\in\mathbb{N}}$ is said to be ***regular*** *if it satisfies the preceding equivalent conditions.*

Proof of Proposition §3.5.1 is given in the lecture. □

§3.5.2 **Remark.** For a negative integrable submartingale (i.e., for a positive integrable supermartingale with its sign changed), the conditions of the proposition hold trivially. Observe that such a submartingale does not converge in mean, although it always converge a.s., and the condition (a) of the preceding proposition is strictly less restrictive than the convergence of the submartingale in L_1. On the other hand it is clear that for a positive submartingale condition (a) gives L_1-convergence of the submartingale. □

§3.5.3 **Corollary.** *For every regular submartingale $(X_n)_{n\in\mathbb{N}}$ and for every stopping time τ, the r.v. X_τ is integrable; for every pair τ_1,τ_2 of stopping times such that $\tau_1\leqslant\tau_2$ a.s., the submartingale inequality $X_{\tau_1}\leqslant\mathbb{E}[X_{\tau_2}|\mathscr{F}_{\tau_1}]$ remains true a.s..*

Proof of Corollary §3.5.3 is given in the lecture. □

§3.5.4 **Remark.** Finally, it is straightforward to extend the regularity of stopping times as given in Proposition §3.4.3 and §3.4.5 to integrable submartingales. The only changes required in the statement of this proposition consist in replacing the word "martingales" by "submartingales" and writing the inequalities $X_{\tau\wedge n}\leqslant\mathbb{E}[X_\tau|\mathscr{F}_n]$ and $X_{\sigma_1}\leqslant\mathbb{E}[X_{\sigma_2}|\mathscr{F}_{\sigma_1}]$ instead of the corresponding equalities. □

3.6 Doob decomposition and square variation

The introduction of the notion of predictable and increasing process as defined in §2.4.4 allows to effect decompositions of {super/sub}martingales. As before, we take once and for all a filtered probability space $(\Omega,\mathscr{A},\mathbb{P},\mathscr{F})$. Let $X=(X_n)_{n\in\mathbb{N}_0}$ be an adapted integrable process. We will decompose X into a sum consisting of an integrable martingale and a predictable process. To this end, define $M_0:=X_0$, $A_0:=0$, $M_n:=X_0+\sum_{k=1}^n\big(X_k-\mathbb{E}[X_k|\mathscr{F}_{k-1}]\big)$ and $A_n:=\sum_{k=1}^n\big(\mathbb{E}[X_k|\mathscr{F}_{k-1}]-X_{k-1}\big)$ for $n\in\mathbb{N}$. Evidently, $X_n=M_n+A_n$. By construction, $M_n-M_{n-1}=X_n-\mathbb{E}[X_n|\mathscr{F}_{n-1}]$ and $A_n-A_{n-1}=\mathbb{E}[X_n|\mathscr{F}_{n-1}]-X_{n-1}$, for $n\in\mathbb{N}$, and, hence A is predictable with $A_0=0$, and M is a martingale since $\mathbb{E}[M_n-M_{n-1}|\mathscr{F}_{n-1}]=\mathbb{E}[X_n-\mathbb{E}[X_n|\mathscr{F}_{n-1}]|\mathscr{F}_{n-1}]=0$.

§3.6.1 **Proposition** (Doob decomposition). *Let $X=(X_n)_{n\in\mathbb{N}_0}$ be an adapted integrable process. Then there exists a unique decomposition $X=M+A$, where A is predictable with $A_0=0$ and M is a martingale. This representation of X is called the* ***Doob decomposition****. X is a submartingale if and only if A is an increasing process.*

Proof of Proposition §3.6.1 is given in the lecture. □

§3.6.2 **Proposition.** *Let $X:=(X_n)_{n\in\mathbb{N}_0}$ be an integrable submartingale and let $X=M+A$ be its Doob decomposition.*

(a) *The condition* $\sup_{n\in\mathbb{N}_0} \mathbb{E}X_n^+ < \infty$ *(which suffices to ensure a.s. convergence of the submartingale) is equivalent to the conjunction of the two conditions* (i) $A_\infty \in L_1$ *and* (ii) $\sup_{n\in\mathbb{N}_0} \mathbb{E}(|M_n|) < \infty$.

(b) *The convergence in* L_1 *of the submartingale* X *is equivalent to the conjunction of the two conditions* (i) M *is a regular martingale and* (ii) $A_\infty \in L_1$.

(c) *For every stopping time* τ *regular for the martingale* M*, the r.v.* X_τ *is integrable if and only if* $\mathbb{E}A_\tau < \infty$*, and then* $\mathbb{E}X_\tau = \mathbb{E}M_0 + \mathbb{E}A_\tau$.

Proof of Proposition §3.6.2 is given in the lecture. □

§3.6.3 **Example.** Let $(X_n)_{n\in\mathbb{N}_0}$ be a square integrable $\mathscr{F}$-martingale, i.e., $X_n \in L_2(\Omega, \mathscr{A}, \mathbb{P})$ for all $n \in \mathbb{N}_0$. By Lemma §3.2.3, $(X_n^2)_{n\in\mathbb{N}_0}$ is a submartingale. Furthermore, $\mathbb{E}[X_{i-1}X_i|\mathscr{F}_{i-1}] = X_{i-1}\mathbb{E}[X_i|\mathscr{F}_{i-1}] = X_{i-1}^2$, hence considering the Doob decomposition of $(X_n^2)_{n\in\mathbb{N}_0}$ we find $A_0 = 0$ and for $n \in \mathbb{N}$,

$$A_n = \sum_{i=1}^{n} \big(\mathbb{E}[X_i^2|\mathscr{F}_{i-1}] - X_{i-1}^2\big)$$
$$= \sum_{i=1}^{n} \big(\mathbb{E}[(X_i - X_{i-1})^2|\mathscr{F}_{i-1}] - 2X_{i-1}^2 + 2\mathbb{E}[X_{i-1}X_i|\mathscr{F}_{i-1}]\big)$$
$$= \sum_{i=1}^{n} \mathbb{E}[(X_i - X_{i-1})^2|\mathscr{F}_{i-1}]. \quad \square$$

§3.6.4 **Definition.** Let $(X_n)_{n\in\mathbb{N}_0}$ be a square integrable $\mathscr{F}$-martingale. The unique increasing process A for which $(X_n^2 - A_n)_{n\in\mathbb{N}_0}$ becomes a martingale is called *square variation process* of X and is denoted by $\langle X\rangle := (\langle X\rangle_n)_{n\in\mathbb{N}_0} := A$. □

§3.6.5 **Proposition.** *Let* X *be as in Definition §3.6.4. Then, for* $n \in \mathbb{N}$, $\langle X\rangle_n = \sum_{i=1}^n \mathbb{E}[(X_i - X_{i-1})^2|\mathscr{F}_{i-1}]$ *and* $\mathbb{E}\langle X\rangle_n = \mathbb{V}\mathrm{ar}(X_n - X_0)$.

Proof of Proposition §3.6.5 is given in the lecture. □

§3.6.6 **Example.** Let $X_1, X_2, \dots$ be independent, square integrable r.v.'s. If $\mathbb{E}(X_n) = 0$, for all $n \in \mathbb{N}$, then $S_n := \sum_{i=1}^n X_i$ defines a square integrable martingale with respect to the filtration $(\sigma(X_1, \dots, X_n))_{n\in\mathbb{N}}$ and we find $\langle S\rangle_n = \sum_{i=1}^n \mathbb{E}[X_i^2|\sigma(X_1, \dots, X_{i-1})] = \sum_{i=1}^n \mathbb{E}[X_i^2]$. Note that in order for $\langle S\rangle$ to have this simple form, it is not enough for the r.v.'s $X_1, X_2, \dots$ to be uncorrelated. On the other hand, if $\mathbb{E}(X_n) = 1$, for all $n \in \mathbb{N}$, then $P_n := \prod_{i=1}^n X_i$ defines a square integrable martingale with respect to the natural filtration $\mathscr{F} = \sigma(P)$ and $\mathbb{E}[(P_n - P_{n-1})^2|\mathscr{F}_{n-1}] = \mathbb{E}[(X_n-1)^2 P_{n-1}^2|\mathscr{F}_{n-1}] = \mathbb{V}\mathrm{ar}(X_n)P_{n-1}^2$. Hence, $\langle P\rangle_n = \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i)P_{i-1}^2$ which is a truly random process. □

§3.6.7 **Lemma.** *Let* $X = (X_n)_{n\in\mathbb{N}_0}$ *be a square integrable martingale with square variation process* $\langle X\rangle$*, and let* τ *be a stopping time. Then the stopped process* X^τ *has square variation process* $\langle X^\tau\rangle = \langle X\rangle^\tau = (\langle X\rangle_{\tau\wedge n})_{n\in\mathbb{N}_0}$.

Proof of Lemma §3.6.7 is given in the lecture. □

§3.6.8 **Proposition.** *Let* $X := (X_n)_{n\in\mathbb{N}_0}$ *be a square integrable martingale with* $X_0 = 0$.

(a) *If $\mathbb{E}\langle X\rangle_\infty < \infty$, then the martingale X converges in L_2 and, hence X is regular; further, $\mathbb{E}(\sup_{n\in\mathbb{N}_0} X_n^2) \leqslant 4\mathbb{E}\langle X\rangle_\infty < \infty$.*

(b) *A stopping time τ is regular for the martingale X whenever $\mathbb{E}\sqrt{\langle X\rangle_\tau} < \infty$ and then $\mathbb{E}\sup_{n\in[\![0,\tau]\!]} |X_n| \leqslant 3\mathbb{E}\sqrt{\langle X\rangle_\tau} < \infty$.*

(c) *in every case the martingale X converges a.s. to a finite limit on the event $\{\langle X\rangle_\infty < \infty\}$.*

Proof of Proposition §3.6.8 is given in the lecture. □

§3.6.9 **Corollary**. *Let $X := (X_n)_{n\in\mathbb{N}_0}$ be a square integrable martingale with square variation process $\langle X\rangle$. Then the following four statements are equivalent:* (i) $\sup_{n\in\mathbb{N}_0}\mathbb{E}(X_n^2) < \infty$, (ii) $\lim_{n\to\infty}\mathbb{E}(\langle X\rangle_n) < \infty$, (iii) *$X$ converges in L_2, and* (iv) *X converges almost surely and in L_2.*

Proof of Corollary §3.6.9 is given in the lecture. □

§3.6.10 **Proposition**. *If X is a square integrable martingale, then for any $\alpha > 1/2$,*

$$(X_n - X_0)/(\langle X\rangle_n)^\alpha \xrightarrow{n\to\infty} 0 \quad a.s. \quad on\ \{\langle X\rangle_\infty = \infty\}.$$

