Probability Theory

Thomas Müller-Gronbach, Klaus Ritter

Darmstadt, WS 2008/09

Literature

In particular,

H. Bauer, *Probability Theory*, de Gruyter, Berlin, 1996.

P. Billingsley, *Probability and Measure*, Wiley, New York, first edition 1979, third edition 1995.

J. Elstrodt, *Maß- und Integrationstheorie*, Springer, Berlin, first edition 1996, fifth edition 2007.

P. Gänssler, W. Stute, *Wahrscheinlichkeitstheorie*, Springer, Berlin, 1977.

A. Klenke, *Wahrscheinlichkeitstheorie*, Springer, Berlin, first edition 2006, second edition 2008.

Contents

Chapter I

Introduction

A *stochastic model*: a probability space $(\Omega, \mathfrak{A}, P)$ together with a collection of random variables (measurable mappings) $\Omega \to \mathbb{R}$, say.

Examples of probability spaces, known from 'Introduction to Stochastics' or 'Analysis':

(i) Given: a countable set Ω and $f : \Omega \to \mathbb{R}_+$ such that $\sum_{\omega\in\Omega} f(\omega) = 1$.
Take the power set $\mathfrak{A} = \mathfrak{P}(\Omega)$ and define

$$P(A) = \sum_{\omega \in A} f(\omega), \qquad A \subset \Omega.$$

(ii) Given: $f : \mathbb{R}^k \to \mathbb{R}_+$ such that $\int_{\mathbb{R}^k} f(\omega)\, d\omega = 1$.
Let $\Omega = \mathbb{R}^k$, take the σ-algebra $\mathfrak{A} = \mathfrak{B}(\mathbb{R}^k)$ of Borel sets in $\mathbb{R}^k$ and define

$$P(A) = \int_A f(\omega)\, d\omega, \qquad A \in \mathfrak{B}(\mathbb{R}^k).$$

Main topics in this course:

(i) construction of probability spaces, including the theory of measure and integration,

(ii) limit theorems,

(iii) conditional probabilities and expectations,

(iv) discrete-time martingales.

Example 1. Limit theorems like the law of large numbers or the central limit theorem deal with sequences $X_1, X_2, \ldots$ of random variables and their partial sums

$$S_n = \sum_{i=1}^{n} X_i$$

(gambling: cumulative gain after n trials; physics: position of a particle after n collisions).

Under which conditions and in which sense does S_n/n or $S_n/\sqrt{n}$ converge, as n tends to infinity?

Example 2. Limit theorems hold in particular for independent and identically distributed (i.i.d.) random variables $X_1, X_2, \ldots$ with $\mathrm{E}(X_i) = 0$ and $\mathrm{Var}(X_i) = 1$. Then S_n/n 'converges' to zero and $S_n/\sqrt{n}$ 'converges' to the standard normal distribution.

In particular, in a simple case of gambling: X_i takes values ± 1 with probability $1/2$. Existence of such a model? Existence for every choice of the distribution of X_i?

Example 3. The fluctuation of a stock price defines a function on the time interval $\mathbb{R}_+$ with values in $\mathbb{R}$ (for simplicity, we admit negative stock prices at this point). What is a reasonable σ-algebra on the space Ω of all mappings $\mathbb{R}_+ \to \mathbb{R}$ or on the subspace of all continuous mappings? How can we define (non-discrete) probability measures on these spaces in order to model the random dynamics of stock prices? Analogously for random perturbations in physics, biology, etc.

More generally, the same questions arise for mappings $I \to S$ with an arbitrary non-empty set I and $S \subset \mathbb{R}^d$ (physics: phase transition in ferromagnetic materials, the orientation of magnetic dipoles on a set I of sites; medicine: spread of diseases, certain biometric parameters for a set I of individuals; environmental science: the concentration of certain pollutants in a region I).

Example 4. Consider two random variables X_1 and X_2. If $P(\{X_2 = v\}) > 0$ then the conditional probability of $\{X_1 \in A\}$ given $\{X_2 = v\}$ is defined by

$$P(\{X_1 \in A\} \,|\, \{X_2 = v\}) = \frac{P(\{X_1 \in A\} \cap \{X_2 = v\})}{P(\{X_2 = v\})}.$$

How can we reasonably extend this definition to the case $P(\{X_2 = v\}) = 0$, e.g., for X_2 being normally distributed? How does the observation $X_2 = v$ change our stochastic model? Cf. Example 3.

Chapter II

Measure and Integral

1 Classes of Sets

Given: a non-empty set Ω and a class $\mathfrak{A} \subset \mathfrak{P}(\Omega)$ of subsets. Put

$$\mathfrak{A}^+ = \Big\{ \bigcup_{i=1}^{n} A_i : n \in \mathbb{N} \wedge A_1, \dots, A_n \in \mathfrak{A} \text{ pairwise disjoint} \Big\}.$$

Definition 1.

(i) $\mathfrak{A}$ *closed w.r.t. intersections* if $A, B \in \mathfrak{A} \Rightarrow A \cap B \in \mathfrak{A}$.

(ii) $\mathfrak{A}$ *closed w.r.t. unions* if $A, B \in \mathfrak{A} \Rightarrow A \cup B \in \mathfrak{A}$.

(iii) $\mathfrak{A}$ *semi-algebra (in* Ω*)* if

  (a) $\Omega \in \mathfrak{A}$,

  (b) $\mathfrak{A}$ closed w.r.t. intersections,

  (c) $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}^+$.

(iv) $\mathfrak{A}$ *algebra (in* Ω*)* if

  (a) $\Omega \in \mathfrak{A}$,

  (b) $\mathfrak{A}$ closed w.r.t. intersections,

  (c) $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$.

(v) $\mathfrak{A}$ σ*-algebra (in* Ω*)* if

  (a) $\Omega \in \mathfrak{A}$,

  (b) $A_1, A_2, \ldots \in \mathfrak{A} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$,

  (c) $A \in \mathfrak{A} \Rightarrow A^c \in \mathfrak{A}$.

Remark 1. Let $\mathfrak{A}$ denote a σ-algebra in Ω. Recall that a *probability measure* P on $(\Omega, \mathfrak{A})$ is a mapping

$$P : \mathfrak{A} \to [0, 1]$$

such that $P(\Omega) = 1$ and

$$A_1, A_2, \ldots \in \mathfrak{A} \text{ pairwise disjoint} \quad \Rightarrow \quad P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i).$$

Moreover, $(\Omega, \mathfrak{A}, P)$ is called a *probability space*, and $P(A)$ is the *probability* of the *event* $A \in \mathfrak{A}$.

Remark 2.

(i) $\mathfrak{A}$ σ-algebra $\Rightarrow$ $\mathfrak{A}$ algebra $\Rightarrow$ $\mathfrak{A}$ semi-algebra.

(ii) $\mathfrak{A}$ closed w.r.t. intersections $\Rightarrow$ $\mathfrak{A}^+$ closed w.r.t. intersections.

(iii) $\mathfrak{A}$ algebra and $A_1, A_2 \in \mathfrak{A} \Rightarrow A_1 \cup A_2,\ A_1 \setminus A_2,\ A_1 \triangle A_2 \in \mathfrak{A}$.

(iv) $\mathfrak{A}$ σ-algebra and $A_1, A_2, \ldots \in \mathfrak{A} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathfrak{A}$.

Example 1.

(i) Let $\Omega = \mathbb{R}$ and consider the class of intervals

$$\mathfrak{A} = \{]a, b] : a, b \in \mathbb{R} \wedge a < b\} \cup \{]-\infty, b] : b \in \mathbb{R}\} \cup \{]a, \infty[: a \in \mathbb{R}\} \cup \{\mathbb{R}, \emptyset\}.$$

Then $\mathfrak{A}$ is a semi-algebra, but not an algebra.

(ii) $\{A \in \mathfrak{P}(\Omega) : A \text{ finite or } A^c \text{ finite}\}$ is an algebra, but not a σ-algebra in general.

(iii) $\{A \in \mathfrak{P}(\Omega) : A \text{ countable or } A^c \text{ countable}\}$ is a σ-algebra.

(iv) $\mathfrak{P}(\Omega)$ is the largest σ-algebra in Ω, $\{\emptyset, \Omega\}$ is the smallest σ-algebra in Ω.

Definition 2. $\mathfrak{A}$ *Dynkin class (in Ω)* if

(i) $\Omega \in \mathfrak{A}$,

(ii) $A_1, A_2 \in \mathfrak{A} \wedge A_1 \subset A_2 \Rightarrow A_2 \setminus A_1 \in \mathfrak{A}$,

(iii) $A_1, A_2, \ldots \in \mathfrak{A}$ pairwise disjoint $\Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$.

Remark 3. $\mathfrak{A}$ σ-algebra $\Rightarrow$ $\mathfrak{A}$ Dynkin class.

Theorem 1. For every Dynkin class $\mathfrak{A}$

$$\mathfrak{A}\ \sigma\text{-algebra} \quad \Leftrightarrow \quad \mathfrak{A} \text{ closed w.r.t. intersections.}$$

Proof. '$\Leftarrow$': For $A \in \mathfrak{A}$ we have $A^c = \Omega \setminus A \in \mathfrak{A}$ since $\mathfrak{A}$ is a Dynkin class. For $A, B \in \mathfrak{A}$ we have

$$A \cup B = A \cup (B \setminus (A \cap B)) \in \mathfrak{A}$$

since $\mathfrak{A}$ is also closed w.r.t. intersections. Thus, for $A_1, A_2, \ldots \in \mathfrak{A}$ and $B_m = \bigcup_{n=1}^m A_n$ we get $B_m \in \mathfrak{A}$ and

$$\bigcup_{n=1}^\infty A_n = \bigcup_{m=1}^\infty (B_m \setminus B_{m-1}) \in \mathfrak{A},$$

where $B_0 = \emptyset$. □

Remark 4. Consider σ-algebras (algebras, Dynkin classes) $\mathfrak{A}_i$ for $i \in I \neq \emptyset$. Then $\bigcap_{i \in I} \mathfrak{A}_i$ is a σ-algebra (algebra, Dynkin class), too. See also Übung 1.2.

Given: a class $\mathfrak{E} \subset \mathfrak{P}(\Omega)$.

Definition 3. The *σ-algebra generated by* $\mathfrak{E}$

$$\sigma(\mathfrak{E}) = \bigcap \{\mathfrak{A} : \mathfrak{A}\ \sigma\text{-algebra in } \Omega \wedge \mathfrak{E} \subset \mathfrak{A}\}.$$

Analogously, $\alpha(\mathfrak{E})$, $\delta(\mathfrak{E})$ the *algebra, Dynkin class,* respectively, *generated by* $\mathfrak{E}$.

Remark 5. For $\gamma \in \{\sigma, \alpha, \delta\}$ and $\mathfrak{E}, \mathfrak{E}_1, \mathfrak{E}_2 \subset \mathfrak{P}(\Omega)$

(i) $\gamma(\mathfrak{E})$ is the smallest 'γ-class' that contains $\mathfrak{E}$,

(ii) $\mathfrak{E}_1 \subset \mathfrak{E}_2 \Rightarrow \gamma(\mathfrak{E}_1) \subset \gamma(\mathfrak{E}_2)$,

(iii) $\gamma(\gamma(\mathfrak{E})) = \gamma(\mathfrak{E})$.

Example 2. Let $\Omega = \mathbb{N}$ and $\mathfrak{E} = \{\{n\} : n \in \mathbb{N}\}$. Then

$$\alpha(\mathfrak{E}) = \{A \in \mathfrak{P}(\Omega) : A \text{ finite or } A^c \text{ finite}\} =: \mathfrak{A}.$$

Proof: $\mathfrak{A}$ is an algebra, see Example 1, and $\mathfrak{E} \subset \mathfrak{A}$. Thus $\alpha(\mathfrak{E}) \subset \mathfrak{A}$. On the other hand, for every finite set $A \subset \Omega$ we have $A = \bigcup_{n \in A}\{n\} \in \alpha(\mathfrak{E})$, and for every set $A \subset \Omega$ with finite complement we have $A = (A^c)^c \in \alpha(\mathfrak{E})$. Thus $\mathfrak{A} \subset \alpha(\mathfrak{E})$.
Moreover,

$$\sigma(\mathfrak{E}) = \mathfrak{P}(\mathbb{N}), \qquad \delta(\mathfrak{E}) = \mathfrak{P}(\mathbb{N}).$$

Theorem 2. $\mathfrak{E}$ closed w.r.t. intersections $\Rightarrow \sigma(\mathfrak{E}) = \delta(\mathfrak{E})$.

Proof. Remark 3 implies

$$\delta(\mathfrak{E}) \subset \sigma(\mathfrak{E}).$$

We claim that

$$\delta(\mathfrak{E}) \text{ is closed w.r.t. intersections.} \tag{1}$$

Then, by Theorem 1,

$$\sigma(\mathfrak{E}) \subset \delta(\mathfrak{E}).$$

Put

$$\mathfrak{C}_B = \{C \subset \Omega : C \cap B \in \delta(\mathfrak{E})\}, \qquad B \in \delta(\mathfrak{E}),$$

so that (1) is equivalent to

$$\forall\, B \in \delta(\mathfrak{E}) : \delta(\mathfrak{E}) \subset \mathfrak{C}_B. \tag{2}$$

It is straightforward to verify that

$$\forall\, B \in \delta(\mathfrak{E}) : \mathfrak{C}_B \text{ Dynkin class.} \tag{3}$$

Moreover, since $\mathfrak{E}$ is closed w.r.t. intersections,

$$\forall\, E \in \mathfrak{E} : \mathfrak{E} \subset \mathfrak{C}_E.$$

Therefore

$$\forall\, E \in \mathfrak{E} : \delta(\mathfrak{E}) \subset \mathfrak{C}_E,$$

which is equivalent to

$$\forall\, B \in \delta(\mathfrak{E}) : \mathfrak{E} \subset \mathfrak{C}_B.$$

Use (3) to obtain (2). □

An algebra $\alpha(\mathfrak{E})$ can be described explicitly, see Gänssler, Stute (1977, p. 14). The corresponding problem for σ-algebras is addressed in Billingsley (1979, p. 24). Here we only state the following fact.

Lemma 1. $\mathfrak{E}$ semi-algebra $\Rightarrow \alpha(\mathfrak{E}) = \mathfrak{E}^+$.

Proof. Clearly $\mathfrak{E} \subset \mathfrak{E}^+ \subset \alpha(\mathfrak{E})$. It remains to show that $\mathfrak{E}^+$ is an algebra. See Gänssler, Stute (1977, p. 14) for details. □

Sometimes it will be convenient to extend the reals as follows. Put

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\},$$

and define for every $a \in \mathbb{R}$

$$(\pm\infty) + (\pm\infty) = a + (\pm\infty) = (\pm\infty) + a = \pm\infty, \qquad a/_{\pm\infty} = 0,$$

$$a \cdot (\pm\infty) = (\pm\infty) \cdot a = \begin{cases} \pm\infty & \text{if } a > 0 \\ 0 & \text{if } a = 0 \\ \mp\infty & \text{if } a < 0 \end{cases}$$

as well as $-\infty < a < \infty$. For instance, the class $\mathfrak{A}$ from Example 1.(i) consists of the sets

$$\{x \in \mathbb{R} : a < x \leq b\}, \qquad a, b \in \overline{\mathbb{R}}.$$

Furthermore, $\lim_{n\to\infty} x_n = \pm\infty$ for a sequence $(x_n)_{n\in\mathbb{N}}$ in $\overline{\mathbb{R}}$ if for all $M \in\,]0, \infty[$ there is an integer n_0 such that $x_n \gtrless \pm M$ for all $n \geq n_0$.

Recall that $(\Omega, \mathfrak{G})$ is a *topological space* if $\mathfrak{G} \subset \mathfrak{P}(\Omega)$ satisfies

(i) $\emptyset, \Omega \in \mathfrak{G}$,

(ii) $\mathfrak{G}$ is closed w.r.t. to intersections,

(iii) for every family $(G_i)_{i \in I}$ with $G_i \in \mathfrak{G}$ we have $\bigcup_{i \in I} G_i \in \mathfrak{G}$.

The elements $G \in \mathfrak{G}$ are called the *open subsets* of Ω, and their complements are called the *closed subsets* of Ω. A set $K \subset \Omega$ is called *compact* if for every family $(G_i)_{i \in I}$ with $G_i \in \mathfrak{G}$ and

$$K \subset \bigcup_{i \in I} G_i$$

there is a finite set $I_0 \subset I$ such that

$$K \subset \bigcup_{i \in I_0} G_i.$$

On $\Omega = \mathbb{R}^k$ and $\Omega = \overline{\mathbb{R}}^k$ we consider the natural topologies, and we use $\mathfrak{G}_k$ to denote the corresponding class of open sets in $\mathbb{R}^k$. In particular, $O \subset \overline{\mathbb{R}}$ is an open set iff $O \cap \mathbb{R} \in \mathfrak{G}_1$ and $]a, \infty] \subset O$ for some $a < \infty$ if $\infty \in O$ and $[-\infty, a[\subset O$ for some $a > -\infty$ if $-\infty \in O$.

Definition 4. For every topological space $(\Omega, \mathfrak{G})$

$$\mathfrak{B}(\Omega) = \sigma(\mathfrak{G})$$

is the *Borel-σ-algebra (in Ω w.r.t. $\mathfrak{G}$)*. In particular,

$$\mathfrak{B}_k = \mathfrak{B}(\mathbb{R}^k), \quad \mathfrak{B} = \mathfrak{B}_1, \quad \overline{\mathfrak{B}}_k = \mathfrak{B}(\overline{\mathbb{R}}^k), \quad \overline{\mathfrak{B}} = \overline{\mathfrak{B}}_1.$$

Remark 6. We have

$$\begin{aligned}\mathfrak{B}_k &= \sigma(\{F \subset \mathbb{R}^k : F \text{ closed}\}) = \sigma(\{K \subset \mathbb{R}^k : K \text{ compact}\}) \\ &= \sigma(\{]-\infty, a] : a \in \mathbb{R}^k\}) = \sigma(\{]-\infty, a] : a \in \mathbb{Q}^k\})\end{aligned}$$

and

$$\overline{\mathfrak{B}} = \{B \subset \overline{\mathbb{R}} : B \cap \mathbb{R} \in \mathfrak{B}\}. \tag{4}$$

Moreover,

$$\mathfrak{B}_k \subsetneq \mathfrak{P}(\mathbb{R}^k)$$

since the cardinalities of $\mathfrak{B}_k$ and $\mathbb{R}^k$ coincide, see Billingsley (1979, Exercise 2.21).

Definition 5. For any σ-algebra $\mathfrak{A}$ in Ω and $\widetilde{\Omega} \subset \Omega$

$$\widetilde{\mathfrak{A}} = \{\widetilde{\Omega} \cap A : A \in \mathfrak{A}\}$$

is the *trace-σ-algebra of $\mathfrak{A}$ in $\widetilde{\Omega}$*, sometimes denoted by $\widetilde{\Omega} \cap \mathfrak{A}$.

Remark 7.

(i) $\widetilde{\mathfrak{A}}$ is a σ-algebra.

(ii) $\widetilde{\mathfrak{A}} \not\subset \mathfrak{A}$ in general, but $\widetilde{\Omega} \in \mathfrak{A} \Rightarrow \widetilde{\mathfrak{A}} = \{A \in \mathfrak{A} : A \subset \widetilde{\Omega}\}$.

(iii) $\mathfrak{A} = \sigma(\mathfrak{E}) \Rightarrow \widetilde{\mathfrak{A}} = \sigma(\{\widetilde{\Omega} \cap E : E \in \mathfrak{E}\})$.

(iv) $\mathfrak{B}_k = \mathbb{R}^k \cap \overline{\mathfrak{B}}_k$, see (4) for $k = 1$.

(v) $[a, b[\cap \mathfrak{B}_k = \sigma(\{[a, c[: a \le c \le b\})$, see (iii).

2 Measurable Mappings

Definition 1. $(\Omega, \mathfrak{A})$ is called *measurable space* if Ω is a non-empty set and $\mathfrak{A}$ is a σ-algebra in Ω. Elements $A \in \mathfrak{A}$ are called *measurable sets*.

Remark 1. Let $f : \Omega_1 \to \Omega_2$.

(i) $f^{-1}(\mathfrak{A}_2) = \{f^{-1}(A) : A \in \mathfrak{A}_2\}$ is a σ-algebra in Ω_1 for every σ-algebra $\mathfrak{A}_2$ in Ω_2.

(ii) $\{A \subset \Omega_2 : f^{-1}(A) \in \mathfrak{A}_1\}$ is a σ-algebra in Ω_2 for every σ-algebra $\mathfrak{A}_1$ in Ω_1.

In the sequel, $(\Omega_i, \mathfrak{A}_i)$ are measurable spaces for $i = 1, 2, 3$.

Definition 2. $f : \Omega_1 \to \Omega_2$ is $\mathfrak{A}_1$-$\mathfrak{A}_2$-*measurable* if $f^{-1}(\mathfrak{A}_2) \subset \mathfrak{A}_1$.

Example 1. Let $f : \Omega_1 \to \Omega_2$.

(i) Every constant mapping f is $\mathfrak{A}_1$-$\mathfrak{A}_2$-measurable.

(ii) Let $\Omega_2 = \{0, 1\}$ and $\mathfrak{A}_2 = \mathfrak{P}(\Omega_2)$. Then f is $\mathfrak{A}_1$-$\mathfrak{A}_2$-measurable iff $f = 1_A$ with $A \in \mathfrak{A}_1$.

How can we prove measurability of a given mapping?

Theorem 1. If $f : \Omega_1 \to \Omega_2$ is $\mathfrak{A}_1$-$\mathfrak{A}_2$-measurable and $g : \Omega_2 \to \Omega_3$ is $\mathfrak{A}_2$-$\mathfrak{A}_3$-measurable, then $g \circ f : \Omega_1 \to \Omega_3$ is $\mathfrak{A}_1$-$\mathfrak{A}_3$-measurable.

Proof. We have $(g \circ f)^{-1}(\mathfrak{A}_3) = f^{-1}(g^{-1}(\mathfrak{A}_3)) \subset f^{-1}(\mathfrak{A}_2) \subset \mathfrak{A}_1$. □

Lemma 1. For $f : \Omega_1 \to \Omega_2$ and $\mathfrak{E} \subset \mathfrak{P}(\Omega_2)$

$$f^{-1}(\sigma(\mathfrak{E})) = \sigma(f^{-1}(\mathfrak{E})).$$

Proof. By $f^{-1}(\mathfrak{E}) \subset f^{-1}(\sigma(\mathfrak{E}))$ and Remark 1.(i) we get $\sigma(f^{-1}(\mathfrak{E})) \subset f^{-1}(\sigma(\mathfrak{E}))$.
Let $\mathfrak{F} = \{A \subset \Omega_2 : f^{-1}(A) \in \sigma(f^{-1}(\mathfrak{E}))\}$. Then $\mathfrak{E} \subset \mathfrak{F}$ and $\mathfrak{F}$ is a σ-algebra, see Remark 1.(ii). Thus we get $\sigma(\mathfrak{E}) \subset \mathfrak{F}$, i.e., $f^{-1}(\sigma(\mathfrak{E})) \subset \sigma(f^{-1}(\mathfrak{E}))$. □

Theorem 2. If $\mathfrak{A}_2 = \sigma(\mathfrak{E})$ with $\mathfrak{E} \subset \mathfrak{P}(\Omega_2)$, then

$$f^{-1}(\mathfrak{E}) \subset \mathfrak{A}_1 \quad \Leftrightarrow \quad f \text{ is } \mathfrak{A}_1\text{-}\mathfrak{A}_2\text{-measurable.}$$

Proof. '⇒': Assume that $f^{-1}(\mathfrak{E}) \subset \mathfrak{A}_1$. By Lemma 1,

$$f^{-1}(\mathfrak{A}_2) = f^{-1}(\sigma(\mathfrak{E})) = \sigma(f^{-1}(\mathfrak{E})) \subset \sigma(\mathfrak{A}_1) = \mathfrak{A}_1.$$

Obviously, '⇐' holds, too. □

Corollary 1. For every pair of topological spaces $(\Omega_1, \mathfrak{G}_1)$ and $(\Omega_2, \mathfrak{G}_2)$ and every mapping $f : \Omega_1 \to \Omega_2$,

$$f \text{ continuous} \quad \Rightarrow \quad f \text{ is } \mathfrak{B}(\Omega_1)\text{-}\mathfrak{B}(\Omega_2)\text{-measurable.}$$

Proof. By assumption,

$$f^{-1}(\mathfrak{G}_2) \subset \mathfrak{G}_1 \subset \sigma(\mathfrak{G}_1) = \mathfrak{B}(\Omega_1).$$

Use Theorem 2. □

Given: measurable spaces $(\Omega_i, \mathfrak{A}_i)$ for $i \in I \neq \emptyset$ and mappings $f_i : \Omega \to \Omega_i$ for $i \in I$ and some non-empty set Ω.

Definition 3. The *σ-algebra generated by $(f_i)_{i \in I}$ (and $(\mathfrak{A}_i)_{i \in I}$)*

$$\sigma(\{f_i : i \in I\}) = \sigma\Big(\bigcup_{i \in I} f_i^{-1}(\mathfrak{A}_i)\Big).$$

Put $\sigma(f) = \sigma(\{f\})$ in the case $|I| = 1$ and $f = f_i$.

Remark 2. $\sigma(\{f_i : i \in I\})$ is the smallest σ-algebra $\mathfrak{A}$ in Ω such that all mappings f_i are $\mathfrak{A}$-$\mathfrak{A}_i$-measurable.

Theorem 3. For every measurable space $(\widetilde{\Omega}, \widetilde{\mathfrak{A}})$ and every mapping $g : \widetilde{\Omega} \to \Omega$,

$$g \text{ is } \widetilde{\mathfrak{A}}\text{-}\sigma(\{f_i : i \in I\})\text{-measurable} \quad \Leftrightarrow \quad \forall\, i \in I : f_i \circ g \text{ is } \widetilde{\mathfrak{A}}\text{-}\mathfrak{A}_i\text{-measurable.}$$

Proof. Use Lemma 1 to obtain

$$g^{-1}(\sigma(\{f_i : i \in I\})) = \sigma\Big(g^{-1}\Big(\bigcup_{i \in I} f_i^{-1}(\mathfrak{A}_i)\Big)\Big) = \sigma\Big(\bigcup_{i \in I} (f_i \circ g)^{-1}(\mathfrak{A}_i)\Big).$$

Therefore

$$g^{-1}\left(\sigma(\{f_i : i \in I\})\right) \subset \widetilde{\mathfrak{A}} \quad \Leftrightarrow \quad \forall\, i \in I : f_i \circ g \text{ is } \widetilde{\mathfrak{A}}\text{-}\mathfrak{A}_i\text{-measurable.}$$

□

Now we turn to the particular case of functions with values in $\mathbb{R}$ or $\overline{\mathbb{R}}$, and we consider the Borel σ-algebra in $\mathbb{R}$ or $\overline{\mathbb{R}}$, respectively. For any measurable space $(\Omega, \mathfrak{A})$ we use the following notation

$$\begin{aligned} \mathfrak{Z}(\Omega, \mathfrak{A}) &= \{f : \Omega \to \mathbb{R} : f \text{ is } \mathfrak{A}\text{-}\mathfrak{B}\text{-measurable}\}, \\ \mathfrak{Z}_+(\Omega, \mathfrak{A}) &= \{f \in \mathfrak{Z}(\Omega, \mathfrak{A}) : f \geq 0\}, \\ \overline{\mathfrak{Z}}(\Omega, \mathfrak{A}) &= \left\{f : \Omega \to \overline{\mathbb{R}} : f \text{ is } \mathfrak{A}\text{-}\overline{\mathfrak{B}}\text{-measurable}\right\}, \\ \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A}) &= \left\{f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A}) : f \geq 0\right\}. \end{aligned}$$

Every function $f : \Omega \to \mathbb{R}$ may also be considered as a function with values in $\overline{\mathbb{R}}$, and in this case $f \in \mathfrak{Z}(\Omega, \mathfrak{A})$ iff $f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$.

Corollary 2. For $\prec \in \{\leq, <, \geq, >\}$ and $f : \Omega \to \overline{\mathbb{R}}$,

$$f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A}) \quad \Leftrightarrow \quad \forall\, a \in \mathbb{R} : \{\omega \in \Omega : f(\omega) \prec a\} \in \mathfrak{A}.$$

Proof. For instance,

$$\{\omega \in \Omega : f(\omega) \leq a\} = f^{-1}([-\infty, a])$$

and $\overline{\mathfrak{B}} = \sigma(\{[-\infty, a] : a \in \mathbb{R}\})$, see Remark 1.6. It remains to apply Theorem 2. □

Theorem 4. For $f, g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ and $\prec \in \{\leq, <, \geq, >, =, \neq\}$,

$$\{\omega \in \Omega : f(\omega) \prec g(\omega)\} \in \mathfrak{A}.$$

Proof. For instance, Corollary 2 yields

$$\begin{aligned}\{\omega \in \Omega : f(\omega) < g(\omega)\} &= \bigcup_{q \in \mathbb{Q}} \{\omega \in \Omega : f(\omega) < q < g(\omega)\} \\ &= \bigcup_{q \in \mathbb{Q}} (\{\omega \in \Omega : f(\omega) < q\} \cap \{\omega \in \Omega : g(\omega) > q\}) \in \mathfrak{A}.\end{aligned}$$

□

As is customary, we use the abbreviation

$$\{f \in A\} = \{\omega \in \Omega : f(\omega) \in A\}$$

for any $f : \Omega \to \widetilde{\Omega}$ and $A \subset \widetilde{\Omega}$.

Theorem 5. For every sequence $f_1, f_2, \ldots \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$,

(i) $\inf_{n \in \mathbb{N}} f_n,\ \sup_{n \in \mathbb{N}} f_n \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$,

(ii) $\liminf_{n \to \infty} f_n,\ \limsup_{n \to \infty} f_n \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$,

(iii) if $(f_n)_{n \in \mathbb{N}}$ converges at every point $\omega \in \Omega$, then $\lim_{n \to \infty} f_n \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$.

Proof. For $a \in \mathbb{R}$

$$\left\{\inf_{n \in \mathbb{N}} f_n < a\right\} = \bigcup_{n \in \mathbb{N}} \{f_n < a\}, \quad \left\{\sup_{n \in \mathbb{N}} f_n \leq a\right\} = \bigcap_{n \in \mathbb{N}} \{f_n \leq a\}.$$

Hence, Corollary 2 yields (i). Since

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

we obtain (ii) from (i). Finally, (iii) follows from (ii). □

By

$$f^+ = \max(0, f), \quad f^- = \max(0, -f)$$

we denote the positive part and the negative part, respectively, of $f : \Omega \to \overline{\mathbb{R}}$.

Remark 3. For $f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ we have $f^+, f^-, |f| \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$.

Theorem 6. For $f, g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$,

$$f \pm g,\ f \cdot g,\ f/g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A}),$$

provided that these functions are well defined.

Proof. The proof is again based on Corollary 2. For simplicity we only consider the case that f and g are real-valued. Clearly $g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ implies $-g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$, too. Furthermore, for every $a \in \mathbb{R}$,

$$\{f + g < a\} = \bigcup_{q \in \mathbb{Q}} \{f < q\} \cap \{g < a - q\},$$

and therefore $f \pm g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$. Clearly $f \cdot g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ if f is constant. Moreover, $x \mapsto x^2$ defines a $\mathfrak{B}$-$\mathfrak{B}$-measurable function, see Corollary 1, and

$$f \cdot g = 1/4 \cdot \big((f+g)^2 - (f-g)^2\big)$$

We apply Theorem 1 to obtain $f \cdot g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ in general. Finally, it is easy to show that $g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ implies $1/g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$. □

Definition 4. $f \in \mathfrak{Z}(\Omega, \mathfrak{A})$ is called *simple function* if $|f(\Omega)| < \infty$. Put

$$\mathfrak{S}(\Omega, \mathfrak{A}) = \{f \in \mathfrak{Z}(\Omega, \mathfrak{A}) : f \text{ simple}\},$$
$$\mathfrak{S}_+(\Omega, \mathfrak{A}) = \{f \in \mathfrak{S}(\Omega, \mathfrak{A}) : f \geq 0\}.$$

Remark 4. $f \in \mathfrak{S}(\Omega, \mathfrak{A})$ iff

$$f = \sum_{i=1}^{n} \alpha_i \cdot 1_{A_i}$$

with $\alpha_1, \ldots \alpha_n \in \mathbb{R}$ pairwise different and $A_1, \ldots, A_n \in \mathfrak{A}$ pairwise disjoint such that $\bigcup_{i=1}^n A_i = \Omega$.

Theorem 7. For every (bounded) function $f \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$ there exists a sequence $f_1, f_2, \ldots \in \mathfrak{S}_+(\Omega, \mathfrak{A})$ such that $f_n \uparrow f$ (with uniform convergence).

Proof. Let $n \in N$ and put

$$f_n = \sum_{k=1}^{n \cdot 2^n} \frac{k-1}{2^n} \cdot 1_{A_{n,k}} + n \cdot 1_{B_n}$$

where

$$A_{n,k} = \{(k-1)/(2^n) \leq f < k/(2^n)\}, \quad B_n = \{f \geq n\}.$$

□

Now we consider a mapping $T : \Omega_1 \to \Omega_2$ and a σ-algebra $\mathfrak{A}_2$ in Ω_2. We characterize measurability of functions with respect to $\sigma(T) = T^{-1}(\mathfrak{A}_2)$.

Theorem 8 (Factorization Lemma)**.** For every function $f : \Omega_1 \to \overline{\mathbb{R}}$

$$f \in \overline{\mathfrak{Z}}(\Omega_1, \sigma(T)) \quad \Leftrightarrow \quad \exists\, g \in \overline{\mathfrak{Z}}(\Omega_2, \mathfrak{A}_2) : f = g \circ T.$$

Proof. '$\Leftarrow$' is trivially satisfied. '$\Rightarrow$': First, assume that $f \in \mathfrak{S}_+(\Omega_1, \sigma(T))$, i.e.,

$$f = \sum_{i=1}^{n} \alpha_i \cdot 1_{A_i}$$

with pairwise disjoint sets $A_1, \dots, A_n \in \sigma(T)$. Take pairwise disjoint sets $B_1, \dots, B_n \in \mathfrak{A}_2$ such that $A_i = T^{-1}(B_i)$ and put

$$g = \sum_{i=1}^{n} \alpha_i \cdot 1_{B_i}.$$

Clearly $f = g \circ T$ and $g \in \overline{\mathfrak{Z}}(\Omega_2, \mathfrak{A}_2)$.

Now, assume that $f \in \overline{\mathfrak{Z}}_+(\Omega_1, \sigma(T))$. Take a sequence $(f_n)_{n\in\mathbb{N}}$ in $\mathfrak{S}_+(\Omega_1, \sigma(T))$ according to Theorem 7. We already know that $f_n = g_n \circ T$ for suitable $g_n \in \overline{\mathfrak{Z}}(\Omega_2, \mathfrak{A}_2)$. Hence

$$f = \sup_n f_n = \sup_n (g_n \circ T) = (\sup_n g_n) \circ T = g \circ T$$

where $g = \sup_n g_n \in \overline{\mathfrak{Z}}(\Omega_2, \mathfrak{A}_2)$.

In the general case, we already know that

$$f^+ = g_1 \circ T, \quad f^- = g_2 \circ T$$

for suitable $g_1, g_2 \in \overline{\mathfrak{Z}}(\Omega_2, \mathfrak{A}_2)$. Put

$$C = \{g_1 = g_2 = \infty\} \in \mathfrak{A}_2,$$

and observe that $T(\Omega_1) \cap C = \emptyset$ since $f = f^+ - f^-$. We conclude that $f = g \circ T$ where

$$g = g_1 \cdot 1_D - g_2 \cdot 1_D \in \overline{\mathfrak{Z}}(\Omega_2, \mathfrak{A}_2)$$

with $D = C^c$. □

Our method of proof for Theorem 8 is sometimes called *algebraic induction.*

3 Product Spaces

Example 1. A stochastic model for coin tossing. For a single trial,

$$\Omega = \{0, 1\}, \quad \mathfrak{A} = \mathfrak{P}(\Omega), \quad \forall\, \omega \in \Omega : P(\{\omega\}) = 1/2. \tag{1}$$

For n 'independent' trials, (1) serves as a building-block,

$$\Omega_i = \{0, 1\}, \quad \mathfrak{A}_i = \mathfrak{P}(\Omega_i), \quad \forall\, \omega_i \in \Omega_i : P_i(\{\omega_i\}) = 1/2,$$

and we define

$$\Omega = \bigtimes_{i=1}^{n} \Omega_i, \quad \mathfrak{A} = \mathfrak{P}(\Omega), \quad \forall\, A \in \mathfrak{A} : P(A) = \frac{|A|}{|\Omega|}.$$

Then

$$P(A_1 \times \cdots \times A_n) = P_1(A_1) \cdot \cdots \cdot P_n(A_n)$$

for all $A_i \in \mathfrak{A}_i$.

Question: How to model an infinite sequence of trials? To this end,

$$\Omega = \bigtimes_{i=1}^{\infty} \Omega_i.$$

How to choose a σ-algebra $\mathfrak{A}$ in Ω and a probability measure P on $(\Omega, \mathfrak{A})$? A reasonable requirement is

$$\begin{aligned} &\forall\, n \in \mathbb{N}\ \forall\, A_i \in \mathfrak{A}_i : \\ &\qquad P(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \ldots) = P_1(A_1) \cdot \cdots \cdot P_n(A_n). \end{aligned} \tag{2}$$

Unfortunately,

$$\mathfrak{A} = \mathfrak{P}(\Omega)$$

is too large, since there exists no probability measure on $(\Omega, \mathfrak{P}(\Omega))$ such that (2) holds. The latter fact follows from a theorem by Banach and Kuratowski, which relies on the continuum hypothesis, see Dudley (2002, p. 526). On the other hand,

$$\mathfrak{A} = \{A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \Omega_{n+2} \cdots : n \in \mathbb{N},\ A_i \in \mathfrak{A}_i \text{ for } i = 1, \ldots, n\} \tag{3}$$

is not a σ-algebra.

