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: April 21, 2016

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

MATHEMATIKON, 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

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 $\mu(\biguplus_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$ 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. □

§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 $\sigma_\Omega(X^{-1}(\mathcal{E})) = X^{-1}(\sigma_{\mathcal{S}}(\mathcal{E}))$. □

§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 $A \in \mathscr{A}$. □

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).

§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 $X = \sum_{i\in\mathcal{I}} x_i \mathbb{1}_{A_i}$. □

§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 $(-X_n) \uparrow (-X)$. □

§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 $A_1, A_2, \ldots$ 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 $A_1, A_2, \ldots$ 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 $\sigma(X_i) = X^{-1}(\mathscr{S}_i)$.

§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 $X_1, X_2, \ldots$ 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 $X, X_1, X_2, \ldots \in \mathcal{M}^+$ 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 $X_1, X_2, \ldots \in \mathcal{M}^+$, then $\mathbb{E}(\liminf_{n\to\infty} X_n) \leqslant \liminf_{n\to\infty} \mathbb{E}(X_n)$.

(c) (Dominated convergence) Assume $X, X_1, X_2, \ldots \in \mathcal{M}$ 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 $C := \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega)$ exists and is finite$\}$. The sequence $(X_n)_{n\geqslant 1}$ *converges almost surely (a.s.)*, if $\mathbb{P}(C) = 1$. We write $X_n \xrightarrow{n\to\infty} X$ $a.s.$, or briefly, $X_n \xrightarrow{a.s.} X$.

(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 $X_n \xrightarrow{n\to\infty} X$ $in\ \mathbb{P}$, or briefly, $X_n \xrightarrow{\mathbb{P}} X$.

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

(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 $X_n \xrightarrow{n\to\infty} X$ $in\ L_p$, or briefly, $X_n \xrightarrow{L_p} X$. □

§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 $X = \limsup_{n\to\infty} X_n$ noting that $X(\omega) = \lim_{n\to\infty} X_n(\omega)$ for $\omega \in C$. □

§1.5.3 **Properties.** *(a) We have $X_n \xrightarrow{a.s.} X$ if and only if $\sup_{m>n}|X_m - X_n| \xrightarrow{n\to\infty} 0$ in $\mathbb{P}$ if and only if $\sup_{j\geqslant n}|X_j - X| \xrightarrow{n\to\infty} 0$ 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 $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{\mathbb{P}} X$.

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

(d) $X_n \xrightarrow{\mathbb{P}} X$ 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 $X_n \xrightarrow{\mathbb{P}} X$, then $g(X_n) \xrightarrow{\mathbb{P}} g(X)$ for any continuous function g.

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

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) *(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 $\|X_n - X\|_p \stackrel{n\to\infty}{\longrightarrow} 0$ we have $\|\mathbb{E}(X_n|\mathscr{F}) - \mathbb{E}(X|\mathscr{F})\|_p \stackrel{n\to\infty}{\longrightarrow} 0$.* □

§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 $\sup_{i\in\mathcal{I}} |X_i| \in L_1(\Omega, \mathscr{A}, \mathbb{P})$. □

§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.* □

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

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

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

(c) *(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 $\mathbb{R}^+$;

(d) *(stationary increments)* $N_t - N_s \overset{d}{=} N_{t-s}$ 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 (c) 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 course. □

§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 $U_{(k+1)} = \min\{U_i\}_{i=1}^n \backslash \{U_{(i)}\}_{i=1}^k$, $k = 2, \dots, n$. 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 (c) in §2.1.3 just uses exponentially distributed inter-arrival times T_k and part (d) 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 course. □

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 course. □

§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 $A \in \tau$.

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 course. □

§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 course. □

§2.3.10 **Theorem**. *Every probability measure on a Polish space is regular.*

Proof of Theorem §2.3.10 is given in the lecture course. □

§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 course. □

§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 course. □

§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 course. □

§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. □

§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 $\mathscr{F} = (\mathscr{F}_t)_{t\in\mathbb{N}_0}$ if X_0 is constant and if, for every $n \in \mathbb{N}$, X_n is $\mathscr{F}_{n-1}$-measurable. □

§2.4.5 **Definition**. A r.v. τ with values in $\mathbb{T} \cup \{\infty\}$ is called a *stopping time* (with respect to the filtration $\mathscr{F}$) if for any $t \in \mathbb{T}$, $\{\tau \leqslant t\} \in \mathscr{F}_t$. □

§2.4.6 **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 §2.4.6 is left as an exercise. □

§2.4.7 **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 *waiting 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 $\tau_\mathcal{S} \equiv 0$. □

§2.4.8 **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 §2.4.8 is left as an exercise. □

§2.4.9 **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. □

§2.4.10 **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 waiting times as in §2.4.7 (b), then (i) $\tau_{S_1\cup S_2} = \tau_{S_1} \wedge \tau_{S_2}$, (ii) $\tau_{S_1\cap S_2} = \tau_{S_1} \vee \tau_{S_2}$ and (iii) if $S_1 \subset S_2$, then $\tau_{S_1} \geqslant \tau_{S_2}$. □

§2.4.11 **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*. □

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

§2.4.13 **Lemma**. *If τ and σ are stopping times then* (i) $\mathscr{F}_{\tau\wedge\sigma} = \mathscr{F}_\tau \cap \mathscr{F}_\sigma$, (ii) $\mathscr{F}_{\tau\vee\sigma} = \mathscr{F}_\tau \vee \mathscr{F}_\sigma$ *and* (iii) *if $\tau \leqslant \sigma$, then $\mathscr{F}_\tau \subset \mathscr{F}_\sigma$.*

Proof of Lemma §2.4.13 is given in the lecture course. □

§2.4.14 **Definition**. If $\tau < \infty$ is a stopping time, then we define $X_\tau(\omega) := X_{\tau(\omega)}(\omega)$. □

§2.4.15 **Lemma**. *Let $\mathbb{T}$ be countable, let X be adapted and let $\tau < \infty$ 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 §2.4.15 is given in the lecture course. □

§2.4.16 **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. □

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

Proof of Corollary §2.4.17 is given in the lecture course. □

§2.4.18 **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* $X^\tau = (X_t^\tau)_{t\in\mathbb{T}}$ by $X_t^\tau = X_{\tau\wedge t}$ for any $t \in \mathbb{T}$ which is adapted to both filtration $\mathscr{F}^\tau = (\mathscr{F}_t^\tau)_{t\in\mathbb{T}} = (\mathscr{F}_{\tau\wedge t})_{t\in\mathbb{T}}$ and $\mathscr{F} = (\mathscr{F}_t)_{t\in\mathbb{T}}$. □

Bibliography

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