Proof of Proposition §3.6.10 is given in the lecture. □

§3.6.11 **Example**. Let $X_1, X_2, \dots$ be independent, square integrable r.v.'s. Consider $S_0 := 0$ and $S_n := \sum_{i=1}^n (X_i - \mathbb{E}X_i)$, $n\in\mathbb{N}$, then $\langle S\rangle_n = \sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i)$ and by Proposition §3.6.10 for any $\alpha > 1/2$ we have $S_n/(\sum_{i=1}^n \mathbb{V}\mathrm{ar}(X_i))^\alpha \xrightarrow{a.s.} 0$ whenever $\sum_{i=1}^\infty \mathbb{V}\mathrm{ar}(X_i) = \infty$. In particular, if $(X_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. square integrable r.v.'s, then $S_n/n^\alpha \xrightarrow{a.s.} 0$. On the other hand, if $(a_n)_{n\in\mathbb{N}}$ is an increasing and diverging sequence in $\mathbb{R}$, then for any sequence $(y_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$ such that $\sum_{n=1}^\infty y_n/a_n < \infty$ by Kronecker's Lemma holds $a_n^{-1}\sum_{i=1}^n y_i \xrightarrow{n\to\infty} 0$. Thereby, if $\sum_{i=1}^\infty \mathbb{V}\mathrm{ar}(X_i)/a_i^2 < \infty$, then by Corollary §3.6.9 the martingale $(\sum_{i=1}^n (X_i - \mathbb{E}(X_i))/a_i)_{n\in\mathbb{N}}$ converges a.s. to a finite limit and, hence due to Kronecker's Lemma $a_n^{-1}\sum_{i=1}^n (X_i - \mathbb{E}X_i) \xrightarrow{a.s.} 0$. In case of i.i.d. r.v.'s we find $n^{-1}\sum_{i=1}^n (X_i - \mathbb{E}X_i) \xrightarrow{a.s.} 0$. □

Chapter 4

Markov chains

4.1 Time-homogeneous Markov chain

In this chapter $X = (X_n)_{n\in\mathbb{N}_0}$ denotes a time-homogeneous Markov chain with at most countable state space $(\mathcal{S},\mathscr{S})$ and transition matrix $P = (P_{ij})_{i,j\in\mathcal{S}}$ as introduced in Section 2.1.2. Considering the transition matrix P and an initial (discrete) probability measure μ on $(\mathcal{S},\mathscr{S})$

$$\mathbb{P}_{[\![0,n]\!]}(B_0\times\cdots\times B_n) := \sum_{j_0\in B_0}\mu(\{j_0\})\sum_{j_1\in B_1}P_{j_0,j_1}\cdots\sum_{j_n\in B_n}P_{j_{n-1},j_n},\text{ for } B_0,B_1,\ldots\in\mathscr{S}$$

defines a consistent family $\{\mathbb{P}_{\mathcal{J}},\mathcal{J}\subset\mathbb{N}_0 \text{ finite}\}$ of probability measures on the product space $(\mathcal{S}^{\mathbb{N}_0},\mathscr{S}^{\otimes\mathbb{N}_0})$ which determines by Kolmogorov's consistency theorem §2.3.11 a probability measure $\mathbb{P}_\mu$ on $(\mathcal{S}^{\mathbb{N}_0},\mathscr{S}^{\otimes\mathbb{N}_0})$. The Markov chain $X = (X_n)_{n\in\mathbb{N}_0}$ realised as a coordinate process, i.e., $X_n = \Pi_n : \mathcal{S}^{\mathbb{N}_0}\to\mathcal{S}$, $(j_m)_{m\in\mathbb{N}_0}\mapsto\Pi_n((j_m)_{m\in\mathbb{N}_0}) = j_n$ as defined in §2.2.3, admits then as image probability measure $\mathbb{P}_\mu$, that is, for $B_0,B_1,\ldots$ in $\mathscr{S}$ we have $\mathbb{P}_\mu(X_0\in B_0,\ldots,X_n\in B_n) = \mathbb{P}_{[\![0,n]\!]}(B_0\times\cdots\times B_n)$ and evidently $\mathbb{P}_\mu(X_0\in B_0) = \mu(B_0)$. When $\mu = \delta_j$, a point mass at $j\in\mathcal{S}$, we use $\mathbb{P}_j$ as an abbreviation for $\mathbb{P}_{\delta_j}$ where for every initial probability measure μ and for every $A\in\mathscr{S}^{\otimes\mathbb{N}_0}$ holds $\mathbb{P}_\mu(A) = \sum_{j\in\mathcal{S}}\mathbb{P}_j(A)\mu(\{j\})$.

§4.1.1 **Definition**. A stochastic process $X = (X_n)_{n\in\mathbb{N}_0}$ with values in an at most countable state space $(\mathcal{S},\mathscr{S})$ is called a time-homogeneous Markov chain with family of probability measures $(\mathbb{P}_j)_{j\in\mathcal{S}}$ on $(\mathcal{S}^{\mathbb{N}_0},\mathscr{S}^{\otimes\mathbb{N}_0})$, if

(i) For every $j\in\mathcal{S}$, $(X_n)_{n\in\mathbb{N}_0}$ is a stochastic process on the probability space $(\mathcal{S}^{\mathbb{N}_0},\mathscr{S}^{\otimes\mathbb{N}_0},\mathbb{P}_j)$ with $\mathbb{P}_j(X_0=j)=1$.

(ii) The map $\kappa : \mathcal{S}\times\mathscr{S}^{\otimes\mathbb{N}_0}\to[0,1]$, $(j,A)\mapsto\mathbb{P}_j(A)$ is a stochastic kernel (a regular conditional distribution). For every $n\in\mathbb{N}_0$, the map $\kappa_n : \mathcal{S}\times\mathscr{S}\to[0,1]$, $(j,B)\mapsto\kappa(j,\Pi_n^{-1}(B)) = \mathbb{P}_j(X_n\in B)$ is a stochastic kernel and the n-step transition matrix $(P^n_{ij})_{i,j\in\mathcal{S}}$ of X is given by $P^n_{ij} = \kappa_n(i,\{j\}) = \mathbb{P}_i(X_n=j)$.

(iii) $X = (X_n)_{n\in\mathbb{N}_0}$ has w.r.t. the natural filtration $\mathscr{F} = (\mathscr{F}_n)_{n\in\mathbb{N}_0}$ with $\mathscr{F}_n = \sigma(X_0,\ldots,X_n)$ the time-homogeneous Markov property: For every $i,j\in\mathcal{S}$ and all $m,n\in\mathbb{N}_0$ we have $\mathbb{P}_i[X_{n+m}=j|\mathscr{F}_m] = \kappa_n(X_m,\{j\}) = \mathbb{P}_{X_m}(X_n=j) = P^n_{X_mj}$, $\mathbb{P}_i$-a.s..

We write $\mathbb{E}_j$ for expectation with respect to $\mathbb{P}_j$, $\mathcal{L}_j(X) = \mathbb{P}_j$, $\mathcal{L}_j(X|\mathscr{A}) = \mathbb{P}_j[X\in\bullet|\mathscr{A}]$ for a regular conditional distribution of X given $\mathscr{A}$ and $\mathbb{E}_j[f(X)|\mathscr{A}]$ for a conditional expectation of $f(X)$ given $\mathscr{A}$. In particular, we use the notation $\mathbb{P}_{X_k} = \kappa(X_k,\bullet)$, that is, we understand X_k as the initial value of a second Markov chain with the same family of probability measures $(\mathbb{P}_j)_{j\in\mathcal{S}}$. □

§4.1.2 **Remark**. The existence of the family $(\kappa_n)_{n\in\mathbb{N}_0}$ of stochastic kernels implies the existence of the kernel κ (cf. Klenke [2008], Theorem 17.8, p.347). Thus, a time-homogeneous

Markov chain is simply a stochastic process with the Markov property and for which the transition probabilities are time-homogeneous. □

§4.1.3 **Definition.** Let $\mathbb{T} \subset \mathbb{R}$ be a set that is closed under addition (for example, $\mathbb{T} = \mathbb{N}_0$). The *shift operator* $\vartheta : \mathcal{S}^{\mathbb{T}} \to \mathcal{S}^{\mathbb{T}}$ is given by $(x_t)_{t\in\mathbb{T}} \mapsto \vartheta((x_t)_{t\in\mathbb{T}}) := (x_{t+1})_{t\in\mathbb{T}}$ and, for $s \in \mathbb{T}$, $\vartheta^s : \mathcal{S}^{\mathbb{T}} \to \mathcal{S}^{\mathbb{T}}$ is given by $(x_t)_{t\in\mathbb{T}} \mapsto \vartheta^s((x_t)_{t\in\mathbb{T}}) = (x_{t+s})_{t\in\mathbb{T}}$. □

§4.1.4 **Property** (Klenke [2008], Theorem 17.9, p.348, Corollary 17.10, p.349). *A stochastic process $X = (X_n)_{n\in\mathbb{N}_0}$ is a time-homogeneous Markov chain if and only if for every $n \in \mathbb{N}_0$ and $j \in \mathcal{S}$, $\mathcal{L}_j[\vartheta^n(X)|\mathscr{F}_n] = \mathcal{L}_{X_n}(X) = \mathbb{P}_{X_n}$ if and only if there exists a stochastic kernel $\kappa : \mathcal{S} \times \mathscr{S}^{\otimes\mathbb{N}_0} \to [0,1]$ such that, for every bounded $\mathscr{S}^{\otimes\mathbb{N}_0}$-measurable function $f : \mathcal{S}^{\mathbb{N}_0} \to \mathbb{R}$ and for every $n \in \mathbb{N}_0$ and $j \in \mathcal{S}$, we have $\mathbb{E}_j[f(\vartheta^n(X))|\mathscr{F}_n] = \mathbb{E}_{X_n}[f(X)] := \int_{\mathcal{S}^{\mathbb{N}_0}} \kappa(X_n, dx)f(x)$.* □

§4.1.5 **Definition.** A time-homogeneous Markov chain $X = (X_n)_{n\in\mathbb{N}_0}$ with family of probability measures $(\mathbb{P}_j)_{j\in\mathcal{S}}$ has the *strong Markov property* if, for every a.s. finite stopping time τ, and every $j \in \mathcal{S}$, $\mathcal{L}_j[\vartheta^\tau(X)|\mathscr{F}_\tau] = \mathcal{L}_{X_\tau}(X) := \kappa(X_\tau, \bullet)$ or equivalently for every bounded $\mathscr{S}^{\otimes\mathbb{N}_0}$-measurable function $f : \mathcal{S}^{\mathbb{N}_0} \to \mathbb{R}$ we have $\mathbb{E}_j[f(\vartheta^\tau(X))|\mathscr{F}_\tau] = \mathbb{E}_{X_\tau}[f(X)] := \int_{\mathcal{S}^{\mathbb{N}_0}} \kappa(X_\tau, dx)f(x)$. □

§4.1.6 **Lemma.** *Every time-homogeneous Markov chain $X = (X_n)_{n\in\mathbb{N}_0}$ has the strong Markov property.*

Proof of Lemma §4.1.6 is given in the lecture. □

4.2 Markov chains: recurrence and transience

§4.2.1 **Definition.** For $i, j \in \mathcal{S}$, $k \in \mathbb{N}$ introduce the *k-th time of return to j* recursively by $\tau_j^k := \inf\{n > \tau_j^{k-1} | X_n = j\}$ and $\tau_j^0 := 0$. We set further $\tau_j := \tau_j^1$ and $\rho_{ij} := \mathbb{P}_i(\tau_j < \infty)$. □

§4.2.2 **Remark.** Note that $\rho_{ij} = \mathbb{P}_i(\text{there is an } k \geqslant 1 \text{ with } X_k = j)$ is the probability of ever going from i to j. In particular, if $\rho_{ij} > 0$ then there exists a $k \in \mathbb{N}$ such that $\mathbb{P}_i(X_k = j) = P_{ij}^k > 0$. Moreover, ρ_{jj} is the return probability (after the first jump) from j to j. Note that $\tau_j^1 > 0$ even if we start the chain at $X_0 = j$. □