Given: a non-empty set I and measurable spaces $(\Omega_i, \mathfrak{A}_i)$ for $i \in I$. Put

$$Y = \bigcup_{i \in I} \Omega_i$$

and define

$$\bigtimes_{i \in I} \Omega_i = \{\omega \in Y^I : \omega(i) \in \Omega_i \text{ for } i \in I\}.$$

Notation: $\omega = (\omega_i)_{i \in I}$ for $\omega \in \bigtimes_{i \in I} \Omega_i$. Moreover, let

$$\mathfrak{P}_0(I) = \{J \subset I : J \text{ non-empty, finite}\}.$$

The following definition is motivated by (3).

Definition 1.

(i) *Measurable rectangle*

$$A = \bigtimes_{j \in J} A_j \times \bigtimes_{i \in I \setminus J} \Omega_i$$

with $J \in \mathfrak{P}_0(I)$ and $A_j \in \mathfrak{A}_j$ for $j \in J$. Notation: $\mathfrak{R}$ class of measurable rectangles.

(ii) *Product (measurable) space* $(\Omega, \mathfrak{A})$ with *components* $(\Omega_i, \mathfrak{A}_i)$, $i \in I$,

$$\Omega = \bigtimes_{i\in I} \Omega_i, \qquad \mathfrak{A} = \sigma(\mathfrak{R}).$$

Notation: $\mathfrak{A} = \bigotimes_{i\in I} \mathfrak{A}_i$, *product σ-algebra.*

Remark 1. The class $\mathfrak{R}$ is a semi-algebra, but not an algebra in general. See Übung 2.3.

Example 2. Obviously, (2) only makes sense if $\mathfrak{A}$ contains the product σ-algebra $\bigotimes_{i=1}^{\infty} \mathfrak{P}(\{0,1\})$. We will show that there exists a uniquely determined probability measure P on the product space $\left(\bigtimes_{i=1}^{\infty}\{0,1\}, \bigotimes_{i=1}^{\infty} \mathfrak{P}(\{0,1\})\right)$ that satisfies (2), see Remark 4.3.(ii). The corresponding probability space yields a stochastic model for the simple case of gambling, which was mentioned in the introductory Example I.2.

We study several classes of mappings or subsets that generate the product σ-algebra. Moreover, we characterize measurability of mappings that take values in a product space.
Put $\Omega = \bigtimes_{i\in I} \Omega_i$. For any $\emptyset \neq S \subset I$ let

$$\pi_S^I : \Omega \to \bigtimes_{i\in S} \Omega_i, \quad (\omega_i)_{i\in I} \mapsto (\omega_i)_{i\in S}$$

denote the *projection* of Ω onto $\bigtimes_{i\in S} \Omega_i$ (restriction of mappings ω). In particular, for $i \in I$ the i-th projection is given by $\pi_{\{i\}}^I$. Sometimes we simply write π_S instead of π_S^I and π_i instead of $\pi_{\{i\}}$.

Theorem 1.

(i) $\bigotimes_{i\in I} \mathfrak{A}_i = \sigma(\{\pi_i : i \in I\})$.

(ii) $\forall\, i \in I : \mathfrak{A}_i = \sigma(\mathfrak{E}_i) \quad \Rightarrow \quad \bigotimes_{i\in I} \mathfrak{A}_i = \sigma\left(\bigcup_{i\in I} \pi_i^{-1}(\mathfrak{E}_i)\right)$.

Proof. Ad (i), '⊃': We show that every projection $\pi_i : \Omega \to \Omega_i$ is $\left(\bigotimes_{i\in I} \mathfrak{A}_i\right)$-$\mathfrak{A}_i$-measurable. For $A_i \in \mathfrak{A}_i$

$$\pi_i^{-1}(A_i) = A_i \times \bigtimes_{k\in I\setminus\{i\}} \Omega_k \in \mathfrak{R}.$$

Ad (i), '⊂': We show that $\mathfrak{R} \subset \sigma(\{\pi_i : i \in I\})$. For $J \in \mathfrak{P}_0(I)$ and $A_j \in \mathfrak{A}_j$ with $j \in J$

$$\bigtimes_{j\in J} A_j \times \bigtimes_{i\in I\setminus J} \Omega_i = \bigcap_{j\in J} \pi_j^{-1}(A_j).$$

Ad (ii): By Lemma 2.1 and (i)

$$\bigotimes_{i\in I} \mathfrak{A}_i = \sigma\Big(\bigcup_{i\in I} \pi_i^{-1}(\mathfrak{A}_i)\Big) = \sigma\Big(\bigcup_{i\in I} \sigma(\pi_i^{-1}(\mathfrak{E}_i))\Big) = \sigma\Big(\bigcup_{i\in I} \pi_i^{-1}(\mathfrak{E}_i)\Big).$$

□

Corollary 1.

(i) For every measurable space $(\widetilde{\Omega}, \widetilde{\mathfrak{A}})$ and every mapping $g : \widetilde{\Omega} \to \Omega$

$$g \text{ is } \widetilde{\mathfrak{A}}\text{-}\bigotimes_{i\in I} \mathfrak{A}_i\text{-measurable} \quad \Leftrightarrow \quad \forall\, i \in I : \pi_i \circ g \text{ is } \widetilde{\mathfrak{A}}\text{-}\mathfrak{A}_i\text{-measurable.}$$

(ii) For every $\emptyset \neq S \subset I$ the projection π_S^I is $\bigotimes_{i\in I} \mathfrak{A}_i$-$\bigotimes_{i\in S} \mathfrak{A}_i$-measurable.

Proof. Ad (i): Follows immediately from Theorem 2.3 and Theorem 1.(i).
Ad (ii): Note that $\pi^S_{\{i\}} \circ \pi^I_S = \pi^I_i$ and use (i). □

Remark 2. From Theorem 1.(i) and Corollary 1 we get

$$\bigotimes_{i\in I} \mathfrak{A}_i = \sigma(\{\pi^I_S : S \in \mathfrak{P}_0(I)\}).$$

The sets

$$\left(\pi^I_S\right)^{-1}(B) = B \times \Big(\bigtimes_{i\in I\setminus S} \Omega_i \Big)$$

with $S \in \mathfrak{P}_0(I)$ and $B \in \bigotimes_{i\in S} \mathfrak{A}_i$ are called *cylinder sets*. Notation: $\mathfrak{C}$ class of cylinder sets. The class $\mathfrak{C}$ is an algebra in Ω, but not a σ-algebra in general. Moreover,

$$\mathfrak{R} \subset \alpha(\mathfrak{R}) \subset \mathfrak{C} \subset \sigma(\mathfrak{R}),$$

where equality does not hold in general.

Every product measurable set is countably determined in the following sense.

Theorem 2. For every $A \in \bigotimes_{i\in I} \mathfrak{A}_i$ there exists a non-empty countable set $S \subset I$ and a set $B \in \bigotimes_{i\in S} \mathfrak{A}_i$ such that

$$A = \left(\pi^I_S\right)^{-1}(B).$$

Proof. Put

$$\widetilde{\mathfrak{A}} = \Big\{ A \in \bigotimes_{i\in I} \mathfrak{A}_i : \exists\, S \subset I \text{ non-empty, countable } \exists\, B \in \bigotimes_{i\in S} \mathfrak{A}_i : A = \left(\pi^I_S\right)^{-1}(B) \Big\}.$$

By definition, $\widetilde{\mathfrak{A}}$ contains every cylinder set and $\widetilde{\mathfrak{A}} \subset \bigotimes_{i\in I} \mathfrak{A}_i$. It remains to show that $\widetilde{\mathfrak{A}}$ is a σ-algebra. See Gänssler, Stute (1977, p. 24) for details. □

Now we study products of Borel-σ-algebras.

Theorem 3.

$$\mathfrak{B}_k = \bigotimes_{i=1}^{k} \mathfrak{B}, \qquad \overline{\mathfrak{B}}_k = \bigotimes_{i=1}^{k} \overline{\mathfrak{B}}.$$

Proof. By Remark 1.6,

$$\mathfrak{B}_k = \sigma\Big(\Big\{ \bigtimes_{i=1}^{k} \,]-\infty, a_i] : a_i \in \mathbb{R} \text{ for } i = 1, \dots, k\Big\}\Big) \subset \bigotimes_{i=1}^{k} \mathfrak{B}.$$

On the other hand, $\pi_i : \mathbb{R}^k \to \mathbb{R}$ is continuous, hence it remains to apply Corollary 2.1 and Theorem 1.(i). Analogously, $\overline{\mathfrak{B}}_k = \bigotimes_{i=1}^{k} \overline{\mathfrak{B}}$ follows. □

Remark 3. More generally, consider a non-empty countable set I and a family of topological spaces $(\Omega_i, \mathfrak{G}_i)$ where $i \in I$. Assume that every space $(\Omega_i, \mathfrak{G}_i)$ has a countable basis and consider the product topology $\mathfrak{G}$ on $\Omega = \bigtimes_{i \in I} \Omega_i$. Then

$$\mathfrak{B}(\Omega) = \bigotimes_{i \in I} \mathfrak{B}(\Omega_i),$$

see Gänssler, Stute (1977, Satz 1.3.12).

Remark 4. Consider a measurable space $(\widetilde{\Omega}, \widetilde{\mathfrak{A}})$ and a mapping

$$f = (f_1, \dots, f_k) : \widetilde{\Omega} \to \overline{\mathbb{R}}^k.$$

Then, according to Theorem 3, f is $\widetilde{\mathfrak{A}}$-$\overline{\mathfrak{B}}_k$-measurable iff all functions f_i are $\widetilde{\mathfrak{A}}$-$\overline{\mathfrak{B}}$-measurable.

We briefly discuss the cardinality of σ-algebras. It is known that

$$2 \leq |V| \leq |\mathbb{R}| \quad \Rightarrow \quad \left|V^{\mathbb{N}}\right| = |\mathbb{R}| \wedge \left|V^{\mathbb{R}}\right| = \left|\{0,1\}^{\mathbb{R}}\right|$$

for every set V, see Hewitt, Stromberg (1965, Exercise 4.34).

Theorem 4. Assume that $\emptyset \in \mathfrak{E} \subset \mathfrak{P}(\Omega)$ and $|\mathfrak{E}| \geq 2$. Then

$$|\sigma(\mathfrak{E})| \leq \left|\mathfrak{E}^{\mathbb{N}}\right|.$$

Proof. See Hewitt, Stromberg (1965, Theorem 10.13). □

Example 3. Let $I = \mathbb{N}$, $\Omega_i = \{0,1\}$, and $\mathfrak{A}_i = \mathfrak{P}(\Omega_i)$, as in Example 1. For the corresponding product space $(\Omega, \mathfrak{A})$ we have $\Omega = \{0,1\}^{\mathbb{N}}$ and

$$|\mathfrak{A}| = |\Omega| = |\mathbb{R}|.$$

Proof: Note that $\{\omega\} \in \mathfrak{A}$ for every $\omega \in \Omega$. Hence $|\mathfrak{A}| \geq |\Omega|$. Conversely, use Theorem 1.(ii) with $\mathfrak{E}_i = \{\{1\}\}$ and Theorem 4 to conclude that $|\mathfrak{A}| \leq |\mathbb{N}^{\mathbb{N}}| = |\mathbb{R}|$.
We add that $|\mathfrak{P}(\Omega)| = |\{0,1\}^{\mathbb{R}}| > |\Omega|$.

Example 4. Let $I = \mathbb{N}$, $\Omega_i = \mathbb{R}$, and $\mathfrak{A}_i = \mathfrak{B}$. For the corresponding product space $(\Omega, \mathfrak{A})$ we have $\Omega = \mathbb{R}^{\mathbb{N}}$ and

$$|\mathfrak{A}| = |\Omega| = |\mathbb{R}|.$$

Proof: As in the previous example, with $\mathfrak{E}_i = \{]-\infty, a] : a \in \mathbb{Q}\}$.
Again we have $|\mathfrak{P}(\Omega)| = |\{0,1\}^{\mathbb{R}}| > |\Omega|$.
The sets $\{(x_n)_{n\in\mathbb{N}} : (x_n)_{n\in\mathbb{N}} \text{ converges}\}$ and $\{(x_n)_{n\in\mathbb{N}} : (x_n)_{n\in\mathbb{N}} \text{ is bounded}\}$ are elements of $\mathfrak{A}$, but they are not cylinder sets.

Example 5. Let $I = \mathbb{R}_+$, $\Omega_i = \mathbb{R}$, and $\mathfrak{A}_i = \mathfrak{B}$. For the corresponding product space $(\Omega, \mathfrak{A})$ we have $\Omega = \mathbb{R}^{\mathbb{R}_+}$ and

$$|\mathfrak{A}| = |\mathbb{R}| < |\Omega|.$$

Proof: Clearly $|\mathbb{R}| \leq |\mathfrak{A}|$ and $|\mathbb{R}| < |\Omega|$. On the other hand, Theorem 2 shows that $\mathfrak{A} = \sigma(\mathfrak{E})$ for some set $\mathfrak{E}$ with $|\mathfrak{E}| = |\mathbb{R}|$. Hence $|\mathfrak{A}| \leq |\mathbb{R}|$ by Theorem 4.

The space $\mathbb{R}^{\mathbb{R}_+}$ already appeared in the introductory Example I.3. The product σ-algebra $\mathfrak{A} = \bigotimes_{i \in \mathbb{R}_+} \mathfrak{B}$ is a proper choice on this space. On the subspace $C(\mathbb{R}_+) \subset \mathbb{R}^{\mathbb{R}_+}$ we can take the trace-σ-algebra. It is important to note, however, that

$$C(\mathbb{R}_+) \notin \mathfrak{A},$$

see Übung 2.4. It turns out that the Borel σ-algebra $\mathfrak{B}(C(\mathbb{R}_+))$ that is generated by the topology of uniform convergence on compact intervals coincides with the trace-σ-algebra of $\mathfrak{A}$ in $C(\mathbb{R}_+)$, see Bauer (1996, Theorem 38.6).

4 Construction of (Probability) Measures

Given: $\Omega \neq \emptyset$ and $\emptyset \neq \mathfrak{A} \subset \mathfrak{P}(\Omega)$.

Definition 1. $\mu : \mathfrak{A} \to \mathbb{R}_+ \cup \{\infty\}$ is called

(i) *additive* if:

$$A, B \in \mathfrak{A} \wedge A \cap B = \emptyset \wedge A \cup B \in \mathfrak{A} \quad \Rightarrow \quad \mu(A \cup B) = \mu(A) + \mu(B),$$

(ii) *σ-additive* if

$$A_1, A_2, \ldots \in \mathfrak{A} \text{ pairwise disjoint } \wedge \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A} \quad \Rightarrow \quad \mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i),$$

(iii) *content (on $\mathfrak{A}$)* if

$$\mathfrak{A} \text{ algebra} \quad \wedge \quad \mu \text{ additive} \quad \wedge \quad \mu(\emptyset) = 0,$$

(iv) *pre-measure (on $\mathfrak{A}$)* if

$$\mathfrak{A} \text{ semi-algebra} \quad \wedge \quad \mu\ \sigma\text{-additive} \quad \wedge \quad \mu(\emptyset) = 0,$$

(v) *measure (on $\mathfrak{A}$)* if

$$\mathfrak{A}\ \sigma\text{-algebra} \quad \wedge \quad \mu \text{ pre-measure},$$

(vi) *probability measure (on $\mathfrak{A}$)* if

$$\mu \text{ measure} \quad \wedge \quad \mu(\Omega) = 1.$$

Definition 2. $(\Omega, \mathfrak{A}, \mu)$ is called a

(i) *measure space*, if μ is a measure on the σ-algebra $\mathfrak{A}$ in Ω,

(ii) *probability space*, if μ is a probability measure on the σ-algebra $\mathfrak{A}$ in Ω.

Example 1.

(i) *Lebesgue pre-measure* λ_1 on the class $\mathfrak{J}_1$ of intervals from Example 1.1.(i): $\lambda_1(A)$ is the length of $A \in \mathfrak{J}_1$, i.e.,

$$\lambda_1(]a, b]) = b - a$$

if $a, b \in \mathbb{R}$ with $a \le b$ and $\lambda_1(A) = \infty$ if $A \in \mathfrak{J}_1$ is unbounded. See Billingsley (1979, p. 22), Elstrodt (1996, §II.2), or Analysis IV.

Analogously for cartesian products of such intervals. Hereby we get the semi-algebra $\mathfrak{J}_k$ of rectangles in $\mathbb{R}^k$. The *Lebesgue pre-measure* λ_k on $\mathfrak{J}_k$ yields the volume $\lambda_k(A)$ of $A \in \mathfrak{J}_k$, i.e., the product of the side-lengths of A. See Elstrodt (1996, §III.2) or Analysis IV.

(ii) for any semi-algebra $\mathfrak{A}$ in Ω and $\omega \in \Omega$

$$\varepsilon_\omega(A) = 1_A(\omega), \qquad A \in \mathfrak{A},$$

defines a pre-measure. If $\mathfrak{A}$ is a σ-algebra, then ε_ω is called the *Dirac measure* at the point ω.

More generally: take sequences $(\omega_n)_{n\in\mathbb{N}}$ in Ω and $(\alpha_n)_{n\in\mathbb{N}}$ in $\mathbb{R}_+$ such that $\sum_{n=1}^\infty \alpha_n = 1$. Then

$$\mu(A) = \sum_{n=1}^\infty \alpha_n \cdot 1_A(\omega_n), \qquad A \in \mathfrak{A},$$

defines a *discrete probability measure* on any σ-algebra $\mathfrak{A}$ in Ω. Note that $\mu = \sum_{n=1}^\infty \alpha_n \cdot \varepsilon_{\omega_n}$.

(iii) *Counting measure* on a σ-algebra $\mathfrak{A}$

$$\mu(A) = |A|, \qquad A \in \mathfrak{A}.$$

Uniform distribution in the case $|\Omega| < \infty$ and $\mathfrak{A} = \mathfrak{P}(\Omega)$

$$\mu(A) = \frac{|A|}{|\Omega|}, \qquad A \subset \Omega.$$

(iv) On the algebra $\mathfrak{A} = \{A \subset \Omega : A \text{ finite or } A^c \text{ finite}\}$ let

$$\mu(A) = \begin{cases} 0 & \text{if } |A| < \infty \\ \infty & \text{if } |A| = \infty. \end{cases}$$

Then μ is a content but not a pre-measure in general.

(v) For the semi-algebra of measurable rectangles in Example 3.1 and $A_i \subset \{0,1\}$

$$\mu(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \cdots) = \frac{|A_1 \times \ldots \times A_n|}{|\{0,1\}^n|}$$

is well defined and yields a pre-measure μ with $\mu(\{0,1\}^{\mathbb{N}}) = 1$.

Remark 1. For every content μ on $\mathfrak{A}$ and $A, B \in \mathfrak{A}$

(i) $A \subset B \Rightarrow \mu(A) \leq \mu(B)$ (*monotonicity*),

(ii) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$,

(iii) $A \subset B \wedge \mu(A) < \infty \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A)$,

(iv) $\mu(A) < \infty \wedge \mu(B) < \infty \Rightarrow |\mu(A) - \mu(B)| \leq \mu(A \triangle B)$,

(v) $\mu(A \cup B) \leq \mu(A) + \mu(B)$ (*subadditivity*).

To proof these facts use, for instance, $A \cup B = A \cup (B \cap A^c)$.

Theorem 1. Consider the following properties for a content μ on $\mathfrak{A}$:

(i) μ pre-measure,

(ii) $A_1, A_2, \ldots \in \mathfrak{A} \wedge \bigcup_{i=1}^{\infty} A_i \in \mathfrak{A} \Rightarrow \mu\big(\bigcup_{i=1}^{\infty} A_i\big) \leq \sum_{i=1}^{\infty} \mu(A_i)$ (*σ-subadditivity*),

(iii) $A_1, A_2, \ldots \in \mathfrak{A} \wedge A_n \uparrow A \in \mathfrak{A} \Rightarrow \lim_{n\to\infty} \mu(A_n) = \mu(A)$ (*σ-continuity from below*),

(iv) $A_1, A_2, \ldots \in \mathfrak{A} \wedge A_n \downarrow A \in \mathfrak{A} \wedge \mu(A_1) < \infty \Rightarrow \lim_{n\to\infty} \mu(A_n) = \mu(A)$ (*σ-continuity from above*),

(v) $A_1, A_2, \ldots \in \mathfrak{A} \wedge A_n \downarrow \emptyset \wedge \mu(A_1) < \infty \Rightarrow \lim_{n\to\infty} \mu(A_n) = 0$ (*σ-continuity at* $\emptyset$).

Then

$$\text{(i)} \Leftrightarrow \text{(ii)} \Leftrightarrow \text{(iii)} \Rightarrow \text{(iv)} \Leftrightarrow \text{(v)}.$$

If $\mu(\Omega) < \infty$, then (iii) $\Leftrightarrow$ (iv).

Proof. '(i) $\Rightarrow$ (ii)': Put $B_m = \bigcup_{i=1}^{m} A_i$ and $B_0 = \emptyset$. Then

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{m=1}^{\infty} (B_m \setminus B_{m-1})$$

with pairwise disjoint sets $B_m \setminus B_{m-1} \in \mathfrak{A}$. Clearly $B_m \setminus B_{m-1} \subset A_m$. Hence, by Remark 1.(i),

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{m=1}^{\infty} \mu(B_m \setminus B_{m-1}) \leq \sum_{m=1}^{\infty} \mu(A_m).$$

'(ii) ⇒ (i)': Let $A_1, A_2, \ldots \in \mathfrak{A}$ be pairwise disjoint with $\bigcup_{i=1}^\infty A_i \in \mathfrak{A}$. Then

$$\mu\Big(\bigcup_{i=1}^\infty A_i\Big) \ge \mu\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n \mu(A_i),$$

and therefore

$$\sum_{i=1}^\infty \mu(A_i) \le \mu\Big(\bigcup_{i=1}^\infty A_i\Big).$$

The reverse estimate holds by assumption.

'(i) ⇒ (iii)': Put $A_0 = \emptyset$ and $B_m = A_m \setminus A_{m-1}$. Then

$$\mu\Big(\bigcup_{i=1}^\infty A_i\Big) = \sum_{m=1}^\infty \mu(B_m) = \lim_{n\to\infty} \sum_{m=1}^n \mu(B_m) = \lim_{n\to\infty} \mu\Big(\bigcup_{m=1}^n B_m\Big) = \lim_{n\to\infty} \mu(A_n).$$

'(iii) ⇒ (i)': Let $A_1, A_2, \ldots \in \mathfrak{A}$ be pairwise disjoint with $\bigcup_{i=1}^\infty A_i \in \mathfrak{A}$, and put $B_m = \bigcup_{i=1}^m A_i$. Then $B_m \uparrow \bigcup_{i=1}^\infty A_i$ and

$$\mu\Big(\bigcup_{i=1}^\infty A_i\Big) = \lim_{m\to\infty} \mu(B_m) = \sum_{i=1}^\infty \mu(A_i).$$

'(iv) ⇒ (v)' trivially holds.

'(v) ⇒ (iv)': Use $B_n = A_n \setminus A \downarrow \emptyset$.

'(i)' ⇒ (v)': Note that $\mu(A_1) = \sum_{i=1}^\infty \mu(A_i \setminus A_{i+1})$. Hence

$$0 = \lim_{k\to\infty} \sum_{i=k}^\infty \mu(A_i \setminus A_{i+1}) = \lim_{k\to\infty} \mu(A_k).$$

'(iv) $\wedge\ \mu(\Omega) < \infty$ ⇒ (iii)': Clearly $A_n \uparrow A$ implies $A_n^c \downarrow A^c$. Thus

$$\mu(A) = \mu(\Omega) - \mu(A^c) = \lim_{n\to\infty} (\mu(\Omega) - \mu(A_n^c)) = \lim_{n\to\infty} \mu(A_n).$$

□

Theorem 2 (Extension: semi-algebra ⇝ algebra)**.** For every semi-algebra $\mathfrak{A}$ and every additive mapping $\mu : \mathfrak{A} \to \mathbb{R}_+ \cup \{\infty\}$ with $\mu(\emptyset) = 0$

$$\exists_1\, \widehat{\mu} \text{ content on } \alpha(\mathfrak{A}) : \quad \widehat{\mu}|_{\mathfrak{A}} = \mu.$$

Moreover, if μ is σ-additive then $\widehat{\mu}$ is σ-additive, too.

Proof. We have $\alpha(\mathfrak{A}) = \mathfrak{A}^+$, see Lemma 1.1. Necessarily

$$\widehat{\mu}\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n \mu(A_i) \tag{1}$$

for $A_1, \ldots, A_n \in \mathfrak{A}$ pairwise disjoint. Use (1) to obtain a well-defined extension of μ onto $\alpha(\mathfrak{A})$. It remains to verify that $\widehat{\mu}$ is additive or even σ-additive. □

Example 2. For the semi-algebra $\mathfrak{A}$ in Example 1.(v) $\alpha(\mathfrak{A})$ is the algebra of cylinder sets, and

$$\widehat{\mu}(A \times \Omega_{n+1} \times \cdots) = \frac{|A|}{|\{0,1\}^n|}, \qquad A \subset \{0,1\}^n .$$

Theorem 3 (Extension: algebra $\rightsquigarrow$ σ-algebra, Carathéodory)**.** For every pre-measure μ on an algebra $\mathfrak{A}$

$$\exists\, \mu^* \text{ measure on } \sigma(\mathfrak{A}): \quad \mu^*|_{\mathfrak{A}} = \mu.$$

Proof. Define $\overline{\mu} : \mathfrak{P}(\Omega) \to \mathbb{R}_+ \cup \{\infty\}$ by

$$\overline{\mu}(A) = \inf\Big\{\sum_{i=1}^{\infty} \mu(A_i) : A_i \in \mathfrak{A},\ A \subset \bigcup_{i=1}^{\infty} A_i\Big\}.$$

Then $\overline{\mu}$ is an *outer measure*, i.e., $\overline{\mu}(\emptyset) = 0$ and $\overline{\mu}$ is monotone and σ-subadditive, see Billingsley (1979, Exmp. 11.1) and compare Analysis IV. Actually it suffices to have $\mu \geq 0$ and $\emptyset \in \mathfrak{A}$ with $\mu(\emptyset) = 0$.
We claim that

(i) $\overline{\mu}|_{\mathfrak{A}} = \mu$,

(ii) $\forall\, A \in \mathfrak{A}\ \forall\, B \in \mathfrak{P}(\Omega) : \quad \overline{\mu}(B) = \overline{\mu}(B \cap A) + \overline{\mu}(B \cap A^c)$.

Ad (i): For $A \in \mathfrak{A}$

$$\overline{\mu}(A) \leq \mu(A) + \sum_{i=2}^{\infty} \mu(\emptyset) = \mu(A),$$

and for $A_i \in \mathfrak{A}$ with $A \subset \bigcup_{i=1}^{\infty} A_i$

$$\mu(A) = \mu\Big(\bigcup_{i=1}^{\infty}(A_i \cap A)\Big) \leq \sum_{i=1}^{\infty} \mu(A_i \cap A) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

follows from Theorem 1.(ii).
Ad (ii): '$\leq$' holds due to subadditivity of $\overline{\mu}$, and '$\geq$' is easily verified.
Consider the class

$$\overline{\mathfrak{A}} = \overline{\mathfrak{A}}_{\overline{\mu}} = \{A \in \mathfrak{P}(\Omega) : \forall\, B \in \mathfrak{P}(\Omega) : \overline{\mu}(B) = \overline{\mu}(B \cap A) + \overline{\mu}(B \cap A^c)\}$$

of so-called $\overline{\mu}$-measurable sets.
We claim that

(iii) $\forall\, A_1, A_2 \in \overline{\mathfrak{A}}\ \forall\, B \in \mathfrak{P}(\Omega) : \quad \overline{\mu}(B) = \overline{\mu}(B \cap (A_1 \cap A_2)) + \overline{\mu}(B \cap (A_1 \cap A_2)^c)$.

(iv) $\overline{\mathfrak{A}}$ algebra,

Ad (iii): We have

$$\begin{aligned}\overline{\mu}(B) &= \overline{\mu}(B \cap A_1) + \overline{\mu}(B \cap A_1^c)\\ &= \overline{\mu}(B \cap A_1 \cap A_2) + \overline{\mu}(B \cap A_1 \cap A_2^c) + \overline{\mu}(B \cap A_1^c)\end{aligned}$$

and

$$\overline{\mu}(B \cap (A_1 \cap A_2)^c) = \overline{\mu}(B \cap A_1^c \cup B \cap A_2^c) = \overline{\mu}(B \cap A_2^c \cap A_1) + \overline{\mu}(B \cap A_1^c).$$

Ad (iv): Cleary $\Omega \in \overline{\mathfrak{A}}$, $A \in \overline{\mathfrak{A}} \Rightarrow A^c \in \overline{\mathfrak{A}}$, and $\overline{\mathfrak{A}}$ is closed w.r.t. intersections by (iii). We claim that

(v) $\forall\, A_1, A_2 \in \overline{\mathfrak{A}}$ disjoint $\forall\, B \in \mathfrak{P}(\Omega)$: $\quad \overline{\mu}(B\cap(A_1\cup A_2)) = \overline{\mu}(B\cap A_1)+\overline{\mu}(B\cap A_2)$.

In fact, since $A_1 \cap A_2 = \emptyset$,

$$\overline{\mu}(B \cap (A_1 \cup A_2)) = \overline{\mu}(B \cap A_1) + \overline{\mu}(B \cap A_2 \cap A_1^c) = \overline{\mu}(B \cap A_1) + \overline{\mu}(B \cap A_2).$$

We claim that

(vi) $\forall\, A_1, A_2, \ldots \in \overline{\mathfrak{A}}$ pairwise disjoint

$$\bigcup_{i=1}^{\infty} A_i \in \overline{\mathfrak{A}} \quad \wedge \quad \overline{\mu}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \overline{\mu}(A_i).$$

Let $B \in \mathfrak{P}(\Omega)$. By (iv), (v), and monotonicity of $\overline{\mu}$

$$\begin{aligned}\overline{\mu}(B) &= \overline{\mu}\Big(B \cap \bigcup_{i=1}^{n} A_i\Big) + \overline{\mu}\Big(B \cap \Big(\bigcup_{i=1}^{n} A_i\Big)^c\Big)\\ &\geq \sum_{i=1}^{n} \overline{\mu}(B \cap A_i) + \overline{\mu}\Big(B \cap \Big(\bigcup_{i=1}^{\infty} A_i\Big)^c\Big).\end{aligned}$$

Use σ-subadditivity of $\overline{\mu}$ to get

$$\begin{aligned}\overline{\mu}(B) &\geq \sum_{i=1}^{\infty} \overline{\mu}(B \cap A_i) + \overline{\mu}\Big(B \cap \Big(\bigcup_{i=1}^{\infty} A_i\Big)^c\Big)\\ &\geq \overline{\mu}\Big(B \cap \bigcup_{i=1}^{\infty} A_i\Big) + \overline{\mu}\Big(B \cap \Big(\bigcup_{i=1}^{\infty} A_i\Big)^c\Big)\\ &\geq \overline{\mu}(B).\end{aligned}$$

Hence $\bigcup_{i=1}^{\infty} A_i \in \overline{\mathfrak{A}}$. Take $B = \bigcup_{i=1}^{\infty} A_i$ to obtain σ-additivity of $\overline{\mu}|_{\overline{\mathfrak{A}}}$.

Conclusions:

- $\overline{\mathfrak{A}}$ is a σ-algebra, see (iv), (vi) and Theorem 1.1,
- $\mathfrak{A} \subset \overline{\mathfrak{A}}$ by (ii), hence $\sigma(\mathfrak{A}) \subset \overline{\mathfrak{A}}$.
- $\overline{\mu}|_{\overline{\mathfrak{A}}}$ is a measure with $\overline{\mu}|_{\mathfrak{A}} = \mu$, see (vi) and (i).

Put $\mu^* = \overline{\mu}|_{\sigma(\mathfrak{A})}$. □

Remark 2. The extension from Theorem 3 is non-unique, in general. For instance, put $\Omega = \mathbb{R}$ and

$$f(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \infty & \text{otherwise} \end{cases}, \qquad A \subset \mathbb{R}.$$

Then $\mu = f|_{\mathfrak{A}}$ defines a pre-measure on the semi-algebra $\mathfrak{A} = \mathfrak{I}_1$ of intervals. Now we have

(i) a unique extension of μ to a pre-measure $\widehat{\mu}$ on $\mathfrak{A}^+$, namely $\widehat{\mu} = f|_{\mathfrak{A}^+}$,

(ii) the outer measure $\overline{\mu} = f$,

(iii) $\sigma(\mathfrak{A}) = \sigma(\mathfrak{A}^+) = \mathfrak{B}$.

For the counting measure μ_1 on $\mathfrak{B}$ and for the measure $\mu_2 = f|_{\mathfrak{B}}$ according to the proof of Theorem 3 we have

$$\mu_1 \neq \mu_2 \wedge \mu_1|_{\mathfrak{A}^+} = \mu_2|_{\mathfrak{A}^+}.$$

Definition 3. $\mu : \mathfrak{A} \to \mathbb{R}_+ \cup \{\infty\}$ is called

(i) *σ-finite*, if

$$\exists\, B_1, B_2, \ldots \in \mathfrak{A} \text{ pairwise disjoint}: \quad \Omega = \bigcup_{i=1}^{\infty} B_i \wedge \forall\, i \in \mathbb{N} : \mu(B_i) < \infty,$$

(ii) *finite*, if $\Omega \in \mathfrak{A}$ and $\mu(\Omega) < \infty$.

Theorem 4 (Uniqueness)**.** For measures μ_1, μ_2 on $\mathfrak{A}$ and $\mathfrak{A}_0 \subset \mathfrak{A}$ with

(i) $\sigma(\mathfrak{A}_0) = \mathfrak{A}$ and $\mathfrak{A}_0$ is closed w.r.t. intersections,

(ii) $\mu_1|_{\mathfrak{A}_0}$ is σ-finite,

(iii) $\mu_1|_{\mathfrak{A}_0} = \mu_2|_{\mathfrak{A}_0}$

we have

$$\mu_1 = \mu_2.$$

Proof. Take B_i according to Definition 3, with $\mathfrak{A}_0$ instead of $\mathfrak{A}$, and put

$$\mathfrak{D}_i = \{A \in \mathfrak{A} : \mu_1(A \cap B_i) = \mu_2(A \cap B_i)\}.$$

Obviously, $\mathfrak{D}_i$ is a Dynkin class and $\mathfrak{A}_0 \subset \mathfrak{D}_i$. Theorem 1.2 yields

$$\mathfrak{D}_i \subset \mathfrak{A} = \sigma(\mathfrak{A}_0) = \delta(\mathfrak{A}_0) \subset \mathfrak{D}_i.$$

Thus $\mathfrak{A} = \mathfrak{D}_i$ and for $A \in \mathfrak{A}$,

$$\mu_1(A) = \sum_{i=1}^{\infty} \mu_1(A \cap B_i) = \sum_{i=1}^{\infty} \mu_2(A \cap B_i) = \mu_2(A).$$

□

Corollary 1. For every semi-algebra $\mathfrak{A}$ and every pre-measure μ on $\mathfrak{A}$ that is σ-finite

$$\underset{1}{\exists}\, \mu^* \text{ measure on } \sigma(\mathfrak{A}) : \quad \mu^*|_{\mathfrak{A}} = \mu.$$

Proof. Use Theorems 2, 3, and 4. □

Remark 3. Applications of Corollary 1:

(i) For $\Omega = \mathbb{R}^k$ and the Lebesgue pre-measure λ_k on $\mathfrak{J}_k$ we get the Lebesgue measure on $\mathfrak{B}_k$. Notation for the latter: λ_k.

(ii) In Example 1.(v) there exists a uniquely determined probability measure P on $\bigotimes_{i=1}^{\infty} \mathfrak{P}(\{0,1\})$ such that

$$P(A_1 \times \dots \times A_n \times \{0,1\} \times \dots) = \frac{|A_1 \times \dots \times A_n|}{|\{0,1\}^n|}$$

for $A_1, \dots, A_n \subset \{0,1\}$. We will study the general construction of product measures in Section 8.

For a pre-measure μ on an algebra $\mathfrak{A}$ the Carathéodory construction yields the extensions

$$\big(\Omega,\ \sigma(\mathfrak{A}),\ \overline{\mu}|_{\sigma(\mathfrak{A})}\big), \qquad \big(\Omega,\ \overline{\mathfrak{A}}_{\overline{\mu}},\ \overline{\mu}|_{\overline{\mathfrak{A}}_{\overline{\mu}}}\big). \tag{2}$$

To what extend is $\overline{\mathfrak{A}}_{\overline{\mu}}$ larger than $\sigma(\mathfrak{A})$?

Definition 4. A measure space $(\Omega, \mathfrak{A}, \mu)$ is *complete* if

$$\mathfrak{N}_\mu \subset \mathfrak{A}$$

for

$$\mathfrak{N}_\mu = \{B \in \mathfrak{P}(\Omega) : \exists\, A \in \mathfrak{A} : B \subset A \wedge \mu(A) = 0\}.$$

Theorem 5. For a measure space $(\Omega, \mathfrak{A}, \mu)$ define

$$\mathfrak{A}^\mu = \{A \cup N : A \in \mathfrak{A}\,,\ N \in \mathfrak{N}_\mu\}$$

and

$$\widetilde{\mu}(A \cup N) = \mu(A), \qquad A \in \mathfrak{A},\ N \in \mathfrak{N}_\mu.$$

Then

(i) $\widetilde{\mu}$ is well defined and $(\Omega, \mathfrak{A}^\mu, \widetilde{\mu})$ is a complete measure space with $\widetilde{\mu}|_{\mathfrak{A}} = \mu$, called the *completion of* $(\Omega, \mathfrak{A}, \mu)$,

(ii) for every complete measure space $(\Omega, \check{\mathfrak{A}}, \check{\mu})$ with $\check{\mathfrak{A}} \supset \mathfrak{A}$ and $\check{\mu}|_{\mathfrak{A}} = \mu$ we have $\check{\mathfrak{A}} \supset \mathfrak{A}^\mu$ and $\check{\mu}|_{\mathfrak{A}^\mu} = \widetilde{\mu}$.

Proof. See Gänssler, Stute (1977, p. 34) or Elstrodt (1996, p. 64). □

Remark 4. It is easy to verify that $\big(\Omega,\ \overline{\mathfrak{A}}_{\overline{\mu}},\ \overline{\mu}|_{\overline{\mathfrak{A}}_{\overline{\mu}}}\big)$ in (2) is complete. However, $\big(\Omega,\ \sigma(\mathfrak{A}),\ \overline{\mu}|_{\sigma(\mathfrak{A})}\big)$ is not complete in general, see Example 3 below.

