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 19, 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 $\mathscr{A} \subset 2^\Omega$ (set of all subsets of $\Omega$) be a class of subsets of $\Omega$. Later, $\Omega$ 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 $\Omega$ and a $\sigma$-algebra $\mathscr{A}$ is called a *measurable space*. The sets $A \in \mathscr{A}$ are called *measurable sets*. If $\Omega$ 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 $\mu$ 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 $\mu$ 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 $A_1, \dots, A_n \in \mathscr{A}$ with $\cup_{i=1}^n A_i \in \mathscr{A}$.

$\mathscr{A}$ 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 $\mu$ is called a

(d) *content*, if $\mu$ is additive,

(e) *premeasure*, if $\mu$ is $\sigma$-additive,

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

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

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

(h) *$\sigma$-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 $\Omega$, the classes $\mathscr{A} = \{\emptyset, \Omega\}$ and $\mathscr{A} = 2^{\Omega}$ are the trivial examples of $\sigma$-algebras.

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

(c) Let $(\Omega, \tau)$ be a topological space with class of open sets $\tau \subset 2^{\Omega}$. The $\sigma$-algebra $\mathscr{B}(\Omega)$ that is generated by the open sets is called the *Borel-$\sigma$-algebra* on $\Omega$. 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-$\sigma$-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 $\delta_\omega$ is a probability measure on any $\sigma$-algebra $\mathscr{A} \subset 2^{\Omega}$. $\delta_\omega$ is called the *Dirac measure* on the point $\omega$.

(e) Let $\Omega$ 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 $\sigma$-finite measure. If $p_\omega = 1$ for every $\omega \in \Omega$, then $\mu$ is called *counting measure* on $\Omega$. If $\Omega$ is finite, then so is $\mu$. □

§1.1.4 **Theorem** (Carathéodory). *Let $\mathscr{A} \subset 2^{\Omega}$ be an algebra and let $\mu$ be a $\sigma$-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 $\sigma$-finite.*

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

§1.1.5 **Remark.** If $\mu$ is a finite content on an algebra $\mathscr{A}$, then *$\sigma$-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 $\sigma$-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 $\Omega$ 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 $\sigma$-algebra on $\Omega$ with respect to which $X$ is measurable. We say that $\sigma(X) := X^{-1}(\mathscr{S})$ is the $\sigma$-algebra on $\Omega$ 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 $\sigma(X_i, i \in \mathcal{I}) := \sigma\big(\cup_{i\in\mathcal{I}} \sigma(X_i)\big) = \sigma\big(\cup_{i\in\mathcal{I}} X_i^{-1}(\mathscr{S}_i)\big)$ is called the $\sigma$-algebra on $\Omega$ that is generated by $(X_i, i \in \mathcal{I})$. This is the the smallest $\sigma$-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 $\sigma$-algebra on $\Omega$ and $\{S \in \mathcal{S} : X^{-1}(S) \in \mathscr{A}\}$ is a $\sigma$-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 $\Omega$. 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 $A_1, A_2, \dots$ 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 $\sigma$-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 $\sigma$-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 $\sigma$-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 $\sigma$-algebra $\bigcap_{n\geqslant 1}\sigma(X_i, i \geqslant n)$ is called the *tail $\sigma$-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 $\mathbb{E}(\liminf_{n\to\infty} X_n) \leqslant \liminf_{n\to\infty} \mathbb{E}(X_n)$.*

*(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 $C := \{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega)$ 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-$\sigma$-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 $\sigma$-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 $\sigma(X)$-$\mathscr{B}$-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 $\sigma$-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-$\sigma$-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-$\sigma$-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 **Corollary**. *Let $p \in [1, \infty]$ and $\mathscr{F} \subset \mathscr{A}$ be a sub-$\sigma$-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 $\sup_{i\in\mathcal{I}} |X_i| \in L_1(\Omega, \mathscr{A}, \mathbb{P})$. □

§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-$\sigma$-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-$\sigma$-algebra of $\mathscr{A}\}$ of r.v.'s in $L_1(\Omega, \mathscr{A}, \mathbb{P})$ is uniformly integrable.* □

# Bibliography

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