§4.2.3 **Definition.** A state $j \in \mathcal{S}$ is called (i) *recurrent* if $\rho_{jj} = 1$, (ii) *positive recurrent* if $\mathbb{E}_j(\tau_j) < \infty$, (iii) *null recurrent* if j is recurrent but not positive recurrent, (iv) *transient* if $\rho_{jj} < 1$, and (v) *absorbing*, if $P_{jj} = 1$. The Markov chain X is called {positive/null} recurrent if every state $j \in \mathcal{S}$ is {positive/null} recurrent and is called transient if every recurrent state is absorbing. □

§4.2.4 **Remark.** Clearly, we have: “absorbing” ⇒ “positive recurrent” ⇒ “recurrent”. □

§4.2.5 **Lemma.** *For $k \in \mathbb{N}$ and $i, j \in \mathcal{S}$ we have $\mathbb{P}_i(\tau_j^k < \infty) = \rho_{ij}\rho_{jj}^{k-1}$.*

Proof of Lemma §4.2.5 is given in the lecture. □

§4.2.6 **Definition.** For $i, j \in \mathcal{S}$ denote by $N_j := \sum_{n=0}^\infty \mathbb{1}_{\{X_n=j\}}$ the total *number of visits* of X to state j and by $G_{ij} = \mathbb{E}_i[N_j] = \sum_{n=0}^\infty \mathbb{P}_i(X_n = j) = \sum_{n=0}^\infty P_{ij}^n$ the *Green function* of X. □

§4.2.7 **Lemma.** (i) *A state $j \in \mathcal{S}$ is recurrent if and only if $G_{jj} = \infty$;*

(ii) *If a state $j \in \mathcal{S}$ is transient then for all $i \in \mathcal{S}$, $G_{ij} < \infty$ with*

$$G_{ij} = \begin{Bmatrix} \frac{\rho_{ij}}{1-\rho_{jj}}, & \text{if } i \neq j \\ \frac{1}{1-\rho_{jj}}, & \text{if } i = j \end{Bmatrix} = \frac{\rho_{ij}}{1-\rho_{jj}} + \mathbb{1}_{\{i=j\}}.$$

Proof of Lemma §4.2.7 is given in the lecture. □

§4.2.8 **Proposition.** *If a state $i \in \mathcal{S}$ is recurrent and $\rho_{ij} > 0$, $j \in \mathcal{S}$, then the state j is recurrent, and $\rho_{ij} = \rho_{ji} = 1$.*

Proof of Proposition §4.2.8 is given in the lecture. □

§4.2.9 **Definition.** A subset $B \subset \mathcal{S}$ of states is *closed* if $\rho_{ij} = 0$ holds for all $i \in B$ and $j \in B^c = \mathcal{S} \backslash B$. A subset $B \subset \mathcal{S}$ is *irreducible* if $\rho_{ij} > 0$ holds for all $i, j \in B$. If the state space $\mathcal{S}$ is irreducible then the Markov chain is called irreducible. □

§4.2.10 **Corollary.** *A irreducible Markov chain is either recurrent or transient. If $|\mathcal{S}| \geqslant 2$, then there is no absorbing state.*

Proof of Corollary §4.2.10 The result is an immediate consequence of Proposition §4.2.8. □

§4.2.11 **Proposition.** *For an irreducible Markov chain on a finite state space $\mathcal{S}$ all states are recurrent.*

Proof of Proposition §4.2.11 is given in the lecture. □

4.3 Invariant distributions

In the following, let $P = (P_{ij})_{i,j \in \mathcal{S}}$ be a transition matrix on a countable state space $\mathcal{S}$ and let $(X_n)_{n \in \mathbb{N}_0}$ be a corresponding Markov chain.

§4.3.1 **Definition.** If μ is a measure on $(\mathcal{S}, \mathscr{S})$ and $f : \mathcal{S} \to \mathbb{R}$ is a map, then we write $\mu P(\{j\}) = \sum_{i \in \mathcal{S}} \mu(\{i\}) P_{ij}$ and $Pf(i) = \sum_{j \in \mathcal{S}} P_{ij} f(j)$ if the sums converge. □

§4.3.2 **Definition.** (i) A σ-finite measure μ on $(\mathcal{S}, \mathscr{S})$ is called an *invariant measure* if $\mu P = \mu$. A probability measure that is an invariant measure is called an *invariant distribution*. Denote by $\mathcal{I}$ the set of invariant distributions.

(ii) A function $f : \mathcal{S} \to \mathbb{R}$ is called *subharmonic* if Pf exists and if $f \leqslant Pf$. f is called *superharmonic* if $f \geqslant Pf$ and *harmonic* if $f = Pf$. □

§4.3.3 **Remark.** In the terminology of linear algebra, an invariant measure is a left eigenvector of P corresponding to the eigenvalue 1. A harmonic function is a right eigenvector corresponding to the eigenvalue 1. □

§4.3.4 **Lemma.** *If f is bounded and {sub/super}harmonic, then $(f(X_n))_{n \in \mathbb{N}_0}$ is a {sub/super} martingale with respect to the natural filtration $\mathscr{F} = \sigma(X)$ generated by X.*

Proof of Lemma §4.3.4 is given in the lecture. □

§4.3.5 **Proposition.** *If X is transient, then an invariant distribution does not exist.*

Proof of Proposition §4.3.5 is given in the lecture. □

§4.3.6 **Theorem**. *Let j be a recurrent state and let $\tau_j = \inf\{n > 0 : X_n = j\}$. Then one invariant measure μ_j is defined by*

$$\mu_j(\{i\}) = \mathbb{E}_j\Big(\sum_{n=0}^{\tau_j - 1} \mathbb{1}_{\{X_n = i\}}\Big) = \sum_{n=0}^{\infty} \mathbb{P}_j(X_n = i; \tau_j > n).$$

Proof of Theorem §4.3.6 is given in the lecture. □

§4.3.7 **Corollary**. *If X is positive recurrent, then $\pi := \mu_j[\mathbb{E}_j(\tau_j)]^{-1}$ is an invariant distribution for any $j \in \mathcal{S}$.*

Proof of Corollary §4.3.7 is given in the lecture. □

§4.3.8 **Theorem**. *If X is irreducible, then X has at most one invariant distribution.*

Proof of Theorem §4.3.8 is given in the lecture. □

§4.3.9 **Remark**. One could in fact show that if X is irreducible and recurrent, then an invariant measure of X is unique up to a multiplicative factor (see Durrett [1996], Theorem 5.4.4). On the other hand, for transient X, there can be more than one invariant measure (see Klenke [2008], Remark 17.50). □

Recall that $\mathcal{I}$ is the set of invariant distributions of X.

§4.3.10 **Theorem**. *Let X be irreducible. X is positive recurrent if and only if $\mathcal{I} \neq \emptyset$. In this case, $\mathcal{I} = \{\pi\}$ with $\pi(\{j\}) = \big[\mathbb{E}_j(\tau_j)\big]^{-1} > 0$ for all $j \in \mathcal{S}$.*

Proof of Theorem §4.3.10 is given in the lecture. □

§4.3.11 **Corollary**. *If X is irreducible, then the following statements are equivalent:* (i) *There exists a positive recurrent state.* (ii) *There exists a invariant distribution.* (iii) *All states are positive recurrent.*

Proof of Corollary §4.3.11 is given in the lecture. □

Chapter 5

Ergodic theory

5.1 Stationary and ergodic processes

Ergodic theory is the study of laws of large numbers for possibly dependent, but stationary, random variables.

§5.1.1 **Definition**. Let $\mathbb{T} \subset \mathbb{R}$ be a set that is closed under addition (e.g., $\mathbb{T} \in \{\mathbb{N}_0, \mathbb{Z}, \mathbb{R}^+, \mathbb{R}\}$) and ϑ be the shift operator as defined in §4.1.3. A stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called stationary if $\mathbb{P}_{\vartheta^t(X)} = \mathbb{P}_X$ for all $t \in \mathbb{T}$. □

§5.1.2 **Remark**. If $\mathbb{T} = \mathbb{N}$ then $\mathbb{P}_{\vartheta^n(X)} = \mathbb{P}_X$ for all $n \in \mathbb{N}$ is equivalent to $\mathbb{P}_{\vartheta(X)} = \mathbb{P}_X$. □

§5.1.3 **Example**. (i) If $X = (X_t)_{t\in\mathbb{T}}$ is i.i.d., then X is stationary. Dismissing the independence assumption, i.e., $\mathbb{P}_{X_t} = \mathbb{P}_{X_0}$ holds for every $t \in \mathbb{T}$, in general X is not stationary. For example, consider $\mathbb{T} = \mathbb{N}_0$ and $X_1 = X_2 = X_3 = \dots$ but $X_0 \neq X_1$. Then X is not stationary.

(ii) Let X be a Markov chain with invariant distribution π. If π is the initial probability measure, i.e., $\mathbb{P}_\pi$ is the distribution of X, then X is stationary.

(iii) Let $X = (X_n)_{n\in\mathbb{Z}}$ be i.i.d. real r.v.'s and let $c_1, \dots, c_k \in \mathbb{R}$. Then $Y_n := \sum_{l=1}^k c_l X_{n-l}$, $n \in \mathbb{Z}$, defines a stationary process Y that is called the moving average with weights $c_1, \dots, c_k$. In fact, Y is stationary if only X is stationary. □

In the sequel, assume that $(\Omega, \mathscr{A}, \mathbb{P})$ is a probability space and $T : \Omega \to \Omega$ is a measurable map.

§5.1.4 **Definition**. T is called *measure preserving* (maßerhaltend) if $\mathbb{P}_T(A) = \mathbb{P}(T^{-1}(A)) = \mathbb{P}(A)$ holds for all $A \in \mathscr{A}$. In this case $(\Omega, \mathscr{A}, \mathbb{P}, T)$ is called a (measure preserving) *dynamical system*. □

§5.1.5 **Example**. Let $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$ be a Polish space equipped with its Borel-σ-algebra.

(i) For a $\mathcal{S}$-valued r.v. Y and a measure preserving map T on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ the process $X_n(\omega) := Y(T^n(\omega))$, $n \in \mathbb{N}_0$, is stationary.

(ii) Let $X = (X_n)_{n\in\mathbb{N}_0}$ be the coordinate process on $(\Omega, \mathscr{A}, \mathbb{P}) = (\mathcal{S}^{\mathbb{N}_0}, \mathscr{B}(\mathcal{S})^{\otimes\mathbb{N}_0}, \mathbb{P})$. If ϑ is the shift operator as defined in §4.1.3, then $X_n(\omega) = X_0(\vartheta^n(\omega))$. X is stationary if and only if $(\Omega, \mathscr{A}, \mathbb{P}, \vartheta)$ is a dynamical system. Moreover, if X is stationary and Y is a $\mathcal{S}$-valued r.v. on $(\Omega, \mathscr{A}, \mathbb{P})$, then $Y_n = Y(\vartheta^n(X))$ is stationary. □

§5.1.6 **Definition**. An event $A \in \mathscr{A}$ is called *strictly invariant* if $T^{-1}(A) = A$ and *(almost) invariant* if $\mathbb{1}_{T^{-1}(A)} = \mathbb{1}_A$ $\mathbb{P}$-a.s., that is $\mathbb{P}(T^{-1}(A)\Delta A) = 0$. The σ-algebra of all (almost) invariant events is denoted by $\mathscr{I}_T$. □

Recall that a σ-algebra $\mathscr{A}$ is called $\mathbb{P}$-trivial if $\mathbb{P}(A) \in \{0, 1\}$ for every $A \in \mathscr{A}$.