Theorem 6. If μ is a σ-finite pre-measure on an algebra $\mathfrak{A}$, then $\big(\Omega,\ \overline{\mathfrak{A}}_{\overline{\mu}},\ \overline{\mu}|_{\overline{\mathfrak{A}}_{\overline{\mu}}}\big)$ is the completion of $\big(\Omega,\ \sigma(\mathfrak{A}),\ \overline{\mu}|_{\sigma(\mathfrak{A})}\big)$.

Proof. See Elstrodt (1996, p. 64). □

Example 3. Consider the completion $\big(\mathbb{R}^k, \mathfrak{L}_k, \widetilde{\lambda}_k\big)$ of $\big(\mathbb{R}^k, \mathfrak{B}_k, \lambda_k\big)$. Here $\mathfrak{L}_k$ is called the σ-algebra of *Lebesgue measurable sets* and $\widetilde{\lambda}_k$ is called the Lebesgue measure on $\mathfrak{L}_k$. Notation: $\lambda_k = \widetilde{\lambda}_k$. We have

$$\mathfrak{B}_k \subsetneq \mathfrak{L}_k,$$

hence $(\mathbb{R}^k, \mathfrak{B}_k, \lambda_k)$ is not complete.
Proof: Assume $k = 1$ for simplicity. For the Cantor set $C \subset \mathbb{R}$

$$C \in \mathfrak{B}_1 \wedge \lambda_1(C) = 0 \wedge |C| = |\mathbb{R}|.$$

By Theorem 3.4, $|\mathfrak{B}_1| = |\mathbb{R}|$, but

$$|\{0,1\}^{\mathbb{R}}| = |\mathfrak{P}(C)| \le |\mathfrak{L}_k| \le |\{0,1\}^{\mathbb{R}}|.$$

We add that $\mathfrak{L}_k \subsetneq \mathfrak{P}(\mathbb{R}^k)$, see Elstrodt (1996, §III.3).

5 Integration

For the proofs, see Analysis IV or Elstrodt (1996, Kap. VI).
Given: a measure space $(\Omega, \mathfrak{A}, \mu)$. Notation: $\mathfrak{S}_+ = \mathfrak{S}_+(\Omega, \mathfrak{A})$ is the class of non-negative simple functions.

Definition 1. *Integral* of $f \in \mathfrak{S}_+$ w.r.t. μ

$$\int f \, d\mu = \sum_{i=1}^n \alpha_i \cdot \mu(A_i)$$

if

$$f = \sum_{i=1}^n \alpha_i \cdot 1_{A_i}$$

with $\alpha_i \ge 0$ and $A_i \in \mathfrak{A}$. (Note that the integral is well defined.)

Lemma 1. For $f, g \in \mathfrak{S}_+$ and $c \in \mathbb{R}_+$

(i) $\int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu$,

(ii) $\int (cf)\,d\mu = c \cdot \int f\,d\mu$,

(iii) $f \le g \Rightarrow \int f\,d\mu \le \int g\,d\mu$ (*monotonicity*).

Notation: $\overline{\mathfrak{Z}}_+ = \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$ is the class of nonnegative $\mathfrak{A}$-$\overline{\mathfrak{B}}$-measurable functions.

Definition 2. *Integral* of $f \in \overline{\mathfrak{Z}}_+$ w.r.t. μ

$$\int f\,d\mu = \sup\Big\{\int g\,d\mu : g \in \mathfrak{S}_+ \wedge g \le f\Big\}.$$

Theorem 1 (Monotone convergence, Beppo Levi)**.** Let $f_n \in \overline{\mathfrak{Z}}_+$ such that

$$\forall\, n \in \mathbb{N} : f_n \le f_{n+1}.$$

Then

$$\int \sup_n f_n\,d\mu = \sup_n \int f_n\,d\mu.$$

Remark 1. For every $f \in \overline{\mathfrak{Z}}_+$ there exists a sequence of functions $f_n \in \mathfrak{S}_+$ such that $f_n \uparrow f$, see Theorem 2.7.

Example 1. Consider

$$f_n = \frac{1}{n} \cdot 1_{[0,n]}$$

on $(\mathbb{R}, \mathfrak{B}, \lambda_1)$. Then

$$\int f_n\,d\lambda_1 = 1, \qquad \lim_{n\to\infty} f_n = 0.$$

Lemma 2. The conclusions from Lemma 1 remain valid on $\overline{\mathfrak{Z}}_+$.

Theorem 2 (Fatou's Lemma)**.** For every sequence $(f_n)_n$ in $\overline{\mathfrak{Z}}_+$

$$\int \liminf_{n\to\infty} f_n\,d\mu \le \liminf_{n\to\infty} \int f_n\,d\mu.$$

Proof. For $g_n = \inf_{k\ge n} f_k$ we have $g_n \in \overline{\mathfrak{Z}}_+$ and $g_n \uparrow \liminf_n f_n$. By Theorem 1 and Lemma 1.(iii)

$$\int \liminf_n f_n\,d\mu = \lim_{n\to\infty} \int g_n\,d\mu \le \liminf_{n\to\infty} \int f_n\,d\mu.$$

□

Theorem 3. Let $f \in \overline{\mathfrak{Z}}_+$. Then

$$\int f\,d\mu = 0 \Leftrightarrow \mu(\{f > 0\}) = 0.$$

Definition 3. A property Π holds *μ-almost everywhere (μ-a.e., a.e.)*, if

$$\exists\, A \in \mathfrak{A} : \{\omega \in \Omega : \Pi \text{ does not hold for } \omega\} \subset A \wedge \mu(A) = 0.$$

In case of a probability measure we say: *μ-almost surely, μ-a.s., with probability one*.

Notation: $\overline{\mathfrak{Z}} = \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ is the class of $\mathfrak{A}$-$\overline{\mathfrak{B}}$-measurable functions.

Definition 4. $f \in \overline{\mathfrak{Z}}$ *quasi-μ-integrable* if

$$\int f_+ \, d\mu < \infty \quad \vee \quad \int f_- \, d\mu < \infty.$$

In this case: *integral* of f (w.r.t. μ)

$$\int f \, d\mu = \int f_+ \, d\mu - \int f_- \, d\mu.$$

$f \in \overline{\mathfrak{Z}}$ *μ-integrable* if

$$\int f_+ \, d\mu < \infty \quad \wedge \quad \int f_- \, d\mu < \infty.$$

Theorem 4.

(i) f μ-integrable $\Rightarrow \mu(\{|f| = \infty\}) = 0$,

(ii) f μ-integrable $\wedge\ g \in \overline{\mathfrak{Z}} \wedge f = g$ μ-a.e. $\Rightarrow g$ μ-integrable $\wedge \int f \, d\mu = \int g \, d\mu$.

(iii) equivalent properties for $f \in \overline{\mathfrak{Z}}$:

  (a) f μ-integrable,

  (b) $|f|$ μ-integrable,

  (c) $\exists\, g : g$ μ-integrable $\wedge\ |f| \leq g$ μ-a.e.,

(iv) for f and g μ-integrable and $c \in \mathbb{R}$

  (a) $f+g$ well-defined μ-a.e. and μ-integrable with $\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu$,

  (b) $c \cdot f$ μ-integrable with $\int (cf) \, d\mu = c \cdot \int f \, d\mu$,

  (c) $f \leq g$ μ-a.e. $\Rightarrow \int f \, d\mu \leq \int g \, d\mu$.

Remark 2. An outlook. Consider an arbitrary set $\Omega \neq \emptyset$ and a vector space $\mathfrak{F} \subset \mathbb{R}^\Omega$ such that

$$f \in \mathfrak{F} \Rightarrow \big(|f| \in \mathfrak{F} \wedge \inf\{f, 1\} \in \mathfrak{F}\big).$$

A monotone linear mapping $I : \mathfrak{F} \to \mathbb{R}$ such that

$$f, f_1, f_2, \ldots \in \mathfrak{F} \wedge f_n \uparrow f \Rightarrow I(f) = \lim_{n\to\infty} I(f_n)$$

is called an *abstract integral*. Note that

$$I(f) = \int f \, d\mu$$

defines an abstract integral on

$$\mathfrak{F} = \{f \in \mathfrak{Z}(\Omega, \mathfrak{A}) : f \ \mu\text{-integrable}\} = \mathfrak{L}^1(\Omega, \mathfrak{A}, \mu).$$

Daniell-Stone-Theorem: for every abstract integral there exists a uniquely determined measure μ on $\mathfrak{A} = \sigma(\mathfrak{F})$ such that

$$\mathfrak{F} \subset \mathfrak{L}^1(\Omega, \mathfrak{A}, \mu) \wedge \forall f \in \mathfrak{F} : I(f) = \int f \, d\mu.$$

See Bauer (1978, Satz 39.4) or Floret (1981).

Application: *Riesz representation theorem.* Here $\mathfrak{F} = C([0,1])$ and $I : \mathfrak{F} \to \mathbb{R}$ linear and monotone. Then I is an abstract integral, which follows from Dini's Theorem, see Floret (1981, p. 45). Hence there exists a uniquely determined measure μ on $\sigma(\mathfrak{F}) = \mathfrak{B}([0,1])$ such that

$$\forall f \in \mathfrak{F} : I(f) = \int f \, d\mu.$$

Theorem 5 (Dominated convergence, Lebesgue)**.** Assume that

(i) $f_n \in \overline{\mathfrak{Z}}$ for $n \in \mathbb{N}$,

(ii) $\exists \, g$ μ-integrable $\forall \, n \in \mathbb{N} : |f_n| \leq g$ μ-a.e.,

(iii) $f \in \overline{\mathfrak{Z}}$ such that $\lim_{n\to\infty} f_n = f$ μ-a.e.

Then f is μ-integrable and

$$\int f \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu.$$

Example 2. Consider

$$f_n = n \cdot 1_{]0,1/n[}$$

on $(\mathbb{R}, \mathfrak{B}, \lambda_1)$. Then

$$\int f_n \, d\lambda_1 = 1, \qquad \lim_{n\to\infty} f_n = 0.$$

6 $\mathfrak{L}^p$-Spaces

Given: a measure space $(\Omega, \mathfrak{A}, \mu)$ and $1 \leq p < \infty$. Put $\mathfrak{Z} = \mathfrak{Z}(\Omega, \mathfrak{A})$.

Definition 1.

$$\mathfrak{L}^p = \mathfrak{L}^p(\Omega, \mathfrak{A}, \mu) = \Big\{ f \in \mathfrak{Z} : \int |f|^p \, d\mu < \infty \Big\}.$$

In particular, for $p = 1$: *integrable functions* and $\mathfrak{L} = \mathfrak{L}^1$, and for $p = 2$: *square-integrable functions*. Put

$$\|f\|_p = \left(\int |f|^p \, d\mu \right)^{1/p}, \qquad f \in \mathfrak{L}^p.$$

Theorem 1 (Hölder inequality)**.** Let $1 < p, q < \infty$ such that $1/p + 1/q = 1$ and let $f \in \mathfrak{L}^p$, $g \in \mathfrak{L}^q$. Then

$$\int |f \cdot g| \, d\mu \leq \|f\|_p \cdot \|g\|_q.$$

In particular, for $p = q = 2$: *Cauchy-Schwarz inequality.*

Proof. See Analysis IV or Elstrodt (1996, §VI.1) as well as Theorem 5.3. □

Theorem 2. $\mathfrak{L}^p$ is a vector space and $\| \cdot \|_p$ is a semi-norm on $\mathfrak{L}^p$. Furthermore,

$$\|f\|_p = 0 \quad \Leftrightarrow \quad f = 0 \ \mu\text{-a.e.}$$

Proof. See Analysis IV or Elstrodt (1996, §VI.2). □

Definition 2. Let $f, f_n \in \mathfrak{L}^p$ for $n \in \mathbb{N}$. $(f_n)_n$ *converges to f in $\mathfrak{L}^p$ (in mean of order p)* if

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

In particular, for $p = 1$: *convergence in mean*, and for $p = 2$: *mean-square convergence*. Notation:

$$f_n \xrightarrow{\mathfrak{L}^p} f.$$

Remark 1. Let $f, f_n \in \overline{\mathfrak{Z}}$ for $n \in \mathbb{N}$. Recall (define) that $(f_n)_n$ *converges to f μ-a.e.* if

$$\mu(A^c) = 0$$

for

$$A = \left\{\lim_{n\to\infty} f_n = f\right\} = \left\{\limsup_{n\to\infty} f_n = \liminf_{n\to\infty} f_n\right\} \cap \left\{\limsup_{n\to\infty} f_n = f\right\} \in \mathfrak{A}.$$

Notation:

$$f_n \xrightarrow{\mu\text{-a.e.}} f.$$

Lemma 1. Let $f, g, f_n \in \mathfrak{L}^p$ for $n \in \mathbb{N}$ such that $f_n \xrightarrow{\mathfrak{L}^p} f$. Then

$$f_n \xrightarrow{\mathfrak{L}^p} g \quad \Leftrightarrow \quad f = g \ \mu\text{-a.e.}$$

Analogously for convergence almost everywhere.

Proof. For convergence in $\mathfrak{L}^p$: '$\Leftarrow$' follows from Theorem 5.4.(ii). Use

$$\|f - g\|_p \leq \|f - f_n\|_p + \|f_n - g\|_p$$

to verify '$\Rightarrow$'.
For convergence almost everywhere: '$\Leftarrow$' trivially holds. Use

$$\left\{\lim_{n\to\infty} f_n = f\right\} \cap \left\{\lim_{n\to\infty} f_n = g\right\} \subset \{f = g\}$$

to verify '$\Rightarrow$'. □

Theorem 3 (Fischer-Riesz)**.** Consider a sequence $(f_n)_n$ in $\mathfrak{L}^p$. Then

(i) $(f_n)_n$ Cauchy sequence $\Rightarrow \exists\, f \in \mathfrak{L}^p :\ f_n \xrightarrow{\mathfrak{L}^p} f$ (*completeness*),

(ii) $f_n \xrightarrow{\mathfrak{L}^p} f \Rightarrow \exists$ subsequence $(f_{n_k})_k : f_{n_k} \xrightarrow{\mu\text{-a.e.}} f$.

Proof. Ad (i): Consider a Cauchy sequence $(f_n)_n$ and a subsequence $(f_{n_k})_k$ such that

$$\forall\, k \in \mathbb{N}\ \forall\, m \geq n_k : \|f_m - f_{n_k}\|_p \leq 2^{-k}.$$

For

$$g_k = f_{n_{k+1}} - f_{n_k} \in \mathfrak{L}^p$$

we have

$$\Big\|\sum_{\ell=1}^{k} |g_\ell|\Big\|_p \leq \sum_{\ell=1}^{k} \|g_\ell\|_p \leq \sum_{\ell=1}^{k} 2^{-\ell} \leq 1.$$

Put $g = \sum_{\ell=1}^{\infty} |g_\ell| \in \overline{\mathfrak{Z}}_+$. By Theorem 5.1

$$\int g^p \, d\mu = \int \sup_k \Big(\sum_{\ell=1}^{k} |g_\ell|\Big)^p \, d\mu = \sup_k \int \Big(\sum_{\ell=1}^{k} |g_\ell|\Big)^p \, d\mu \leq 1. \tag{1}$$

Thus, in particular, $\sum_{\ell=1}^{\infty} |g_\ell|$ and $\sum_{\ell=1}^{\infty} g_\ell$ converge μ-a.e., see Theorem 5.4.(i). Since

$$f_{n_{k+1}} = \sum_{\ell=1}^{k} g_\ell + f_{n_1},$$

we have

$$f = \lim_{k\to\infty} f_{n_k} \ \mu\text{-a.e.}$$

for some $f \in \mathfrak{Z}$. Furthermore,

$$|f - f_{n_k}| \leq \sum_{\ell=k}^{\infty} |g_\ell| \leq g \ \mu\text{-a.e.},$$

so that, by Theorem 5.5 and (1),

$$\lim_{k\to\infty} \int |f - f_{n_k}|^p \, d\mu = 0.$$

It follows that

$$\lim_{n\to\infty} \|f - f_n\|_p = 0,$$

too. Finally, by Theorem 2, $f \in \mathfrak{L}^p$.
Ad (ii): Assume that

$$f_n \xrightarrow{\mathfrak{L}^p} f.$$

According to the proof of (i) there exists $\widetilde{f} \in \mathfrak{L}^p$ and a subsequence $(f_{n_k})_k$ such that

$$f_{n_k} \xrightarrow{\mu\text{-a.e.}} \widetilde{f} \wedge f_{n_k} \xrightarrow{\mathfrak{L}^p} \widetilde{f}.$$

Use Lemma 1. □

Example 1. Let $(\Omega, \mathfrak{A}, \mu) = \big([0,1], \mathfrak{B}([0,1]), \lambda_1|_{\mathfrak{B}([0,1])}\big)$. (By Remark 1.7.(ii) we have $\mathfrak{B}([0,1]) \subset \mathfrak{B}_1$). Define

$$\begin{gathered} A_1 = [0,1] \\ A_2 = [0,1/2], \quad A_3 = [1/2,1] \\ A_4 = [0,1/3], \quad A_5 = [1/3,2/3], \quad A_6 = [2/3,1] \\ \text{etc.} \end{gathered}$$

Put $f_n = 1_{A_n}$. Then

$$\lim_{n\to\infty} \|f_n - 0\|_p = \lim_{n\to\infty} \|f_n\|_p = 0 \tag{2}$$

but

$$\{(f_n)_n \text{ converges}\} = \emptyset.$$

Remark 2. Define

$$\mathfrak{L}^\infty = \mathfrak{L}^\infty(\Omega, \mathfrak{A}, P) = \{f \in \mathfrak{Z} : \exists\, c \in \mathbb{R}_+ : |f| \le c \ \mu\text{-a.e.}\}$$

and

$$\|f\|_\infty = \inf\{c \in \mathbb{R}_+ : |f| \le c \ \mu\text{-a.e.}\}, \qquad f \in \mathfrak{L}^\infty.$$

$f \in \mathfrak{L}^\infty$ is called *essentially bounded* and $\|f\|_\infty$ is called the *essential supremum of* $|f|$. Use Theorem 4.1.(iii) to verify that

$$|f| \le \|f\|_\infty \ \mu\text{-a.e.}$$

The definitions and results of this section, except (2), extend to the case $p = \infty$, where $q = 1$ in Theorem 1. In Theorem 3.(ii) we even have $f_n \xrightarrow{\mathfrak{L}^\infty} f \Rightarrow f_n \xrightarrow{\mu\text{-a.e.}} f$.

Remark 3. Put

$$\mathfrak{N}^p = \{f \in \mathfrak{L}^p : f = 0 \ \mu\text{-a.e.}\}.$$

Then the quotient space $L^p = \mathfrak{L}^p/\mathfrak{N}^p$ is a Banach space. In particular, for $p = 2$, L^2 is a Hilbert space, with semi-inner product on $\mathfrak{L}^2$ given by

$$\langle f, g\rangle = \int f \cdot g \, d\mu, \qquad f, g \in \mathfrak{L}^2.$$

Theorem 4. If μ is finite and $1 \le p < q \le \infty$ then

$$\mathfrak{L}^q \subset \mathfrak{L}^p$$

and

$$\|f\|_p \le \mu(\Omega)^{1/p-1/q} \cdot \|f\|_q, \qquad f \in \mathfrak{L}^q.$$

Proof. The result trivially holds for $q = \infty$. In the sequel, $q < \infty$. Use $|f|^p \le 1 + |f|^q$ and Theorem 5.4.(iii) to obtain $\mathfrak{L}^q \subset \mathfrak{L}^p$. Put $r = q/p$ and define s by $1/r + 1/s = 1$. Theorem 1 yields

$$\int |f|^p \, d\mu \le \left(\int |f|^{p\cdot r} \, d\mu\right)^{1/r} \cdot \big(\mu(\Omega)\big)^{1/s}.$$

□

Example 2. Let $1 \le p < q \le \infty$. With respect to the counting measure on $\mathfrak{P}(\mathbb{N})$, $\mathfrak{L}^p \subset \mathfrak{L}^q$. With respect to the Lebesgue measure on $\mathfrak{B}_k$ neither $\mathfrak{L}^q \subset \mathfrak{L}^p$ nor $\mathfrak{L}^p \subset \mathfrak{L}^q$.

7 The Radon-Nikodym-Theorem

Given: a measure space $(\Omega, \mathfrak{A}, \mu)$. Put $\overline{\mathfrak{Z}}_+ = \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$.

Definition 1. For f (quasi-)μ-integrable and $A \in \mathfrak{A}$, the *integral of f over A* is

$$\int_A f \, d\mu = \int 1_A \cdot f \, d\mu.$$

(Note: $|1_A \cdot f| \leq |f|$.)

Theorem 1. Let $f \in \overline{\mathfrak{Z}}_+$ and put

$$\nu(A) = \int_A f \, d\mu, \qquad A \in \mathfrak{A}.$$

Then ν is a measure on $\mathfrak{A}$.

Proof. Clearly $\nu(\emptyset) = 0$ and $\nu \geq 0$. For $A_1, A_2, \ldots \in \mathfrak{A}$ pairwise disjoint

$$\begin{aligned}\nu\Big(\bigcup_{i=1}^{\infty} A_i\Big) &= \int \sum_{i=1}^{\infty} 1_{A_i} \cdot f \, d\mu = \int \lim_{n\to\infty} \Big(\sum_{i=1}^{n} 1_{A_i} \cdot f\Big) \, d\mu \\ &= \lim_{n\to\infty} \int \sum_{i=1}^{n} 1_{A_i} \cdot f \, d\mu = \sum_{i=1}^{\infty} \int 1_{A_i} \cdot f \, d\mu \\ &= \sum_{i=1}^{\infty} \nu(A_i)\end{aligned}$$

follows from Theorem 5.1. □

Definition 2. The mapping ν in Theorem 1 is called *measure with μ-density f*. Notation: $\nu = f \cdot \mu$. If $\int f \, d\mu = 1$ then f is called *probability density*.

Example 1. The introductory examples of probability spaces were defined by means of probability densities.

(i) Let $(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}^k, \mathfrak{B}_k, \lambda_k)$. For

$$f(x) = (2\pi)^{-k/2} \cdot \exp\left(-\tfrac{1}{2} \textstyle\sum_{i=1}^{k} x_i^2\right)$$

we get the *k-dimensional standard normal distribution ν*.

For $B \in \mathfrak{B}_k$ such that $0 < \lambda_k(B) < \infty$ and

$$f = \frac{1}{\lambda_k(B)} \cdot 1_B$$

we get the *uniform distribution on B*.

(ii) Suppose that Ω is countable, $\mathfrak{A} = \mathfrak{P}(\Omega)$, and μ is the counting measure on $\mathfrak{A}$. Take $f : \Omega \to \mathbb{R}_+ \cup \{\infty\}$ and use Theorem 5.1 to obtain

$$\forall A \in \mathfrak{A} : \nu(A) = \int_A f \, d\mu = \sum_{\omega \in A} f(\omega). \tag{1}$$

Conversely, for any measure ν on $\mathfrak{A}$ put $f(\omega) = \nu(\{\omega\})$. Then we have (1).

Theorem 2. Let $\nu = f \cdot \mu$ with $f \in \overline{\mathfrak{Z}}_+$ and $g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$. Then

$$g \text{ (quasi)-}\nu\text{-integrable} \quad \Leftrightarrow \quad g \cdot f \text{ (quasi)-}\mu\text{-integrable},$$

in which case

$$\int g \, d\nu = \int g \cdot f \, d\mu$$

Proof. First, assume that $g = 1_A$ with $A \in \mathfrak{A}$. Then the statements hold by definition. For $g \in \mathfrak{S}_+(\Omega, \mathfrak{A})$ we now use linearity of the integral. For $g \in \overline{\mathfrak{Z}}_+$ we take a sequence $(g_n)_{n \in \mathbb{N}}$ in $\mathfrak{S}_+(\Omega, \mathfrak{A})$ such that $g_n \uparrow g$. Then $g_n \cdot f \in \overline{\mathfrak{Z}}_+$ and $g_n \cdot f \uparrow g \cdot f$. Hence, by Theorem 5.1 and the previous part of the proof

$$\int g \, d\nu = \lim_{n \to \infty} \int g_n \, d\nu = \lim_{n \to \infty} \int g_n \cdot f \, d\mu = \int g \cdot f \, d\mu.$$

Finally, for $g \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ we already know that

$$\int g^{\pm} d\nu = \int g^{\pm} \cdot f \, d\mu = \int (g \cdot f)^{\pm} \, d\mu.$$

Use linearity of the integral. □

Remark 1.

$$f, g \in \overline{\mathfrak{Z}}_+ \wedge f = g \ \mu\text{-a.e.} \quad \Rightarrow \quad f \cdot \mu = g \cdot \mu.$$

Theorem 3 (Uniqueness of densities)**.** Let $f, g \in \overline{\mathfrak{Z}}_+$ such that $f \cdot \mu = g \cdot \mu$. Then

(i) f μ-integrable $\Rightarrow f = g$ μ-a.e.,

(ii) μ σ-finite $\Rightarrow f = g$ μ-a.e.

Proof. Ad (i): It suffices to verify that

$$f, g \ \mu\text{-integrable} \wedge \left(\forall A \in \mathfrak{A} : \int_A f \, d\mu \leq \int_A g \, d\mu \right) \quad \Rightarrow \quad f \leq g \ \mu\text{-a.e.}$$

To this end, take $A = \{f > g\}$. By assumption,

$$-\infty < \int_A f \, d\mu \leq \int_A g \, d\mu < \infty$$

and therefore $\int_A (f - g) \, d\mu \leq 0$. However,

$$1_A \cdot (f - g) \geq 0,$$

hence $\int_A (f-g)\,d\mu \geq 0$. Thus

$$\int 1_A \cdot (f-g)\,d\mu = 0.$$

Theorem 5.3 implies $1_A \cdot (f-g) = 0$ μ-a.e., and by definition of A we get $\mu(A) = 0$.
Ad (ii): see Elstrodt (1996, p. 141). □

Remark 2. Let $(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}^k, \mathfrak{B}_k, \lambda_k)$ and $x \in \mathbb{R}^k$. There is no density $f \in \overline{\mathfrak{Z}}_+$ w.r.t. λ_k such that $\varepsilon_x = f \cdot \lambda_k$. This follows from $\varepsilon_x(\{x\}) = 1$ and

$$(f \cdot \lambda_k)(\{x\}) = \int_{\{x\}} f\,d\lambda_k = 0.$$

Definition 3. A measure ν on $\mathfrak{A}$ is *absolutely continuous w.r.t.* μ if

$$\forall A \in \mathfrak{A} : \mu(A) = 0 \Rightarrow \nu(A) = 0.$$

Notation: $\nu \ll \mu$.

Remark 3.

(i) $\nu = f \cdot \mu \Rightarrow \nu \ll \mu$.

(ii) In Remark 2 neither $\varepsilon_x \ll \lambda_1$ nor $\lambda_1 \ll \varepsilon_x$.

(iii) Let μ denote the counting measure on $\mathfrak{A}$. Then $\nu \ll \mu$ for every measure ν on $\mathfrak{A}$.

(iv) Let μ denote the counting measure on $\mathfrak{B}_1$. Then there is no density $f \in \overline{\mathfrak{Z}}_+$ such that $\lambda_1 = f \cdot \mu$.

Lemma 1. Let $f_n \xrightarrow{\mathfrak{L}^p} f$ and $A \in \mathfrak{A}$. If $p = 1$ or $\mu(A) < \infty$ then

$$\int_A f_n\,d\mu \to \int_A f\,d\mu.$$

Proof. Übung 5.4. See also L^p and its dual space in Elstrodt (1996, §VII.3), e.g. □

Theorem 4 (Radon, Nikodym)**.** For every σ-finite measure μ and every measure ν on $\mathfrak{A}$ we have

$$\nu \ll \mu \quad \Rightarrow \quad \exists\, f \in \overline{\mathfrak{Z}}_+ : \nu = f \cdot \mu.$$

Proof. See Elstrodt (1996, §VII.2).
Here we consider the particular case

$$\forall A \in \mathfrak{A} : \nu(A) \leq \mu(A) \wedge \mu(\Omega) < \infty.$$

A class $\mathfrak{U} = \{A_1, \dots, A_n\}$ is called a (finite measurable) partition of Ω if $A_1, \dots, A_n \in \mathfrak{A}$ are pairwise disjoint and $\bigcup_{i=1}^n A_i = \Omega$. The set of all partitions is partially ordered by

$$\mathfrak{U} \sqsubset \mathfrak{V} \quad \text{if} \quad \forall A \in \mathfrak{U}\ \exists B \in \mathfrak{V} : A \subset B.$$

The infimum of two partitions is given by

$$\mathfrak{U} \wedge \mathfrak{V} = \{A \cap B : A \in \mathfrak{U}, B \in \mathfrak{V}\}.$$

For any partition $\mathfrak{U}$ we define

$$f_{\mathfrak{U}} = \sum_{A \in \mathfrak{U}} \alpha_A \cdot 1_A$$

with

$$\alpha_A = \begin{cases} \nu(A)/\mu(A) & \text{if } \mu(A) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Clearly $f_{\mathfrak{U}} \in \mathfrak{S}_+(\Omega, \sigma(\mathfrak{U})) \subset \mathfrak{S}_+(\Omega, \mathfrak{A})$, $\sigma(\mathfrak{U}) = \mathfrak{U}^+ \cup \{\emptyset\}$, and

$$\forall\, A \in \sigma(\mathfrak{U}) : \nu(A) = \int_A f_{\mathfrak{U}}\, d\mu.$$

(Thus we have $\nu|_{\sigma(\mathfrak{U})} = f_{\mathfrak{U}} \cdot \mu|_{\sigma(\mathfrak{U})}$.) Let $\mathfrak{U} \sqsubset \mathfrak{V}$ and $A \in \mathfrak{V}$. Then

$$\nu(A) = \int_A f_{\mathfrak{V}}\, d\mu = \int_A f_{\mathfrak{U}}\, d\mu,$$

since $A \in \sigma(\mathfrak{U})$. Hence

$$\int_A f_{\mathfrak{V}}^2\, d\mu = \int_A f_{\mathfrak{V}} \cdot f_{\mathfrak{U}}\, d\mu,$$

since $f_{\mathfrak{V}}|_A$ is constant, and therefore

$$0 \leq \int (f_{\mathfrak{U}} - f_{\mathfrak{V}})^2\, d\mu = \int f_{\mathfrak{U}}^2\, d\mu - \int f_{\mathfrak{V}}^2\, d\mu. \tag{2}$$

Put

$$\beta = \sup \left\{ \int f_{\mathfrak{U}}^2\, d\mu : \mathfrak{U} \text{ partition} \right\},$$

and note that $0 \leq \beta \leq \mu(\Omega) < \infty$, since $f_{\mathfrak{U}} \leq 1$. Consider a sequence of functions $f_n = f_{\mathfrak{U}_n}$ such that

$$\lim_{n\to\infty} \int f_n^2\, d\mu = \beta.$$

Due to (2) we may assume that $\mathfrak{U}_{n+1} \sqsubset \mathfrak{U}_n$. Then, by (2), $(f_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathfrak{L}^2$, so that there exists $f \in \mathfrak{L}^2$ with

$$\lim_{n\to\infty} \|f_n - f\|_2 = 0 \quad \wedge \quad 0 \leq f \leq 1 \ \mu\text{-a.e.},$$

see Theorem 6.3.

We claim that $\nu = f \cdot \mu$. Let $A \in \mathfrak{A}$. Put

$$\widetilde{\mathfrak{U}}_n = \mathfrak{U}_n \wedge \{A, A^c\}$$

and

$$\widetilde{f}_n = f_{\widetilde{\mathfrak{U}}_n}.$$

Then

$$\nu(A) = \int_A \widetilde{f}_n\, d\mu = \int_A f_n\, d\mu + \int_A (\widetilde{f}_n - f_n)\, d\mu,$$

and (2) yields $\lim_{n\to\infty} \|\widetilde{f}_n - f_n\|_2 = 0$. It remains to apply Lemma 1. $\square$

8 Kernels and Product Measures

Given: measurable spaces $(\Omega_1, \mathfrak{A}_1)$ and $(\Omega_2, \mathfrak{A}_2)$.
Motivation: two-stage experiment. Output $\omega_1 \in \Omega_1$ of the first stage determines probabilistic model for the second stage.

Example 1. Choose one out of n coins and throw it once. Parameters $a_1, \dots, a_n \geq 0$ such that $\sum_{i=1}^n a_i = 1$ and $b_1, \dots, b_n \in [0, 1]$.
Let

$$\Omega_1 = \{1, \dots, n\}, \qquad \mathfrak{A}_1 = \mathfrak{P}(\Omega_1)$$

and define

$$\mu(\{i\}) = a_i, \qquad i \in \Omega_1,$$

to be the probability of choosing the i-th coin. Moreover, let

$$\Omega_2 = \{\mathrm{H}, \mathrm{T}\}, \qquad \mathfrak{A}_2 = \mathfrak{P}(\Omega_2)$$

and define

$$K(i, \{\mathrm{H}\}) = b_i$$

to be the probability for obtaining H when throwing the i-th coin. Thus, for $A_2 \in \mathfrak{A}_2$,

$$K(i, A_2) = b_i \cdot \varepsilon_{\mathrm{H}}(A_2) + (1 - b_i) \cdot \varepsilon_{\mathrm{T}}(A_2).$$

Definition 1. $K : \Omega_1 \times \mathfrak{A}_2 \to \overline{\mathbb{R}}$ is a *kernel (from* $(\Omega_1, \mathfrak{A}_1)$ *to* $(\Omega_2, \mathfrak{A}_2)$*)*, if

(i) $K(\omega_1, \cdot)$ is a measure on $\mathfrak{A}_2$ for every $\omega_1 \in \Omega_1$,

(ii) $K(\cdot, A_2)$ is $\mathfrak{A}_1$-$\overline{\mathfrak{B}}$-measurable for every $A_2 \in \mathfrak{A}_2$.

K is a *Markov (transition) kernel*, if, additionally, $K(\omega_1, \Omega_2) = 1$ for every $\omega_1 \in \Omega_1$.
K is a *σ-finite* kernel if, additionally,

$$\exists\, A_{2,1}, A_{2,2}, \ldots \in \mathfrak{A}_2 \text{ pairwise disjoint}:$$
$$\Omega_2 = \bigcup_{i=1}^{\infty} A_{2,i} \quad \wedge \quad \forall\, i \in \mathbb{N}: \sup_{\omega_1 \in \Omega_1} K(\omega_1, A_{2,i}) < \infty.$$

Example 2. Extremal cases, non-disjoint.

(i) Model for the second stage not influenced by output of the first stage, i.e., for a (probability) measure ν on $\mathfrak{A}_2$

$$\forall\, \omega_1 \in \Omega_1 : \quad K(\omega_1, \cdot) = \nu.$$

In Example 1 this means $b_1 = \cdots = b_n$.

(ii) Output of the first stage determines the output of the second stage, i.e., for a $\mathfrak{A}_1$-$\mathfrak{A}_2$-measurable mapping $f : \Omega_1 \to \Omega_2$

$$\forall\, \omega_1 \in \Omega_1 : \quad K(\omega_1, \cdot) = \varepsilon_{f(\omega_1)}.$$

In Example 1 this means $b_1, \dots, b_n \in \{0, 1\}$.

Notation: $\int f \, d\mu = \int_\Omega f(\omega) \, \mu(d\omega)$.

Given: a (probability) measure μ on $\mathfrak{A}_1$ and a (Markov) kernel K from $(\Omega_1, \mathfrak{A}_1)$ to $(\Omega_2, \mathfrak{A}_2)$. Question: stochastic model $(\Omega, \mathfrak{A}, P)$ for a compound experiment? Reasonable, and assumed in the sequel,

$$\Omega = \Omega_1 \times \Omega_2, \qquad \mathfrak{A} = \mathfrak{A}_1 \otimes \mathfrak{A}_2.$$

Question: How to define P?

Example 3. In Example 1, a reasonable requirement for P is

$$P(\{i\} \times \Omega_2) = a_i, \qquad P(\{i\} \times \{\mathrm{H}\}) = a_i \cdot b_i$$

for every $i \in \Omega_1$. Consequently, for $A_2 \subset \Omega_2$

$$P(\{i\} \times A_2) = K(i, A_2) \cdot a_i$$

and for $A \subset \Omega$

$$\begin{aligned} P(A) &= \sum_{i=1}^n P(\{(\omega_1, \omega_2) \in A : \omega_1 = i\}) = \sum_{i=1}^n P(\{i\} \times \{\omega_2 \in \Omega_2 : (i, \omega_2) \in A\}) \\ &= \sum_{i=1}^n K(i, \{\omega_2 \in \Omega_2 : (i, \omega_2) \in A\}) \cdot a_i \\ &= \int_{\Omega_1} K(i, \{\omega_2 \in \Omega_2 : (i, \omega_2) \in A\}) \, \mu(di). \end{aligned}$$

May we generally use the right-hand side integral for the definition of P?

Lemma 1. Let $f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$. Then, for $\omega_1 \in \Omega_1$, the ω_1-*section*

$$f(\omega_1, \cdot) : \Omega_2 \to \overline{\mathbb{R}}$$

of f is $\mathfrak{A}_2$-$\overline{\mathfrak{B}}$-measurable, and for $\omega_2 \in \Omega_2$ the ω_2-*section*

$$f(\cdot, \omega_2) : \Omega_1 \to \overline{\mathbb{R}}$$

of f is $\mathfrak{A}_1$-$\overline{\mathfrak{B}}$-measurable.

Proof. In the case of an ω_1-section. Fix $\omega_1 \in \Omega_1$. Then $\Omega_2 \to \Omega_1 \times \Omega_2 : \ \omega_2 \mapsto (\omega_1, \omega_2)$ is $\mathfrak{A}_2$-$\mathfrak{A}$-measurable due to Corollary 3.1.(i). Apply Theorem 2.1. □

Remark 1. In particular, for $A \in \mathfrak{A}$ and $f = 1_A$

$$f(\omega_1, \cdot) = 1_A(\omega_1, \cdot) = 1_{A(\omega_1)}$$

where[1]

$$A(\omega_1) = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in A\}$$

[1] poor notation

is the ω_1-*section* of A. By Lemma 1

$$\forall \omega_1 \in \Omega_1 : \quad A(\omega_1) \in \mathfrak{A}_2.$$

Analogously for the ω_2-*section*

$$A(\omega_2) = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in A\}$$

of A.