§5.1.7 **Definition**. If T is measure preserving and the σ-algebra $\mathscr{I}_T$ of (almost) invariant events is $\mathbb{P}$-trivial, then $(\Omega, \mathscr{A}, \mathbb{P}, T)$ is called *ergodic*. □

§5.1.8 **Remark**. For every (almost) invariant event $A \in \mathscr{I}_T$ there exists a strictly invariant event A^* such that $\mathbb{P}(A\Delta A^*) = 0$. Thereby, if the σ-algebra $\mathscr{I}_T^*$ of all strictly invariant events is $\mathbb{P}$-trivial, then $(\Omega, \mathscr{A}, \mathbb{P}, T)$ is ergodic. □

§5.1.9 **Lemma**. *(i) A measurable map $f : (\Omega, \mathscr{A}) \to (\mathbb{R}, \mathscr{B})$ is $\mathscr{I}_T$-measurable if and only if $f \circ T = f$.*

(ii) $(\Omega, \mathscr{A}, \mathbb{P}, T)$ is ergodic if and only if any $\mathscr{I}_T$-measurable $f : (\Omega, \mathscr{I}_T) \to (\mathbb{R}, \mathscr{B})$ is $\mathbb{P}$-a.s. constant.

Proof of Lemma §5.1.9 is given in the lecture. □

§5.1.10 **Definition**. If $(\mathcal{S}^{\mathbb{N}_0}, \mathscr{B}(\mathcal{S})^{\otimes\mathbb{N}_0}, \mathbb{P}, \vartheta)$ is ergodic, then the coordinate process $(X_n)_{n\in\mathbb{N}_0}$ (as in Example §5.1.5 (ii)) is called *ergodic*. □

§5.1.11 **Example**. Consider $X = (X_n)_{n\in\mathbb{N}_0}$ and $Y = (Y_n)_{n\in\mathbb{N}_0}$ as in Example §5.1.5 (ii).

(i) If X is ergodic, then Y is ergodic.

(ii) Let $(X_n)_{n\in\mathbb{N}_0}$ be i.i.d. If $A \in \mathscr{I}_\vartheta$, then, $A = (\vartheta^n)^{-1}(A) = \{\omega : \vartheta^n(\omega) \in A\} \in \sigma(\vartheta^n(X)) = \sigma(X_n, X_{n+1}, \dots)$ for every $n \in \mathbb{N}_0$. Hence, if we let $\mathscr{T} := \cap_{n=1}^{\infty}\sigma(\vartheta^n(X))$ be the tail σ-algebra of $(X_n)_{n\in\mathbb{N}}$ then $\mathscr{I}_T \subset \mathscr{T}$. By Kolmogorov's 0-1 law (Theorem §1.3.8), $\mathscr{T}$ is $\mathbb{P}$-trivial. Hence, $\mathscr{I}_T$ is also $\mathbb{P}$-trivial and therefore $(X_n)_{n\in\mathbb{N}_0}$ is ergodic. □

5.2 Ergodic theorems

In this section, $(\Omega, \mathscr{A}, \mathbb{P}, T)$ always denotes a measure preserving dynamical system. Further let $f : \Omega \to \mathbb{R}$ be measurable and

$$X_n(\omega) = f \circ T^n(\omega) \qquad \text{for all } n \in \mathbb{N}_0.$$

Hence $X = (X_n)_{n\in\mathbb{N}_0}$ is a stationary real-valued stochastic process. Let

$$S_n := \sum_{k=0}^{n-1} X_k \quad (S_0 := 0)$$

denote the nth partial sum. Ergodic theorems are laws of large numbers for $(S_n)_{n\in\mathbb{N}}$. We start with a preliminary lemma.

§5.2.1 **Lemma** (Hopf's maximal-ergodic lemma). *Let $f = X_0 \in L_1(\mathbb{P})$. Define $M_n := \max\{S_k, k \in [\![0, n]\!]\}$, $n \in \mathbb{N}$, and $M_\infty := \sup\{S_k, k \in \mathbb{N}_0\}$. Then $\mathbb{E}(X_0\mathbb{1}_{\{M_n>0\}}) \geqslant 0$ for every $n \in \mathbb{N}$ and by dominated convergence $\mathbb{E}(X_0\mathbb{1}_{\{M_\infty>0\}}) \geqslant 0$.*

Proof of Lemma §5.2.1 is given in the lecture. □

§5.2.2 **Theorem** (Birkhoff's ergodic theorem). *Let $X_0 \in L_1(\mathbb{P})$. Then*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k = \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \xrightarrow{n\to\infty} \mathbb{E}[X_0|\mathscr{S}_T] \quad \mathbb{P}\text{-a.s.}.$$

In particular, if T is ergodic, then $\frac{1}{n}\sum_{k=0}^{n-1} X_k \xrightarrow{n\to\infty} \mathbb{E}[X_0]$ $\mathbb{P}$-a.s..

Proof of Theorem §5.2.2 is given in the lecture. □

§5.2.3 **Lemma**. *Let $p \geqslant 1$ and let $(X_n)_{n\in\mathbb{N}_0}$ be identically distributed, real r.v.'s with $\mathbb{E}(|X_0|^p) < \infty$. Define $Y_n := \left|\frac{1}{n}\sum_{k=0}^{n-1} X_k\right|^p$ for $n \in \mathbb{N}$. Then $(Y_n)_{n\in\mathbb{N}}$ is uniformly integrable.*

Proof of Lemma §5.2.3 is given in the lecture. □

§5.2.4 **Theorem** (von Neumann's ergodic theorem). *Let $(\Omega, \mathscr{A}, \mathbb{P}, T)$ be a measure preserving dynamical system, $p \geqslant 1$, $X_0 \in L_p(\mathbb{P})$ and $X_n = X_0 \circ T^n$. Then*

$$\frac{1}{n}\sum_{k=0}^{n-1} X_k \xrightarrow{n\to\infty} \mathbb{E}[X_0|\mathscr{S}_T] \quad \text{in } L_p(\mathbb{P}).$$

In particular, if T is ergodic, then $\frac{1}{n}\sum_{k=0}^{n-1} X_k \xrightarrow{n\to\infty} \mathbb{E}[X_0]$ in $L_p(\mathbb{P})$.

Proof of Theorem §5.2.4 is given in the lecture. □

§5.2.5 **Theorem**. *Let X be a positive recurrent, irreducible Markov chain on a countable state space $\mathcal{S}$. Let π be the invariant distribution of X given in Theorem §4.3.10. If π is the initial probability measure of X, then the Markov chain is ergodic.*

Proof of Theorem §5.2.5 is given in the lecture. □

§5.2.6 **Remark**. By Corollary §4.3.11 for a irreducible Markov chain are equivalent: (i) There exists a positive recurrent state. (ii) There exists a invariant distribution π. (iii) All states are positive recurrent. Thereby, an irreducible Markov chain with some positive-recurrent state j is ergodic under the invariant initial distribution π or, if an irreducible Markov chain has an invariant distribution, then it is ergodic. □

Chapter 6

Weak convergence

6.1 Fundamental properties

In the sequel, $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$ denotes a metric space $(\mathcal{S}, d)$ equipped with Borel σ-algebra $\mathscr{B}(\mathcal{S})$. The space of all bounded continuous and real-valued functions on $\mathcal{S}$ is denoted by $\mathcal{C}_b(\mathcal{S})$. If μ is a measure on $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$ and $f \in L_1(\mathcal{S}, \mathscr{B}(\mathcal{S}), \mu)$ we write $\mu f = \int_{\mathcal{S}} f d\mu$.

§6.1.1 **Definition**. Let $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$ be probability measures on $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$. We say that $(\mathbb{P}_n)_{n\in\mathbb{N}}$ *converges weakly* to $\mathbb{P}$, if $\lim_{n\to\infty} \mathbb{P}_n f = \mathbb{P} f$ for all $f \in \mathcal{C}_b(\mathcal{S})$, and we write formally $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ or $\mathbb{P} = w\text{-}\lim\limits_{n\to\infty} \mathbb{P}_n$. □

§6.1.2 **Remark**. Weak convergence induces on the space of finite measures on $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$ the weak topology (or weak*-topology in functional analysis). This is the coarsest topology such that for all $f \in \mathcal{C}_b(\mathcal{S})$, the map $\mu \mapsto \mu f$ is continuous. If $\mathcal{S}$ is separable, then it can be shown that the weak topology is metrisable; for example, by virtue of the so-called Prohorov metric (see, for example, Billingsley [1999], Appendix III). □

§6.1.3 **Example**. Let $x, x_1, x_2, \dots$ be elements of $\mathcal{S}$ such that $d(x_n, x) \xrightarrow{n\to\infty} 0$. Consider $\mathbb{P}_n := \delta_{x_n}$, $n \in \mathbb{N}$ and $\mathbb{P} := \delta_x$. Then by definition $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ since $f(x_n) \xrightarrow{n\to\infty} f(x)$ for all $f \in \mathcal{C}_b(\mathcal{S})$. For open $O \in \mathscr{B}(\mathcal{S})$ with $x_n \in O$, $n \in \mathbb{N}$, and $x \in \partial O$ we have $\lim_{n\to\infty} \mathbb{P}_n(O) = 1$ while $\mathbb{P}(O) = 0$. For events $B \in \mathscr{B}(\mathcal{S})$ with $x \notin \partial B$ and $x \in B^{\circ}$ it follows $x_n \in B^{\circ}$ for all $n \geqslant n_o$ and thus $\lim_{n\to\infty} \mathbb{P}_n(B) = 1 = \mathbb{P}(B)$. □

For measurable $g : \mathcal{S} \to \mathbb{R}$, let U_g be the set of points of discontinuity of g where U_g is Borel measurable.

§6.1.4 **Theorem** (Portemanteau). *For probability measures $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$ on $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$ the following are equivalent:*

(i) $\mathbb{P} = w\text{-}\lim\limits_{n\to\infty} \mathbb{P}_n$;

(ii) $\lim\limits_{n\to\infty} \mathbb{P}_n f = \mathbb{P} f$ *for all bounded Lipschitz continuous* f*;*

(iii) $\lim\limits_{n\to\infty} \mathbb{P}_n f = \mathbb{P} f$ *for all bounded measurable* f *with* $\mathbb{P}(U_f) = 0$*;*

(iv) $\liminf\limits_{n\to\infty} \mathbb{P}_n(U) \geqslant \mathbb{P}(U)$ *for all open* $U \subset \mathcal{S}$*;*

(v) $\limsup\limits_{n\to\infty} \mathbb{P}_n(F) \leqslant \mathbb{P}(F)$ *for all closed* $F \subset \mathcal{S}$*;*

(vi) $\lim_{n\to\infty} \mathbb{P}_n(B) = \mathbb{P}(B)$ *for all measurable* B *with* $\mathbb{P}(\partial B) = 0$.