Given:

- a σ-finite kernel K from $(\Omega_1, \mathfrak{A}_1)$ to $(\Omega_2, \mathfrak{A}_2)$,
- a σ-finite measure μ on $\mathfrak{A}_1$.

Lemma 2. Let $f \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$. Then

$$g : \ \Omega_1 \to \mathbb{R}_+ \cup \{\infty\}, \quad \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)$$

is $\mathfrak{A}_1$-$\mathfrak{B}([0, \infty])$-measurable.

Proof. First we additionally assume

$$\forall \omega_1 \in \Omega_1 : K(\omega_1, \Omega_2) < \infty. \tag{1}$$

Let $\mathfrak{F}$ denote the set of all functions $f \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$ with the measurability property as claimed. We show that

$$\forall A_1 \in \mathfrak{A}_1,\ A_2 \in \mathfrak{A}_2 : \quad 1_{A_1 \times A_2} \in \mathfrak{F}. \tag{2}$$

Indeed,

$$\int_{\Omega_2} 1_{A_1 \times A_2}(\omega_1, \omega_2) K(\omega_1, d\omega_2) = 1_{A_1}(\omega_1) K(\omega_1, A_2).$$

Furthermore, we show that

$$\forall A \in \mathfrak{A} : \quad 1_A \in \mathfrak{F}. \tag{3}$$

To this end let

$$\mathfrak{D} = \{A \in \mathfrak{A} : 1_A \in \mathfrak{F}\}$$

and

$$\mathfrak{E} = \{A_1 \times A_2 : A_1 \in \mathfrak{A}_1 \wedge A_2 \in \mathfrak{A}_2\}.$$

Then $\mathfrak{E} \subset \mathfrak{D}$ by (2), $\mathfrak{E}$ is closed w.r.t. intersections, and $\sigma(\mathfrak{E}) = \mathfrak{A}$. From (1) it easily follows that $\mathfrak{D}$ is a Dynkin class. Hence Theorem 1.2 yields

$$\mathfrak{A} = \sigma(\mathfrak{E}) = \delta(\mathfrak{E}) \subset \mathfrak{D} \subset \mathfrak{A},$$

which implies (3). From Lemma 5.2 and Theorem 2.6 we get

$$f_1, f_2 \in \mathfrak{F} \wedge \alpha \in \mathbb{R}_+ \quad \Rightarrow \quad \alpha f_1 + f_2 \in \mathfrak{F}. \tag{4}$$

Finally, Theorem 5.1 and Theorem 2.5.(iii) imply that

$$f_n \in \mathfrak{F} \wedge f_n \uparrow f \quad \Rightarrow \quad f \in \mathfrak{F}. \tag{5}$$

Use Theorem 2.7 together with (3)–(5) to conclude that $\mathfrak{F} = \overline{\mathfrak{Z}}_+$.

In the general case we take $A_{2,1}, A_{2,2}, \ldots \in \mathfrak{A}_2$ pairwise disjoint such that

$$\bigcup_{i=1}^{\infty} A_{2,i} = \Omega_2 \quad \wedge \quad \forall\, i \in \mathbb{N} : \sup_{\omega_1 \in \Omega_1} K(\omega_1, A_{2,i}) < \infty,$$

and we define

$$K_i(\omega_1, \cdot) = K(\omega_1, \cdot \cap A_{2,i}) = 1_{A_{2,i}} \cdot K(\omega_1, \cdot).$$

Then, using Theorems 5.1 and 7.2,

$$\begin{aligned} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) &= \sum_{i=1}^{\infty} \int_{\Omega_2} 1_{A_{2,i}}(\omega_2)\, f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2) \\ &= \sum_{i=1}^{\infty} \int_{\Omega_2} f(\omega_1, \omega_2)\, K_i(\omega_1, d\omega_2). \end{aligned}$$

Since $K_i(\omega_1, \Omega_2) < \infty$ for every $\omega_1 \in \Omega_1$, we conclude that $\int_{\Omega_2} f(\cdot, \omega_2)\, K_i(\cdot, d\omega_2)$ is $\mathfrak{A}_1$-$\mathfrak{B}([0,\infty])$-measurable. Apply Theorems 2.5 and 2.6. □

Theorem 1.

$$\begin{aligned} &\underset{1}{\exists} \text{ measure } \mu \times K \text{ on } \mathfrak{A} \;\; \forall\, A_1 \in \mathfrak{A}_1 \;\; \forall\, A_2 \in \mathfrak{A}_2 : \\ &\qquad \mu \times K(A_1 \times A_2) = \textstyle\int_{A_1} K(\omega_1, A_2)\, \mu(d\omega_1). \end{aligned} \tag{6}$$

Moreover, $\mu \times K$ is σ-finite, and

$$\forall\, A \in \mathfrak{A} : \quad \mu \times K(A) = \int_{\Omega_1} K(\omega_1, A(\omega_1))\, \mu(d\omega_1). \tag{7}$$

If μ is a probability measure and K is a Markov kernel then $\mu \times K$ is a probability measure, too.

Proof. 'Existence': For $A \in \mathfrak{A}$ and $\omega_1 \in \Omega_1$

$$K(\omega_1, A(\omega_1)) = \int_{\Omega_2} 1_{A(\omega_1)}(\omega_2)\, K(\omega_1, d\omega_2) = \int_{\Omega_2} 1_A(\omega_1, \omega_2)\, K(\omega_1, d\omega_2).$$

According to Lemma 8.2 $\mu \times K$ is well-defined via (7). Using Theorem 5.1, it is easy to verify that $\mu \times K$ is a measure on $\mathfrak{A}$.

For $A_1 \in \mathfrak{A}_1$ and $A_2 \in \mathfrak{A}_2$

$$K(\omega_1, (A_1 \times A_2)(\omega_1)) = \begin{cases} K(\omega_1, A_2) & \text{if } \omega_1 \in A_1 \\ 0 & \text{otherwise.} \end{cases}$$

Hence $\mu \times K$ satisfies (6).

By assumption we have $A_{1,1}, A_{1,2}, \ldots \in \mathfrak{A}_1$ pairwise disjoint such that

$$\bigcup_{i=1}^{\infty} A_{1,i} = \Omega_1 \quad \wedge \quad \forall\, i \in \mathbb{N} : \mu(A_{1,i}) < \infty$$

and $A_{2,1}, A_{2,2}, \ldots \in \mathfrak{A}_2$ pairwise disjoint such that

$$\bigcup_{j=1}^{\infty} A_{2,j} = \Omega_2 \quad \wedge \quad \forall\, j \in \mathbb{N} : \sup_{\omega_1 \in \Omega_1} K(\omega_1, A_{2,j}) < \infty.$$

Consider the sets $A_{1,i} \times A_{2,j}$ with $i, j \in \mathbb{N}$ and note that

$$\begin{aligned}(\mu \times K)(A_{1,i} \times A_{2,j}) &= \int_{A_{1,i}} K(\omega_1, A_{2,j})\, \mu(d\omega_1) \\ &\le \sup_{\omega_1 \in \Omega_1} K(\omega_1, A_{2,j})\, \mu(A_{1,i}) < \infty,\end{aligned}$$

to conclude that $\mu \times K$ ist σ-finite.
'Uniqueness': Apply Theorem 4.4 with $\mathfrak{A}_0 = \{A_1 \times A_2 : A_i \in \mathfrak{A}_i\}$. □

Example 4. In Example 3 we have $P = \mu \times K$.

Remark 2. Particular case of Theorem 1 with

$$\mu = \mu_1, \qquad \forall\, \omega_1 \in \Omega_1 : \; K(\omega_1, \cdot) = \mu_2$$

for σ-finite measures μ_i on $(\Omega_i, \mathfrak{A}_i)$:

$$\begin{aligned}&\underset{1}{\exists}\, \text{measure } \mu_1 \times \mu_2 \text{ on } \mathfrak{A} \;\; \forall\, A_1 \in \mathfrak{A}_1 \;\; \forall\, A_2 \in \mathfrak{A}_2 : \\ &\quad \mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2).\end{aligned} \tag{8}$$

Moreover, $\mu_1 \times \mu_2$ is σ-finite and satisfies

$$\forall\, A \in \mathfrak{A} : \quad \mu_1 \times \mu_2(A) = \int_{\Omega_1} \mu_2(A(\omega_1))\, \mu(d\omega_1). \tag{9}$$

We add that σ-finiteness is used for the definition (9) and the uniqueness in (8). In general, we only have existence of a measure $\mu_1 \times \mu_2$ with (8). See Elstrodt (1996, §V.1).

Definition 2. $\mu = \mu_1 \times \mu_2$ is called the *product measure* corresponding to μ_1 and μ_2, and $(\Omega, \mathfrak{A}, \mu)$ is called the *product measure space* corresponding to $(\Omega_1, \mathfrak{A}_1, \mu_1)$ and $(\Omega_2, \mathfrak{A}_2, \mu_2)$.

Example 5.

(i) In Example 3 with $b = b_1 = \cdots = b_n$ and $\nu = b \cdot \varepsilon_{\mathrm{H}} + (1 - b) \cdot \varepsilon_{\mathrm{T}}$ we have $P = \mu \times \nu$.

(ii) For countable spaces Ω_i and σ-algebras $\mathfrak{A}_i = \mathfrak{P}(\Omega_i)$ we get

$$\mu_1 \times \mu_2(A) = \sum_{\omega_1 \in \Omega_1} \mu_2(A(\omega_1)) \cdot \mu_1(\{\omega_1\}), \qquad A \subset \Omega.$$

In particular, for uniform distributions μ_i on finite spaces, $\mu_1 \times \mu_2$ is the uniform distribution on Ω. Cf. Example 3.1 in the case $n = 2$.

(iii) The multi-dimensional Lebesgue measure is a product measure. Namely, for $k, \ell \in \mathbb{N}$ and $A_1 \in \mathfrak{I}_k$, $A_2 \in \mathfrak{I}_\ell$ we have

$$\lambda_{k+\ell}(A_1 \times A_2) = \lambda_k(A_1) \cdot \lambda_\ell(A_2) = \lambda_k \times \lambda_\ell(A_1 \times A_2),$$

see Example 4.1.(i). Corollary 4.1 yields

$$\lambda_{k+\ell} = \lambda_k \times \lambda_\ell.$$

From (9) we get

$$\lambda_{k+\ell}(A) = \int_{\mathbb{R}^k} \lambda_\ell(A(\omega_1))\, \lambda_k(d\omega_1), \qquad A \in \mathfrak{B}_{k+\ell},$$

cf. *Cavalieri's Principle.*

Theorem 2 (Fubini's Theorem)**.**

(i) For $f \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$

$$\int_\Omega f\, d(\mu \times K) = \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu(d\omega_1).$$

(ii) For f $(\mu \times K)$-integrable and

$$A_1 = \{\omega_1 \in \Omega_1 : f(\omega_1, \cdot)\ K(\omega_1, \cdot)\text{-integrable}\}$$

we have

(a) $A_1 \in \mathfrak{A}_1$ and $\mu(A_1^c) = 0$,

(b) $A_1 \to \mathbb{R} : \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \cdot)\, dK(\omega_1, \cdot)$ is integrable w.r.t. $\mu|_{A_1 \cap \mathfrak{A}_1}$,

(c)
$$\int_\Omega f\, d(\mu \times K) = \int_{A_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu|_{A_1 \cap \mathfrak{A}_1}(d\omega_1).$$

Proof. Ad (i): algebraic induction. Ad (ii): consider f^+ and f^- and use (i). □

Remark 3. For brevity, we write

$$\int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu(d\omega_1) = \int_{A_1} \int_{\Omega_2} f(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu|_{A_1 \cap \mathfrak{A}_1}(d\omega_1),$$

if f is $(\mu \times K)$-integrable. For $f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$

$$f \text{ is } (\mu \times K)\text{-integrable} \quad \Leftrightarrow \quad \int_{\Omega_1} \int_{\Omega_2} |f|(\omega_1, \omega_2)\, K(\omega_1, d\omega_2)\, \mu(d\omega_1) < \infty.$$

Corollary 1 (Fubini's Theorem)**.** For σ-finite measures μ_i on $\mathfrak{A}_i$ and a $(\mu_1 \times \mu_2)$-integrable function f

$$\begin{aligned}\int_\Omega f \, d(\mu_1 \times \mu_2) &= \int_{\Omega_1} \int_{\Omega_2} f(\omega_1, \omega_2) \, \mu_2(d\omega_2) \, \mu_1(d\omega_1) \\ &= \int_{\Omega_2} \int_{\Omega_1} f(\omega_1, \omega_2) \, \mu_1(d\omega_1) \, \mu_2(d\omega_2).\end{aligned}$$

Proof. Theorem 2 yields the first equality. For the second equality, put $\widetilde{f}(\omega_2, \omega_1) = f(\omega_1, \omega_2)$ and note that $\int_\Omega f \, d(\mu_1 \times \mu_2) = \int_\Omega \widetilde{f} \, d(\mu_2 \times \mu_1)$. □

Corollary 2. For every measurable space $(\Omega, \mathfrak{A})$, every σ-finite measure μ on $\mathfrak{A}$, and every $f \in \overline{\mathfrak{Z}}_+(\Omega, \mathfrak{A})$

$$\int_\Omega f \, d\mu = \int_{]0,\infty[} \mu(\{f > x\}) \, \lambda_1(dx).$$

Proof. Übung 6.4. □

Now we construct a stochastic model for a series of experiments, where the outputs of the first $i-1$ stages determine the model for the ith stage.

Given: measurable spaces $(\Omega_i, \mathfrak{A}_i)$ for $i \in I$, where $I = \{1, \dots, n\}$ or $I = \mathbb{N}$. Put

$$(\Omega'_i, \mathfrak{A}'_i) = \left(\bigtimes_{j=1}^{i} \Omega_j, \bigotimes_{j=1}^{i} \mathfrak{A}_j \right),$$

and note that

$$\bigtimes_{j=1}^{i} \Omega_j = \Omega'_{i-1} \times \Omega_i \qquad \wedge \qquad \bigotimes_{j=1}^{i} \mathfrak{A}_j = \mathfrak{A}'_{i-1} \otimes \mathfrak{A}_i$$

for $i \in I \setminus \{1\}$. Furthermore, let

$$\Omega = \bigtimes_{i \in I} \Omega_i, \qquad \mathfrak{A} = \bigotimes_{i \in I} \mathfrak{A}_i. \tag{10}$$

Given:

- σ-finite kernels K_i from $(\Omega'_{i-1}, \mathfrak{A}'_{i-1})$ to $(\Omega_i, \mathfrak{A}_i)$ for $i \in I \setminus \{1\}$,
- a σ-finite measure μ on $\mathfrak{A}_1$.

Theorem 3. For $I = \{1, \dots, n\}$

$\underset{1}{\exists}$ measure ν on $\mathfrak{A}$ $\ \forall A_1 \in \mathfrak{A}_1 \dots \forall A_n \in \mathfrak{A}_n :$

$$\begin{aligned}&\nu(A_1 \times \cdots \times A_n) \\ &= \int_{A_1} \cdots \int_{A_{n-1}} K_n((\omega_1, \dots, \omega_{n-1}), A_n) \, K_{n-1}((\omega_1, \dots, \omega_{n-2}), d\omega_{n-1}) \cdots \mu(d\omega_1).\end{aligned}$$

Moreover, ν is σ-finite and for f ν-integrable (the short version)

$$\int_\Omega f \, d\nu = \int_{\Omega_1} \cdots \int_{\Omega_n} f(\omega_1, \dots, \omega_n) K_n((\omega_1, \dots, \omega_{n-1}), d\omega_n) \cdots \mu(d\omega_1). \tag{11}$$

Notation: $\nu = \mu \times K_2 \times \cdots \times K_n$.

Proof. Induction, using Theorems 1 and 2. $\square$

Remark 4. Particular case of Theorem 3 with

$$\mu = \mu_1, \qquad \forall\, i \in I \setminus \{1\}\ \forall\, \omega'_{i-1} \in \Omega'_{i-1} :\ K_i(\omega'_{i-1}, \cdot) = \mu_i \tag{12}$$

for σ-finite measures μ_i on $\mathfrak{A}_i$:

$$\underset{1}{\exists}\, \text{measure } \mu_1 \times \cdots \times \mu_n \text{ on } \mathfrak{A}\ \ \forall\, A_1 \in \mathfrak{A}_1 \ldots \forall\, A_n \in \mathfrak{A}_n :$$
$$\mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdot \cdots \cdot \mu_n(A_n).$$

Moreover, $\mu_1 \times \cdots \times \mu_n$ is σ-finite and for every $\mu_1 \times \cdots \times \mu_n$-integrable function f

$$\int_\Omega f\, d(\mu_1 \times \cdots \times \mu_n) = \int_{\Omega_1} \ldots \int_{\Omega_n} f(\omega_1, \ldots, \omega_n)\ \mu_n(d\omega_n) \cdots \mu_1(d\omega_1).$$

Definition 3. $\mu = \mu_1 \times \cdots \times \mu_n$ is called the *product measure* corresponding to μ_i for $i = 1, \ldots, n$, and $(\Omega, \mathfrak{A}, \mu)$ is called the *product measure space* corresponding to $(\Omega_i, \mathfrak{A}_i, \mu_i)$ for $i = 1, \ldots, n$.

Example 6.

(i) For uniform distributions μ_i on finite spaces Ω_i, $\mu_1 \times \cdots \times \mu_n$ is the uniform distribution on Ω. Cf. Example 3.1 in the case $n \in \mathbb{N}$.

(ii)
$$\lambda_n = \lambda_1 \times \cdots \times \lambda_1.$$

Theorem 4 (Ionescu-Tulcea)**.** Assume that μ is a probability measure and that K_i are Markov kernels for $i \in \mathbb{N} \setminus \{1\}$. Then, for $I = \mathbb{N}$,

$$\underset{1}{\exists}\, \text{probability measure } P \text{ on } \mathfrak{A}\ \ \forall\, n \in \mathbb{N}\ \ \forall\, A_1 \in \mathfrak{A}_1 \ldots \forall\, A_n \in \mathfrak{A}_n :$$
$$P\Big(A_1 \times \cdots \times A_n \times \bigtimes_{i=n+1}^{\infty} \Omega_i\Big) = (\mu \times K_2 \times \cdots \times K_n)(A_1 \times \cdots \times A_n). \tag{13}$$

Proof. 'Existence': Consider σ-algebras

$$\bigotimes_{i=1}^{n} \mathfrak{A}_i, \qquad \widetilde{\mathfrak{A}}_n = \sigma\left(\pi^{\mathbb{N}}_{\{1,\ldots,n\}}\right)$$

on $\bigtimes_{i=1}^n \Omega_i$ and $\bigtimes_{i=1}^\infty \Omega_i$, respectively. Define a probability measure $\widetilde{P}_n$ on $\widetilde{\mathfrak{A}}_n$ by

$$\widetilde{P}_n\Big(A \times \bigtimes_{i=n+1}^{\infty} \Omega_i\Big) = (\mu \times K_2 \times \cdots \times K_n)(A), \qquad A \in \bigotimes_{i=1}^{n} \mathfrak{A}_i.$$

Then (11) yields the following *consistency property*

$$\widetilde{P}_{n+1}\Big(A \times \Omega_{n+1} \times \bigtimes_{i=n+2}^{\infty} \Omega_i\Big) = \widetilde{P}_n\Big(A \times \bigtimes_{i=n+1}^{\infty} \Omega_i\Big), \qquad A \in \bigotimes_{i=1}^{n} \mathfrak{A}_i.$$

Thus

$$\widetilde{P}(\widetilde{A}) = \widetilde{P}_n(\widetilde{A}), \qquad \widetilde{A} \in \widetilde{\mathfrak{A}}_n,$$

yields a well-defined mapping on the algebra

$$\widetilde{\mathfrak{A}} = \bigcup_{n\in\mathbb{N}} \widetilde{\mathfrak{A}}_n$$

of cylinder sets. Obviously, $\widetilde{P}$ is a content and (13) holds for $P = \widetilde{P}$.
Claim: $\widetilde{P}$ is σ-continuous at $\emptyset$.
It suffices to show that for every sequence of sets

$$A^{(n)} = B^{(n)} \times \bigtimes_{i=n+1}^{\infty} \Omega_i$$

with $B^{(n)} \in \bigotimes_{i=1}^n \mathfrak{A}_i$ and $A^{(n)} \downarrow \emptyset$ we have

$$\lim_{n\to\infty} (\mu \times K_2 \times \ldots \times K_n)(B^{(n)}) = 0.$$

Assume that $\inf_{n\in\mathbb{N}}(\mu \times K_2 \times \ldots \times K_n)(B^{(n)}) > 0$. Put

$$K^{\omega_1,\ldots,\omega_m}(A) = K_{m+1}((\omega_1,\ldots,\omega_m), A)$$

for $\omega_i \in \Omega_i$, $m \geq 1$, and $A \in \mathfrak{A}_{m+1}$ as well as

$$K_n^{\omega_1,\ldots,\omega_m}((\omega_{m+1},\ldots,\omega_{n-1}), A) = K_n((\omega_1,\ldots,\omega_{n-1}), A)$$

for $\omega_i \in \Omega_i$, $n \geq m+2$, and $A \in \mathfrak{A}_n$. Then $K^{\omega_1,\ldots,\omega_m}$ is a probability measure on $(\Omega_{m+1}, \mathfrak{A}_{m+1})$ and $K_n^{\omega_1,\ldots,\omega_m}$ is a Markov kernel from $\left(\bigtimes_{i=m+1}^{n-1} \Omega_i, \bigotimes_{i=m+1}^{n-1} \mathfrak{A}_i\right)$ to $(\Omega_n, \mathfrak{A}_n)$. Let $n \geq m+1$. Put

$$f_{m,n}(\omega_1,\ldots,\omega_m) = K^{\omega_1,\ldots,\omega_m} \times K_{m+2}^{\omega_1,\ldots,\omega_m} \times \ldots \times K_n^{\omega_1,\ldots,\omega_m}(B^{(n)}(\omega_1,\ldots,\omega_m)).$$

Since

$$B^{(n)} \times \Omega_{n+1} \supset B^{(n+1)}$$

we get

$$f_{m,n}(\omega_1,\ldots,\omega_m) \geq f_{m,n+1}(\omega_1,\ldots,\omega_m).$$

Furthermore $0 \leq f_{m,n} \leq 1$. Hence

$$\int_{\Omega_1} \inf_{n\geq 2} f_{1,n}(\omega)\mu(d\omega_1) = \inf_{n\geq 2} \int_{\Omega_1} f_{1,n}(\omega)\mu(d\omega_1) = \inf_{n\geq 2} (\mu \times K_2 \times \ldots \times K_n)(B^{(n)}) > 0.$$

Consequently

$$\exists\, \omega_1^* \in \Omega_1 : \inf_{n\geq 2} f_{1,n}(\omega_1^*) > 0.$$

Inductively we get a sequence of point $\omega_m^* \in \Omega_m$ such that

$$\inf_{n\geq m+1} f_{m,n}(\omega_1^*,\ldots,\omega_m^*) > 0.$$

In particular $B^{(m+1)}(\omega_1^*, \dots, \omega_m^*) \neq \emptyset$, which implies $(\omega_1^*, \dots, \omega_m^*) \in B^{(m)}$. We conclude that

$$(\omega_1^*, \omega_2^*, \dots) \in \bigcap_{m=1}^{\infty} A^{(m)},$$

contradicting $A^{(m)} \downarrow \emptyset$.

By Theorem 4.1, $\widetilde{P}$ is σ-additive and it remains to apply Theorem 4.3.

'Uniqueness': By (13), P is uniquely determined on the class of measurable rectangles. Apply Theorem 4.4. □

Example 7. The queueing model, see Übung 7.1. Here $K_i((\omega_1, \dots, \omega_{i-1}), \cdot)$ only depends on ω_{i-1}. Outlook: Markov processes.

Given: a non-empty set I and probability spaces $(\Omega_i, \mathfrak{A}_i, \mu_i)$ for $i \in I$. Recall the definition (10).

Theorem 5.

$$\underset{1}{\exists} \text{probability measure } P \text{ on } \mathfrak{A} \;\; \forall S \in \mathfrak{P}_0(I) \;\; \forall A_i \in \mathfrak{A}_i, \; i \in S:$$

$$P\Big(\bigtimes_{i \in S} A_i \times \bigtimes_{i \in I \setminus S} \Omega_i\Big) = \prod_{i \in S} \mu_i(A_i). \tag{14}$$

Notation: $P = \bigtimes_{i \in I} \mu_i$.

Proof. See Remark 4 in the case of a finite set I.

If $|I| = |\mathbb{N}|$, assume $I = \mathbb{N}$ without loss of generality. The particular case of Theorem 4 with (12) for probability measures μ_i on $\mathfrak{A}_i$ shows

$$\underset{1}{\exists} \text{probability measure } P \text{ on } \mathfrak{A} \;\; \forall n \in \mathbb{N} \;\; \forall A_1 \in \mathfrak{A}_1 \dots \forall A_n \in \mathfrak{A}_n :$$

$$P\Big(A_1 \times \dots \times A_n \times \bigtimes_{i=n+1}^{\infty} \Omega_i\Big) = \mu_1(A_1) \cdot \dots \cdot \mu_n(A_n).$$

If I is uncountable, we use Theorem 3.2. For $S \subset I$ non-empty and countable and for $B \in \bigotimes_{i \in S} \mathfrak{A}_i$ we put

$$P\big((\pi_S^I)^{-1} B\big) = \bigtimes_{i \in S} \mu_i(B).$$

Hereby we get a well-defined mapping $P : \mathfrak{A} \to \mathbb{R}$, which clearly is a probability measure and satisfies (14). Use Theorem 4.4 to obtain the uniqueness result. □

Definition 4. $P = \bigtimes_{i \in I} \mu_i$ is called the *product measure* corresponding to μ_i for $i \in I$, and $(\Omega, \mathfrak{A}, P)$ is called the *product measure space* corresponding to $(\Omega_i, \mathfrak{A}_i, \mu_i)$ for $i \in I$.

Remark 5. Theorem 5 answers the question that is posed in Example 3.1 in full generality. Moreover, it is the basis for a positive answer to the question from the introductory Example I.2, see Theorem III.5.2.

9 Image Measures

Given: a measure space $(\Omega, \mathfrak{A}, \mu)$, a measurable space $(\Omega', \mathfrak{A}')$, and an $\mathfrak{A}$-$\mathfrak{A}'$-measurable mapping $f : \Omega \to \Omega'$.

Lemma 1.

$$\begin{aligned} f(\mu) : \mathfrak{A}' &\to \mathbb{R}_+ \cup \{\infty\} \\ A' &\mapsto \mu(f^{-1}(A')) = \mu(\{f \in A'\}) \end{aligned}$$

defines a measure on $\mathfrak{A}'$.

Proof. $f(\mu)$ is well-defined, since $f^{-1}(A') \in \mathfrak{A}$ for any $A' \in \mathfrak{A}'$. The respective properties of $f(\mu)$ are easy to verify. □

Definition 1. $f(\mu)$ is called the *image measure* of μ under f.

Example 1. Let

$$(\Omega, \mathfrak{A}, \mu) = (\mathbb{R}^k, \mathfrak{B}_k, \lambda_k), \qquad (\Omega', \mathfrak{A}') = (\mathbb{R}^k, \mathfrak{B}_k).$$

(i) Fix $a \in \mathbb{R}^k$. For $f(\omega) = \omega + a$ we get

$$f(\lambda_k)(A') = \lambda_k(A' - a) = \lambda_k(A'),$$

see Analysis IV ('or' verify this identity for measurable rectangles and apply Theorem 4.4). Thus

$$f(\lambda_k) = \lambda_k.$$

(ii) Fix $r \in \mathbb{R} \setminus \{0\}$. For $f(\omega) = r \cdot \omega$ we get

$$f(\lambda_k)(A') = \lambda_k(1/r \cdot A') = \frac{1}{|r|^k} \cdot \lambda_k(A'),$$

see Analysis IV ('or' verify this identity for measurable rectangles and apply Theorem 4.4). Thus

$$f(\lambda_k) = \frac{1}{|r|^k} \cdot \lambda_k.$$

Theorem 1 (Transformation 'Theorem')**.**

(i) for $g \in \overline{\mathfrak{Z}}_+(\Omega', \mathfrak{A}')$

$$\int_{\Omega'} g \, df(\mu) = \int_{\Omega} g \circ f \, d\mu \tag{1}$$

(ii) for $g \in \overline{\mathfrak{Z}}(\Omega', \mathfrak{A}')$

$$g \text{ is } f(\mu)\text{-integrable} \quad \Leftrightarrow \quad g \circ f \text{ is } \mu\text{-integrable},$$

in which case (1) holds.

Proof. Algebraic induction. □

Example 2. Consider open sets $U, V \subset \mathbb{R}^k$ and a $\mathfrak{C}^1$-diffeomorphism $f : U \to V$ as well as

$$(\Omega, \mathfrak{A}, \mu) = (U, U \cap \mathfrak{B}_k, \lambda_k|_{U \cap \mathfrak{B}_k}), \qquad (\Omega', \mathfrak{A}', \mu') = (V, V \cap \mathfrak{B}_k, \lambda_k|_{V \cap \mathfrak{B}_k}).$$

Then

$$f^{-1}(\nu)(A) = \nu(f(A)) = \int_A |\det Df| \, d\mu,$$

see Analysis IV for the case of an open set $A \subset U$. Thus

$$f^{-1}(\nu) = |\det Df| \cdot \mu,$$

and therefore

$$\int_V g \, d\nu = \int_U g \circ f \, df^{-1}(\nu) = \int_U g \circ f \cdot |\det Df| \, d\mu$$

if g is ν-integrable.

Chapter III

Basic Concepts of Probability Theory

Context for probability theoretical concepts: a probability space $(\Omega, \mathfrak{A}, P)$.
Terminology: $A \in \mathfrak{A}$ *event*, $P(A)$ *probability of the event* $A \in \mathfrak{A}$.

1 Random Variables and Distributions

Given: a probability space $(\Omega, \mathfrak{A}, P)$ and a measurable space $(\Omega', \mathfrak{A}')$.

Definition 1. $X : \Omega \to \Omega'$ *random element* if X is $\mathfrak{A}$-$\mathfrak{A}'$-measurable. Particular cases:

(i) X *(real) random variable* if $(\Omega', \mathfrak{A}') = (\mathbb{R}, \mathfrak{B})$,

(ii) X *numerical random variable* if $(\Omega', \mathfrak{A}') = (\overline{\mathbb{R}}, \overline{\mathfrak{B}})$,

(iii) X *k-dimensional (real) random vector* if $(\Omega', \mathfrak{A}') = (\mathbb{R}^k, \mathfrak{B}_k)$,

(iv) X *k-dimensional numerical random vector* if $(\Omega', \mathfrak{A}') = (\overline{\mathbb{R}}^k, \overline{\mathfrak{B}}_k)$.

Definition 2.

(i) *Distribution (probability law)* of a random element $X : \Omega \to \Omega'$ *(with respect to P)*

$$P_X = X(P).$$

Notation: $X \sim Q$ if $P_X = Q$.

(ii) Given: probability spaces $(\Omega_1, \mathfrak{A}_1, P_1)$, $(\Omega_2, \mathfrak{A}_2, P_2)$ and random elements

$$X_1 : \Omega_1 \to \Omega', \qquad X_2 : \Omega_2 \to \Omega'.$$

X_1 and X_2 are *identically distributed* if

$$(P_1)_{X_1} = (P_2)_{X_2}.$$

Remark 1.

(i) $P_X(A') = P(\{X \in A'\})$ for every $A' \in \mathfrak{A}'$.

(ii) For random elements $X, Y : \Omega \to \Omega'$

$$X = Y \ P\text{-a.s.} \quad \Rightarrow \quad P_X = P_Y,$$

but the converse is not true in general. For instance, let P be the uniform distribution on $\Omega = \{0, 1\}$ and define $X(\omega) = \omega$ and $Y(\omega) = 1 - \omega$.

(iii) For every probability measure Q on $(\Omega', \mathfrak{A}')$ there exists a probability space $(\Omega, \mathfrak{A}, P)$ and a random element $X : \Omega \to \Omega'$ such that $X \sim Q$. Take $(\Omega, \mathfrak{A}, P) = (\Omega', \mathfrak{A}', Q)$ and $X = \mathrm{id}_\Omega$.

(iv) A major part of probability theory deals with properties of random elements that can be formulated in terms of their distributions.

Example 1.

(i) *Discrete distributions*, specified by a countable set $\emptyset \neq D \subset \Omega'$ and a mapping $p : D \to \mathbb{R}$ such that

$$\forall\, r \in D : p(r) \geq 0 \qquad \wedge \qquad \sum_{r \in D} p(r) = 1,$$

namely,

$$P_X = \sum_{r \in D} p(r) \cdot \varepsilon_r.$$

Thus, if $\{r\} \in \mathfrak{A}'$ for every $r \in D$,

$$P(\{X = r\}) = p(r).$$

If $|D| < \infty$ then $p(r) = \frac{1}{|D|}$ yields the *uniform distribution on D*.

For $(\Omega', \mathfrak{A}') = (\mathbb{R}, \mathfrak{B})$

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

is the *binomial distribution* with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$. In particular, for $n = 1$ we get the *Bernoulli distribution*

$$B(1, p) = (1 - p) \cdot \varepsilon_0 + p \cdot \varepsilon_1.$$

Further examples include the *geometric distribution* with parameter $p \in]0, 1]$,

$$G(p) = \sum_{k=1}^{\infty} p \cdot (1 - p)^{k-1} \cdot \varepsilon_k,$$

and the *Poisson distribution* with parameter $\lambda > 0$,

$$\pi(\lambda) = \sum_{k=0}^{\infty} \exp(-\lambda) \cdot \frac{\lambda^k}{k!} \cdot \varepsilon_k.$$

(ii) *Distributions on* $(\mathbb{R}^k, \mathfrak{B}_k)$ *that are absolutely continuous w.r.t.* λ_k, namely, due to the Radon-Nikodym-Theorem

$$P_X = f \cdot \lambda_k,$$

where

$$f \in \overline{\mathfrak{Z}}_+(\mathbb{R}^k, \mathfrak{B}_k) \qquad \wedge \qquad \int f \, d\lambda_k = 1.$$

Thus

$$P(\{X \in A'\}) = \int_{A'} f \, d\lambda_k$$

for every $A' \in \mathfrak{B}_k$.

We present some examples in the case $k = 1$. The *normal distribution*

$$N(\mu, \sigma^2) = f \cdot \lambda_1 \,,$$

with parameters $\mu \in \mathbb{R}$ and σ^2, where $\sigma > 0$, is obtained by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right), \qquad x \in \mathbb{R}.$$

The *exponential distribution* with parameter $\lambda > 0$ is obtained by

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda \cdot \exp(-\lambda\, x) & \text{if } x \geq 0. \end{cases}$$

The *uniform distribution* on $D \in \mathfrak{B}$ with $\lambda_1(D) \in \,]0, \infty[$ is obtained by

$$f = \frac{1}{\lambda_1(D)} \cdot 1_D.$$

(iii) Distributions on product spaces can be constructed by means of the results from Section II.8.

Remark 2. Define $\infty^r = \infty$ for $r > 0$. For $1 \leq p < q < \infty$ and $X \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$

$$\int |X|^p \, dP \leq \left(\int |X|^q \, dP\right)^{p/q},$$

due to Hölder's inequality.

Notation:

$$\mathfrak{L} = \mathfrak{L}(\Omega, \mathfrak{A}, P) = \left\{X \in \mathfrak{Z}(\Omega, \mathfrak{A}) : \int |X| \, dP < \infty\right\}$$

is the class of P-integrable random variables, and analogously

$$\overline{\mathfrak{L}} = \overline{\mathfrak{L}}(\Omega, \mathfrak{A}, P) = \left\{X \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A}) : \int |X| \, dP < \infty\right\}$$

is the class of P-integrable numerical random variables. We consider P_X as a distribution on $(\mathbb{R}, \mathfrak{B})$ if $P(\{X \in \mathbb{R}\}) = 1$ for a numerical random variable X, and we consider $\mathfrak{L}$ as a subspace of $\overline{\mathfrak{L}}$.

Definition 3. For $X \in \overline{\mathfrak{L}}$

$$\mathrm{E}(X) = \int X \, dP$$

is the *expectation* of X. For $X \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$ such that $X^2 \in \overline{\mathfrak{L}}$

$$\mathrm{Var}(X) = \int (X - \mathrm{E}(X))^2 \, dP$$

and $\sqrt{\mathrm{Var}(X)}$ are the *variance* and the *standard deviation* of X, respectively.

Remark 3. Theorem II.9.1 implies

$$\int_\Omega |X|^p \, dP < \infty \qquad \Leftrightarrow \qquad \int_{\mathbb{R}} |x|^p \, P_X(dx) < \infty$$

for $X \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$, in which case, for $p = 1$

$$\mathrm{E}(X) = \int_{\mathbb{R}} x \, P_X(dx),$$

and for $p = 2$

$$\mathrm{Var}(X) = \int_{\mathbb{R}} (x - \mathrm{E}(X))^2 \, P_X(dx).$$

Thus $\mathrm{E}(X)$ and $\mathrm{Var}(X)$ depend only on P_X.

Example 2.

$$\begin{array}{lll} X \sim B(n,p) & \mathrm{E}(X) = n \cdot p & \mathrm{Var}(X) = n \cdot p \cdot (1-p) \\ X \sim G(p) & \mathrm{E}(X) = \dfrac{1}{p} & \mathrm{Var}(X) = \dfrac{1-p}{p^2} \\ X \sim \pi(\lambda) & \mathrm{E}(X) = \lambda & \mathrm{Var}(X) = \lambda, \end{array}$$

see Introduction to Stochastics.

X is *Cauchy distributed* with parameter $\alpha > 0$ if $X \sim f \cdot \lambda_1$ where

$$f(x) = \frac{\alpha}{\pi(\alpha^2 + x^2)}, \qquad x \in \mathbb{R}.$$

Since $\int_0^t \frac{x}{1+x^2} dx = \frac{1}{2} \log(1 + t^2)$ neither $\mathrm{E}(X^+) < \infty$ nor $\mathrm{E}(X^-) < \infty$, and therefore $X \notin \mathfrak{L}$.