Proof of Theorem §6.1.4 is given in the lecture. □

§6.1.5 **Definition**. Let $X, X_1, X_2, \dots$ be r.v.'s with values in $\mathcal{S}$. We say that $(X_n)_{n\in\mathbb{N}}$ *converges in distribution* to X, formally $X_n \xrightarrow{d} X$ or $X_n \xrightarrow{d} \mathbb{P}_X$, if the distributions converge weakly, and hence if $\mathbb{P}_X = \underset{n\to\infty}{w\text{-lim}}\, \mathbb{P}_{X_n}$. □

§6.1.6 **Theorem** (Slutzky's theorem). *Let $X, X_1, X_2, \dots$ and $Y_1, Y_2, \dots$ be r.v.'s with values in $\mathcal{S}$. Assume $X_n \xrightarrow{d} X$ and $d(X_n, Y_n) \xrightarrow{\mathbb{P}} 0$. Then $Y_n \xrightarrow{d} X$.*

Proof of Theorem §6.1.6 is given in the lecture. □

§6.1.7 **Corollary**. *If $X_n \xrightarrow{\mathbb{P}} X$, then $X_n \xrightarrow{d} X$. The converse is false in general.*

Proof of Corollary §6.1.7 is given in the lecture. □

§6.1.8 **Example**. If $X, X_1, X_2, \dots$ are i.i.d. (with nontrivial distribution), then trivially $X_n \xrightarrow{d} X$ but not $X_n \xrightarrow{\mathbb{P}} X$. □

§6.1.9 **Definition**. Let $F, F_1, F_2, \dots$ be distribution functions of probability measures on $\mathbb{R}$. We say that $(F_n)_{n\in\mathbb{N}}$ *converges weakly* to F, formally $F_n \xrightarrow{d} F$ or $F = \underset{n\to\infty}{w\text{-lim}}\, F_n$, if $F(x) = \lim_{n\to\infty} F_n(x)$ for all points of continuity x of F. □

§6.1.10 **Example**. If F is the distribution function of a probability measure on $\mathbb{R}$ and $F_n(x) := F(x+n)$ for $x \in \mathbb{R}$, then $(F_n)_{n\in\mathbb{N}}$ converges pointwise to 1. However, this is not a distribution function, as 1 does not converge to 0 for $x \to -\infty$. On the other hand, if $G_n(x) := F(x-n)$, then $(G_n)_{n\in\mathbb{N}}$ converges pointwise to $G \equiv 0$. However, $G(\infty) = 0 < \limsup_{n\to\infty} G_n(\infty) = 1$; hence we do not have weak convergence here either. Indeed, in each case, there is a mass defect in the limit (in the case of the F_n on the left and in the case of the G_n on the right). However, the definition of weak convergence of distribution functions is constructed so that no mass defect occurs in the limit. □

§6.1.11 **Theorem** (Helly-Bray). *Let $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$ probability measures on $\mathbb{R}$ with corresponding distribution functions $F, F_1, F_2, \dots$. The following are equivalent:* (i) $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ *and* (ii) $F_n \xrightarrow{d} F$.

Proof of Theorem §6.1.11 is given in the lecture. □

§6.1.12 **Corollary**. *Let $X, X_1, X_2, \dots$ be real r.v.'s with distribution functions $F, F_1, F_2, \dots$. Then the following are equivalent:* (i) $X_n \xrightarrow{d} X$; (ii) $\mathbb{E}[f(X_n)] \xrightarrow{n\to\infty} \mathbb{E}[f(X)]$ *for all* $f \in \mathcal{C}_b(\mathbb{R})$ *and* (iii) $F_n \xrightarrow{d} F$. □

§6.1.13 **Theorem** (Continuous mapping theorem). *Let $(\mathcal{S}_1, d_1)$ and $(\mathcal{S}_2, d_2)$ be metric spaces and let $\varphi : \mathcal{S}_1 \to \mathcal{S}_2$ be measurable. Denote by U_φ the set of points of discontinuity of φ.*

(i) *If $\mathbb{P}, \mathbb{P}_1, \mathbb{P}_2, \dots$ be probability measures on $\mathcal{S}$ with $\mathbb{P}(U_\varphi) = 0$ and $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$, then $\mathbb{P}_n \circ \varphi^{-1} \xrightarrow{w} \mathbb{P} \circ \varphi^{-1}$.*

(ii) *If $X, X_1, X_2, \dots$ are $\mathcal{S}_1$-valued r.v.'s with $\mathbb{P}(X \in U_\varphi) = 0$ and $X_n \xrightarrow{d} X$, then $\varphi(X_n) \xrightarrow{d} \varphi(X)$.*

Proof of Theorem §6.1.13 is given in the lecture. □

6.2 Prohorov's theorem

§6.2.1 **Example**. Let $(\mathcal{C}([0,1]), \|\cdot\|_{\sup})$ denote the metric space of continuous and real-valued functions on $[0,1]$ equipped with the topology of uniform convergence using the metric $d(f,g) = \|f-g\|_{\sup} = \sup_{t\in[0,1]} |f(t)-g(t)|$. Recall that the canonical projections $\Pi_{\{t_1,\dots,t_k\}} : \mathcal{C}([0,1]) \to \mathbb{R}^k$, $f \mapsto (f(t_1),\dots,f(t_k))$ are continuous and thus measurable. Moreover, any probability measure on $(\mathcal{C}([0,1]), \mathscr{B}(\mathcal{C}([0,1])))$ is uniquely determined by its finite dimensional distributions. However, weak converges of the finite dimensional distributions does generally not imply weak convergence in $(\mathcal{C}([0,1]), \mathscr{B}(\mathcal{C}([0,1])))$. For example, let $\mathbb{P}_n := \delta_{x_n}$, $n \in \mathbb{N}$, with $x_n(t) = nt\mathbb{1}_{[0,1/n]} + (2-nt)\mathbb{1}_{[1/n,2/n]}$, $t \in [0,1]$, and $\mathbb{P} := \delta_{x_0}$ with $x_0 \equiv 0$. Obviously, $(\mathbb{P}_n)_{n\in\mathbb{N}}$ would converges weakly to $\mathbb{P}$, if $\mathbb{P}_n f = f(x_n)$ converges to $\mathbb{P} f = f(x_0)$ for all $f \in \mathcal{C}_b(\mathcal{C}([0,1]))$. Consider $f(x) := \min(\|x\|_{\sup}, 1)$ then $f(x_n) = 1$, $n \in \mathbb{N}$, and $f(x_0) = 0$, hence $(\mathbb{P}_n)_{n\in\mathbb{N}}$ does not converge weakly to $\mathbb{P}$. On the other hand, $x_n(t) \overset{n\to\infty}{\longrightarrow} 0 = x_0(t)$ for all $t \in [0,1]$ and, thus the finite dimensional distributions converge weakly. □

§6.2.2 **Definition**. Let $(\mathcal{S}, d)$ be a metric space equipped with its Borel-σ-algebra $\mathscr{B}(\mathcal{S})$ and let $\mathcal{P}(\mathcal{S})$ denote the space of all probability measures on $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$. A family $\mathcal{F} \subset \mathcal{P}(\mathcal{S})$ of probability measures is called

(a) *weakly relatively sequentially compact* if each sequence $(\mathbb{P}_n)_{n\in\mathbb{N}}$ in $\mathcal{F}$ has a weakly convergent subsequence with limit in $\mathcal{P}(\mathcal{S})$:

(b) *weakly sequentially compact* if each sequence $(\mathbb{P}_n)_{n\in\mathbb{N}}$ in $\mathcal{F}$ has a weakly convergent subsequence with limit in $\mathcal{F}$. □

§6.2.3 **Remark**. If $(\mathcal{S}, d)$ is separable, then the weak topology is metrisable (Remark §6.1.2), and thus the notions compact and sequentially compact coincide (Remark §2.3.2). □

§6.2.4 **Proposition**. *Let $\mathcal{S}$ be a compact metric space. Then the set $\mathcal{P}(\mathcal{S})$ is weakly (sequentially) compact.*

Proof of Proposition §6.2.4 is given in the lecture. □

§6.2.5 **Theorem**. *Let $(\mathcal{S}, d)$ be Polish and let μ be a measure on the Borel-σ-algebra $\mathscr{B}(\mathcal{S})$. Then there is a compact metric space $(\mathcal{S}^*, d^*)$ and a measure μ^* on $\mathscr{B}(\mathcal{S}^*)$ satisfying*

(i) *$\mathcal{S}$ is a subset of $\mathcal{S}^*$;*

(ii) *$\mathscr{B}(\mathcal{S})$ is a subset of $\mathscr{B}(\mathcal{S}^*)$ and $\mu(B) = \mu^*(B)$ for all $B \in \mathscr{B}(\mathcal{S})$;*

(iii) $\mu^*(\mathcal{S}^* \backslash \mathcal{S}) = 0$.

In particular, $\mathcal{S}$ is G_δ (a countable intersection of open sets in $\mathscr{B}(\mathcal{S}^*)$) and hence $\mathcal{S}$ is $\mathscr{B}(\mathcal{S}^*)$-measurable.

Proof of Theorem §6.2.5 An outline of the proof is given in the lecture. □

§6.2.6 **Definition**. A family $\mathcal{F} \subset \mathcal{P}(\mathcal{S})$ of probability measures on $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$ is called *(uniformly) tight* (straff) if, for any $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset \mathcal{S}$ such that $\mathbb{P}(K_\varepsilon) \geqslant 1-\varepsilon$ for all $\mathbb{P} \in \mathcal{F}$. □

§6.2.7 **Remark**. If $\mathcal{S}$ is Polish, then by Proposition §2.3.9, every singleton, $\{\mathbb{P}\}$ with $\mathbb{P} \in \mathcal{P}(\mathcal{S})$, is tight and thus so is every finite family. □

§6.2.8 **Theorem** (Prohorov's theorem). *Let $(\mathcal{S}, d)$ be a metric space and let $\mathcal{F} \subset \mathcal{P}(\mathcal{S})$ be a family of probability measures on $(\mathcal{S}, \mathscr{B}(\mathcal{S}))$.*

(i) *If $\mathcal{F}$ is tight then $\mathcal{F}$ is weakly relatively sequentially compact.*

(ii) *If $\mathcal{S}$ is Polish, then also the converse holds: If $\mathcal{F}$ is weakly relatively sequentially compact then $\mathcal{F}$ is tight.*

Proof of Theorem §6.2.8 In the lecture a proof of (i) and (ii) is given assuming $\mathcal{S}$ is Polish. In case $\mathcal{S}$ is not Polish the proof of (i) is far more involved and we refer to Billingsley [1999] (Theorem 6.1 and 6.2) or Klenke [2008] (Theorem 13.29). □

Let X and $(X^n)_{n\in\mathbb{N}}$ be r.v.'s with values in $(\mathcal{C}([0,1]), \|\cdot\|_{\sup})$ equipped with its Borel-σ-algebra $\mathscr{B}(\mathcal{C}([0,1]))$ (i.e., continuous stochastic processes) with distributions $\mathbb{P}_X$ and $(\mathbb{P}_{X^n})_{n\in\mathbb{N}}$.

§6.2.9 **Definition**. We say that the *finite-dimensional distributions* of (X^n) *converge* to those of X if, for every $k \in \mathbb{N}$ and $t_1, \dots, t_k \in [0,1]$, we have $(X^n_{t_1}, \dots, (X^n_{t_k}) \xrightarrow{d} (X_{t_1}, \dots, (X_{t_k})$. In this case, we write $X^n \xrightarrow{fdd} X$ or $\mathbb{P}_{X^n} \xrightarrow{fdd} \mathbb{P}_X$. □

§6.2.10 **Remark**. The finite dimensional distributions determine uniquely a probability measure on $(\mathcal{C}([0,1]), \mathscr{B}(\mathcal{C}([0,1])))$. Consequently, $\mathbb{P}_n \xrightarrow{fdd} \mathbb{P}$ and $\mathbb{P}_n \xrightarrow{fdd} \mathbb{Q}$ imply $\mathbb{P} = \mathbb{Q}$. □

§6.2.11 **Proposition**. *Weak convergence implies convergence of the finite-dimensional distributions: $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$ implies $\mathbb{P}_n \xrightarrow{fdd} \mathbb{P}$.*

Proof of Proposition §6.2.11 is given in the lecture. □

§6.2.12 **Theorem**. *Let $(\mathbb{P}_n)_{n\in\mathbb{N}}$ and $\mathbb{P}$ be probability measures on $(\mathcal{C}([0,1]), \mathscr{B}(\mathcal{C}([0,1])))$. Then the following are equivalent:*

(i) $\mathbb{P}_n \xrightarrow{fdd} \mathbb{P}$ *and $(\mathbb{P}_n)_{n\in\mathbb{N}}$ is (uniformly) tight.*

(ii) $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$.