If $X \sim N(\mu, \sigma^2)$ then

$$\mathrm{E}(X) = \mu \qquad \wedge \qquad \mathrm{Var}(X) = \sigma^2,$$

see Introduction to Stochastics.

If X is exponentially distributed with parameter $\lambda > 0$ then

$$\mathrm{E}(X) = \frac{1}{\lambda} \qquad \wedge \qquad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Definition 4. Let $X = (X_1, \dots, X_k)$ be a random vector. Then

$$F_X : \mathbb{R}^k \to [0,1]$$
$$(x_1, \dots, x_k) \mapsto P_X\Big(\bigtimes_{i=1}^{k}]-\infty\,,\ x_i]\Big) = P\Big(\bigcap_{i=1}^{k} \{X_i \le x_i\}\Big)$$

is called the *distribution function* of X.

Theorem 1. Given: probability spaces $(\Omega_1, \mathfrak{A}_1, P_1)$, $(\Omega_2, \mathfrak{A}_2, P_2)$ and random vectors

$$X^1 : \Omega_1 \to \mathbb{R}^k, \qquad X^2 : \Omega_2 \to \mathbb{R}^k.$$

Then

$$(P_1)_{X^1} = (P_2)_{X^2} \qquad \Leftrightarrow \qquad F_{X^1} = F_{X^2}.$$

Proof. '$\Rightarrow$' holds trivially. '$\Leftarrow$': By Remark II.1.6, $\mathfrak{B}_k = \sigma(\mathfrak{E})$ for

$$\mathfrak{E} = \Big\{ \bigtimes_{i=1}^{k}]-\infty, x_i] : x_1, \dots, x_k \in \mathbb{R} \Big\}.$$

Use Theorem II.4.4. □

For notational convenience, we consider the case $k = 1$ in the sequel.

Theorem 2.

(i) F_X is non-decreasing,

(ii) F_X is right-continuous,

(iii) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$,

(iv) F_X is continuous at x iff $P(\{X = x\}) = 0$.

Proof. Übung 3.4.a. □

Theorem 3. For every function F that satisfies (i)–(iii) from Theorem 2,

$$\underset{1}{\exists}\, Q \text{ probability measure on } \mathfrak{B} : \forall\, x \in \mathbb{R} : Q(]-\infty, x]) = F(x).$$

Proof. Analogously to the construction of the Lebesgue measure, see Übung 3.4.b. □

2 Convergence in Probability

Motivated by the Examples II.5.2 and II.6.1 we introduce a notion of convergence that is weaker than convergence in mean and convergence almost surely.

In the sequel, X, X_n, etc. random variables on a common probability space $(\Omega, \mathfrak{A}, P)$.

Definition 1. $(X_n)_n$ *converges* to X *in probability* if

$$\forall \varepsilon > 0 : \lim_{n\to\infty} P(\{|X_n - X| > \varepsilon\}) = 0.$$

Notation: $X_n \xrightarrow{P} X$.

Theorem 1 (Chebyshev-Markov Inequality)**.** Let $(\Omega, \mathfrak{A}, \mu)$ be a measure space and $f \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A})$. For every $\varepsilon > 0$ and every $1 \leq p < \infty$

$$\mu(\{|f| \geq \varepsilon\}) \leq \frac{1}{\varepsilon^p} \cdot \int |f|^p \, d\mu.$$

Proof. We have

$$\int_{\{|f|\geq\varepsilon\}} \varepsilon^p \, d\mu \leq \int_\Omega |f|^p \, d\mu.$$

□

Corollary 1. If $\mathrm{E}(X^2) < \infty$, then

$$P(\{|X - \mathrm{E}(X)| \geq \varepsilon\}) \leq \frac{1}{\varepsilon^2} \cdot \mathrm{Var}(X).$$

Theorem 2.

$$d(X, Y) = \int \min(1, |X - Y|) \, dP$$

defines a semi-metric on $\mathfrak{Z}(\Omega, \mathfrak{A})$, and

$$X_n \xrightarrow{P} X \quad \Leftrightarrow \quad \lim_{n\to\infty} d(X_n, X) = 0.$$

Proof. '⇒' For $\varepsilon > 0$

$$\begin{aligned} &\int \min(1,\, |X_n - X|) \, dP \\ &\quad= \int_{\{|X_n - X| > \varepsilon\}} \min(1,\, |X_n - X|) \, dP + \int_{\{|X_n - X| \leq \varepsilon\}} \min(1,\, |X_n - X|) \, dP \\ &\quad\leq P(\{|X_n - X| > \varepsilon\}) + \min(1, \varepsilon). \end{aligned}$$

'⇐': Let $0 < \varepsilon < 1$. Use Theorem 1 to obtain

$$\begin{aligned} P(\{|X_n - X| > \varepsilon\}) &= P(\{\min(1, |X_n - X|) > \varepsilon\}) \\ &\leq \frac{1}{\varepsilon} \cdot \int \min(1,\, |X_n - X|) \, dP = \frac{1}{\varepsilon} \cdot d(X_n, X). \end{aligned}$$

□

Remark 1. By Theorem 2,

$$X_n \xrightarrow{\mathfrak{L}^p} X \quad \Rightarrow \quad X_n \xrightarrow{P} X.$$

Example II.5.2 shows that '⇐' does not hold in general.

Remark 2. By Theorems 2 and II.5.5,

$$X_n \xrightarrow{P\text{-a.s.}} X \quad \Rightarrow \quad X_n \xrightarrow{P} X.$$

Example II.6.1 shows that '⇐' does not hold in general. The Law of Large Numbers deals with convergence almost surely or convergence in probability, see the introductory Example I.1 and Sections IV.2 and IV.3.

Corollary 2.

$$X_n \xrightarrow{P} X \quad \Rightarrow \quad \exists\,\text{subsequence } (X_{n_k})_{k\in\mathbb{N}} : \; X_{n_k} \xrightarrow{P\text{-a.s.}} X.$$

Proof. Due to Theorems II.6.3 and 2 there exists a subsequence $(X_{n_k})_{k\in\mathbb{N}}$ such that

$$\min(1, |X_{n_k} - X|) \xrightarrow{P\text{-a.s.}} 0.$$

□

Remark 3. In any semi-metric space (M, d) a sequence $(a_n)_{n\in\mathbb{N}}$ converges to a iff

$$\forall\,\text{subsequence } (a_{n_k})_{k\in\mathbb{N}} \;\exists\,\text{subsequence } (a_{n_{k_\ell}})_{\ell\in\mathbb{N}} : \; \lim_{\ell\to\infty} d(a_{n_{k_\ell}}, a) = 0.$$

Corollary 3. $X_n \xrightarrow{P} X$ iff

$$\forall\,\text{subsequence } (X_{n_k})_{k\in\mathbb{N}} \;\exists\,\text{subsequence } (X_{n_{k_\ell}})_{\ell\in\mathbb{N}} : \; X_{n_{k_\ell}} \xrightarrow{P\text{-a.s.}} X.$$

Proof. '⇒': Corollary 2. '⇐': Remarks 2 and 3 together with Theorem 2. □

Remark 4. We conclude that, in general, there is no semi-metric on $\mathfrak{Z}(\Omega, \mathfrak{A})$ that defines a.s.-convergence. However, if Ω is countable, then

$$X_n \xrightarrow{P\text{-a.s.}} X \quad \Leftrightarrow \quad X_n \xrightarrow{P} X.$$

Proof: Übung 7.4.

Lemma 1. Let $\longrightarrow$ denote convergence almost everywhere or convergence in probability. If $X_n^{(i)} \longrightarrow X^{(i)}$ for $i = 1, \dots, k$ and $f : \mathbb{R}^k \to \mathbb{R}$ is continuous, then

$$f \circ (X_n^{(1)}, \dots, X_n^{(k)}) \longrightarrow f \circ (X^{(1)}, \dots, X^{(k)}).$$

Proof. Trivial for convergence almost everywhere, and by Corollary 3 the conclusion holds for convergence in probability, too. □

Corollary 4. Let $X_n \xrightarrow{P} X$. Then

$$X_n \xrightarrow{P} Y \quad \Leftrightarrow \quad X = Y \;\; P\text{-a.s.}$$

Proof. Corollary 3 and Lemma II.6.1. □

3 Convergence in Distribution

Given: a metric space (M, ρ). Put

$$C^b(M) = \{f : M \to \mathbb{R} : f \text{ bounded, continuous}\},$$

and consider the Borel-σ-algebra $\mathfrak{B}(M)$ in M. Moreover, let $\mathfrak{M}(M)$ denote the set of all probability measures on $\mathfrak{B}(M)$.

Definition 1.

(i) A sequence $(Q_n)_{n\in\mathbb{N}}$ in $\mathfrak{M}(M)$ *converges weakly* to $Q \in \mathfrak{M}(M)$ if

$$\forall\, f \in C^b(M) : \lim_{n\to\infty} \int f\, dQ_n = \int f\, dQ.$$

Notation: $Q_n \xrightarrow{\text{w}} Q$.

(ii) A sequence $(X_n)_{n\in\mathbb{N}}$ of random elements with values in M *converges in distribution* to a random element X with values in M if $Q_n \xrightarrow{\text{w}} Q$ for the distributions Q_n of X_n and Q of X, respectively.
Notation: $X_n \xrightarrow{\text{d}} X$.

Remark 1. For convergence in distribution the random elements need not be defined on a common probability space.

In the sequel: Q_n, $Q \in \mathfrak{M}(M)$ for $n \in \mathbb{N}$.

Example 1.

(i) For $x_n,\ x \in M$

$$\varepsilon_{x_n} \xrightarrow{\text{w}} \varepsilon_x \quad \Leftrightarrow \quad \lim_{n\to\infty} \rho(x_n, x) = 0.$$

For the proof of '$\Leftarrow$', note that

$$\int f\, d\varepsilon_{x_n} = f(x_n), \qquad \int f\, d\varepsilon_x = f(x).$$

For the proof of '$\Rightarrow$', suppose that $\limsup_{n\to\infty} \rho(x_n, x) > 0$. Take

$$f(y) = \min(\rho(y, x), 1), \qquad y \in M,$$

and observe that $f \in C^b(M)$ and

$$\limsup_{n\to\infty} \int f\, d\varepsilon_{x_n} = \limsup_{n\to\infty} \min(\rho(x_n, x), 1) > 0$$

while $\int f\, d\varepsilon_x = 0$.

(ii) For the euclidean distance ρ on $M = \mathbb{R}^k$

$$(M, \mathfrak{B}(M)) = (\mathbb{R}^k, \mathfrak{B}_k).$$

Now, in particular, $k = 1$ and

$$Q_n = N(\mu_n, \sigma_n^2)$$

where $\sigma_n > 0$. For $f \in C^b(\mathbb{R})$

$$\int f\, dQ_n = 1/\sqrt{2\pi} \cdot \int_{\mathbb{R}} f(\sigma_n \cdot x + \mu_n) \cdot \exp(-1/2 \cdot x^2)\, \lambda_1(dx).$$

Put $N(\mu, 0) = \varepsilon_\mu$. Then

$$\lim_{n\to\infty} \mu_n = \mu \ \wedge \ \lim_{n\to\infty} \sigma_n = \sigma \quad \Rightarrow \quad Q_n \xrightarrow{\text{w}} N(\mu, \sigma^2).$$

Otherwise $(Q_n)_{n\in\mathbb{N}}$ does not converge weakly. Übung 8.2.

(iii) For $M = C([0,T])$ let $\rho(x,y) = \sup_{t\in[0,T]} |x(t) - y(t)|$. Cf. the introductory Example I.3.

Remark 2. Note that $Q_n \xrightarrow{\text{w}} Q$ does not imply

$$\forall\, A \in \mathfrak{B}(M) : \lim_{n\to\infty} Q_n(A) = Q(A).$$

For instance, assume $\lim_{n\to\infty} \rho(x_n, x) = 0$ with $x_n \neq x$ for every $n \in \mathbb{N}$. Then

$$\varepsilon_{x_n}(\{x\}) = 0, \qquad \varepsilon_x(\{x\}) = 1.$$

Theorem 1 (Portmanteau Theorem)**.** The following properties are equivalent:

(i) $Q_n \xrightarrow{\text{w}} Q$,

(ii) $\forall\, f \in C^b(M)$ uniformly continuous : $\lim_{n\to\infty} \int f\, dQ_n = \int f\, dQ$,

(iii) $\forall\, A \subset M$ closed : $\limsup_{n\to\infty} Q_n(A) \leq Q(A)$,

(iv) $\forall\, A \subset M$ open : $\liminf_{n\to\infty} Q_n(A) \geq Q(A)$,

(v) $\forall\, A \in \mathfrak{B}(M)$: $Q(\partial A) = 0 \ \Rightarrow\ \lim_{n\to\infty} Q_n(A) = Q(A)$.

Proof. See Gänssler, Stute (1977, Satz 8.4.9). □

In the sequel, we study the particular case $(M, \mathfrak{B}(M)) = (\mathbb{R}, \mathfrak{B})$, i.e., convergence in distribution for random variables. The Central Limit Theorem deals with this notion of convergence, see the introductory Example I.1 and Section IV.5.
Notation: for any $Q \in \mathfrak{M}(\mathbb{R})$

$$F_Q(x) = Q(]-\infty, x]), \qquad x \in \mathbb{R},$$

and for any function $F : \mathbb{R} \to \mathbb{R}$

$$\mathrm{Cont}(F) = \{x \in \mathbb{R} : F \text{ continuous at } x\}.$$

Theorem 2.

$$Q_n \xrightarrow{\mathrm{w}} Q \quad \Leftrightarrow \quad \forall\, x \in \mathrm{Cont}(F_Q) : \lim_{n\to\infty} F_{Q_n}(x) = F_Q(x).$$

Moreover, if $Q_n \xrightarrow{\mathrm{w}} Q$ and $\mathrm{Cont}(F_Q) = \mathbb{R}$ then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_{Q_n}(x) - F_Q(x)| = 0.$$

Proof. '$\Rightarrow$': If $x \in \mathrm{Cont}(F_Q)$ and $A =]-\infty, x]$ then $Q(\partial A) = Q(\{x\}) = 0$, see Theorem 1.2. Hence Theorem 1 implies

$$\lim_{n\to\infty} F_{Q_n}(x) = \lim_{n\to\infty} Q_n(A) = Q(A) = F_Q(x).$$

'$\Leftarrow$': Consider a non-empty open set $A \subset \mathbb{R}$. Take pairwise disjoint open intervals $A_1, A_2, \dots$ such that $A = \bigcup_{i=1}^{\infty} A_i$. Fatou's Lemma implies

$$\liminf_{n\to\infty} Q_n(A) = \liminf_{n\to\infty} \sum_{i=1}^{\infty} Q_n(A_i) \geq \sum_{i=1}^{\infty} \liminf_{n\to\infty} Q_n(A_i).$$

Note that $\mathbb{R} \setminus \mathrm{Cont}(F_Q)$ is countable. Fix $\varepsilon > 0$, and take

$$A'_i =]a'_i, b'_i] \subset A_i$$

for $i \in \mathbb{N}$ such that

$$a'_i, b'_i \in \mathrm{Cont}(F_Q) \;\wedge\; Q(A_i) \leq Q(A'_i) + \varepsilon \cdot 2^{-i}.$$

Then

$$\liminf_{n\to\infty} Q_n(A_i) \geq \liminf_{n\to\infty} Q_n(A'_i) = Q(A'_i) \geq Q(A_i) - \varepsilon \cdot 2^{-i}.$$

We conclude that

$$\liminf_{n\to\infty} Q_n(A) \geq Q(A) - \varepsilon,$$

and therefore $Q_n \xrightarrow{\mathrm{w}} Q$ by Theorem 1.

Uniform convergence, see Einführung in die Stochastik. □

Corollary 1.

$$Q_n \xrightarrow{\mathrm{w}} Q \;\wedge\; Q_n \xrightarrow{\mathrm{w}} \widetilde{Q} \quad \Rightarrow \quad Q = \widetilde{Q}.$$

Proof. By Theorem 2 $F_Q(x) = F_{\widetilde{Q}}(x)$ if $x \in D = \mathrm{Cont}(F_Q) \cap \mathrm{Cont}(F_{\widetilde{Q}})$. Since D is dense in $\mathbb{R}$ and F_Q as well as $F_{\widetilde{Q}}$ are right-continuous, we get $F_Q = F_{\widetilde{Q}}$. Apply Theorem 1.3. □

Given: random variables X_n, X on $(\Omega, \mathfrak{A}, P)$ for $n \in \mathbb{N}$.

Theorem 3.

$$X_n \xrightarrow{P} X \quad \Rightarrow \quad X_n \xrightarrow{\mathrm{d}} X$$

and

$$X_n \xrightarrow{\mathrm{d}} X \;\wedge\; X \text{ constant a.s.} \quad \Rightarrow \quad X_n \xrightarrow{P} X.$$

Proof. Assume $X_n \xrightarrow{P} X$. For $\varepsilon > 0$ and $x \in \mathbb{R}$

$$\begin{aligned} &P(\{X \le x - \varepsilon\}) - P(\{|X - X_n| > \varepsilon\}) \\ &\quad \le P(\{X \le x - \varepsilon\} \cap \{|X - X_n| \le \varepsilon\}) \\ &\quad \le P(\{X_n \le x\}) \\ &\quad = P(\{X_n \le x\} \cap \{X \le x + \varepsilon\}) + P(\{X_n \le x\} \cap \{X > x + \varepsilon\}) \\ &\quad \le P(\{X \le x + \varepsilon\}) + P(\{|X - X_n| > \varepsilon\}). \end{aligned}$$

Thus

$$F_X(x - \varepsilon) \le \liminf_{n\to\infty} F_{X_n}(x) \le \limsup_{n\to\infty} F_{X_n}(x) \le F_X(x + \varepsilon).$$

For $x \in \mathrm{Cont}(F_X)$ we get $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$. Apply Theorem 2.

Now, assume that $X_n \xrightarrow{\mathrm{d}} X$ and $P_X = \varepsilon_x$. Let $\varepsilon > 0$ and take $f \in C^b(\mathbb{R})$ such that $f \ge 0$, $f(x) = 0$, and $f(y) = 1$ if $|x - y| \ge \varepsilon$. Then

$$P(\{|X - X_n| > \varepsilon\}) = P(\{|x - X_n| > \varepsilon\}) = \int 1_{\mathbb{R}\setminus[x-\varepsilon,x+\varepsilon]}\, dP_{X_n} \le \int f\, dP_{X_n}$$

and

$$\lim_{n\to\infty} \int f\, dP_{X_n} = \int f\, dP_X = 0.$$

□

Example 2. Consider the uniform distribution P on $\Omega = \{0, 1\}$. Put

$$X_n(\omega) = \omega, \qquad X(\omega) = 1 - \omega.$$

Then $P_{X_n} = P_X$ and therefore

$$X_n \xrightarrow{\mathrm{d}} X.$$

However, $\{|X_n - X| < 1/2\} = \emptyset$ and therefore

$$X_n \xrightarrow{P} X \text{ does not hold.}$$

Theorem 4 (Skorohod)**.** There exists a probability space $(\Omega, \mathfrak{A}, P)$ with the following property. If

$$Q_n \xrightarrow{\mathrm{w}} Q,$$

then there exist X_n, $X \in \mathfrak{Z}(\Omega, \mathfrak{A})$ for $n \in \mathbb{N}$ such that

$$\forall\, n \in \mathbb{N} : \; Q_n = P_{X_n} \quad \wedge \quad Q = P_X \quad \wedge \quad X_n \xrightarrow{P\text{-a.s.}} X.$$

Proof. Take $\Omega = \,]0, 1[$, $\mathfrak{A} = \mathfrak{B}(\Omega)$, and consider the uniform distribution P on Ω. Define

$$X_Q(\omega) = \inf\{z \in \mathbb{R} : \omega \le F_Q(z)\}, \qquad \omega \in \,]0, 1[\,,$$

for any $Q \in \mathfrak{M}(\mathbb{R})$. Since X_Q is non-decreasing, we have $X_Q \in \mathfrak{Z}(\Omega, \mathfrak{A})$. Furthermore,

$$P_{X_Q} = Q, \tag{1}$$

see Einführung in die Stochastik.
Assuming $Q_n \xrightarrow{\text{w}} Q$ we define $X_n = X_{Q_n}$ and $X = X_Q$. Since X is non-decreasing, we conclude that $\Omega \setminus \text{Cont}(X)$ is countable. Thus it suffices to show

$$\forall\, \omega \in \text{Cont}(X): \lim_{n\to\infty} X_n(\omega) = X(\omega).$$

Let $\omega \in \text{Cont}(X)$ and $\varepsilon > 0$. Put $x = X(\omega)$ and take $x_i \in \text{Cont}(F_Q)$ such that

$$x - \varepsilon < x_1 < x < x_2 < x + \varepsilon.$$

Hence

$$F_Q(x_1) < \omega < F_Q(x_2).$$

By assumption there exists $n_0 \in \mathbb{N}$ such that

$$F_{Q_n}(x_1) < \omega < F_{Q_n}(x_2)$$

for $n \geq n_0$. Hence $X_n(\omega) \in [x_1, x_2]$, i.e. $|X_n(\omega) - X(\omega)| \leq \varepsilon$. □

Remark 3. By (1) we have a general method to transform uniformly distributed 'random numbers' from $]0,1[$ into 'random numbers' with distribution Q.

Remark 4.

(i) Put

$$C^{(r)} = \{f : \mathbb{R} \to \mathbb{R} : f, f^{(1)}, \ldots, f^{(r)} \text{ bounded, uniformly continuous}\}.$$

Then

$$Q_n \xrightarrow{\text{w}} Q \quad \Leftrightarrow \quad \exists\, r \in \mathbb{N}_0\; \forall\, f \in C^{(r)}: \lim_{n\to\infty} \int f\, dQ_n = \int f\, dQ,$$

see Gänssler, Stute (1977, p. 66).

(ii) The *Lévy distance*

$$d(Q,R) = \inf\{h \in\,]0,\infty[\, : \forall\, x \in \mathbb{R} : F_Q(x-h) - h \leq F_R(x) \leq F_Q(x+h) + h\}$$

defines a metric on $\mathfrak{M}(\mathbb{R})$, and

$$Q_n \xrightarrow{\text{w}} Q \quad \Leftrightarrow \quad \lim_{n\to\infty} d(Q_n, Q) = 0,$$

see Chow, Teicher (1978, Thm. 8.1.3).

(iii) Suppose that (M, ρ) is a complete separable metric space. Then there exists a metric d on $\mathfrak{M}(M)$ such that $(\mathfrak{M}(M), d)$ is complete and separable as well, and

$$Q_n \xrightarrow{\text{w}} Q \quad \Leftrightarrow \quad \lim_{n\to\infty} d(Q_n, Q) = 0,$$

see Parthasarathy (1967, Sec. II.6).

Finally, we present a compactness criterion, which is very useful for construction of probability measures on $\mathfrak{B}(M)$.

Lemma 1. Let $x_{n,\ell} \in \mathbb{R}$ for $n, \ell \in \mathbb{N}$ with

$$\forall \ell \in \mathbb{N} : \sup_{n \in \mathbb{N}} |x_{n,\ell}| < \infty.$$

Then there exists an increasing sequence $(n_i)_{i \in \mathbb{N}}$ in $\mathbb{N}$ such that

$$\forall \ell \in \mathbb{N} : (x_{n_i,\ell})_{i \in \mathbb{N}} \text{ converges.}$$

Proof. See Billingsley (1979, Thm. 25.13). □

Definition 2.

(i) $\mathfrak{P} \subset \mathfrak{M}(M)$ *tight* if

$$\forall \varepsilon > 0 \; \exists K \subset M \text{ compact } \forall P \in \mathfrak{P} : \quad P(K) \geq 1 - \varepsilon.$$

(ii) $\mathfrak{P} \subset \mathfrak{M}(M)$ *relatively compact* if every sequence in $\mathfrak{P}$ contains a subsequence that converges weakly.

Theorem 5 (Prohorov)**.** Assume that M is a complete separable metric space and $\mathfrak{P} \subset \mathfrak{M}(M)$. Then

$$\mathfrak{P} \text{ relatively compact} \quad \Leftrightarrow \quad \mathfrak{P} \text{ tight.}$$

Proof. See Parthasarathy (1967, Thm. II.6.7). Here: $M = \mathbb{R}$.

'$\Rightarrow$': Suppose that $\mathfrak{P}$ is not tight. Then, for some $\varepsilon > 0$, there exists a sequence $(P_n)_{n \in \mathbb{N}}$ in $\mathfrak{P}$ such that

$$P_n([-n, n]) < 1 - \varepsilon.$$

For a suitable subsequence, $P_{n_k} \xrightarrow{\text{w}} P \in \mathfrak{M}(\mathbb{R})$. Take $m > 0$ such that

$$P(]-m, m[) > 1 - \varepsilon.$$

Theorem 1 implies

$$P(]-m, m[) \leq \liminf_{k \to \infty} P_{n_k}(]-m, m[) \leq \liminf_{k \to \infty} P_{n_k}([-n_k, n_k]) < 1 - \varepsilon,$$

which is a contradiction.

'$\Leftarrow$': Consider any sequence $(P_n)_{n \in \mathbb{N}}$ in $\mathfrak{P}$ and the corresponding sequence $(F_n)_{n \in \mathbb{N}}$ of distribution functions. Use Lemma 1 to obtain a subsequence $(F_{n_i})_{i \in \mathbb{N}}$ and a non-decreasing function $G : \mathbb{Q} \to [0, 1]$ with

$$\forall q \in \mathbb{Q} : \lim_{i \to \infty} F_{n_i}(q) = G(q).$$

Put

$$F(x) = \inf\{G(q) : q \in \mathbb{Q} \wedge x < q\}, \qquad x \in \mathbb{R}.$$

Claim (*Helly's Theorem*):

(i) F is non-decreasing and right-continuous,

(ii) $\forall\, x \in \mathrm{Cont}(F):\ \lim_{i\to\infty} F_{n_i}(x) = F(x)$.

Proof: Ad (i): Obviously F is non-decreasing. For $x \in \mathbb{R}$ and $\varepsilon > 0$ take $\delta_2 > 0$ such that

$$\forall\, q \in \mathbb{Q} \cap \,]x, x+\delta_2[\,: \quad G(q) \leq F(x) + \varepsilon.$$

Thus, for $z \in \,]x, x+\delta_2[$,

$$F(x) \leq F(z) \leq F(x) + \varepsilon.$$

Ad (ii): If $x \in \mathrm{Cont}(F)$ and $\varepsilon > 0$ take $\delta_1 > 0$ such that

$$F(x) - \varepsilon \leq F(x - \delta_1).$$

Thus, for $q_1, q_2 \in \mathbb{Q}$ with

$$x - \delta_1 < q_1 < x < q_2 < x + \delta_2,$$

we get

$$F(x) - \varepsilon \leq F(x-\delta_1) \leq G(q_1) \leq \liminf_{i\to\infty} F_{n_i}(x) \leq \limsup_{i\to\infty} F_{n_i}(x)$$
$$\leq G(q_2) \leq F(x) + \varepsilon.$$

Claim:

$$\lim_{x\to-\infty} F(x) = 0 \ \wedge\ \lim_{x\to\infty} F(x) = 1.$$

Proof: For $\varepsilon > 0$ take $m \in \mathbb{Q}$ such that

$$\forall\, n \in \mathbb{N}:\ P_n(]-m, m]) \geq 1 - \varepsilon.$$

Thus

$$G(m) - G(-m) = \lim_{i\to\infty} \big(F_{n_i}(m) - F_{n_i}(-m)\big) = \lim_{i\to\infty} P_{n_i}(]-m, m]) \geq 1 - \varepsilon.$$

Since $F(m) \geq G(m)$ and $F(-m-1) \leq G(-m)$, we obtain

$$F(m) - F(-m-1) \geq 1 - \varepsilon.$$

It remains to apply Theorems 1.3 and 2. □

4 Uniform Integrability

In the sequel: X_n, X random variables on a common probability space $(\Omega, \mathfrak{A}, P)$.

Definition 1. $(X_n)_{n\in\mathbb{N}}$ *uniformly integrable (u.i.)* if

$$\lim_{\alpha\to\infty} \sup_{n\in\mathbb{N}} \int_{\{|X_n|\geq\alpha\}} |X_n|\, dP = 0.$$

Remark 1.

(i) $(X_n)_{n\in\mathbb{N}}$ u.i. $\Rightarrow (\forall\, n \in \mathbb{N}:\; X_n \in \mathfrak{L}^1) \wedge \sup_{n\in\mathbb{N}} \|X_n\|_1 < \infty$.

(ii) $\exists\, Y \in \mathfrak{L}^1\; \forall\, n \in \mathbb{N}:\; |X_n| \le Y \Rightarrow (X_n)_{n\in\mathbb{N}}$ u.i.

(iii) $\exists\, p > 1\;\; (\forall\, n \in \mathbb{N} : X_n \in \mathfrak{L}^p) \wedge \sup_{n\in\mathbb{N}} \|X_n\|_p < \infty \Rightarrow (X_n)_{n\in\mathbb{N}}$ u.i.
Proof: $\int_{\{|X_n|\ge\alpha\}} |X_n|\, dP = 1/\alpha^{p-1} \cdot \int_{\{|X_n|\ge\alpha\}} \alpha^{p-1} |X_n|\, dP \le 1/\alpha^{p-1} \cdot \|X_n\|_p^p$.

Example 1. For the uniform distribution P on $[0,1]$ and

$$X_n = n \cdot 1_{[0,1/n]}$$

we have $X_n \in \mathfrak{L}^1$ and $\|X_n\|_1 = 1$, but for any $\alpha > 0$ and $n \ge \alpha$

$$\int_{\{|X_n|\ge\alpha\}} |X_n|\, dP = n \cdot P([0,1/n]) = 1,$$

so that $(X_n)_{n\in\mathbb{N}}$ is not u.i.

Lemma 1. $(X_n)_{n\in\mathbb{N}}$ u.i. iff

$$\sup_{n\in\mathbb{N}} \mathrm{E}(|X_n|) < \infty \tag{1}$$

and

$$\forall\, \varepsilon > 0\; \exists\, \delta > 0\; \forall\, A \in \mathfrak{A}: \quad P(A) < \delta \;\Rightarrow\; \sup_{n\in\mathbb{N}} \int_A |X_n|\, dP < \varepsilon. \tag{2}$$

Proof. '$\Rightarrow$': For (1), see Remark 1.(i). Moreover,

$$\begin{aligned}\int_A |X_n|\, dP &= \int_{A\cap\{|X_n|\ge\alpha\}} |X_n|\, dP + \int_{A\cap\{|X_n|<\alpha\}} |X_n|\, dP \\ &\le \int_{\{|X_n|\ge\alpha\}} |X_n|\, dP + \alpha \cdot P(A).\end{aligned}$$

For $\varepsilon > 0$ take $\alpha > 0$ with

$$\sup_{n\in\mathbb{N}} \int_{\{|X_n|\ge\alpha\}} |X_n|\, dP < \varepsilon/2$$

and $\delta = \varepsilon/(2\alpha)$ to obtain (2).
'$\Leftarrow$': Put $M = \sup_{n\in\mathbb{N}} \mathrm{E}(|X_n|)$. Then

$$M \ge \int_{\{|X_n|\ge\alpha\}} |X_n|\, dP \ge \alpha \cdot P(\{|X_n| \ge \alpha\}).$$

Hence $P(\{|X_n| \ge \alpha\}) \le M/\alpha$. Let $\varepsilon > 0$, take $\delta > 0$ according to (2) to obtain for $\alpha > M/\delta$

$$\sup_{n\in\mathbb{N}} \int_{\{|X_n|\ge\alpha\}} |X_n|\, dP < \varepsilon.$$

□

Theorem 1. Let $1 \leq p < \infty$, and assume $X_n \in \mathfrak{L}^p$ for every $n \in \mathbb{N}$. Then

$$(X_n)_{n\in\mathbb{N}} \text{ converges in } \mathfrak{L}^p$$

iff

$$(X_n)_{n\in\mathbb{N}} \text{ converges in probability } \ \wedge\ (|X_n|^p)_{n\in\mathbb{N}} \text{ is u.i.}$$

Proof. '$\Rightarrow$': Assume $X_n \xrightarrow{\mathfrak{L}^p} X$. From Remark 2.1 we get $X_n \xrightarrow{P} X$. For every $A \in \mathfrak{A}$

$$\|1_A \cdot X_n\|_p \leq \|1_A \cdot (X_n - X)\|_p + \|1_A \cdot X\|_p.$$

Take $A = \Omega$ to obtain $\sup_{n\in\mathbb{N}} \mathrm{E}(|X_n|^p) < \infty$. Let $\varepsilon > 0$, take $k \in \mathbb{N}$ such that

$$\sup_{n>k} \|X_n - X\|_p < \varepsilon. \tag{3}$$

Put $X_0 = 0$. Note that

$$\sup_{0\leq n\leq k} |X_n - X|^p \leq \sum_{n=0}^{k} |X_n - X|^p \in \mathfrak{L}^1.$$

Hence, by Remark 1.(ii),

$$(|X_1 - X|^p, \dots, |X_k - X|^p, |X|^p, |X|^p, \dots) \text{ u.i.}$$

By Lemma 1

$$P(A) < \delta \quad \Rightarrow \quad \sup_{0\leq n\leq k} \|1_A \cdot (X_n - X)\|_p < \varepsilon.$$

for a suitable $\delta > 0$. Together with (3) this implies

$$P(A) < \delta \quad \Rightarrow \quad \sup_{n\in\mathbb{N}} \|1_A \cdot X_n\|_p < 2 \cdot \varepsilon.$$

'$\Leftarrow$': Let $\varepsilon > 0$, put $A = A_{m,n} = \{|X_m - X_n| > \varepsilon\}$. Then

$$\begin{aligned} \|X_m - X_n\|_p &\leq \|1_A \cdot (X_m - X_n)\|_p + \|1_{A^c} \cdot (X_m - X_n)\|_p \\ &\leq \|1_A \cdot X_m\|_p + \|1_A \cdot X_n\|_p + \varepsilon. \end{aligned}$$

By assumption $X_n \xrightarrow{P} X$ for some $X \in \mathfrak{Z}(\Omega, \mathfrak{A})$. Take $\delta > 0$ according to (2) for $(|X_n|^p)_{n\in\mathbb{N}}$, and note that

$$A_{m,n} \subset \{|X_m - X| > \varepsilon/2\} \cup \{|X_n - X| > \varepsilon/2\}.$$

Hence, for m, n sufficiently large,

$$P(A_{m,n}) < \delta,$$

which implies

$$\|X_m - X_n\|_p \leq 2 \cdot \varepsilon^{1/p} + \varepsilon.$$

Apply Theorem II.6.3. □

Remark 2.

(i) Theorem 1 yields a generalization of Lebesgue's convergence theorem: If $X_n \in \mathfrak{L}^1$ for every $n \in \mathbb{N}$ and $X_n \xrightarrow{P\text{-a.s.}} X$, then

$$(X_n)_{n\in\mathbb{N}} \text{ u.i.} \quad \Leftrightarrow \quad X \in \mathfrak{L}^1 \wedge X_n \xrightarrow{\mathfrak{L}^1} X.$$

(ii) Uniform integrability is a property of the distributions only.

Theorem 2.

$$X_n \xrightarrow{d} X \quad \Rightarrow \quad \mathrm{E}(|X|) \leq \liminf_{n\to\infty} \mathrm{E}(|X_n|).$$

Proof. From Skorohod's Theorem 3.4 we get a probability space $(\widetilde{\Omega}, \widetilde{\mathfrak{A}}, \widetilde{P})$ with random variables $\widetilde{X}_n$, $\widetilde{X}$ such that

$$\widetilde{X}_n \xrightarrow{\widetilde{P}\text{-a.s.}} \widetilde{X} \quad \wedge \quad \widetilde{P}_{\widetilde{X}_n} = P_{X_n} \quad \wedge \quad \widetilde{P}_{\widetilde{X}} = P_X.$$

Thus $\mathrm{E}(|X|) = \mathrm{E}(|\widetilde{X}|)$ and $\mathrm{E}(|X_n|) = \mathrm{E}(|\widetilde{X}_n|)$. Apply Fatou's Lemma II.5.2. □

Theorem 3. If

$$X_n \xrightarrow{\mathrm{d}} X \quad \wedge \quad (X_n)_{n\in\mathbb{N}} \text{ u.i.}$$

then

$$X \in \mathfrak{L}^1 \quad \wedge \quad \lim_{n\to\infty} \mathrm{E}(X_n) = \mathrm{E}(X).$$

Proof. Notation as previously. Now $(|\widetilde{X}_n|)_{n\in\mathbb{N}}$ is u.i., see Remark 2.(ii). Hence, by Remark 2.(i), $\widetilde{X} \in \mathfrak{L}^1$ and $\widetilde{X}_n \xrightarrow{\mathfrak{L}^1} \widetilde{X}$. Thus $\mathrm{E}(|X|) < \infty$ and

$$\lim_{n\to\infty} \mathrm{E}(X_n) = \lim_{n\to\infty} \mathrm{E}(\widetilde{X}_n) = \mathrm{E}(\widetilde{X}) = \mathrm{E}(X).$$

□

Example 2. Example 1 continued. With $X = 0$ we have $X_n \xrightarrow{P\text{-a.s.}} X$, and therefore $X_n \xrightarrow{\mathrm{d}} X$. But $\mathrm{E}(X_n) = 1 > 0 = \mathrm{E}(X)$.

5 Independence

'…the concept of independence …plays a central role in probability theory; it is precisely this concept that distinguishes probability theory from the general theory of measure spaces', see Shiryayev (1984, p. 27).

In the sequel, $(\Omega, \mathfrak{A}, P)$ denotes a probability space and I is a non-empty set.

Definition 1. Let $A_i \in \mathfrak{A}$ for $i \in I$. Then $(A_i)_{i\in I}$ is *independent* if

$$P\Big(\bigcap_{i\in S} A_i\Big) = \prod_{i\in S} P(A_i) \tag{1}$$

for every $S \in \mathfrak{P}_0(I)$. Elementary case: $|I| = 2$.

In the sequel, $\mathfrak{E}_i \subset \mathfrak{A}$ for $i \in I$.

Definition 2. $(\mathfrak{E}_i)_{i\in I}$ is *independent* if (1) holds for every $S \in \mathfrak{P}_0(I)$ and all $A_i \in \mathfrak{E}_i$ for $i \in S$.