Proof of Theorem §6.2.12 is given in the lecture. □

§6.2.13 **Definition**. For $\delta > 0$ and $f \in \mathcal{C}([0,1])$ the *modulus of continuity* is defined by $w_f(\delta) := \sup\{|f(t) - f(s)| : |t - s| \leqslant \delta, t, s \in [0,1]\}$. □

§6.2.14 **Remark**. Since any $f \in \mathcal{C}([0,1])$ is uniformly continuous, it follows $\lim_{\delta\to 0} w_f(\delta) = 0$ and moreover $|w_f(\delta) - w_g(\delta)| \leqslant 2\|f - g\|_{\sup}$. Thereby, for fixed δ, $w_\bullet(\delta)$ is continuous on $(\mathcal{C}([0,1]), \|\cdot\|_{\sup})$ and thus $\mathscr{B}(\mathcal{C}([0,1]))$-measurable. □

§6.2.15 **Theorem** (Arzelà-Ascoli). *Let (K, d) be a compact metric space and let $(\mathcal{C}(K), \|\cdot\|_{\sup})$ be the metric space of continuous and real-valued functions on K. A subset $B \subset \mathcal{C}(K)$ is relatively compact if and only if the following two conditions hold.*

(i) *There exists $x \in K$ and $c > 0$ such that $|f(x)| \leqslant c < \infty$ for all $f \in B$.*

(ii) *We have $\lim_{\delta\downarrow 0} \sup\{w_\delta(f) : f \in B\} = 0$.*

Proof of Theorem §6.2.8 We refer to, e.g., Dudley [2002] (Theorem 2.4.7). □

§6.2.16 **Remark**. Due to (ii) the condition (i) can be replaced by: B is bounded in $(\mathcal{C}(K), \|\cdot\|_{\sup})$, that is, there is $c > 0$, $\|f\|_{\sup} \leqslant c < \infty$ for all $f \in B$. □

§6.2.17 **Theorem**. *A sequence $(\mathbb{P}_n)$ of probability measures on $\mathcal{C}([0,1])$ is (uniformly) tight if and only if the following two conditions hold.*

(i) *For every $\eta > 0$, there is $a > 0$ such that $\sup_{n\in\mathbb{N}} \mathbb{P}_n(\{f : |f(0)| \geqslant a\}) \leqslant \eta$.*

(ii) *For all $\varepsilon, \eta > 0$ there is $\delta \in (0,1)$ such that $\sup_{n\in\mathbb{N}} \mathbb{P}_n(\{f : w_f(\delta) \geqslant \varepsilon\}) \leqslant \eta$.*

Thereby, a sequence of r.v.'s $(X^n)_{n\in\mathbb{N}}$ with values in $\mathcal{C}([0,1])$ is (uniformly) tight, if the sequence $(X_0^n)_{n\in\mathbb{N}}$ is (uniformly) tight and

(iii) *For all $\varepsilon, \eta > 0$ there is $\delta \in (0,1)$ such that $\mathbb{P}\big(\sup_{|t-s|\leqslant\delta} |X_s^n - X_t^n| \geqslant \varepsilon\big) \leqslant \eta$ for all $n \in \mathbb{N}$.*

Proof of Theorem §6.2.17 is given in the lecture. □

§6.2.18 **Corollary**. *Let $(X^n)_{n\in\mathbb{N}}$ and $(Y^n)_{n\in\mathbb{N}}$ be families of r.v.'s with values in $\mathcal{C}([0,1])$. Assume that $(\mathbb{P}_{X^n})_{n\in\mathbb{N}}$ and $(\mathbb{P}_{Y^n})_{n\in\mathbb{N}}$ are tight. Then $(\mathbb{P}_{X^n+Y^n})_{n\in\mathbb{N}}$ is tight.* □

Proof of Corollary §6.2.18 Apply the triangle inequality in order to check (i) and (ii) in Theorem §6.2.17. □

§6.2.19 **Example**. Let $\mathcal{C}([0,\infty))$ denote the space of continuous and real-valued functions on $[0,\infty)$ equipped with the topology of uniform convergence on compact sets. For $f, g \in \mathcal{C}([0,\infty))$ and $N \in \mathbb{N}$, let $d_N(f,g) := \big\|(f-g)|_{[0,N]}\big\|_{\sup} \wedge 1$ and $d(f,g) := \sum_{N=1}^{\infty} 2^{-N} d_N(f,g)$. Consider $\Omega = \mathcal{C}([0,\infty)) \subset \mathbb{R}^{[0,\infty)}$. Define the evaluation map $\Pi_t : \Omega \to \mathbb{R}$, $\omega \mapsto \omega(t)$, that is, with a slight abuse of notations the restriction of the canonical projection $\mathbb{R}^{[0,\infty)} \to \mathbb{R}$ to Ω. Then d is a complete metric on Ω that induces the topology of uniform convergence on compact sets. The space (Ω, d) is separable and Polish (c.f. Klenke [2008], Theorem 21.30). Moreover, with respect to the Borel-σ-algebra $\mathscr{B}(\Omega)$, the canonical projections Π_t, $t \in [0,\infty)$ are measurable. On the other hand, the Π_t generate $\mathscr{B}(\Omega)$. Hence, $\mathscr{B}(\mathbb{R})^{\otimes[0,\infty)}|_\Omega = \sigma\big(\Pi_t, t \in [0,\infty)\big) = \mathscr{B}(\Omega)$ (c.f. Klenke [2008], Theorem 21.31). Thereby, the map $\mathcal{C}([0,\infty)) \to [0,\infty]$, $f \mapsto \sup\{f(t) : t \in [0,\infty)\}$ is $\sigma\big(\Pi_t, t \in [0,\infty)\big)$ measurable. Let X and $(X^n)_{n\in\mathbb{N}}$ be r.v.'s with values in $(\mathcal{C}([0,\infty)), d)$ equipped with its Borel-σ-algebra $\mathscr{B}(\mathcal{C}([0,\infty)))$ (i.e., continuous stochastic processes) with distributions $\mathbb{P}_X$ and $(\mathbb{P}_{X^n})_{n\in\mathbb{N}}$. As in Proposition §6.2.11 we have that $\mathbb{P}_{X^n} \xrightarrow{w} \mathbb{P}_X$ implies $\mathbb{P}_{X^n} \xrightarrow{fdd} \mathbb{P}_X$. Moreover, the following are equivalent: (i) $\mathbb{P}_{X^n} \xrightarrow{fdd} \mathbb{P}_X$ and $(\mathbb{P}_{X^n})_{n\in\mathbb{N}}$ is tight, (ii) $\mathbb{P}_{X^n} \xrightarrow{w} \mathbb{P}_X$ (compare with Theorem §6.2.12). Introduce further the modulus of continuity on compact sets, that is, for $N > 0$, $w_f^N(\delta) := \sup\{|f(t) - f(s)| : |t-s| \leqslant \delta, t, s \in [0,N]\}$. A family $(\mathbb{P}_t)_{t\in\mathbb{T}}$ of probability measures on $\mathcal{C}([0,\infty))$ is (uniformly) tight if and only if the following two conditions hold. (i) $(\mathbb{P}_t \circ \Pi_0^{-1}, t \in \mathbb{T})$ is tight. (ii) For all $\varepsilon, \eta > 0$ and $N \in \mathbb{N}$ there is $\delta \in (0,1)$ such that $\sup_{t\in\mathbb{T}} \mathbb{P}_t\big(\{f : w_f^N(\delta) \geqslant \varepsilon\}\big) \leqslant \eta$. (compare with Theorem §6.2.17). □

Chapter 7
Brownian motion

7.1 Continuous versions

In section §2.1.3, we defined a process $(X_t)_{t\in[0,\infty)}$ with independent stationary normally distributed increments. We are interested in properties of this process X that cannot be described in terms of finite-dimensional distributions but reflect the whole path $t \to X_t$. We first investigate continuity properties of paths of stochastic processes and show how they ensure measurability of some path functionals. A priori the paths of a canonical process are of course not continuous since every map $[0,\infty) \to \mathbb{R}$ is possible. Hence, it will be important to find out which paths are a.s. negligible. Therefore, recall the notions of modification (version) and indistinguishable as given in definition §2.2.11.

§7.1.1 **Definition.** Let $(\mathcal{S}_1, d_1)$ and $(\mathcal{S}_2, d_2)$ be metric spaces and $\gamma \in (0, 1]$.

(i) A map $f : \mathcal{S}_1 \to \mathcal{S}_2$ is called *Hölder-continuous of order γ* (briefly, *Hölder-γ-continuous*) *at the point $s \in \mathcal{S}_1$* if there exist $\varepsilon > 0$ and $C < \infty$ such that, for any $r \in \mathcal{S}_1$ with $d_1(s, r) < \varepsilon$, we have

$$d_2(f(s), f(r)) \leqslant C d_1^\gamma(s, r). \tag{7.1}$$

(ii) f is called *locally Hölder-γ-continuous* if, for every $t \in \mathcal{S}_1$, there exists $\varepsilon > 0$ and $C := C(t, \varepsilon) > 0$ such that, for all $s, r \in \mathcal{S}_1$ with $d_1(s, t) < \varepsilon$ and $d_1(r, t) < \varepsilon$, the inequality (7.1) holds.

(iii) f is called *Hölder-γ-continuous* if there exist a C such that (7.1) holds for all $s, r \in \mathcal{S}_1$. □

§7.1.2 **Remark.** In the case $\gamma = 1$, Hölder-γ-continuity is Lipschitz continuity. Furthermore, for $\mathcal{S} = \mathbb{R}$ and $\gamma > 1$, every locally Hölder-γ-continuous function is constant. Evidently, a locally Hölder-γ-continuous map is Hölder-γ-continuous at every point. On the other hand, for a function f that is Hölder-γ-continuous at a given point s, there need not exist an open neighbourhood in which f is continuous. In particular, f need not be locally Hölder-γ-continuous. □

§7.1.3 **Lemma.** *Let $\mathbb{T} \subset \mathbb{R}$ and $f : \mathbb{T} \to \mathbb{R}$ be locally Hölder-continuous of order $\gamma \in (0, 1]$. Then the following statements hold.*

(i) *f is locally Hölder-continuous of order γ' for every $\gamma' \in (0, \gamma)$.*

(ii) *If $\mathbb{T}$ is compact, then f is Hölder-γ-continuous.*

(iii) *Let $\mathbb{T}$ be a bounded interval of length $T > 0$. Assume that there exists an $\varepsilon > 0$ and an $C_\varepsilon > 0$ such that, for all $s, r \in \mathbb{T}$ with $|r - s| \leqslant \varepsilon$, we have $|f(s) - f(r)| \leqslant C_\varepsilon |s - r|^\gamma$. Then f is Hölder-γ-continuous with constant $C := C_\varepsilon \lceil T/\varepsilon \rceil^{1-\gamma}$.*

Proof of Lemma §7.1.3 We refer to Klenke [2008], Lemma 21.5. □

§7.1.4 **Definition**. Let $\mathbb{T} \subset \mathbb{R}$ and let $X = (X_t)_{t\in\mathbb{T}}$ be a stochastic process on some probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with values in a metric space $(\mathcal{S}, d)$. Let $\gamma \in (0, 1]$.