Remark 1.

(i) $(\mathfrak{E}_i)_{i\in I}$ independent $\wedge\ \forall\, i \in I:\ \widetilde{\mathfrak{E}}_i \subset \mathfrak{E}_i \ \Rightarrow\ (\widetilde{\mathfrak{E}}_i)_{i\in I}$ independent.

(ii) $(\mathfrak{E}_i)_{i\in I}$ independent $\ \Leftrightarrow\ \forall\, S \in \mathfrak{P}_0(I):\ (\mathfrak{E}_i)_{i\in S}$ independent.

Lemma 1.

$$(\mathfrak{E}_i)_{i\in I} \text{ independent} \quad \Rightarrow \quad (\delta(\mathfrak{E}_i))_{i\in I} \text{ independent.}$$

Proof. Without loss of generality, $I = \{1,\dots,n\}$ and $n \geq 2$, see Remark 1.(ii). Put

$$\mathfrak{D}_1 = \{A \in \delta(\mathfrak{E}_1) : (\{A\}, \mathfrak{E}_2, \dots, \mathfrak{E}_n) \text{ independent}\}.$$

Then $\mathfrak{D}_1$ is a Dynkin class and $\mathfrak{E}_1 \subset \mathfrak{D}_1$, hence $\delta(\mathfrak{E}_1) = \mathfrak{D}_1$. Thus

$$(\delta(\mathfrak{E}_1), \mathfrak{E}_2, \dots, \mathfrak{E}_n) \text{ independent.}$$

Repeat this step for $2,\dots,n$. □

Theorem 1. If

$$(\mathfrak{E}_i)_{i\in I} \text{ independent } \wedge\ \forall\, i \in I:\ \mathfrak{E}_i \text{ closed w.r.t. intersections} \tag{2}$$

then

$$(\sigma(\mathfrak{E}_i))_{i\in I} \text{ independent.}$$

Proof. Use Theorem II.1.2 and Lemma 1. □

Corollary 1. Assume that $I = \bigcup_{j\in J} I_j$ for pairwise disjoint sets $I_j \neq \emptyset$. If (2) holds, then

$$\left(\sigma\left(\bigcup_{i\in I_j} \mathfrak{E}_i\right)\right)_{j\in J} \text{ independent.}$$

Proof. Let

$$\widetilde{\mathfrak{E}}_j = \left\{\bigcap_{i\in S} A_i : S \in \mathfrak{P}_0(I_j) \wedge A_i \in \mathfrak{E}_i \text{ for } i \in S\right\}.$$

Then $\widetilde{\mathfrak{E}}_j$ is closed w.r.t. intersections and $\left(\widetilde{\mathfrak{E}}_j\right)_{j\in J}$ is independent. Finally

$$\sigma\left(\bigcup_{i\in I_j} \mathfrak{E}_i\right) = \sigma(\widetilde{\mathfrak{E}}_j).$$

□

In the sequel, $(\Omega_i, \mathfrak{A}_i)$ denotes a measurable space for $i \in I$, and $X_i : \Omega \to \Omega_i$ is $\mathfrak{A}$-$\mathfrak{A}_i$-measurable for $i \in I$.

Definition 3. $(X_i)_{i\in I}$ is *independent* if $(\sigma(X_i))_{i\in I}$ is independent.

Example 1. Actually, the essence of independence. Assume that

$$(\Omega, \mathfrak{A}, P) = \left(\bigtimes_{i\in I} \Omega_i, \bigotimes_{i\in I} \mathfrak{A}_i, \bigtimes_{i\in I} P_i \right)$$

for probability measures P_i on $\mathfrak{A}_i$. Let

$$X_i = \pi_i.$$

Then, for $S \in \mathfrak{P}_0(I)$ and $A_i \in \mathfrak{A}_i$ for $i \in S$

$$P\left(\bigcap_{i\in S} \{X_i \in A_i\} \right) = P\left(\bigtimes_{i\in S} A_i \times \bigtimes_{i\in I\setminus S} \Omega_i \right) = \prod_{i\in S} P_i(A_i) = \prod_{i\in S} P(\{X_i \in A_i\}).$$

Hence $(X_i)_{i\in I}$ is independent. Furthermore, $P_{X_i} = P_i$.

Recall the question that was posed in the introductory Example I.2.

Theorem 2. Given: probability spaces $(\Omega_i, \mathfrak{A}_i, P_i)$ for $i \in I$. Then there exist

(i) a probability space $(\Omega, \mathfrak{A}, P)$ and

(ii) $\mathfrak{A}$-$\mathfrak{A}_i$-measurable mappings $X_i : \Omega \to \Omega_i$ for $i \in I$

such that

$$(X_i)_{i\in I} \text{ independent} \quad \wedge \quad \forall\, i \in I: \; P_{X_i} = P_i.$$

Proof. See Example 1. □

Theorem 3. Let $\mathfrak{F}_i \subset \mathfrak{A}_i$ for $i \in I$. If

$$\forall\, i \in I: \; \sigma(\mathfrak{F}_i) = \mathfrak{A}_i \; \wedge \; \mathfrak{F}_i \text{ closed w.r.t. intersections}$$

then

$$(X_i)_{i\in I} \text{ independent} \quad \Leftrightarrow \quad (X_i^{-1}(\mathfrak{F}_i))_{i\in I} \text{ independent}.$$

Proof. Recall that $\sigma(X_i) = X_i^{-1}(\mathfrak{A}_i) = \sigma(X_i^{-1}(\mathfrak{F}_i))$. '⇒': See Remark 1.(i). '⇐': Note that $X_i^{-1}(\mathfrak{F}_i)$ is closed w.r.t. intersections. Use Theorem 1. □

Example 2. Independence of a family of random variables X_i, i.e., $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}, \mathfrak{B})$ for $i \in I$. In this case $(X_i)_{i\in I}$ is independent iff

$$\forall\, S \in \mathfrak{P}_0(I)\; \forall\, c_i \in \mathbb{R}, i \in S : P\left(\bigcap_{i\in S} \{X_i \le c_i\} \right) = \prod_{i\in S} P(\{X_i \le c_i\}).$$

Theorem 4. Let

(i) $I = \bigcup_{j\in J} I_j$ for pairwise disjoint sets $I_j \neq \emptyset$,

(ii) $(\widetilde{\Omega}_j, \widetilde{\mathfrak{A}}_j)$ be measurable spaces for $j \in J$,

(iii) $f_j : \times_{i\in I_j} \Omega_i \to \widetilde{\Omega}_j$ be $\left(\bigotimes_{i\in I_j} \mathfrak{A}_i\right)$-$\widetilde{\mathfrak{A}}_j$ measurable mappings for $j \in J$.

Put

$$Y_j = (X_i)_{i\in I_j} : \Omega \to \bigtimes_{i\in I_j} \Omega_i.$$

Then

$$(X_i)_{i\in I} \text{ independent} \quad \Rightarrow \quad (f_j \circ Y_j)_{j\in J} \text{ independent.}$$

Proof.

$$\begin{aligned}\sigma(f_j \circ Y_j) &= Y_j^{-1}(f_j^{-1}(\widetilde{\mathfrak{A}}_j)) \subset Y_j^{-1}\left(\bigotimes_{i\in I_j} \mathfrak{A}_i\right) \\ &= \sigma(\{X_i : i \in I_j\}) = \sigma\left(\bigcup_{i\in I_j} X_i^{-1}(\mathfrak{A}_i)\right).\end{aligned}$$

Use Corollary 1 and Remark 1.(i). □

Example 3. For an independent sequence $(X_i)_{i\in\mathbb{N}}$ of random variables

$$\left(\max(X_1, X_3),\ 1_{\mathbb{R}_+}(X_2),\ \limsup_{n\to\infty} 1/n \sum_{i=1}^{n} X_i\right)$$

are independent.

Remark 2. Consider the mapping

$$X : \Omega \to \bigtimes_{i\in I} \Omega_i : \omega \mapsto (X_i(\omega))_{i\in I}.$$

Clearly X is $\mathfrak{A}$-$\bigotimes_{i\in I} \mathfrak{A}_i$-measurable. By definition, $P_X(A) = P(\{X \in A\})$ for $A \in \bigotimes_{i\in I} \mathfrak{A}_i$. In particular, for measurable rectangles $A \in \bigotimes_{i\in I} \mathfrak{A}_i$, i.e.,

$$A = \bigtimes_{i\in S} A_i \times \bigtimes_{i\in I\setminus S} \Omega_i \tag{3}$$

with $S \in \mathfrak{P}_0(I)$ and $A_i \in \mathfrak{A}_i$,

$$P_X(A) = P\left(\bigcap_{i\in S} \{X_i \in A_i\}\right). \tag{4}$$

Definition 4.

(i) P_X is called the *joint distribution* of random elements X_i, $i \in I$.

(ii) Let P denote a probability measure on $(\times_{i\in I}\Omega_i, \bigotimes_{i\in I}\mathfrak{A}_i)$, and let $i \in I$. Then P_{π_i} is called a *(one-dimensional) marginal distribution* of P.

Example 4. Let $\Omega = \{1,\dots,6\}^2$ and consider the uniform distribution P on $\mathfrak{A} = \mathfrak{P}(\Omega)$, which is a model for rolling a die twice.
Moreover, let $\Omega_i = \mathbb{N}$ and $\mathfrak{A}_i = \mathfrak{P}(\Omega_i)$ such that $\bigotimes_{i=1}^2 \mathfrak{A}_i = \mathfrak{P}(\mathbb{N}^2)$. Consider the random variables

$$X_1(\omega_1,\omega_2) = \omega_1, \qquad X_2(\omega_1,\omega_2) = \omega_1 + \omega_2.$$

Then

$$P_X(A) = \frac{|A \cap M|}{36}, \qquad A \subset \mathbb{N}^2,$$

where

$$M = \{(k,\ell) \in \mathbb{N}^2 : 1 \le k \le 6 \wedge k+1 \le \ell \le k+6\}$$

Claim: (X_1, X_2) are not independent. Proof:

$$P(\{X_1 = 1\} \cap \{X_2 = 3\}) = P_X(\{(1,3)\}) = P(\{(1,2)\}) = 1/36$$

but

$$P(\{X_1 = 1\}) \cdot P(\{X_2 = 3\}) = 1/6 \cdot P(\{(1,2),(2,1)\}) = 1/3 \cdot 1/36.$$

We add that

$$P_{X_1} = \sum_{k=1}^{6} 1/6 \cdot \varepsilon_k, \qquad P_{X_2} = \sum_{\ell=2}^{12} (6 - |\ell - 7|)/36 \cdot \varepsilon_\ell.$$

Theorem 5.

$$(X_i)_{i\in I} \text{ independent} \quad \Leftrightarrow \quad P_X = \bigtimes_{i\in I} P_{X_i}.$$

Proof. For A given by (3)

$$\Big(\bigtimes_{i\in I} P_{X_i}\Big)(A) = \prod_{i\in S} P_{X_i}(A_i) = \prod_{i\in S} P(\{X_i \in A_i\}).$$

On the other hand, we have (4). Thus '$\Leftarrow$' hold trivially. Use Theorem II.4.4 to obtain '$\Rightarrow$'. $\square$

In the sequel, we consider random variables X_i, i.e., $(\Omega_i, \mathfrak{A}_i) = (\mathbb{R}, \mathfrak{B})$ for $i \in I$.

Theorem 6. Let $I = \{1,\dots,n\}$. If

$$(X_1,\dots,X_n) \text{ independent} \ \wedge\ \forall\, i \in I:\ X_i \ge 0 \ (X_i \text{ integrable})$$

then ($\prod_{i=1}^n X_i$ is integrable and)

$$\mathrm{E}\Big(\prod_{i=1}^n X_i\Big) = \prod_{i=1}^n \mathrm{E}(X_i).$$

Proof. Use Fubini's Theorem and Theorem 5 to obtain

$$\begin{aligned}
\mathrm{E}\Bigg(\bigg|\prod_{i=1}^{n} X_i\bigg|\Bigg) &= \int_{\mathbb{R}^n} |x_1 \cdot \dots \cdot x_n| \, P_{(X_1,\dots,X_n)}(d(x_1,\dots,x_n)) \\
&= \int_{\mathbb{R}^n} |x_1 \cdot \dots \cdot x_n| \, (P_{X_1} \times \dots \times P_{X_n})(d(x_1,\dots,x_n)) \\
&= \prod_{i=1}^{n} \int_{\mathbb{R}} |x_i| \, P_{X_i}(dx_i) = \prod_{i=1}^{n} \mathrm{E}(|X_i|).
\end{aligned}$$

Drop $|\cdot|$ if the random variables are integrable. □

Definition 5. $X_1, X_2 \in \mathfrak{L}^2$ are *uncorrelated* if

$$\mathrm{E}(X_1 \cdot X_2) = \mathrm{E}(X_1) \cdot \mathrm{E}(X_2).$$

Theorem 7 (Bienaymé)**.** Let $X_1, \dots, X_n \in \mathfrak{L}^2$ be pairwise uncorrelated. Then

$$\mathrm{Var}\bigg(\sum_{i=1}^{n} X_i\bigg) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Proof. We have

$$\begin{aligned}
\mathrm{Var}\bigg(\sum_{i=1}^{n} X_i\bigg) &= \mathrm{E}\bigg(\sum_{i=1}^{n} (X_i - \mathrm{E}(X_i))\bigg)^2 \\
&= \sum_{i=1}^{n} \mathrm{E}(X_i - \mathrm{E}(X_i))^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \mathrm{E}((X_i - \mathrm{E}(X_i)) \cdot (X_j - \mathrm{E}(X_j))).
\end{aligned}$$

Moreover,

$$\mathrm{E}((X_i - \mathrm{E}(X_i)) \cdot (X_j - \mathrm{E}(X_j))) = \mathrm{E}(X_i \cdot X_j) - \mathrm{E}(X_i) \cdot \mathrm{E}(X_j).$$

(The latter quantity is called the *covariance* between X_i and X_j.) □

Definition 6. The *convolution product* of probability measures $P_1, \dots, P_n$ on $\mathfrak{B}$ is defined by

$$P_1 * \dots * P_n = s(P_1 \times \dots \times P_n)$$

where

$$s(x_1, \dots, x_n) = x_1 + \dots + x_n.$$

Theorem 8. Let $(X_1, \dots, X_n)$ be independent and $S = \sum_{i=1}^{n} X_i$. Then

$$P_S = P_{X_1} * \dots * P_{X_n}.$$

Proof. Put $X = (X_1, \dots, X_n)$. Since $S = s \circ (X_1, \dots, X_n)$ we get

$$P_S = s(P_X) = s(P_{X_1} \times \dots \times P_{X_n}).$$

□

Remark 3. The class of probability measure on $\mathfrak{B}$ forms an abelian semi-group w.r.t. $*$, and $P * \varepsilon_0 = P$.

Theorem 9. For all probability measures P_1, P_2 on $\mathfrak{B}$ and every $P_1 * P_2$-integrable function f

$$\int_{\mathbb{R}} f \, d(P_1 * P_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x+y) \, P_1(dx) \, P_2(dy).$$

If $P_1 = h_1 \cdot \lambda_1$ then $P_1 * P_2 = h \cdot \lambda_1$ with

$$h(x) = \int_{\mathbb{R}} h_1(x-y) \, P_2(dy).$$

If $P_2 = h_2 \cdot \lambda_1$, additionally, then

$$h(x) = \int_{\mathbb{R}} h_1(x-y) \cdot h_2(y) \, \lambda(dy).$$

Proof. Use Fubini's Theorem and the transformation theorem. See Billingsley (1979, p. 230). □

Example 5.

(i) Put $N(\mu, 0) = \varepsilon_\mu$. By Theorem 9

$$N(\mu_1, \sigma_1^2) * N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

for $\mu_i \in \mathbb{R}$ and $\sigma_i \geq 0$.

(ii) Consider n independent Bernoulli trials, i.e., $(X_1, \dots, X_n)$ independent with

$$P_{X_i} = p \cdot \varepsilon_1 + (1-p) \cdot \varepsilon_0$$

for every $i \in \{1, \dots, n\}$, where $p \in [0, 1]$. Inductively, we get for $k \in \{1, \dots, n\}$

$$\sum_{i=1}^{k} X_i \sim B(k, p),$$

see also Übung 7.3. Thus, for any $n, m \in \mathbb{N}$,

$$B(n, p) * B(m, p) = B(n + m, p).$$

Chapter IV

Limit Theorems

Given: a sequence of random variables X_n, $n \in \mathbb{N}$, on a probability space $(\Omega, \mathfrak{A}, P)$. Put

$$S_n = \sum_{i=1}^{n} X_i, \qquad n \in \mathbb{N}.$$

For instance, S_n might be the cumulative gain after n trials or (one of the coordinates of) the position of a particle after n collisions.

Question: Convergence of S_n/a_n for suitable weights $0 < a_n \uparrow \infty$ in a suitable sense? Particular case: $a_n = n$.

1 Zero-One Laws

Definition 1. For σ-algebras $\mathfrak{A}_n \subset \mathfrak{A}$, $n \in \mathbb{N}$, the corresponding *tail σ-algebra* is

$$\mathfrak{A}_\infty = \bigcap_{n \in \mathbb{N}} \sigma\Big(\bigcup_{m \geq n} \mathfrak{A}_m \Big),$$

and $A \in \mathfrak{A}_\infty$ is called a *tail (terminal) event*.

Example 1. Let $\mathfrak{A}_n = \sigma(X_n)$. Put $\mathfrak{C} = \bigotimes_{i=1}^{\infty} \mathfrak{B}$. Then

$$\mathfrak{A}_\infty = \bigcap_{n \in \mathbb{N}} \sigma(\{X_m : m \geq n\})$$

and

$$A \in \mathfrak{A}_\infty \quad \Leftrightarrow \quad \forall\, n \in \mathbb{N}\ \exists\, C \in \mathfrak{C} : \ A = \{(X_n, X_{n+1}, \dots) \in C\}.$$

For instance,

$$\{(S_n)_{n \in \mathbb{N}} \text{ converges}\},\ \{(S_n/a_n)_{n \in \mathbb{N}} \text{ converges}\} \in \mathfrak{A}_\infty,$$

and the function $\liminf_{n\to\infty} S_n/a_n$ is $\mathfrak{A}_\infty$-$\overline{\mathfrak{B}}$-measurable. However, S_n as well as $\liminf_{n\to\infty} S_n$ are not $\mathfrak{A}_\infty$-$\overline{\mathfrak{B}}$-measurable, in general. Analogously for the $\limsup$'s.

Theorem 1 (Kolmogorov's Zero-One Law)**.** Let $(\mathfrak{A}_n)_{n\in\mathbb{N}}$ be an independent sequence of σ-algebras $\mathfrak{A}_n \subset \mathfrak{A}$. Then

$$\forall\, A \in \mathfrak{A}_\infty : \ P(A) \in \{0,1\}\,.$$

Proof. We show that $\mathfrak{A}_\infty$ and $\mathfrak{A}_\infty$ are independent (*terminology*), which implies $P(A) = P(A)\cdot P(A)$ for every $A \in \mathfrak{A}_\infty$. Put

$$\overline{\mathfrak{A}}_n = \sigma(\mathfrak{A}_1 \cup \dots \cup \mathfrak{A}_n).$$

Note that $\mathfrak{A}_\infty \subset \sigma(\mathfrak{A}_{n+1} \cup \dots)$. By Corollary III.5.1 and Remark III.5.1.(i)

$$\overline{\mathfrak{A}}_n, \mathfrak{A}_\infty \text{ independent,}$$

and therefore $\bigcup_{n\in\mathbb{N}} \overline{\mathfrak{A}}_n$ and $\mathfrak{A}_\infty$ are independent, too. Thus, by Theorem III.5.1,

$$\sigma\Big(\bigcup_{n\in\mathbb{N}} \overline{\mathfrak{A}}_n\Big), \mathfrak{A}_\infty \text{ independent.}$$

Finally,

$$\mathfrak{A}_\infty \subset \sigma\Big(\bigcup_{n\in\mathbb{N}} \mathfrak{A}_n\Big) = \sigma\Big(\bigcup_{n\in\mathbb{N}} \overline{\mathfrak{A}}_n\Big).$$

□

Corollary 1. Let $X \in \overline{\mathfrak{Z}}(\Omega, \mathfrak{A}_\infty)$. Under the assumptions of Theorem 1, X is constant P-a.s.

Remark 1. Assume that $(X_n)_{n\in\mathbb{N}}$ is independent. Then

$$P(\{(S_n)_{n\in\mathbb{N}} \text{ converges}\}),\ P(\{(S_n/a_n)_{n\in\mathbb{N}} \text{ converges}\}) \in \{0,1\}.$$

In case of convergence P-a.s., $\lim_{n\to\infty} S_n/a_n$ is constant P-a.s.

Definition 2. Let $A_n \in \mathfrak{A}$ for $n \in \mathbb{N}$. Then

$$\liminf_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} A_m, \qquad \limsup_{n\to\infty} A_n = \bigcap_{n\in\mathbb{N}} \bigcup_{m\geq n} A_m.$$

Remark 2.

(i) $\Big(\liminf_{n\to\infty} A_n\Big)^c = \limsup_{n\to\infty} A_n^c$.

(ii) $P\Big(\liminf_{n\to\infty} A_n\Big) \leq \liminf_{n\to\infty} P(A_n) \leq \limsup_{n\to\infty} P(A_n) \leq P\Big(\limsup_{n\to\infty} A_n\Big)$.

(iii) If $(A_n)_{n\in\mathbb{N}}$ is independent, then $P\Big(\limsup_{n\to\infty} A_n\Big) \in \{0,1\}$ (*Borel's Zero-One Law*).

Proof: Übung 10.1

Theorem 2 (Borel-Cantelli Lemma)**.** Let $A = \limsup_{n\to\infty} A_n$ with $A_n \in \mathfrak{A}$.

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

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

Proof. Ad (i):

$$P(A) \le P\Big(\bigcup_{m\ge n} A_m\Big) \le \sum_{m=n}^{\infty} P(A_m).$$

By assumption, the right-hand side tends to zero as n tends to ∞.
Ad (ii): We have

$$P(A^c) = P(\liminf_{n\to\infty} A_n^c) \le \sum_{n=1}^{\infty} P\Big(\bigcap_{m\ge n} A_m^c\Big).$$

Use $1 - x \le \exp(-x)$ for $x \ge 0$ to obtain

$$P\Big(\bigcap_{m=n}^{\ell} A_m^c\Big) = \prod_{m=n}^{\ell}(1 - P(A_m)) \le \prod_{m=n}^{\ell} \exp(-P(A_m)) = \exp(-\sum_{m=n}^{\ell} P(A_m)).$$

By assumption, the right-hand side tends to zero as ℓ tends to ∞. Thus $P(A^c) = 0$. $\square$

Example 2. A fair coin is tossed an infinite number of times. Determine the probability that 0 occurs twice in a row infinitely often. Model: $(X_n)_{n\in\mathbb{N}}$ is independent and

$$P(\{X_n = 0\}) = P(\{X_n = 1\}) = 1/2, \qquad n \in \mathbb{N}.$$

Put

$$A_n = \{X_n = X_{n+1} = 0\}.$$

Then $(A_{2n})_{n\in\mathbb{N}}$ is independent and $P(A_{2n}) = 1/4$. Thus $P(\limsup_{n\to\infty} A_n) = 1$.

Remark 3. A stronger version of Theorem 2.(ii) requires only pairwise independence, see Bauer (1996, p. 70).

Example 3. Let $(X_n)_{n\in\mathbb{N}}$ be independent with

$$P(\{X_n = 1\}) = p = 1 - P(\{X_n = -1\}), \qquad n \in \mathbb{N},$$

with some constant $p \in [0,1]$. Put

$$A = \limsup_{n\to\infty} \{S_n = 0\},$$

and note that

$$A \notin \mathfrak{A}_\infty = \bigcap_{n\in\mathbb{N}} \sigma(\{X_m : m \ge n\}).$$

Clearly

$$1/2 \cdot (S_n + n) \sim B(n,p).$$

Use Stirling's Formula

$$n! \approx \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

to obtain

$$P(\{S_{2n} = 0\}) = \binom{2n}{n} \cdot p^n \cdot (1-p)^n \approx \frac{r^n}{\sqrt{\pi n}},$$

where $r = 4p \cdot (1-p) \in [0,1]$.
Suppose that

$$p \neq 1/2.$$

Then $r < 1$, and therefore

$$\sum_{n=0}^{\infty} P(\{S_n = 0\}) = \sum_{n=0}^{\infty} P(\{S_{2n} = 0\}) < \infty.$$

The Borel-Cantelli Lemma implies

$$P(A) = 0.$$

Suppose that

$$p = 1/2.$$

Then

$$\sum_{n=0}^{\infty} P(\{S_n = 0\}) = \sum_{n=0}^{\infty} P(\{S_{2n} = 0\}) = \infty,$$

but $(\{S_n = 0\})_{n\in\mathbb{N}}$ is not independent. Using the Central Limit Theorem (De Moivre-Laplace), one can show that $P(A) = 1$, see Übung 10.2.

2 Strong Law of Large Numbers

Definition 1. $(X_n)_{n\in\mathbb{N}}$ *independent and identically distributed (i.i.d.)* if $(X_n)_{n\in\mathbb{N}}$ is independent and

$$\forall\, n \in \mathbb{N} : \ P_{X_n} = P_{X_1}.$$

Throughout this section: $(X_n)_{n\in\mathbb{N}}$ independent.
Put

$$C = \big\{(S_n)_{n\in\mathbb{N}} \text{ converges in } \mathbb{R}\big\}.$$

By Remark 1, $P(C) \in \{0,1\}$.
First we provide sufficient conditions for $P(C) = 1$ to hold.

Theorem 1 (Hajek-Rényi inequality)**.** If

$$b_1 \geq b_2 \geq \cdots \geq b_n > 0$$

and

$$\forall\, i \in \{1,\dots,n\} : \ X_i \in \mathfrak{L}^2 \wedge \mathrm{E}(X_i) = 0,$$

then

$$P\Big(\Big\{\sup_{1\le k\le n} b_k \cdot |S_k| \ge 1\Big\}\Big) \le \sum_{i=1}^{n} b_i^2 \cdot \mathrm{Var}(X_i).$$

In particular, for $b_1 = \cdots = b_n = 1/\varepsilon > 0$ *(Kolmogorov's inequality)*

$$P\Big(\Big\{\sup_{1\le k\le n} |S_k| \ge \varepsilon\Big\}\Big) \le \frac{1}{\varepsilon^2} \cdot \mathrm{Var}(S_n).$$

Proof. See Gänssler, Stute (1977, p. 98) for the Hajek-Rényi inequality. Here: the Kolmogorov inequality. Let $1 \le k \le n$. We show that

$$\forall\, B \in \sigma(\{X_1, \dots, X_k\}) : \int_B S_k^2 \, dP \le \int_B S_n^2 \, dP. \tag{1}$$

Note that

$$S_n^2 = (S_n - S_k)^2 + 2S_nS_k - S_k^2 = (S_n - S_k)^2 + 2S_k(S_n - S_k) + S_k^2.$$

Moreover, for $B \in \sigma(\{X_1, \dots, X_k\})$,

$$1_B \cdot S_k \text{ is } \sigma(\{X_1, \dots, X_k\})\text{-}\mathfrak{B}\text{-measurable,}$$
$$S_n - S_k \text{ is } \sigma(\{X_{k+1}, \dots, X_n\})\text{-}\mathfrak{B}\text{-measurable,}$$

see Theorem II.2.8. Use Theorem III.5.4 to obtain

$$1_B \cdot S_k,\ S_n - S_k \text{ independent.}$$

Hence Theorem III.5.6 yields

$$\mathrm{E}(1_B \cdot S_k \cdot (S_n - S_k)) = \mathrm{E}(1_B \cdot S_k) \cdot \mathrm{E}(S_n - S_k) = 0,$$

and thereby

$$\mathrm{E}(1_B \cdot S_n^2) \ge 2 \cdot \mathrm{E}(1_B \cdot S_k \cdot (S_n - S_k)) + \mathrm{E}(1_B \cdot S_k^2) = \mathrm{E}(1_B \cdot S_k^2).$$

This completes the proof of (1).

Put

$$A_k = \bigcap_{\ell=1}^{k-1} \{|S_\ell| < \varepsilon\} \cap \{|S_k| \ge \varepsilon\}.$$

Then $A_k \in \sigma(\{X_1, \dots, X_k\})$, and by (1)

$$\begin{aligned}\varepsilon^2 \cdot P\Big(\Big\{\sup_{1\le k\le n} |S_k| \ge \varepsilon\Big\}\Big) &= \varepsilon^2 \cdot \sum_{k=1}^{n} P(A_k) \le \sum_{k=1}^{n} \int_{A_k} S_k^2 \, dP \\ &\le \sum_{k=1}^{n} \int_{A_k} S_n^2 \, dP \le \int_\Omega S_n^2 \, dP = \mathrm{Var}(S_n).\end{aligned}$$

□

Theorem 2. If

$$\forall\, n \in \mathbb{N}:\ X_n \in \mathfrak{L}^2 \wedge \mathrm{E}(X_n) = 0$$

and

$$\sum_{i=1}^{\infty} \mathrm{Var}(X_i) < \infty,$$

then

$$P(C) = 1.$$

Proof. Clearly

$$\omega \in C \quad \Leftrightarrow \quad \forall\, \varepsilon > 0\ \exists\, n \in \mathbb{N}\ \forall\, k \in \mathbb{N}\ |S_{n+k}(\omega) - S_n(\omega)| < \varepsilon.$$

Put

$$M = \inf_{n\in\mathbb{N}} \sup_{k\in\mathbb{N}} |S_{n+k} - S_n|.$$

Then

$$C = \{M = 0\}.$$

Let $\varepsilon > 0$. For every $n \in \mathbb{N}$

$$\{M > \varepsilon\} \subset \Big\{ \sup_{k\in\mathbb{N}} |S_{n+k} - S_n| > \varepsilon \Big\},$$

and

$$\Big\{ \sup_{1\le k\le r} |S_{n+k} - S_n| > \varepsilon \Big\} \uparrow \Big\{ \sup_{k\in\mathbb{N}} |S_{n+k} - S_n| > \varepsilon \Big\}$$

as r tends to ∞. Hence

$$P(\{M > \varepsilon\}) \le \lim_{r\to\infty} P\Big(\Big\{ \sup_{1\le k\le r} |S_{n+k} - S_n| > \varepsilon \Big\}\Big),$$

and Kolmogorov's inequality yields

$$P\Big(\Big\{ \sup_{1\le k\le r} |S_{n+k} - S_n| > \varepsilon \Big\}\Big) \le \frac{1}{\varepsilon^2} \cdot \sum_{i=n+1}^{n+r} \mathrm{Var}(X_i) \le \frac{1}{\varepsilon^2} \cdot \sum_{i=n+1}^{\infty} \mathrm{Var}(X_i).$$

Thus $P(\{M > \varepsilon\}) = 0$ for every $\varepsilon > 0$, which implies $P(\{M > 0\}) = 0$. □

Example 1. Let $(Y_n)_{n\in\mathbb{N}}$ be i.i.d. with $P_{Y_1} = 1/2 \cdot (\varepsilon_1 + \varepsilon_{-1})$. Then $\mathrm{E}(Y_n) = 0$ and $\mathrm{Var}(Y_n) = 1$, so that $\sum_{i=1}^{\infty} Y_i \cdot \frac{1}{i}$ converges P-a.s.

In the sequel: $0 < a_n \uparrow \infty$.
We now study convergence almost surely of $(S_n/a_n)_{n\in\mathbb{N}}$.

Lemma 1 (Kronecker's Lemma)**.** For every sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$

$$\sum_{i=1}^{\infty} \frac{x_i}{a_i} \text{ converges} \quad \Rightarrow \quad \lim_{n\to\infty} \frac{1}{a_n} \cdot \sum_{i=1}^{n} x_i = 0.$$

Proof. Put $c = \sum_{i=1}^{\infty} x_i/a_i$ and $c_n = \sum_{i=1}^{n} x_i/a_i$. It is straightforward to verify that

$$\frac{1}{a_n} \cdot \sum_{i=1}^{n} x_i = c_n - \frac{1}{a_n} \cdot \sum_{i=2}^{n} (a_i - a_{i-1}) \cdot c_{i-1}.$$

Moreover, since $a_{i-1} \leq a_i$ and $\lim_{i\to\infty} a_i = \infty$,

$$c = \lim_{n\to\infty} \frac{1}{a_n} \cdot \sum_{i=2}^{n} (a_i - a_{i-1}) \cdot c_{i-1}.$$

□

Theorem 3 (Strong Law of Large Numbers, $\mathfrak{L}^2$ Case)**.** If

$$\forall\, n \in \mathbb{N}: \ X_n \in \mathfrak{L}^2 \quad \wedge \quad \sum_{i=1}^{\infty} \frac{1}{a_i^2} \cdot \mathrm{Var}(X_i) < \infty \tag{2}$$

then

$$\frac{1}{a_n} \cdot \sum_{i=1}^{n} (X_i - \mathrm{E}(X_i)) \overset{P\text{-a.s.}}{\longrightarrow} 0.$$

Proof. Put $Y_n = 1/a_n \cdot (X_n - \mathrm{E}(X_n))$. Then $\mathrm{E}(Y_n) = 0$ and $(Y_n)_{n\in\mathbb{N}}$ is independent. Moreover,

$$\sum_{i=1}^{\infty} \mathrm{Var}(Y_i) = \sum_{i=1}^{\infty} \frac{1}{a_i^2} \cdot \mathrm{Var}(X_i) < \infty.$$

Thus $\sum_{i=1}^{\infty} Y_i$ converges P-a.s. due to Theorem 2. Apply Lemma 1. □

Remark 1. In particular, if $(X_n)_{n\in\mathbb{N}}$ is i.i.d. and $X_1 \in \mathfrak{L}^2$, then Theorem 3 with $a_n = n$ implies

$$\frac{1}{n} \cdot \sum_{i=1}^{n} X_i \overset{P\text{-a.s.}}{\longrightarrow} \mathrm{E}(X_1),$$

see Einführung in die Stochastik. In fact, this conclusion already holds if $X_1 \in \mathfrak{L}^1$, see Theorem 4 below.

Remark 2. Assume

$$\sup_{n\in\mathbb{N}} \mathrm{Var}(X_n) < \infty.$$

Then another possible choice of a_n in Theorem 3 is

$$a_n = \sqrt{n} \cdot (\log n)^{1/2+\varepsilon}$$

for any $\varepsilon > 0$, and we have

$$\lim_{n\to\infty} \frac{S_n - \mathrm{E}(S_n)}{a_n} = 0 \ P\text{-a.s.}$$

Note that $\lim_{n\to\infty} a_n/n = 0$. Precise description of the fluctuation of $S_n(\omega)$ for P-a.e. $\omega \in \Omega$: law of the iterated logarithm, see Section 6. See also Übung 10.2.

Lemma 2. Let $U_i,\ V_i,\ W \in \mathfrak{Z}(\Omega, \mathfrak{A})$ such that

$$\sum_{i=1}^{\infty} P(\{U_i \neq V_i\}) < \infty.$$

Then

$$\frac{1}{n} \cdot \sum_{i=1}^{n} U_i \xrightarrow{P\text{-a.s.}} W \quad \Leftrightarrow \quad \frac{1}{n} \cdot \sum_{i=1}^{n} V_i \xrightarrow{P\text{-a.s.}} W.$$

Proof. The Borel-Cantelli Lemma implies $P(\limsup_{i\to\infty}\{U_i \neq V_i\}) = 0$. □

Lemma 3. For $X \in \mathfrak{Z}_+(\Omega, \mathfrak{A})$

$$\mathrm{E}(X) \leq \sum_{k=0}^{\infty} P(\{X > k\}) \leq \mathrm{E}(X) + 1.$$

(Cf. Corollary II.8.2.)

Proof. We have

$$\mathrm{E}(X) = \sum_{k=1}^{\infty} \int_{\{k-1<X\leq k\}} X \, dP,$$

and therefore

$$\mathrm{E}(X) \leq \sum_{k=1}^{\infty} k \cdot P(\{k-1 < X \leq k\}) = \sum_{k=0}^{\infty} P(\{X > k\})$$

as well as

$$\mathrm{E}(X) \geq \sum_{k=1}^{\infty} (k-1) \cdot P(\{k-1 < X \leq k\}) \geq \sum_{k=0}^{\infty} P(\{X > k\}) - 1.$$

□

Theorem 4 (Strong Law of Large Numbers, i.i.d. Case)**.** Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. Then

$$\exists\, Z \in \mathfrak{Z}(\Omega, \mathfrak{A}) : \ \frac{1}{n} \cdot S_n \xrightarrow{P\text{-a.s.}} Z \quad \Leftrightarrow \quad X_1 \in \mathfrak{L}^1,$$

in which case $Z = \mathrm{E}(X_1)$ P-a.s.

Proof. '⇒': Clearly

$$P(\{|X_1| > n\}) = P(A_n)$$

where

$$A_n = \{|X_n| > n\}.$$

Note that

$$\frac{1}{n} \cdot X_n = \frac{1}{n} \cdot S_n - \frac{n-1}{n} \cdot \frac{1}{n-1} \cdot S_{n-1} \xrightarrow{P\text{-a.s.}} 0.$$

Hence

$$P(\limsup_{n\to\infty} A_n) = 0.$$

Since $(A_n)_{n\in\mathbb{N}}$ is independent, the Borel-Cantelli Lemma implies

$$\sum_{n=1}^{\infty} P(A_n) < \infty.$$

Use Lemma 3 to obtain $\mathrm{E}(|X_1|) < \infty$.
'$\Leftarrow$': Consider the truncated random variables

$$Y_n = \begin{cases} X_n & \text{if } |X_n| < n \\ 0 & \text{otherwise.} \end{cases}$$

We have

$$\sum_{i=1}^{\infty} \frac{1}{i^2} \cdot \mathrm{Var}(Y_i) < \infty. \tag{3}$$

Proof: Observe that

$$\begin{aligned} \mathrm{Var}(Y_i) \le \mathrm{E}(Y_i^2) &= \sum_{k=1}^{i} \mathrm{E}(Y_i^2 \cdot 1_{[k-1,k[} \circ |Y_i|) \\ &= \sum_{k=1}^{i} \mathrm{E}(X_i^2 \cdot 1_{[k-1,k[} \circ |X_i|) \le \sum_{k=1}^{i} k^2 \cdot P(\{k-1 \le |X_1| < k\}). \end{aligned}$$

Thus

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{1}{i^2} \cdot \mathrm{Var}(Y_i) &\le \sum_{k=1}^{\infty} k^2 \cdot P(\{k-1 \le |X_1| < k\}) \cdot \sum_{i=k}^{\infty} \frac{1}{i^2} \\ &\le 2 \cdot \sum_{k=1}^{\infty} k \cdot P(\{k-1 \le |X_1| < k\}) \le 2 \cdot (\mathrm{E}(|X_1|) + 1) < \infty, \end{aligned}$$

cf. the proof of Lemma 3.
Moreover,

$$\sum_{i=1}^{\infty} P(\{X_i \neq Y_i\}) < \infty, \tag{4}$$

since, by Lemma 3,

$$\sum_{i=1}^{\infty} P(\{X_i \neq Y_i\}) = \sum_{i=1}^{\infty} P(\{|X_i| \ge i\}) \le \sum_{i=0}^{\infty} P(\{|X_1| > i\}) \le \mathrm{E}(|X_1|) + 1 < \infty.$$

Furthermore,

$$\lim_{n\to\infty} \mathrm{E}(Y_n) = \mathrm{E}(X_1), \tag{5}$$

according to the dominated convergence theorem.
We obtain

$$\frac{1}{n} \cdot \sum_{i=1}^{n} (Y_i - \mathrm{E}(Y_i)) \xrightarrow{P\text{-a.s.}} 0$$

from Theorem 3 and (3). Due to (5)

$$\lim_{n\to\infty} \frac{1}{n} \cdot \sum_{i=1}^{n} (\mathrm{E}(Y_i) - \mathrm{E}(X_i)) = 0.$$

Thus

$$\frac{1}{n} \cdot \sum_{i=1}^{n} (Y_i - \mathrm{E}(X_i)) \stackrel{P\text{-a.s.}}{\longrightarrow} 0.$$

Finally, by Lemma 2 and (4)

$$\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - \mathrm{E}(X_i)) \stackrel{P\text{-a.s.}}{\longrightarrow} 0.$$

□

Theorem 5. Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d.

(i) If $\mathrm{E}(X_1^-) < \infty \wedge \mathrm{E}(X_1^+) = \infty$ then

$$\frac{1}{n} \cdot S_n \stackrel{P\text{-a.s.}}{\longrightarrow} \infty.$$

(ii) If $\mathrm{E}(|X_1|) = \infty$ then

$$\limsup_{n\to\infty} \left| \frac{1}{n} \cdot S_n \right| = \infty \ P\text{-a.s.}$$

Proof. (i) follows from Theorem 4, and (ii) is an application of the Borel-Cantelli Lemma, see Gänssler, Stute (1977, p. 131). □

Remark 3. We already have $S_n/n \stackrel{P\text{-a.s.}}{\longrightarrow} \mathrm{E}(X_1)$ if the random variables X_n are identically distributed, P-integrable, and pairwise independent. See Bauer (1996, §12).

Remark 4. The basic idea of *Monte-Carlo algorithms*: to compute a quantity $a \in \mathbb{R}$

(i) find a probability measure μ on $(\mathbb{R}, \mathfrak{B})$ such that $\int_{\mathbb{R}} x \, \mu(dx) = a$,

(ii) take an i.i.d. sequence $(X_n)_{n\in\mathbb{N}}$ with $P_{X_1} = \mu$ and approximate a by $1/n \cdot S_n(\omega)$.

Clearly S_n/n is an *unbiased estimator* for a, i.e.,

$$E\Big(\frac{1}{n} \cdot S_n\Big) = a.$$

Due to the Strong Law of Large Numbers S_n/n converges almost surely to a. If $X_1 \in \mathfrak{L}^2$, then

$$\mathrm{E}\Big(\frac{1}{n} \cdot S_n - a\Big)^2 = \mathrm{Var}\Big(\frac{1}{n} \cdot S_n - a\Big) = \mathrm{Var}\Big(\frac{1}{n} \cdot \sum_{i=1}^{n} (X_i - a)\Big) = \frac{1}{n} \cdot \mathrm{Var}(X_1),$$

i.e., the variance of X_1 is the key quantity for the error of the Monte Carlo algorithm in the mean square sense. Moreover,

$$\frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - S_n/n)^2 \xrightarrow{P\text{-a.s.}} \mathrm{Var}(X_1)$$

provides a simple estimator for this variance, see Einführung in die Stochastik.

Applications: see, e.g., Übung 10.3 and 10.4.

Remark 5. Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with $\mu = P_{X_1}$ and corresponding distribution function $F = F_{X_1}$. Suppose that μ is unknown, but observations $X_1(\omega), \dots, X_n(\omega)$ are available for 'estimation of μ'.

Fix $C \in \mathfrak{B}$. Due to Theorem 4

$$\frac{1}{n} \cdot \sum_{i=1}^{n} 1_C \circ X_i \xrightarrow{P\text{-a.s.}} \mu(C).$$

The particular case $C =]-\infty, x]$ leads to the definitions

$$F_n(x,\omega) = \frac{1}{n} \cdot |\{i \in \{1,\dots,n\} : X_i(\omega) \le x\}|, \qquad x \in \mathbb{R},$$

and

$$\mu_n(\cdot,\omega) = \frac{1}{n} \cdot \sum_{i=1}^{n} \varepsilon_{X_i(\omega)}$$

of the *empirical distribution function* $F_n(\cdot,\omega)$ and the *empirical distribution* $\mu_n(\cdot,\omega)$, resp. We obtain

$$\forall\, x \in \mathbb{R}\ \exists\, A \in \mathfrak{A} :\ P(A) = 1 \wedge \Big(\forall\, \omega \in A :\ \lim_{n\to\infty} F_n(x,\omega) = F(x)\Big).$$

Therefore

$$\exists\, A \in \mathfrak{A} :\ P(A) = 1 \wedge \Big(\forall\, q \in \mathbb{Q}\ \forall\, \omega \in A :\ \lim_{n\to\infty} F_n(q,\omega) = F(q)\Big),$$

which implies

$$\exists\, A \in \mathfrak{A} :\ P(A) = 1 \wedge \Big(\forall\, \omega \in A :\ \mu_n(\cdot,\omega) \xrightarrow{\mathrm{w}} \mu\Big),$$

see Helly's Theorem (ii), p. 61, and Theorem III.3.2.

A refined analysis yields the *Glivenko-Cantelli Theorem*

$$\exists\, A \in \mathfrak{A} :\ P(A) = 1 \wedge \Big(\forall\, \omega \in A :\ \lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x,\omega) - F(x)| = 0\Big),$$

see Einführung in die Stochastik.

3 Weak Law of Large Numbers

As previously: $0 < a_n \uparrow \infty$.

In the sequel: $(X_n)_{n\in\mathbb{N}}$ pairwise uncorrelated.

Particular case, $(X_n)_{n\in\mathbb{N}}$ pairwise independent and $X_n \in \mathfrak{L}^2$ for every $n \in \mathbb{N}$, see Theorem III.5.6.

Theorem 1 (Khintchine)**.** If

$$\lim_{n\to\infty} \frac{1}{a_n^2} \cdot \sum_{i=1}^n \mathrm{Var}(X_i) = 0,$$

then

$$\frac{1}{a_n} \cdot \sum_{i=1}^n (X_i - \mathrm{E}(X_i)) \xrightarrow{P} 0.$$

Proof. Without loss of generality $\mathrm{E}(X_n) = 0$ for every $n \in \mathbb{N}$. For $\varepsilon > 0$ the Chebyshev-Markov inequality and Bienaymé's Theorem yield

$$P\Big(\Big\{\Big|\frac{1}{a_n} \cdot S_n\Big| \geq \varepsilon\Big\}\Big) \leq \frac{1}{\varepsilon^2} \cdot \mathrm{Var}\Big(\frac{1}{a_n} \cdot S_n\Big) = \frac{1}{\varepsilon^2} \cdot \frac{1}{a_n^2} \cdot \sum_{i=1}^n \mathrm{Var}(X_i). \tag{1}$$

□

Remark 1. Assume that $\sup_{n\in\mathbb{N}} \mathrm{Var}(X_n) < \infty$. Then Theorem 1 is applicable for any sequence $(a_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} a_n/\sqrt{n} = \infty$.

Example 1. Consider an independent sequence $(X_n)_{n\in\mathbb{N}}$ with

$$P(\{X_n = 0\}) = 1 - \frac{1}{n\log(n+1)}, \qquad P(\{X_n = \pm n\}) = \frac{1}{2n\log(n+1)}.$$

Hence

$$\mathrm{E}(X_n) = 0, \qquad \mathrm{Var}(X_n) = \frac{n}{\log(n+1)},$$

and

$$\frac{1}{n^2} \cdot \sum_{i=1}^n \mathrm{Var}(X_i) \leq \frac{1}{\log(n+1)}.$$

Thus $1/n \cdot S_n \xrightarrow{P} 0$ due to Theorem 1, but $1/n \cdot S_n \xrightarrow{P\text{-a.s.}} 0$ does not hold, see Übung 11.2.

4 Characteristic Functions

We use the notation $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ for the euclidean inner product and norm. Recall that $\mathfrak{M}(\mathbb{R}^k)$ denotes the class of all probability measures on $(\mathbb{R}^k, \mathfrak{B}_k)$.

Given: a probability measure $\mu \in \mathfrak{M}(\mathbb{R}^k)$.

Definition 1. $f : \mathbb{R}^k \to \mathbb{C}$ is *μ-integrable* if $\Re f$ and $\Im f$ are μ-integrable, in which case

$$\int f\, d\mu = \int \Re f\, d\mu + \imath \cdot \int \Im f\, d\mu.$$

Definition 2. The mapping $\widehat{\mu} : \mathbb{R}^k \to \mathbb{C}$ with

$$\widehat{\mu}(y) = \int \exp(\imath\langle x, y\rangle)\, \mu(dx), \qquad y \in \mathbb{R}^k,$$

is called the *Fourier transform* of μ.

Example 1.

(i) For a discrete probability measure

$$\mu = \sum_{j=1}^{\infty} \alpha_j \cdot \varepsilon_{x_j}$$

we have

$$\widehat{\mu}(y) = \sum_{j=1}^{\infty} \alpha_j \cdot \exp(\imath\langle x_j, y\rangle).$$

For instance, if $\mu = \pi(\lambda)$ is the Poisson distribution with parameter $\lambda > 0$, then

$$\begin{aligned}\widehat{\mu}(y) &= \exp(-\lambda) \cdot \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} \cdot \exp(\imath j y) = \exp(-\lambda) \cdot \exp(\lambda \cdot \exp(\imath y)) \\ &= \exp(\lambda \cdot (\exp(\imath y) - 1)).\end{aligned}$$

(ii) If $\mu = f \cdot \lambda_k$ then

$$\widehat{\mu}(y) = \int \exp(\imath\langle x, y\rangle) \cdot f(x)\, \lambda_k(dx).$$

For any λ_k-integrable function f, the right-hand side defines its *Fourier transform*, see also Analysis or Funktionalanalysis. For instance, if μ is the k-dimensional standard normal distribution, i.e.,

$$f(x) = (2\pi)^{-k/2} \cdot \exp(-\|x\|^2/2),$$

then

$$\widehat{\mu}(y) = \exp(-\|y\|^2/2).$$

See Bauer (1996, p. 187) for the case $k = 1$. Use Fubini's Theorem for $k > 1$.

Theorem 1.

(i) $\widehat{\mu}$ is uniformly continuous on $\mathbb{R}^k$,

(ii) $|\widehat{\mu}(y)| \le 1 = \widehat{\mu}(0)$ for $y \in \mathbb{R}^k$,

(iii) for $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{C}$, and $y_1, \dots, y_n \in \mathbb{R}^k$,

$$\sum_{j,\ell=1}^{n} a_j \cdot \overline{a_\ell} \cdot \widehat{\mu}(y_j - y_\ell) \geq 0$$

(*positive semi-definite*).

Proof. Ad (i): Observe that

$$\left|\exp(\imath\langle x, y_1\rangle) - \exp(\imath\langle x, y_2\rangle)\right| \leq \|x\| \cdot \|y_1 - y_2\|. \tag{1}$$

For $\varepsilon > 0$ take $r > 0$ such that $\mu(B) \geq 1 - \varepsilon$, where $B = \{x \in \mathbb{R}^k : \|x\| \leq r\}$. Then

$$\begin{aligned} \left|\widehat{\mu}(y_1) - \widehat{\mu}(y_2)\right| &\leq \int_B \left|\exp(\imath\langle x, y_1\rangle) - \exp(\imath\langle x, y_2\rangle)\right| \mu(dx) + 2 \cdot \varepsilon \\ &\leq r \cdot \|y_1 - y_2\| + 2 \cdot \varepsilon. \end{aligned}$$

Properties (ii) and (iii) are easily verified. □

Remark 1. *Bochner's Theorem* states that every continuous, positive semi-definite function $\varphi : \mathbb{R}^k \to \mathbb{C}$ with $\varphi(0) = 1$ is the Fourier transform of a probability measure on $(\mathbb{R}^k, \mathfrak{B}_k)$. See Bauer (1996, p. 184) for references.

In the sequel: $X, Y, \dots$ are k-dimensional random vectors on a probability space $(\Omega, \mathfrak{A}, P)$.

Definition 3. The *characteristic function* of X is given by

$$\varphi_X = \widehat{P_X}.$$

Remark 2. Due to Theorem II.9.1

$$\varphi_X(y) = \int_{\mathbb{R}^k} \exp(\imath\langle x, y\rangle)\, P_X(dx) = \int_\Omega \exp(\imath\langle X(\omega), y\rangle)\, P(d\omega).$$

Theorem 2.

(i) For every linear mapping $T : \mathbb{R}^k \to \mathbb{R}^\ell$

$$\varphi_{T \circ X} = \varphi_X \circ T^t.$$

(ii) For independent random vectors X and Y

$$\varphi_{X+Y} = \varphi_X \cdot \varphi_Y.$$

In particular, for $a \in \mathbb{R}^k$,

$$\varphi_{X+a} = \exp(\imath\langle a, \cdot\rangle) \cdot \varphi_X.$$

Proof. Ad (i): Let $z \in \mathbb{R}^\ell$. Use $P_{T\circ X} = T(P_X)$ to obtain

$$\varphi_{T\circ X}(z) = \int_{\mathbb{R}^k} \exp(\imath\langle T(x), z\rangle)\, P_X(dx) = \varphi_X(T^t(z)).$$

Ad (ii): Let $z \in \mathbb{R}^k$. Fubini's Theorem and Theorem III.5.5 imply

$$\varphi_{X+Y}(z) = \int_{\mathbb{R}^{2k}} \exp(\imath\langle x+y, z\rangle)\, P_{(X,Y)}(d(x,y)) = \varphi_X(z)\cdot\varphi_Y(z).$$

□

Corollary 1 (Convolution Theorem)**.** For probability measures $\mu_j \in \mathfrak{M}(\mathbb{R})$,

$$\widehat{\mu_1 * \mu_2} = \widehat{\mu}_1 \cdot \widehat{\mu}_2.$$

Proof. Use Theorem 2.(ii) and Theorem III.5.8. □

Example 2. For $\mu = N(m,\sigma^2)$ with $\sigma \geq 0$ and $m \in \mathbb{R}$

$$\widehat{\mu}(y) = \exp(\imath m y)\cdot\exp(-\sigma^2 y^2/2).$$

See Example 1.(ii) and Theorem 2.

Lemma 1. For $z \in \mathbb{R}$ and $\sigma > 0$

$$\int \exp(-\imath y z)\cdot\widehat{\mu}(y)\, N(0,\sigma^{-2})(dy) = \int \exp\bigl(-(z-x)^2/(2\sigma^2)\bigr)\,\mu(dx).$$

Proof. See Gänssler, Stute (1977, p. 92). □

Lemma 2. For $\sigma_n > 0$ with $\lim_{n\to\infty}\sigma_n = 0$,

$$N(0,\sigma_n^2) * \mu \xrightarrow{\text{w}} \mu.$$

Proof. Consider independent random variables X_n and Y such that $X_n \sim N(0,\sigma_n^2)$ and $Y \sim \mu$. Then $X_n \xrightarrow{\mathfrak{L}^2} 0$, and therefore $X_n + Y \xrightarrow{\mathfrak{L}^2} Y$, which implies

$$X_n + Y \xrightarrow{\text{d}} Y.$$

□

Theorem 3 (Uniqueness Theorem)**.** For probability measures $\mu_j \in \mathfrak{M}(\mathbb{R}^k)$,

$$\mu_1 = \mu_2 \qquad \Leftrightarrow \qquad \widehat{\mu}_1 = \widehat{\mu}_2.$$

Proof. '⇒' holds by definition. '⇐': See Bauer (1996, Thm. 23.4) or Billingsley (1979, Sec. 29) for the case $k > 1$. Here: the case $k = 1$. For $\sigma > 0$ and $A \in \mathfrak{B}$

$$N(0,\sigma^2) * \mu_j(A) = \int\int 1_A(z+x)\, N(0,\sigma^2)(dz)\,\mu_j(dx),$$

and

$$\int 1_A(z+x)\,N(0,\sigma^2)(dz) = (2\pi\sigma^2)^{-1/2} \cdot \int 1_A(z+x)\cdot \exp(-z^2/(2\sigma^2))\,\lambda_1(dz)$$
$$= (2\pi\sigma^2)^{-1/2} \cdot \int_A \exp(-(z-x)^2/(2\sigma^2))\,\lambda_1(dz).$$

Therefore

$$N(0,\sigma^2) * \mu_j(A) = (2\pi\sigma^2)^{-1/2} \cdot \int_A \int \exp(-(z-x)^2/(2\sigma^2))\,\mu_j(dx)\,\lambda_1(dz).$$

Use Lemma 1 to conclude that

$$\forall\, \sigma > 0 :\ N(0,\sigma^2) * \mu_1 = N(0,\sigma^2) * \mu_2.$$

Then, by Lemma 2 and Corollary III.3.1, $\mu_1 = \mu_2$. □

Example 3. For independent random variables X_1 and X_2 with $X_j \sim \pi(\lambda_j)$ we have $X_1 + X_2 \sim \pi(\lambda_1 + \lambda_2)$.
Proof: Theorem 2 and Example 1.(i) yield

$$\varphi_{X_1+X_2}(y) = \exp(\lambda_1 \cdot (\exp(\imath y) - 1)) \cdot \exp(\lambda_2 \cdot (\exp(\imath y) - 1))$$
$$= \exp((\lambda_1 + \lambda_2) \cdot (\exp(\imath y) - 1)).$$

Use Theorem 3.

Lemma 3. For every $\varepsilon > 0$ and every probability measure $\mu \in \mathfrak{M}(\mathbb{R})$,

$$\mu(\{x \in \mathbb{R} : |x| \geq 1/\varepsilon\}) \leq 7/\varepsilon \cdot \int_0^\varepsilon (1 - \Re\widehat{\mu}(y))\,dy.$$

Proof. Clearly

$$\Re\widehat{\mu}(y) = \int_{\mathbb{R}} \cos(xy)\,\mu(dx).$$

Hence, with the convention $\sin(0)/0 = 1$,

$$1/\varepsilon \cdot \int_0^\varepsilon (1 - \Re\widehat{\mu}(y))\,dy = 1/\varepsilon \cdot \int_{[0,\varepsilon]} \int_{\mathbb{R}} (1 - \cos(xy))\,\mu(dx)\,\lambda_1(dy)$$
$$= \int_{\mathbb{R}} \left(1/\varepsilon \cdot \int_0^\varepsilon (1 - \cos(xy))\,dy \right) \mu(dx)$$
$$= \int_{\mathbb{R}} (1 - \sin(\varepsilon x)/(\varepsilon x))\,\mu(dx)$$
$$\geq \inf_{|z|\geq 1} (1 - \sin(z)/z) \cdot \mu(\{x \in \mathbb{R} : |\varepsilon x| \geq 1\}).$$

Finally,

$$\inf_{|z|\geq 1} (1 - \sin(z)/z) \geq 1/7.$$

□

Theorem 4 (Continuity Theorem, Lévy)**.**

(i) Let $\mu, \mu_n \in \mathfrak{M}(\mathbb{R}^k)$ for $n \in \mathbb{N}$. Then

$$\mu_n \xrightarrow{\text{w}} \mu \qquad \Rightarrow \qquad \forall\, y \in \mathbb{R}^k : \lim_{n\to\infty} \widehat{\mu_n}(y) = \widehat{\mu}(y).$$

(ii) Let $\mu_n \in \mathfrak{M}(\mathbb{R}^k)$ for $n \in \mathbb{N}$, and let $\varphi : \mathbb{R}^k \to \mathbb{C}$ be continuous at 0. Then

$$\forall\, y \in \mathbb{R}^k : \lim_{n\to\infty} \widehat{\mu_n}(y) = \varphi(y) \qquad \Rightarrow \qquad \exists\, \mu \in \mathfrak{M}(\mathbb{R}^k) : \widehat{\mu} = \varphi \wedge \mu_n \xrightarrow{\text{w}} \mu.$$

Proof. Ad (i): Note that $x \mapsto \exp(\imath\langle x, y\rangle)$ is bounded and continuous on $\mathbb{R}^k$.
Ad (ii): See Bauer (1996, Thm. 23.8) or Billingsley (1979, Sec. 29) for the case $k > 1$. Here: the case $k = 1$.
We first show that

$$\{\mu_n : n \in \mathbb{N}\} \text{ is tight.} \tag{2}$$

By Lemma 3

$$\mu_n(\{x \in \mathbb{R} : |x| \geq 1/\varepsilon\}) \leq c_n(\varepsilon)$$

with

$$c_n(\varepsilon) = 7/\varepsilon \cdot \int_0^\varepsilon (1 - \Re\widehat{\mu_n}(y))\, dy.$$

The dominated convergence theorem and the continuity of φ at 0 yield

$$\lim_{n\to\infty} c_n(\varepsilon) = c(\varepsilon)$$

with

$$c(\varepsilon) = 7/\varepsilon \cdot \int_0^\varepsilon (1 - \Re\widehat{\varphi}(y))\, dy,$$

if ε is sufficiently small. Given $\delta > 0$ take $\varepsilon > 0$ such that

$$c(\varepsilon) \leq \delta/2.$$

Furthermore, take $n_0 \in \mathbb{N}$ such that, for every $n \geq n_0$,

$$|c_n(\varepsilon) - c(\varepsilon)| \leq \delta/2.$$

Hence, for $n \geq n_0$,

$$\mu_n(\{x \in \mathbb{R} : |x| \geq 1/\varepsilon\}) \leq \delta,$$

and hereby we get (2).
Thus, by Prohorov's Theorem,

$$\{\mu_n : n \in \mathbb{N}\} \text{ is relatively compact.} \tag{3}$$

We fix a probability measure $\mu \in \mathfrak{M}(\mathbb{R})$ such that $\mu_{n_k} \xrightarrow{\text{w}} \mu$ for a suitable subsequence of $(\mu_n)_{n\in\mathbb{N}}$. By assumption and (i), we get $\widehat{\mu} = \varphi$ as well as the following fact:

$$\text{if } \mu_{n_k} \xrightarrow{\text{w}} \nu \text{ for any subsequence } (\mu_{n_k})_{k\in\mathbb{N}}, \text{ then } \nu = \mu, \tag{4}$$

see Theorem 3.
We claim that $\mu_n \xrightarrow{\text{w}} \mu$. Due to Remarks III.2.3 and III.3.4.(ii) it suffices to show that every subsequence of $(\mu_n)_{n\in\mathbb{N}}$ contains a subsequence that converges weakly to μ. The latter property follows from (3) and (4). $\square$

Example 4. For the uniform distribution μ_n on $[-n, n]$ we have

$$\widehat{\mu_n}(y) = \begin{cases} 1 & \text{if } y = 0 \\ \frac{\sin(n\,y)}{n\,y} & \text{otherwise.} \end{cases}$$

Hence

$$\lim_{n\to\infty} \widehat{\mu_n}(y) = \begin{cases} 1 & \text{if } y = 0 \\ 0 & \text{otherwise,} \end{cases}$$

but $(\mu_n)_{n\in\mathbb{N}}$ does not converge weakly.

Corollary 2. Weak convergence in $\mathfrak{M}(\mathbb{R}^k)$ is equivalent to pointwise convergence of Fourier transforms.

Example 5. Let $\mu_n = B(n, p_n)$ and assume that

$$\lim_{n\to\infty} n \cdot p_n = \lambda > 0.$$

Then

$$\mu_n \xrightarrow{\text{w}} \pi(\lambda).$$

Proof: `Übung` 11.4.

5 The Central Limit Theorem

Given: a triangular array of random variables X_{nk}, where $n \in \mathbb{N}$ and $k \in \{1, \dots, r_n\}$ with $r_n \in \mathbb{N}$.
Assumptions:

(i) $X_{nk} \in \mathfrak{L}^2$ for every $n \in \mathbb{N}$ and $k \in \{1, \dots, r_n\}$,

(ii) $(X_{n1}, \dots, X_{nr_n})$ independent for every $n \in \mathbb{N}$.

Put

$$S_n = \sum_{k=1}^{r_n} (X_{nk} - \mathrm{E}(X_{nk}))$$

and

$$\sigma_{nk}^2 = \mathrm{Var}(X_{nk}), \qquad s_n^2 = \mathrm{Var}(S_n) = \sum_{k=1}^{r_n} \sigma_{nk}^2.$$

Additional assumption:

(iii) $s_n^2 > 0$ for every $n \in \mathbb{N}$.

Normalization

$$S_n^* = \frac{1}{s_n} \cdot S_n = \sum_{k=1}^{r_n} \frac{X_{nk} - \mathrm{E}(X_{nk})}{s_n}$$

for $n \in \mathbb{N}$. Clearly

$$\mathrm{E}(S_n^*) = 0 \qquad \wedge \qquad \mathrm{Var}(S_n^*) = 1.$$

Question: convergence in distribution of $(S_n^*)_{n\in\mathbb{N}}$?

For notational convenience: all random variables X_{nk} are defined on a common probability space $(\Omega, \mathfrak{A}, P)$.

Example 1. $(X_n)_{n\in\mathbb{N}}$ i.i.d. with $X_1 \in \mathfrak{L}^2$ and $\mathrm{Var}(X_1) = \sigma^2 > 0$. Put $m = \mathrm{E}(X_1)$, take

$$r_n = n, \qquad\qquad X_{nk} = X_k.$$

Then

$$S_n^* = \frac{\sum_{k=1}^n X_k - n \cdot m}{\sqrt{n} \cdot \sigma}.$$

See Einführung in die Stochastik.

In the sequel we assume, without loss of generality,

$$\mathrm{E}(X_{nk}) = 0 \quad \wedge \quad s_n = 1$$

for $n \in \mathbb{N}$ and $k \in \{1, \dots, r_n\}$, hence

$$S_n^* = \sum_{k=1}^{r_n} X_{nk}$$

(Otherwise, consider the random variables $(X_{nk} - \mathrm{E}(X_{nk}))/s_n$.)

Definition 1.

(i) *Lyapunov condition*

$$\exists\, \delta > 0 : \lim_{n\to\infty} \sum_{k=1}^{r_n} \mathrm{E}(|X_{nk}|^{2+\delta}) = 0.$$

(ii) *Lindeberg condition*

$$\forall\, \varepsilon > 0 : \lim_{n\to\infty} \sum_{k=1}^{r_n} \int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2 \, dP = 0.$$

(iii) *Feller condition*

$$\lim_{n\to\infty} \max_{1\leq k\leq r_n} \mathrm{Var}(X_{nk}) = 0.$$

(iv) The random variables X_{nk} are *asymptotically negligible* if

$$\forall\, \varepsilon > 0 : \lim_{n\to\infty} \max_{1\leq k\leq r_n} P(\{|X_{nk}| > \varepsilon\}) = 0.$$

Lemma 1. The conditions from Definition 1 satisfy

$$\text{(i)} \Rightarrow \text{(ii)} \Rightarrow \text{(iii)} \Rightarrow \text{(iv)}.$$

Moreover, (iii) implies $\lim_{n\to\infty} r_n = \infty$.

Proof. From

$$\int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2\,dP \leq \frac{1}{\varepsilon^\delta}\cdot\int_{\{|X_{nk}|\geq\varepsilon\}} |X_{nk}|^{2+\delta}\,dP \leq \frac{1}{\varepsilon^\delta}\cdot \mathrm{E}(|X_{nk}|^{2+\delta})$$

we get '(i) ⇒ (ii)'. From

$$\mathrm{Var}(X_{nk}) \leq \varepsilon^2 + \int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2\,dP$$

we get '(ii) ⇒ (iii)'. The Chebyshev-Markov inequality yields '(iii) ⇒ (iv)'. Finally,

$$1 = \mathrm{Var}(S_n^*) \leq r_n \cdot \max_{1\leq k\leq r_n} \mathrm{Var}(X_{nk}),$$

so that (iii) implies $\lim_{n\to\infty} r_n = \infty$. □

Example 2. Example 1 continued in the case $m = 0$. We take $r_n = n$ and

$$X_{nk} = \frac{X_k}{\sqrt{n}\cdot\sigma}$$

to obtain

$$\sum_{k=1}^{n}\int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2\,dP = \frac{1}{\sigma^2}\cdot\int_{\{|X_1|\geq\varepsilon\cdot\sqrt{n}\cdot\sigma\}} X_1^2\,dP.$$

Hence the Lindeberg condition is satisfied.

In the sequel

$$\varphi_{nk} = \varphi_{X_{nk}}$$

denotes the characteristic function of X_{nk}. Use (4.1) and Theorem II.5.5 to conclude that

$$\varphi'(0) = \imath\cdot \mathrm{E}(X_{nk}) = 0$$

and

$$\varphi''(0) = -\,\mathrm{E}(X_{nk}^2) = -\sigma_{nk}.$$

Lemma 2. For $y\in\mathbb{R}$ and $\varepsilon > 0$

$$\left|\varphi_{nk}(y) - (1-\sigma_{nk}^2/2\cdot y^2)\right| \leq y^2\cdot\Big(\varepsilon\cdot|y|\cdot\sigma_{nk}^2 + \int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2\,dP\Big).$$

Proof. For $u\in\mathbb{R}$

$$\left|\exp(\imath u) - (1+\imath u - u^2/2)\right| \leq \min(u^2,\ |u|^3/6),$$

see Billingsley (1979, Eqn. (26.4)). Hence

$$\begin{aligned}
&\left|\varphi_{nk}(y) - (1-\sigma_{nk}^2/2\cdot y^2)\right| \\
&\quad= \left|\mathrm{E}(\exp(\imath\cdot X_{nk}\cdot y)) - \mathrm{E}\Big(1+\imath\cdot X_{nk}\cdot y - X_{nk}^2\cdot y^2/2)\right| \\
&\quad\leq \mathrm{E}\big(\min(y^2\cdot X_{nk}^2\,,\ |y|^3\cdot|X_{nk}|^3)\big) \\
&\quad\leq |y|^3\cdot\int_{\{|X_{nk}|<\varepsilon\}} \varepsilon\cdot X_{nk}^2\,dP + y^2\cdot\int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2\,dP \\
&\quad\leq \varepsilon\cdot|y|^3\cdot\sigma_{nk}^2 + y^2\cdot\int_{\{|X_{nk}|\geq\varepsilon\}} X_{nk}^2\,dP.
\end{aligned}$$

□

Lemma 3. Put

$$\Delta_n(y) = \prod_{k=1}^{r_n} \varphi_{nk}(y) - \exp(-y^2/2), \qquad y \in \mathbb{R}.$$

If the Lindeberg condition is satisfied, then

$$\forall\, y \in \mathbb{R} : \lim_{n\to\infty} \Delta_n(y) = 0.$$

Proof. Since $|\varphi_{nk}(y)| \le 1$ and $|\exp(-\sigma_{nk}^2/2 \cdot y^2)| \le 1$, we get

$$\begin{aligned} |\Delta_n(y)| &= \Big|\prod_{k=1}^{r_n} \varphi_{nk}(y) - \prod_{k=1}^{r_n} \exp(-\sigma_{nk}^2/2 \cdot y^2)\Big| \\ &\le \sum_{k=1}^{r_n} \big|\varphi_{nk}(y) - \exp(-\sigma_{nk}^2/2 \cdot y^2)\big| \end{aligned}$$

by induction, see Billingsley (1979, Lemma 27.1).
We assume

$$\max_{1\le k\le r_n} \sigma_{nk}^2 \cdot y^2 \le 1 \,,$$

which holds for fixed $y \in \mathbb{R}$ if n is sufficiently large, see Lemma 1. Using

$$0 \le u \le 1/2 \quad \Rightarrow \quad |\exp(-u) - (1-u)| \le u^2$$

and Lemma 2 we obtain

$$\begin{aligned} |\Delta_n(y)| &\le \sum_{k=1}^{r_n} |\varphi_{nk}(y) - (1 - \sigma_{nk}^2/2 \cdot y^2)| + \sum_{k=1}^{r_n} \sigma_{nk}^4/4 \cdot y^4 \\ &\le y^2 \cdot \Big(\varepsilon \cdot |y| + \sum_{k=1}^{r_n} \int_{\{|X_{nk}|\ge\varepsilon\}} X_{nk}^2 \, dP\Big) + y^4/4 \cdot \max_{1\le k\le r_n} \sigma_{nk}^2 \end{aligned}$$

for every $\varepsilon > 0$. Thus Lemma 1 yields

$$\limsup_{n\to\infty} |\Delta_n(y)| \le |y|^3 \cdot \varepsilon.$$

□

Theorem 1 (Central Limit Theorem)**.** The following properties are equivalent:

(i) $(X_{nk})_{n,k}$ satisfies the Lindeberg condition.

(ii) $P_{S_n^*} \xrightarrow{\text{w}} N(0,1)$ and $(X_{nk})_{n,k}$ satisfies the Feller condition.

(iii) $P_{S_n^*} \xrightarrow{\text{w}} N(0,1)$ and the random variables X_{nk} are asymptotically negligible.

Proof. ‘(i) $\Rightarrow$ (ii)’: Due to Lemma 1 we only have to prove the weak convergence. Recall that $\widehat{\mu}(y) = \exp(-y^2/2)$ for the standard normal distribution μ. Consider the characteristic function $\varphi_n = \varphi_{S_n^*}$ of S_n^*. By Theorem 4.2.(ii)

$$\varphi_n = \prod_{k=1}^{r_n} \varphi_{nk},$$

and therefore Lemma 3 implies

$$\forall\, y \in \mathbb{R} : \lim_{n\to\infty} \varphi_n(y) = \widehat{\mu}(y).$$

It remains to apply Corollary 4.2.

Lemma 1 yields ‘(ii) $\Rightarrow$ (iii)’. See Billingsley (1979, p. 314–315) for the proof of ‘(iii) $\Rightarrow$ (i)’. □

Corollary 1. Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with $X_1 \in \mathfrak{L}^2$ and $\sigma^2 = \mathrm{Var}(X_1) > 0$. Then

$$\frac{\sum_{k=1}^{n} X_k - n \cdot \mathrm{E}(X_1)}{\sqrt{n} \cdot \sigma} \xrightarrow{\mathrm{d}} Z$$

where $Z \sim N(0, 1)$.

Proof. Theorem 1 and Example 2. □

Example 3. Example 2 continued, and Corollary 1 reformulated. Let

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x} \exp(-u^2/2)\, du, \qquad x \in \mathbb{R},$$

denote the distribution function of the standard normal distribution, and let

$$\delta_n = \sup_{x\in\mathbb{R}} \bigl|P(\{S_n \le x \cdot \sqrt{n} \cdot \sigma\}) - \Phi(x)\bigr| = \sup_{x\in\mathbb{R}} \bigl|P(\{S_n \le x\}) - \Phi(x/(\sqrt{n} \cdot \sigma))\bigr|. \tag{1}$$

Due to the Central Limit Theorem and Theorem III.3.2

$$\lim_{n\to\infty} \delta_n = 0.$$

Theorem 2 (Berry-Esséen)**.** Let $(X_n)_{n\in\mathbb{N}}$ be i.i.d. with $X_1 \in \mathfrak{L}^3$, $\mathrm{E}(X_1) = 0$, and $\mathrm{Var}(X_1) = \sigma^2 > 0$. For δ_n given by (1)

$$\forall\, n \in \mathbb{N} : \delta_n \le \frac{6 \cdot \mathrm{E}(|X_1|^3)}{\sigma^3} \cdot \frac{1}{\sqrt{n}}.$$

Proof. See Gänssler, Stute (1977, Section 4.2). □

Example 4. Example 3 continued with

$$P_{X_1} = \frac{1}{2} \cdot (\varepsilon_1 + \varepsilon_{-1}). \tag{2}$$

Since $(-X_n)_{n\in\mathbb{N}}$ is i.i.d. as well, and since $P_{-X_1} = P_{X_1}$, we have

$$P(\{S_{2n} \leq 0\}) = P(\{S_{2n} \geq 0\}),$$

which yields

$$P(\{S_{2n} \leq 0\}) = \frac{1}{2} \cdot (1 + P(\{S_{2n} = 0\})).$$

From Example 1.3 we know that

$$P(\{S_{2n} = 0\}) \approx \frac{1}{\sqrt{\pi n}}\,,$$

and therefore

$$\delta_{2n} \geq P(\{S_{2n} \leq 0\}) - \frac{1}{2} = \frac{1}{2} \cdot P(\{S_{2n} = 0\}) \approx \frac{1}{2\sqrt{\pi n}}.$$

Hence the upper bound from Theorem 2 cannot be improved in terms of powers of n.

Example 5. Example 3 continued, i.e., $(X_n)_{n\in\mathbb{N}}$ is i.i.d. with $X_1 \in \mathfrak{L}^2$, $\mathrm{E}(X_1) = 0$, and $\mathrm{Var}(X_1) = \sigma^2 > 0$. Recall that $S_n = \sum_{i=1}^n X_i$.
Let

$$B_c = \{\limsup_{n\to\infty} S_n/\sqrt{n} \geq c\}, \qquad c > 0.$$

Using Remark 1.2.(ii) we get

$$P(B_c) \geq P(\limsup_{n\to\infty}\{S_n/\sqrt{n} > c\}) \geq \limsup_{n\to\infty} P(\{S_n/\sqrt{n} > c\}) = 1 - \Phi(c/\sigma) > 0.$$

Kolmogorov's Zero-One Law yields

$$P(B_c) = 1,$$

and therefore

$$P(\{\limsup_{n\to\infty} S_n/\sqrt{n} = \infty\}) = P\Big(\bigcap_{c\in\mathbb{N}} B_c\Big) = 1.$$

By symmetry

$$P(\{\liminf_{n\to\infty} S_n/\sqrt{n} = -\infty\}) = 1.$$

In particular, for P_{X_1} given by (2),

$$P(\limsup_{n\to\infty}\{S_n = 0\}) = 1,$$

see also Example 1.3 and `Übung` 10.2.