(i) For every $\omega \in \Omega$, we say that the map $\mathbb{T} \to \mathcal{S}$, $t \mapsto X_t(\omega)$ is a path of X.

(ii) We say that X has almost surely continuous paths, or briefly that X is a.s. continuous, if for almost all $\omega \in \Omega$, the path $t \mapsto X_t(\omega)$ is continuous. Similarly, we define locally Hölder-γ-continuous paths and so on. □

§7.1.5 **Lemma**. *Let X and Y be modifications of each other. If one of the following conditions*

(a) *$\mathbb{T}$ is countable,*

(b) *$\mathbb{T} \subset \mathbb{R}$ is a (possibly unbounded) interval and X and Y are a.s. right continuous,*

holds, then X and Y are indistinguishable.

Proof of Lemma §7.1.5 is left as an exercise (sheet 2). □

§7.1.6 **Theorem** (Kolmogorov-Chentsov). *Let $X = (X_t)_{t\in[0,\infty)}$ be a real-valued process. Assume for every $T > 0$, there are numbers $\alpha, \beta, C > 0$ such that*

$$\mathbb{E}[|X_t - X_s|^\alpha] \leqslant C|t - s|^{1+\beta} \quad \textit{for all } s, t \in [0, T]. \tag{7.2}$$

Then the following statements hold.

(i) *There is a modification $\tilde{X} = (\tilde{X}_t)_{t\in[0,\infty)}$ of X whose paths are locally Hölder-continuous of every order $\gamma \in (0, \beta/\alpha)$.*

(ii) *Let $\gamma \in (0, \beta/\alpha)$. For every $\varepsilon > 0$ and $T < \infty$, there exists a number $K < \infty$ that depends only on $\varepsilon, T, \alpha, \beta, C, \gamma$ such that*

$$\mathbb{P}\big(|\tilde{X}_t - \tilde{X}_s| \leqslant K|t - s|^\gamma, s, t \in [0, T]\big) \geqslant 1 - \varepsilon. \tag{7.3}$$

Proof of Theorem §7.1.6 is given in the lecture. □

§7.1.7 **Theorem** (Kolmogorov's criterion for weak relative compactness). *Let $(X^n)_{n\in\mathbb{N}}$ be a sequence of continuous stochastic processes. Assume that the following conditions are satisfied.*

(i) *The family $(\mathbb{P}_{X_0^n})_{n\in\mathbb{N}}$ of initial distribution is tight.*

(ii) *There are numbers $\alpha, \beta, C > 0$ such that, for all $s, t \in [0, \infty)$ and every $n \in \mathbb{N}$, we have $\mathbb{E}[|X_s^n - X_t^n|^\alpha] \leqslant C|s - t|^{1+\beta}$.*

Then the family $(\mathbb{P}_{X^n})_{n\in\mathbb{N}}$ of distributions of X^n is a weakly relatively compact family of probability measures on $(C([0, \infty)), \mathscr{B}(C([0, \infty))))$.

Proof of Theorem §7.1.7 is given in the lecture. □

7.2 Construction and path properties

§7.2.1 **Definition**. A real-valued stochastic process $B = (B_t)_{t\in[0,\infty)}$ is called a *Brownian motion* if

(a) $B_0 = 0$,

(b) B has independent, stationary increments (compare §2.1.7 (b) and (c)),

(c) $B_t \sim \mathfrak{N}(0, t)$ for all $T > 0$, and

(d) $t \mapsto B_t$ is a.s. continuous. □

§7.2.2 **Remark.** It can be easily checked that (a) - (c) defines a consistent family of probability measures $\{\mathbb{P}_{\mathcal{J}}, \mathcal{J} \subset \mathcal{I} = [0,\infty) \text{ finite}\}$ on the product space $(\mathbb{R}^{[0,\infty)}, \mathscr{B}^{\otimes[0,\infty)})$. Thus, by Kolmogorov's consistency theorem (Theorem §2.3.11), there exists a probability measure $\mathbb{P}$ (called Wiener measure) on $(\mathbb{R}^{[0,\infty)}, \mathscr{B}^{\otimes[0,\infty)})$ satisfying (a) - (c). It remains to check if there exists a version of $(B_t)_{t\in[0,\infty)}$ that fulfils (d). We will give below a constructive proof of the existence of the Brownian motion. □

§7.2.3 **Definition.** A real-valued stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called a *Gaussian process* if, for every $n \in \mathbb{N}$ and for all $t_1, \dots, t_n \in \mathbb{T}$, we have that $X_{t_1}, \dots, X_{t_n}$ is n-dimensional normally distributed. X is called *centred* if $\mathbb{E}(X_t) = 0$ for every $t \in \mathbb{T}$, The map $\operatorname{cov}(s,t) = \mathbb{C}\mathrm{ov}(X_s, X_t)$ for $s,t \in \mathbb{T}$ is called the *covariance function* of X. □

§7.2.4 **Remark.** The covariance function determines the finite-dimensional distributions of a centred Gaussian process since a multidimensional normal distribution is determined by the vector of expectations and by the covariance matrix. □

§7.2.5 **Proposition.** *Let $X = (X_t)_{t\in[0,\infty)}$ be a stochastic process. The following are equivalent:*

(i) *X is a Brownian motion.*

(ii) *X is a continuous centred Gaussian process with $\operatorname{cov}(t,s) = s \wedge t$ for all $s,t \geqslant 0$.*

Proof of Proposition §7.2.5 is given in the lecture. □

§7.2.6 **Corollary** (Scaling property of Brownian motion). *If B is a Brownian motion and if $K \neq 0$, then $(K^{-1}B_{K^2 t})_{t\in[0,\infty)}$ is also a Brownian motion. In particular, $-B$ is a Brownian motion.* □

§7.2.7 **Proposition.** *There exists a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ and a Brownian motion B on $(\Omega, \mathscr{A}, \mathbb{P})$.*

Proof of Proposition §7.2.7 is given in the lecture. □

§7.2.8 **Example.** Let $B = (B_t)_{t\in[0,1]}$ be a Brownian motion. The stochastic process $X_t := B_t - tB_1$ which is a centred Gaussian process with continuous paths and covariance function $\operatorname{cov}(s,t) = s \wedge t - st$, is called a *Brownian bridge*. Obviously, if $L_t := tB_1$ then $B_t = L_t + X_t$. Moreover, it is easily seen that L and X are uncorrelated and hence independent. As a consequence, given the event $\{B_1 = a\}$ it holds $B_t = at + X_t$, that is, the Brownian motion is obtained by adding the Brownian bridge X to the linear Process $(at)_{t\in[0,1]}$, The distribution $\mathbb{P}_a$ of $(at + X_t)_{t\in[0,1]}$ depends in the weak topology of measures on $\mathscr{B}(\mathcal{C}([0,1]))$ continuously on a and thus, it is a version of the conditional distribution $\mathbb{P}_{B|B_1=a}$ of B given $B_1 = a$. In this sense, the conditional distribution of the Brownian motion $(B_t)_{t\in[0,1]}$ given $B_1 = 0$ is the distribution $\mathbb{P}_0$ of the Brownian bridge $(B_t - tB_1)_{t\in[0,1]}$. □

§7.2.9 **Proposition.** *Let $B = (B_t)_{t\in[0,\infty)}$ be a Brownian motion and let $Z = (Z_t)_{t\in[0,\infty)}$ be given point-wise by $Z_t = tB_{1/t}$ for $t > 0$ and $Z_0 = 0$. Then Z is a Brownian motion.*

Proof of Proposition §7.2.9 is given in the lecture. □

§7.2.10 **Theorem.** *The paths of B are a.s. locally Hölder-γ-continuous for every $\gamma < 1/2$.*

Proof of Theorem §7.2.10 is given in the lecture. □

§7.2.11 **Theorem** (Blumenthal's 0-1 law). *Let B be a Brownian motion and let $\sigma(B) = \mathscr{F}^B = (\mathscr{F}_t^B)_{t\in[0,\infty)}$ be the natural filtration generated by B. Then $\mathscr{F}_{0+}^B := \cap_{t>0}\mathscr{F}_t^B$ is a $\mathbb{P}$-trivial σ-algebra.*

Proof of Theorem §7.2.11 is given in the lecture. □

§7.2.12 **Example**. Let B be a Brownian motion. Define $A_s := \big\{\inf\{t > 0 : B_t \geqslant K\sqrt{t}\} < s\big\}$ and $A := \big\{\inf\{t > 0 : B_t \geqslant K\sqrt{t}\} = 0\big\} = \cap_{s>0}A_s \in \mathscr{F}_{0+}^B$. From §7.2.11 follows $P(A) \in \{0, 1\}$. By the scaling property §7.2.6 of Brownian motion, $\mathbb{P}(A) = \inf_{s>0}\mathbb{P}(A_s) \geqslant \mathbb{P}(B_1 \geqslant K) > 0$ and thus for every $K > 0$

$$\mathbb{P}\big(\inf\{t > 0 : B_t \geqslant K\sqrt{t}\} = 0\big) = 1. \tag{7.4}$$

This shows that, for every $t \geqslant 0$, a.s. B is not Hölder-$1/2$-continuous at t. Note that the order of quantifiers is subtle. It does not show that a.s. B was not Hölder-$1/2$-continuous at any $t \geqslant 0$. However, it can be shown that a.s. B is not locally Hölder-$1/2$-continuous (see Klenke [2008] Remark 22.4 and, e.g., Revuz and Yor [2005] Theorem I.2.5 for a proof). □

§7.2.13 **Theorem** (Paley-Wiener-Zygmund). *For every $\gamma > 1/2$, a.s. the paths of Brownian motion $(B_t)_{t\in[0,\infty)}$ are not Hölder-continuous of order γ at any point. In particular, the paths are almost surely nowhere differentiable.*

Proof of Theorem §7.2.13 is given in the lecture. □

§7.2.14 **Remark**. In addition, it can been shown that a.s. there is not an interval on which Brownian motion is monotone, and the set of it local minima is dense in $[0, 1]$. □

7.3 Markov properties

Denote by $\mathbb{P}_a$ the probability measure such that $B = (B_t)_{[0,\infty)}$ is a Brownian motion started at $a \in \mathbb{R}$. To put it differently, under $\mathbb{P}_a$, the process $B - a = (B_t - a)_{[0,\infty)}$ is a standard Brownian motion satisfying the Definition §7.2.1, i.e., a Brownian motion started at 0. We write $\mathbb{E}_a$ for expectation with respect to $\mathbb{P}_a$, $\mathcal{L}_a(B) = \mathbb{P}_a$, $\mathcal{L}_a(B|\mathscr{A}) = \mathbb{P}_a[B \in \bullet|\mathscr{A}]$ for a regular conditional distribution of B given $\mathscr{A}$ and $\mathbb{E}_a[f(B)|\mathscr{A}]$ for a conditional expectation of $f(B)$ given $\mathscr{A}$. In particular, for $s \geqslant 0$ we understand $\mathbb{P}_{B_s}$ as the distribution of a second Brownian motion started at B_s and use the notation $\mathbb{E}_{B_s}$ for the expectation with respect to $\mathbb{P}_{B_s}$. Given the shift operator ϑ^s as in definition §4.1.3 we consider the shifted Brownian motion $\vartheta^s(B) = (B_{s+t})_{t\in[0,\infty)}$ and observe that the process $\vartheta^s(B) - B_s = (B_{t+s} - B_s)_{t\in[0,\infty)}$ is independent of $\mathscr{F}_s^B$ for all $s \geqslant 0$ where $\sigma(B) = \mathscr{F}^B = (\mathscr{F}_t^B)_{t\in[0,\infty)}$ is the natural filtration generated by B. A Brownian motion has the Markov property if $\mathbb{P}_{B_s}$ is a version of the conditional distribution of $\vartheta^s(B)$ given $\mathscr{F}_s^B$ or equivalently $\mathbb{E}_a[h(\vartheta^s(B))|\mathscr{F}_s^B] = \mathbb{E}_{B_s}[h(B)]$ for every bounded measurable function $h : \mathcal{C}([0, \infty)) \to \mathbb{R}$.

§7.3.1 **Theorem** (Markov property). *Brownian motion B with distributions $(\mathbb{P}_a)_{a\in\mathbb{R}}$ has the Markov property.*

Proof of Theorem §7.3.1 is given in the lecture. □

We show below the strong Markov property of a Brownian motion. To this end, given an arbitrary filtration $\mathscr{F} = (\mathscr{F}_t)_{t\in[0,\infty)}$ we define a new filtration $\mathscr{F}^+ = (\mathscr{F}_t^+)_{t\in[0,\infty)}$ by

$\mathscr{F}_t^+ = \mathscr{F}_{t+} := \cap_{s>t}\mathscr{F}_s$. We say that $\mathscr{F}$ is right-continuous if $\mathscr{F}^+ = \mathscr{F}$. In particular, $\mathscr{F}^+$ is right-continuous for any filtration $\mathscr{F}$. In general the natural filtration of a process is not right-continuous, however, Blumenthal's 0-1 law given in §7.2.11 shows that $\mathscr{F}^{B+}$ and $\mathscr{F}^B$ for a Brownian motion B differ only by sets of measure zero. Keeping in mind Definition §2.4.6 a r.v. τ with values in $[0,\infty]$ is a $\mathscr{F}$-stopping time, if for any $t \geqslant 0$, $\{\tau \leqslant t\} \in \mathscr{F}_t$. We say that τ is a weak $\mathscr{F}$-stopping time, if for any $t > 0$, $\{\tau < t\} \in \mathscr{F}_t$. In that case $\tau + s$ is a $\mathscr{F}$-stopping time for any $s > 0$ noting that $\{\tau \leqslant t - s\} \in \mathscr{F}_{t-s+} \subset \mathscr{F}_t$ for any $t \geqslant 0$. Considering as in §2.4.12 for a $\mathscr{F}$-stopping time τ the σ-algebra of τ-past $\mathscr{F}_\tau := \{A \in \mathscr{A} : A \cap \{\tau \leqslant t\} \in \mathscr{F}_t \text{ for any } t \geqslant 0\}$, we define for a (weak) $\mathscr{F}$-stopping time τ the σ-algebra $\mathscr{F}_{\tau+} := \cap\{\mathscr{F}_\sigma : \sigma \text{ is } \mathscr{F}\text{-stopping time with } \sigma > \tau\}$. Due to Lemma §2.4.16 for a $\mathscr{F}$-stopping time τ we have $\mathscr{F}_\tau \subset \mathscr{F}_{\tau+}$.

§7.3.2 **Lemma**. *A r.v. τ is a weak $\mathscr{F}$-stopping time if and only if it is a $\mathscr{F}^+$-stopping time, in which case, $\mathscr{F}_{\tau+} = \mathscr{F}_\tau^+ = \{A \in \mathscr{A} : A \cap \{\tau < t\} \in \mathscr{F}_t \text{ for any } t > 0\}$.*

Proof of Lemma §7.3.2 is given in the lecture. □

The last result shows that the notions of $\mathscr{F}$-stopping time and weak $\mathscr{F}$-stopping time agree when $\mathscr{F}$ is right-continuous.

§7.3.3 **Lemma**. *For any r.v.'s $\tau_1, \tau_2, \dots$ with values in $[0,\infty]$ and filtration $\mathscr{F}$, we have*

(a) *If the τ_n are $\mathscr{F}$-stopping times, then so is $\tau = \sup_{n\in\mathbb{N}} \tau_n$.*

(b) *If the τ_n are weak $\mathscr{F}$-stopping times, then so is $\tau = \inf_{n\in\mathbb{N}} \tau_n$, and we have $\mathscr{F}_\tau^+ = \cap_{n\in\mathbb{N}}\mathscr{F}_{\tau_n}^+$.*

Moreover, for any weak $\mathscr{F}$-stopping time τ the r.v.'s $\tau_n := 2^{-n}\lfloor 2^n\tau + 1\rfloor$ taking values in $\{k2^{-n}, k \in \mathbb{N}\} \cup \{\infty\}$ for $n \in \mathbb{N}$ are $\mathscr{F}$-stopping times satisfying $\tau_n \downarrow \tau$.

Proof of Lemma §7.3.3 is given in the lecture. □

A Brownian motion has the strong Markov property if for every (a.s.) finite $\mathscr{F}$-stopping time τ and for every bounded measurable function $h : \mathcal{C}([0,\infty)) \to \mathbb{R}$ it holds $\mathbb{E}_a[h(\vartheta^\tau(B))|\mathscr{F}_\tau^B] = \mathbb{E}_{B_\tau}[h(B)]$. We will show that $\mathbb{E}_a[h(\vartheta^\tau(B))|\mathscr{F}_{\tau+}^B] = \mathbb{E}_{B_\tau}[h(B)]$ holds for any finite weak $\mathscr{F}$-stopping time τ which implies the claim by exploiting that $\mathscr{F}_\tau^B \subset \mathscr{F}_{\tau+}^B$ and that $\mathbb{E}_{B_\tau}[h(B)]$ is $\mathscr{F}_\tau^B$-measurable.

§7.3.4 **Theorem** (Strong Markov property). *Brownian motion B with distributions $(\mathbb{P}_a)_{a\in\mathbb{R}}$ has the strong Markov property.*

Proof of Theorem §7.3.4 is given in the lecture. □

§7.3.5 **Corollary**. *If τ is a (a.s.) finite weak $\mathscr{F}^B$-stopping time, then $\vartheta^\tau(B) - B_\tau$ is a Brownian motion that is independent of $\mathscr{F}_{\tau+}^B$.*

Proof of Corollary §7.3.5 is given in the lecture. □

§7.3.6 **Theorem** (Reflection principle). *For any $\mathscr{F}^B$-stopping time τ, a Brownian motion B has the same distribution as the reflected process $\widetilde{B} = (\widetilde{B}_t)_{t\in[0,\infty)}$ with $\widetilde{B}_t = B_{\tau\wedge t} - (B_t - B_{\tau\wedge t})$.*

Proof of Theorem §7.3.6 is given in the lecture. □

§7.3.7 **Corollary**. *For every $a > 0$ and $T > 0$ we have $\mathbb{P}_0(\sup_{t\in[0,T]} B_t > a) = 2\mathbb{P}_0(B_T > a)$.*

Proof of Theorem §7.3.6 is given in the lecture. □

7.4 Donsker's theorem

Let $Y_1, Y_2, \ldots$ be i.i.d. r.v.'s with $\mathbb{E}(Y_1) = 0$ and $\mathbb{V}\mathrm{ar}(Y_1) = \sigma^2 > 0$. For $t > 0$, let $S_t^n := \sum_{i=1}^{\lfloor nt \rfloor} Y_i$ with $\lfloor nt \rfloor = \max\{k \in \mathbb{N} : k \leqslant nt\}$ and $\widetilde{S}_t^n := \frac{1}{\sigma\sqrt{n}} S_t^n$. By the central limit theorem, we have $\widetilde{S}_t^n \xrightarrow{d} \mathfrak{N}(0,t)$ as $n \to \infty$. Let $B = (B_t)_{t\in[0,1]}$ be a Brownian motion and denote by $\mathbb{P}_W$ the Wiener measure, that is, the probability measure on $\mathcal{C}([0,1])$ with respect to which the canonical process is a Brownian motion. Then $\widetilde{S}_t^n \xrightarrow{d} \mathcal{L}(B_t)$ as $n \to \infty$ for any $t > 0$. By the multidimensional central limit theorem for $k \in \mathbb{N}$ and $t_1, \ldots, t_k \in [0,1]$ we also have $(\widetilde{S}_{t_1}^n, \ldots, \widetilde{S}_{t_k}^n) \xrightarrow{d} \mathcal{L}(B_{t_1}, \ldots, B_{t_k})$ as $n \to \infty$. We now define $\overline{S}^n = (\overline{S}_t^n)_{t\in[0,1]}$ as $\widetilde{S}^n$ but linearly interpolated

$$\overline{S}_t^n = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} Y_i + \frac{(tn - \lfloor tn \rfloor)}{\sigma\sqrt{n}} Y_{\lfloor nt \rfloor + 1}. \tag{7.5}$$

Then for $\varepsilon > 0$,

$$\mathbb{P}(|\widetilde{S}_t^n - \overline{S}_t^n| > \varepsilon) \leqslant \varepsilon^2 \mathbb{E}(|\widetilde{S}_t^n - \overline{S}_t^n|^2) \leqslant \frac{1}{\varepsilon^2 n} \frac{1}{\sigma^2} \mathbb{E}(Y_1^2) = \frac{1}{\varepsilon^2 n} \xrightarrow{n\to\infty} 0.$$

By Slutzky's theorem we thus have convergence of the finite-dimensional distributions to the Wiener measure $\mathbb{P}_W$, that is, $\mathbb{P}_{\overline{S}^n} \xrightarrow{fdd} \mathbb{P}_W$ as $n \to \infty$. The aim of this section is to strengthen this convergence statement to weak convergence of probability measures on $\mathcal{C}([0,1])$. The main theorem of this section is the functional central limit theorem, which goes back to Donsker [1951]. Theorems of this type are also called invariance principles since the limiting distribution is the same for all distributions Y_i with expectation 0 and the same variance.

§7.4.1 **Theorem** (Donsker's invariance principle). *In the sense of weak convergence on $\mathcal{C}([0,1])$, the distributions of $\overline{S}^n$ converge to the Wiener measure, $\mathbb{P}_{\overline{S}^n} \xrightarrow{w} \mathbb{P}_W$ as $n \to \infty$.*

Proof of Theorem §7.4.1 is given in the lecture. □

Bibliography

P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 2nd ed. edition, 1999.

M. D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 6, 1951.

R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, revised reprint of the 1989 original. edition, 2002.

R. Durrett. *Probability: Theory and examples.* Duxbury Press, Belmont, CA, 2nd edition edition, 1996.

O. Kallenberg. *Foundations of Modern Probability*. Springer, Berlin, Heidelberg, 2002.

A. Klenke. *Probability theory. A comprehensive course.* London: Springer, 2008.

J. Neveu. *Discrete-parameter martingales.*, volume 10 of *North-Holland mathematical library*. Elsevier, 1975.

D. Revuz and M. Yor. *Continuous martingales and Brownian motion.*, volume 293 of *Grundlehren der Mathematischen Wissenschaften*. Berlin: Springer, 3rd edition edition, 2005.