6 Law of the Iterated Logarithm

Given: an i.i.d. sequence $(X_n)_{n\in\mathbb{N}}$ of random variables on $(\Omega, \mathfrak{A}, P)$.

Definition 1. $(S_n)_{n\in\mathbb{N}}$ with $S_n = \sum_{k=1}^n X_k$ is called the associated *random walk*.

In the sequel we assume

$$X_1 \in \mathfrak{L}^2 \quad \wedge \quad \mathrm{E}(X_1) = 0 \quad \wedge \quad \mathrm{Var}(X_1) = \sigma^2 > 0.$$

Remark 1. For every $\varepsilon > 0$, with probability one,

$$\lim_{n\to\infty} \frac{S_n}{\sqrt{n} \cdot (\log n)^{1/2+\varepsilon}} = 0,$$

see Remark 2.2. On the other hand, with probability one,

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{n}} = \infty \quad \wedge \quad \liminf_{n\to\infty} \frac{S_n}{\sqrt{n}} = -\infty,$$

see Example 5.5.

Question: precise description of the fluctuation of $(S_n(\omega))_{n\in\mathbb{N}}$ for P-almost every ω? In particular: existence of a deterministic sequence $(\gamma(n))_{n\in\mathbb{N}}$ of positive reals such that, with probability one,

$$\limsup_{n\to\infty} \frac{S_n}{\gamma(n)} = 1 \quad \wedge \quad \liminf_{n\to\infty} \frac{S_n}{\gamma(n)} = -1?$$

Notation: $L((u_n)_{n\in\mathbb{N}})$ is the set of all limit points in $\overline{\mathbb{R}}$ of a sequence $(u_n)_{n\in\mathbb{N}}$ in $\mathbb{R}$.

Let

$$\gamma(n) = \sqrt{2n \cdot \log(\log n) \cdot \sigma^2}, \qquad n \geq 3,$$

where log denotes the logarithm with base e.

Theorem 1 (Strassen's Law of the Iterated Logarithm)**.**
With probability one,

$$L\left(\left(\frac{S_n}{\gamma(n)}\right)_{n\in\mathbb{N}}\right) = [-1,\ 1].$$

Proof. See Bauer (1996, §33). □

Corollary 1 (Hartman and Wintner's Law of the Iterated Logarithm)**.**
With probability one,

$$\limsup_{n\to\infty} \frac{S_n}{\gamma(n)} = 1 \quad \wedge \quad \liminf_{n\to\infty} \frac{S_n}{\gamma(n)} = -1.$$

Chapter V

Conditional Expectations and Martingales

1 Conditional Expectations

'Access to the martingale concept is afforded by one of the truely basic ideas of probability theory, that of conditional expectation.', see Bauer (1996, p. 109).

Soweit nichts anderes gesagt, betrachten wir die Borelsche σ-Algebra $\mathfrak{B}$ auf $\mathbb{R}$; $\mathfrak{G}$-$\mathfrak{B}$-meßbare Abbildungen werden kurz $\mathfrak{G}$-meßbar genannt.

Erinnerung: *Elementare bedingte Wahrscheinlichkeit*

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad A, B \in \mathfrak{A}, \quad P(B) > 0.$$

Das Wahrscheinlichkeitsmaß $P(\cdot \mid B)$ besitzt die P-Dichte $1/P(B) \cdot 1_B$.

Gegeben: Zufallsvariable X auf $(\Omega, \mathfrak{A}, P)$ mit $X \in \mathfrak{L}^1$. *Elementare bedingte Erwartung*

$$\mathrm{E}(X \mid B) = \frac{1}{P(B)} \cdot \mathrm{E}(1_B \cdot X) = \int X \, dP(\cdot \mid B).$$

Klar für $A \in \mathfrak{A}$

$$\mathrm{E}(1_A \mid B) = P(A \mid B).$$

Nun etwas allgemeiner: Sei I höchstens abzählbar. Betrachte Partititon

$$B_i \in \mathfrak{A}, \qquad i \in I,$$

von Ω mit

$$\forall\, i \in I: \quad P(B_i) > 0.$$

Durch

$$\mathfrak{G} = \left\{ \bigcup_{j \in J} B_j : J \subset I \right\}$$

ist eine σ-Algebra $\mathfrak{G} \subset \mathfrak{A}$ gegeben. In der Tat gilt $\mathfrak{G} = \sigma(\{B_i : i \in I\})$. Definiere Zufallsvariable $\mathrm{E}(X \mid \mathfrak{G})$ durch

$$\mathrm{E}(X \mid \mathfrak{G})(\omega) = \sum_{i \in I} \mathrm{E}(X \mid B_i) \cdot 1_{B_i}(\omega), \qquad \omega \in \Omega. \tag{1}$$

Dann ist $\mathrm{E}(X \mid \mathfrak{G})$ $\mathfrak{G}$-meßbar und gehört zu $\mathfrak{L}^1$. Ferner

$$\int_{B_j} \mathrm{E}(X \mid \mathfrak{G})\, dP = \mathrm{E}(X \mid B_j) \cdot P(B_j) = \int_{B_j} X\, dP,$$

und somit für jedes $G \in \mathfrak{G}$

$$\int_G \mathrm{E}(X \mid \mathfrak{G})\, dP = \int_G X\, dP.$$

Der Übergang von X zu $\mathrm{E}(X \mid \mathfrak{G})$ ist eine 'Vergröberung'. Man vergleiche diese Konstruktion mit dem Beweis der ersten Variante des Satzes von Radon und Nikodym.

Example 1. Extremfälle. Einerseits $|I| = 1$, also $\mathfrak{G} = \{\emptyset, \Omega\}$ und

$$\mathrm{E}(X \mid \mathfrak{G}) = \mathrm{E}(X).$$

Andererseits Ω höchstens abzählbar und $I = \Omega$ mit $B_i = \{i\}$, also $\mathfrak{G} = \mathfrak{P}(\Omega)$ und

$$\mathrm{E}(X \mid \mathfrak{G}) = X.$$

Allgemein: Gegeben: Zufallsvariable X auf $(\Omega, \mathfrak{A}, P)$ mit $X \in \mathfrak{L}^1$ und σ-Algebra $\mathfrak{G} \subset \mathfrak{A}$.

Definition 1. Jede Zufallsvariable Z auf $(\Omega, \mathfrak{A}, P)$ mit $Z \in \mathfrak{L}^1$ sowie

(i) Z ist $\mathfrak{G}$-meßbar,

(ii) $\forall\, G \in \mathfrak{G} : \int_G Z\, dP = \int_G X\, dP$

heißt *(Version der) bedingte(n) Erwartung von X gegeben $\mathfrak{G}$*. Bez.: $Z = \mathrm{E}(X \mid \mathfrak{G})$.
Im Falle $X = 1_A$ mit $A \in \mathfrak{A}$ heißt Z *(Version der) bedingte(n) Wahrscheinlichkeit von A gegeben $\mathfrak{G}$*. Bez.: $Z = P(A \mid \mathfrak{G})$.

Theorem 1. Die bedingte Erwartung existiert und ist $P|_{\mathfrak{G}}$-f.s. eindeutig bestimmt.

Proof. Spezialfall: $X \geq 0$. Durch

$$Q(G) = \int_G X\, dP, \qquad G \in \mathfrak{G},$$

wird gemäß Theorem II.7.1 ein endliches Maß auf $(\Omega, \mathfrak{G})$ definiert. Es gilt $Q \ll P|_{\mathfrak{G}}$. Nach dem Satz von Radon-Nikodym existiert eine $\mathfrak{G}$-meßbare Abbildung $Z : \Omega \to [0, \infty[$ mit

$$\forall\, G \in \mathfrak{G} : \quad Q(G) = \int_G Z\, dP.$$

Der allgemeine Fall wird durch Zerlegung in Positiv- und Negativteil erledigt.
Zur Eindeutigkeit: Aus $\int_G Z_1\, dP|_{\mathfrak{G}} = \int_G Z_2\, dP|_{\mathfrak{G}}$ für alle $G \in \mathfrak{G}$ folgt $Z_1 = Z_2$ $P|_{\mathfrak{G}}$-f.s. Siehe Beweis von Theorem II.7.3. □

Im folgenden oft kurz $X = Y$ oder $X \geq Y$, falls diese Eigenschaften f.s. gelten. Ebenso identifizieren wir Abbildungen, die f.s. übereinstimmen.

Die 'explizite' Bestimmung von bedingten Erwartungen ist i.a. nicht-trivial.

Remark 1. Klar: $\mathrm{E}(X \mid \mathfrak{G}) = X$, falls X $\mathfrak{G}$-meßbar. Ferner:

$$\mathrm{E}(X) = \int_\Omega \mathrm{E}(X \mid \mathfrak{G})\, dP = \mathrm{E}(\mathrm{E}(X \mid \mathfrak{G})).$$

Im Spezialfall (1) für $X = 1_A$ mit $A \in \mathfrak{A}$ ist dies die Formel von der totalen Wahrscheinlichkeit, also

$$P(A) = \sum_{i \in I} P(A \mid B_i) \cdot P(B_i).$$

Lemma 1. Die bedingte Erwartung ist positiv und linear.

Proof. Folgt aus den entsprechenden Eigenschaften des Integrals und den Abschlußeigenschaften von Mengen meßbarer Abbildungen. □

Lemma 2. Sei Y $\mathfrak{G}$-meßbar mit $X \cdot Y \in \mathfrak{L}^1$. Dann

$$\mathrm{E}(X \cdot Y \mid \mathfrak{G}) = Y \cdot \mathrm{E}(X \mid \mathfrak{G}).$$

Proof. Klar: $Y \cdot \mathrm{E}(X \mid \mathfrak{G})$ ist $\mathfrak{G}$-meßbar.

Spezialfall: $Y = 1_C$ mit $C \in \mathfrak{G}$. Dann gilt für $G \in \mathfrak{G}$:

$$\int_G Y \cdot \mathrm{E}(X \mid \mathfrak{G})\, dP = \int_{G \cap C} \mathrm{E}(X \mid \mathfrak{G})\, dP = \int_{G \cap C} X\, dP = \int_G X \cdot Y\, dP.$$

Jetzt algebraische Induktion. □

Lemma 3. Für σ-Algebren $\mathfrak{G}_1 \subset \mathfrak{G}_2 \subset \mathfrak{A}$ gilt

$$\mathrm{E}(\mathrm{E}(X \mid \mathfrak{G}_1) \mid \mathfrak{G}_2) = \mathrm{E}(X \mid \mathfrak{G}_1) = \mathrm{E}(\mathrm{E}(X \mid \mathfrak{G}_2) \mid \mathfrak{G}_1).$$

Proof. Die erste Identität folgt mit Remark 1. Zur zweiten Identität beachte man für $G \in \mathfrak{G}_1 \subset \mathfrak{G}_2$

$$\int_G \mathrm{E}(\mathrm{E}(X \mid \mathfrak{G}_2) \mid \mathfrak{G}_1)\, dP = \int_G \mathrm{E}(X \mid \mathfrak{G}_2)\, dP = \int_G X\, dP.$$

□

Terminologie: X und $\mathfrak{G}$ heißen unabhängig, falls $(\sigma(X), \mathfrak{G})$ unabhängig ist.

Lemma 4. Seien X und $\mathfrak{G}$ unabhängig. Dann

$$\mathrm{E}(X \mid \mathfrak{G}) = \mathrm{E}(X).$$

Proof. Klar: $\mathrm{E}(X)$ $\mathfrak{G}$-meßbar. Sei $G \in \mathfrak{G}$. Nach Voraussetzung sind X und 1_G unabhängig. Also

$$\int_G \mathrm{E}(X)\, dP = \mathrm{E}(X) \cdot \mathrm{E}(1_G) = \mathrm{E}(X \cdot 1_G) = \int_G X\, dP.$$

□

Theorem 2 (Jensensche Ungleichung)**.** Sei $J \subset \mathbb{R}$ ein Intervall, und gelte $X(\omega) \in J$ für alle $\omega \in \Omega$. Sei $\varphi : J \to \mathbb{R}$ konvex mit $\varphi \circ X \in \mathfrak{L}^1$. Dann:

$$\varphi \circ \mathrm{E}(X \mid \mathfrak{G}) \leq \mathrm{E}(\varphi \circ X \mid \mathfrak{G}).$$

Proof. Gänssler, Stute (1977, Kap. V.4). □

Remark 2. Spezialfall: $J = \mathbb{R}$ und $\varphi(u) = |u|^{p/q}$ mit $1 \leq q \leq p < \infty$. Dann:

$$\left(\mathrm{E}(|X|^q \mid \mathfrak{G})\right)^{1/q} \leq \left(\mathrm{E}(|X|^p \mid \mathfrak{G})\right)^{1/p}$$

für $X \in \mathfrak{L}^p$ sowie

$$\left(\int_\Omega |\,\mathrm{E}(X \mid \mathfrak{G})|^p\, dP\right)^{1/p} \leq \left(\int_\Omega |X|^p\, dP\right)^{1/p}. \qquad (2)$$

Also ist $\mathrm{E}(\cdot \mid \mathfrak{G})$ ein idempotenter beschränkter linearer Operator auf $L^p(\Omega, \mathfrak{A}, P)$ mit Norm 1. Speziell für $p = 2$: orthogonale Projektion auf den Unterraum $L^2(\Omega, \mathfrak{G}, P)$.

Oft liegt folgende Situation vor. Gegeben: Meßraum $(\Omega', \mathfrak{A}')$ und Zufallselement $Y : \Omega \to \Omega'$. Betrachte die von Y erzeugte σ-Algebra $\mathfrak{G} = \sigma(Y)$.

Definition 2. *Bedingte Erwartung von X gegeben Y:*

$$\mathrm{E}(X \mid Y) = \mathrm{E}(X \mid \sigma(Y)).$$

Anwendung von Theorem II.2.8 auf obige Situation: Faktorisierung der bedingten Erwartung: Es existiert eine $\mathfrak{A}'$-meßbare Abbildung $g : \Omega' \to \mathbb{R}$ mit

$$\mathrm{E}(X \mid Y) = g \circ Y.$$

Je zwei solche Abbildungen stimmen P_Y-f.s. überein.

Definition 3. *Bedingte Erwartung von X gegeben $Y = y$:*

$$\mathrm{E}(X \mid Y = y) = g(y),$$

wobei g wie oben gewählt.

Analoge Begriffsbildung für *bedingte Wahrscheinlichkeiten*, wobei $X = 1_A$ mit $A \in \mathfrak{A}$.

Example 2. Gelte $(\Omega, \mathfrak{A}, P) = ([0,1], \mathfrak{B}([0,1]), \lambda)$ und $(\Omega', \mathfrak{A}') = (\mathbb{R}, \mathfrak{B})$. Ferner

$$X(\omega) = \omega^2, \qquad Y(\omega) = \begin{cases} 1, & \text{falls } \omega \in [0,1/2], \\ \omega - 1/2, & \text{falls } \omega \in \,]1/2,1]. \end{cases}$$

Dann

$$\sigma(Y) = \{A \cup B : A \in \{\emptyset, [0,1/2]\},\ B \subset \,]1/2,1]\,,\ B \in \mathfrak{A}\}$$

sowie

$$\mathrm{E}(X \,|\, Y)(\omega) = \begin{cases} 1/12, & \text{falls } \omega \in [0,1/2], \\ \omega^2, & \text{falls } \omega \in \,]1/2,1] \end{cases}$$

und

$$\mathrm{E}(X \,|\, Y = y) = \begin{cases} 1/12, & \text{falls } y = 1, \\ (y+1/2)^2, & \text{falls } y \in \,]0,1/2]. \end{cases}$$

Beachte, daß $P(\{Y = y\}) = 0$ für alle $y \in \,]0,1/2]$.

Remark 3. Klar: für $A' \in \mathfrak{A}'$ gilt

$$\int_{\{Y \in A'\}} X \, dP = \int_{A'} \mathrm{E}(X \,|\, Y = y)\, P_Y(dy) \tag{3}$$

und insbesondere

$$P(A \cap \{Y \in A'\}) = \int_{A'} P(A \,|\, Y = y)\, P_Y(dy)$$

für $A \in \mathfrak{A}$. Durch (3) für alle $A' \in \mathfrak{A}'$ und die Forderung der $\mathfrak{A}'$-Meßbarkeit ist $\mathrm{E}(X \,|\, Y = \cdot)$ P_Y-f.s. eindeutig bestimmt.

Wie das folgende Theorem zeigt, ist die bedingte Erwartung die beste Vorhersage im Quadratmittel. Vgl. Einführung in die Stochastik und Lemma 4.

Theorem 3. Gelte $X \in \mathfrak{L}^2$. Dann gilt für jede $\mathfrak{A}'$-meßbare Abbildung $\varphi : \Omega' \to \mathbb{R}$

$$\int_\Omega (X - \mathrm{E}(X \,|\, Y))^2 \, dP \le \int_\Omega (X - \varphi \circ Y)^2 \, dP$$

mit Gleichheit gdw. $\varphi = \mathrm{E}(X \,|\, Y = \cdot)$ P_Y-f.s.

Proof. Setze $Z^* = \mathrm{E}(X \,|\, Y)$ und $Z = \varphi \circ Y$. Die Jensensche Ungleichung liefert $Z^* \in \mathfrak{L}^2$, siehe (2) mit $p = 2$. OBdA: $Z \in \mathfrak{L}^2$. Dann

$$\mathrm{E}(X - Z)^2 = \mathrm{E}(X - Z^*)^2 + \underbrace{\mathrm{E}(Z^* - Z)^2}_{\ge 0} + 2 \cdot \mathrm{E}((X - Z^*)(Z^* - Z)).$$

Mit Lemma 1 und 2 folgt

$$\begin{aligned} \mathrm{E}((X - Z^*)(Z^* - Z)) &= \int_\Omega \mathrm{E}((X - Z^*)(Z^* - Z) \,|\, Y)\, dP \\ &= \int_\Omega (Z^* - Z) \cdot \mathrm{E}((X - Z^*) \,|\, Y)\, dP \\ &= \int_\Omega (Z^* - Z) \cdot \underbrace{(\mathrm{E}(X \,|\, Y) - Z^*)}_{=0}\, dP. \end{aligned}$$

□

Nun: der Zusammenhang zwischen Markov-Kernen und bedingten Wahrscheinlichkeiten.

Gegeben: Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$ sowie Meßräume $(\Omega', \mathfrak{A}')$ und $(\Omega'', \mathfrak{A}'')$. Ferner eine $\mathfrak{A}$-$\mathfrak{A}'$ meßbare Abbildung $Y : \Omega \to \Omega'$ und eine $\mathfrak{A}$-$\mathfrak{A}''$ meßbare Abbildung $X : \Omega \to \Omega''$. Spezialfall

$$(\Omega'', \mathfrak{A}'') = (\Omega, \mathfrak{A}), \qquad X = \mathrm{id}\,. \tag{4}$$

Lemma 5. Für jede Abbildung $P_{X\,|\,Y} : \Omega' \times \mathfrak{A}'' \to \mathbb{R}$ sind äquivalent:

(i) $P_{X\,|\,Y}$ Markov-Kern von $(\Omega', \mathfrak{A}')$ nach $(\Omega'', \mathfrak{A}'')$ und

$$P_{(Y,X)} = P_Y \times P_{X\,|\,Y}, \tag{5}$$

(ii) für jedes $y \in \Omega'$ ist $P_{X\,|\,Y}(y, \cdot)$ ein Wahrscheinlichkeitsmaß auf $(\Omega'', \mathfrak{A}'')$ und für alle $A'' \in \mathfrak{A}''$ gilt

$$P_{X\,|\,Y}(\cdot, A'') = P(\{X \in A''\} \,|\, Y = \cdot).$$

Gilt (i), so sind X und Y genau dann unabhängig, wenn

$$P_{X\,|\,Y}(y, \cdot) = P_X$$

für P_Y-f.a. $y \in \mathbb{R}$ gilt.

Proof. Definitionsgemäß gilt für $A' \in \mathfrak{A}'$ und $A'' \in \mathfrak{A}''$

$$P_Y \times P_{X\,|\,Y}(A' \times A'') = \int_{A'} P_{X\,|\,Y}(y, A'')\, P_Y(dy)$$

und

$$P_{(Y,X)}(A' \times A'') = \int_{A'} P(\{X \in A''\} \,|\, Y = y)\, P_Y(dy).$$

Dies zeigt die Äquivalenz von (i) und (ii). Charakterisierung der Unabhängigkeit: Übung 14.2. □

Example 3. Betrachte ein Wahrscheinlichkeitsmaß μ auf $(\Omega', \mathfrak{A}')$ und einen Markov-Kern $P_{X\,|\,Y}$ von $(\Omega', \mathfrak{A}')$ nach $(\Omega'', \mathfrak{A}'')$. Auf dem Produktraum

$$(\Omega, \mathfrak{A}) = (\Omega' \times \Omega'', \mathfrak{A}' \otimes \mathfrak{A}'')$$

betrachten wir das Wahrscheinlichkeitsmaß

$$P = \mu \times P_{X\,|\,Y}$$

und die Projektionen

$$Y(\omega', \omega'') = \omega', \qquad X(\omega', \omega'') = \omega''.$$

Unmittelbar aus den Definitionen folgt $\mu = P_Y$ und $P = P_{(Y,X)}$. Lemma 5 sichert für jedes $A'' \in \mathfrak{A}''$, daß

$$P_{X\,|\,Y}(y, A'') = P(\{X \in A''\} \,|\, Y = y)$$

für P_Y-f.a. $y \in \Omega'$ gilt. Analog für Folgen von Markov-Kernen.

Example 4. Sei $(\Omega', \mathfrak{A}') = (\Omega'', \mathfrak{A}'') = (\mathbb{R}, \mathfrak{B})$ und gelte

$$P_{(Y,X)} = f \cdot \lambda_2$$

mit einer Wahrscheinlichkeitsdichte f auf $(\mathbb{R}^2, \mathfrak{B}_2)$. Für $A', A'' \in \mathfrak{B}$ sichert der Satz von Fubini (beachte: Meßbarkeit als Teilaussage)

$$P_{(Y,X)}(A' \times A'') = \int_{A' \times A''} f \, d(\lambda_1 \times \lambda_1) = \int_{A'} \int_{A''} f(y,x) \, \lambda_1(dx) \, \lambda_1(dy).$$

Die Wahl von $A'' = \mathbb{R}$ zeigt

$$P_Y = h \cdot \lambda_1$$

mit der Wahrscheinlichkeitsdichte

$$h(y) = \int_{\mathbb{R}} f(y, \cdot) \, d\lambda_1, \qquad y \in \mathbb{R}.$$

Definiere für $y, x \in \mathbb{R}$

$$f(x \,|\, y) = \begin{cases} f(y,x)/h(y) & \text{falls } h(y) > 0 \\ 1_{[0,1]}(x) & \text{sonst} \end{cases}$$

und einen Markov-Kern $P_{X \,|\, Y} : \Omega' \times \mathfrak{A}'' \to [0,1]$ durch

$$P_{X \,|\, Y}(y, A'') = \int_{A''} f(x \,|\, y) \, \lambda_1(dx),$$

also $P_{X \,|\, Y}(y, \cdot) = f(\cdot \,|\, y) \cdot \lambda_1$. Man erhält

$$\begin{aligned} P_{(Y,X)}(A' \times A'') &= \int_{A'} \int_{A''} f(x \,|\, y) \, \lambda_1(dx) \cdot h(y) \, \lambda_1(dy) \\ &= \int_{A'} P_{X \,|\, Y}(y, A'') \, P_Y(dy). \end{aligned}$$

Lemma 5 zeigt

$$P(\{X \in A''\} \,|\, Y = y) = \int_{A''} f(x \,|\, y) \, \lambda_1(dy)$$

für P_Y-f.a. $y \in \Omega'$. Die Abbildung $f(\cdot \,|\, \cdot)$ heißt auch *bedingte Dichte von X gegeben Y*.

Definition 4. Jeder Markov-Kern $P_{X \,|\, Y}$ von $(\Omega', \mathfrak{A}')$ nach $(\Omega'', \mathfrak{A}'')$ mit der Eigenschaft (5) heißt eine *reguläre bedingte Verteilung von X gegeben Y* und im Spezialfall (4) auch *reguläre bedingte Wahrscheinlichkeit gegeben Y*. Ferner heißt (5) *Desintegration der gemeinsamen Verteilung von Y und X*.

Example 5. In Example 4 ist $P_{X \,|\, Y}$ eine reguläre bedingte Verteilung von X gegeben Y. Wesentlich für die Richtigkeit dieser Sachverhalts ist nur, daß die gemeinsame Verteilung von X und Y eine Dichte bzgl. des Produktes zweier σ-endlicher Maße besitzt. Schließlich gilt für P_Y-f.a. $y \in \mathbb{R}$

$$\mathrm{E}(X \,|\, Y = y) = \int_{\mathbb{R}} x \, P_{X \,|\, Y}(y, dx) = \int_{\mathbb{R}} x \cdot f(x \,|\, y) \, \lambda_1(dx). \tag{6}$$

Beweis: Übung 14.2.

Example 6. In der Situation von Example 2 gilt

$$P_{X\,|\,Y}(y,\cdot) = \begin{cases} g \cdot \lambda_1 & \text{falls } y = 1, \\ \varepsilon_{(y+1/2)^2} & \text{falls } y \in \,]0,1/2], \end{cases}$$

wobei die Dichte g durch

$$g(x) = 1/\sqrt{x} \cdot 1_{]0,1/4]}(x)$$

gegeben ist. Ferner gilt

$$P_{\mathrm{id}\,|\,Y}(y,\cdot) = \begin{cases} \nu & \text{falls } y = 1, \\ \varepsilon_{y+1/2} & \text{falls } y \in \,]0,1/2], \end{cases}$$

wobei ν die Gleichverteilung auf $[0,1/2]$ bezeichnet. Beweis: Übung 14.1.

Remark 4. Betrachte im Spezialfall (4) paarweise disjunkte Mengen $A_1, A_2, \ldots \in \mathfrak{A}$. Für jede Menge $A' \in \mathfrak{A}'$ gilt

$$\begin{aligned} \int_{A'} P\left(\bigcup_{i=1}^{\infty} A_i \,|\, Y = y\right) P_Y(dy) &= P\left(\bigcup_{i=1}^{\infty} A_i \cap \{Y \in A'\}\right) \\ &= \sum_{i=1}^{\infty} P(A_i \cap \{Y \in A'\}) \\ &= \sum_{i=1}^{\infty} \int_{A'} P(A_i \,|\, Y = y)\, P_Y(dy) \\ &= \int_{A'} \sum_{i=1}^{\infty} P(A_i \,|\, Y = y)\, P_Y(dy). \end{aligned}$$

Es folgt P_Y-f.s.

$$P\left(\bigcup_{i=1}^{\infty} A_i \,|\, Y = \cdot\right) = \sum_{i=1}^{\infty} P(A_i \,|\, Y = \cdot).$$

Beachte: die entsprechende Nullmenge in $\mathfrak{A}'$ kann von der Wahl der Mengen A_i abhängen.

Theorem 4. Gelte $(\Omega', \mathfrak{A}') = (M, \mathfrak{B}(M))$ mit einem vollständigen und separablen metrischen Raum (M, ρ). Dann existiert eine reguläre bedingte Verteilung von X gegeben Y. Für je zwei solche Verteilungen $P^{(i)}_{X\,|\,Y}$ existiert eine Menge $A' \in \mathfrak{A}'$ mit $P_Y(A') = 1$ und

$$\forall\, y \in A' \;\forall\, A'' \in \mathfrak{A}'': \quad P^{(1)}_{X\,|\,Y}(y, A'') = P^{(2)}_{X\,|\,Y}(y, A'').$$

Proof. Siehe Gänssler, Stute (1977, Kap. V.3) oder Yeh (1995, App. C). □

Im folgenden sei $V \in \mathfrak{L}^1(\Omega, \mathfrak{A}, P)$ und $P_{\mathrm{id}\,|\,Y}$ eine reguläre bedingte Wahrscheinlichkeit gegeben Y.

Theorem 5.

(i) $\int_\Omega V(\omega)\, P_{\mathrm{id}\,|\,Y}(\cdot, d\omega)$ $\mathfrak{A}'$-meßbar,

(ii) $\int_{A'}\int_\Omega V(\omega)\, P_{\mathrm{id}\,|\,Y}(y, d\omega)\, P_Y(dy) = \int_{\{Y\in A'\}} V\, dP$ für $A' \in \mathfrak{A}'$.

Also für P_Y-f.a. $y \in \Omega'$

$$\mathrm{E}(V\,|\,Y = y) = \int_\Omega V(\omega)\, P_{\mathrm{id}\,|\,Y}(y, d\omega).$$

Proof. Algebraische Induktion. □

Theorem 6. Für P_Y-f.a. $y \in \Omega'$

$$P_{\mathrm{id}\,|\,Y}(y, Y^{-1}(\{y\})) = 1.$$

Proof. Siehe Yeh (1995, p. 486). □

Die Ergebnisse dieses Abschnittes beantworten die im einführenden Example I.4 gestellten Fragen.

2 Discrete-Time Martingales

Gegeben: Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$.

Definition 1. Folge $\widetilde{\mathfrak{A}} = (\mathfrak{A}_n)_{n\in\mathbb{N}_0}$ von σ-Algebren $\mathfrak{A}_n \subset \mathfrak{A}$ mit

$$\forall\, n \in \mathbb{N}_0 : \; \mathfrak{A}_n \subset \mathfrak{A}_{n+1}$$

heißt *Filtration*.

Gegeben: Folge $\widetilde{X} = (X_n)_{n\in\mathbb{N}_0}$ von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$.

Example 1. *Kanonische Filtration*:

$$\mathfrak{A}_n = \sigma(\{X_0, \dots, X_n\}), \qquad n \in \mathbb{N}_0, \tag{1}$$

definiert die 'kleinste' Filtration $\widetilde{\mathfrak{A}}$, so daß X_n $\mathfrak{A}_n$-meßbar ist. Bei dieser Wahl ist $Y : \Omega \to \mathbb{R}$ genau dann $\mathfrak{A}_n$-meßbar, wenn eine $\mathfrak{B}_{n+1}$-meßbare Abbildung $g : \mathbb{R}^{n+1} \to \mathbb{R}$ existiert, die $Y = g \circ (X_0, \dots, X_n)$ erfüllt. Siehe Theorem II.2.8 und Corollary II.3.1.(i).

Definition 2. $\widetilde{X}$ heißt *Martingal (bzgl. $\widetilde{\mathfrak{A}}$)*, falls $X_n \in \mathfrak{L}^1$ für alle $n \in \mathbb{N}_0$ und

$$\forall\, n, m \in \mathbb{N}_0 : \quad n < m \quad \Rightarrow \quad \mathrm{E}(X_m\,|\,\mathfrak{A}_n) = X_n.$$

Zur Interpretation: Theorem 1.3.

Remark 1. Für jedes Martingal $\widetilde{X}$ und $n < m$

$$\mathrm{E}(X_m) = \mathrm{E}(\mathrm{E}(X_m \,|\, \mathfrak{A}_n)) = \mathrm{E}(X_n).$$

Klar: Aus der Konstanz der Erwartungswerte folgt i.a. nicht die Martingaleigenschaft.

Remark 2. $\widetilde{X}$ Martingal gdw.

$$\forall\, n \in \mathbb{N}_0 : \quad \mathrm{E}(X_{n+1} \,|\, \mathfrak{A}_n) = X_n.$$

Zum Beweis verwende man Lemma 1.3.

Example 2. Sei $(Y_i)_{i\in\mathbb{N}}$ eine unabhängige Folge von Zufallsvariablen mit $\mathrm{E}(Y_i) = a$ für alle $i \in \mathbb{N}$. Ferner sei $\mathfrak{A}_0 = \{\emptyset, \Omega\}$ und $\mathfrak{A}_n = \sigma(\{Y_1, \dots, Y_n\})$ für $n \geq 1$. Setze $X_0 = 0$ und

$$X_n = \sum_{i=1}^{n} Y_i, \qquad n \in \mathbb{N}.$$

Es gilt (1) sowie

$$\mathrm{E}(X_{n+1} \,|\, \mathfrak{A}_n) = \mathrm{E}(X_n \,|\, \mathfrak{A}_n) + \mathrm{E}(Y_{n+1} \,|\, \mathfrak{A}_n) = X_n + \mathrm{E}(Y_{n+1}) = X_n + a.$$

Somit ist $\widetilde{X}$ genau im Falle $a = 0$ ein Martingal. Bsp: random walk, coin tossing.

Mögliche Interpretation: Y_i Gewinn bei einfachem Spiel in Runde i und X_n akkumulierter Gewinn nach n Runden. Martingal heißt: 'faires Spiel'.

Frage: Kann man im Martingalfall durch

(i) eine geeignete Wahl der Einsätze und

(ii) einen geeigneten Abbruch des Spiels

im Mittel einen positiven Gesamtgewinn erreichen?

Example 3. Das *Cox-Ross-Rubinstein-Modell* für Aktienkurse X_n zu Zeiten $n \in \mathbb{N}_0$. Wähle reelle Zahlen

$$X_0 > 0, \quad 0 < p < 1, \quad 0 < d < u,$$

und betrachte $(Y_i)_{i\in\mathbb{N}}$ i.i.d. mit

$$P(\{Y_i = u\}) = p = 1 - P(\{Y_i = d\}).$$

Setze $\mathfrak{A}_0 = \{\emptyset, \Omega\}$ und definiere

$$X_n = X_0 \cdot \prod_{i=1}^{n} Y_i$$

sowie $\mathfrak{A}_n = \sigma(\{Y_1, \dots, Y_n\})$ für $n \in \mathbb{N}$. Für $n < m$ zeigen Lemma 1.2 und Lemma 1.4

$$\mathrm{E}(X_m \,|\, \mathfrak{A}_n) = X_n \cdot \mathrm{E}\left(\prod_{\ell=n+1}^{m} Y_\ell\right) = X_n \cdot \mathrm{E}(Y_1)^{m-n}.$$

Also

$$\widetilde{X} \text{ Martingal} \quad \Leftrightarrow \quad \mathrm{E}(Y_1) = 1,$$

bzw., da $\mathrm{E}(Y_1) = pu + (1-p)d$,

$$\widetilde{X} \text{ Martingal} \quad \Leftrightarrow \quad d < 1 < u \;\wedge\; p = \frac{1-d}{u-d}.$$

Frage: Wie in Example 2 ('Handelsstrategie', 'Verkaufsstrategie').

Im folgenden:

(i) $\widetilde{X} = (X_n)_{n\in\mathbb{N}_0}$ Martingal bzgl. $\widetilde{\mathfrak{A}}$,

(ii) $\widetilde{H} = (H_n)_{n\in\mathbb{N}_0}$ Folge von Zufallsvariablen, so daß

$$\forall\, n \in \mathbb{N}_0 : \;H_n \;\mathfrak{A}_n\text{-meßbar} \wedge \;H_n \cdot (X_{n+1} - X_n) \in \mathfrak{L}^1.$$

Definition 3. Die Folge $\widetilde{Z} = (Z_n)_{n\in\mathbb{N}_0}$ von Zufallsvariablen $Z_0 = 0$ und

$$Z_n = \sum_{i=0}^{n-1} H_i \cdot (X_{i+1} - X_i), \qquad n \geq 1,$$

heißt *Martingaltransformation von* $\widetilde{X}$ *mittels* $\widetilde{H}$. Bez.: $\widetilde{Z} = \widetilde{H} \bullet \widetilde{X}$.

Example 4. In Example 2: H_n Einsatz im $(n+1)$-ten Spiel und Z_n akkumulierter Gewinn nach n Runden mit Einsätzen $H_0, \dots, H_{n-1}$. Spezialfall: $H_n \in \{\pm 1\}$ bei coin tossing.

Theorem 1. $\widetilde{Z} = \widetilde{H} \bullet \widetilde{X}$ ist Martingal bzgl. $\widetilde{\mathfrak{A}}$.

Proof. Klar für $n \in \mathbb{N}_0$: Z_n ist $\mathfrak{A}_n$-meßbar und $Z_n \in \mathfrak{L}^1$. Somit

$$\mathrm{E}(Z_{n+1} \,|\, \mathfrak{A}_n) = Z_n + \mathrm{E}(H_n \cdot (X_{n+1} - X_n) \,|\, \mathfrak{A}_n),$$

und weiter

$$\mathrm{E}(H_n \cdot (X_{n+1} - X_n) \,|\, \mathfrak{A}_n) = H_n \cdot \mathrm{E}((X_{n+1} - X_n) \,|\, \mathfrak{A}_n) = 0.$$

□

3 Optional Sampling

Gegeben: Filtration $\widetilde{\mathfrak{A}}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$.

Definition 1. $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ heißt *Stoppzeit (bzgl.* $\widetilde{\mathfrak{A}}$*)*, falls

$$\forall\, n \in \mathbb{N}_0 : \;\{\tau \leq n\} \in \mathfrak{A}_n.$$

Lemma 1.

$$\tau \text{ Stoppzeit} \quad \Leftrightarrow \quad \forall\, n \in \mathbb{N}_0 : \;\{\tau = n\} \in \mathfrak{A}_n.$$

Proof. Verwende

$$\{\tau \leq n\} = \bigcup_{i=0}^{n} \{\tau = i\}, \qquad \{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\}.$$

□

Example 1. Verkaufsstrategien für eine Aktie mit Preis X_n zur Zeit $n \in \mathbb{N}_0$:

(i) Verkaufe, sobald der Preis a erreicht oder überschritten ist, spätestens jedoch zur Zeit N.

(ii) Verkaufe beim ersten Eintreten des Maximum von $X_0, \dots, X_N$.

Formal heißt (i)

$$\tau = \inf\left(\{i \in \{0, \dots, N\} : X_i \geq a\} \cup \{N\}\right).$$

Dann: τ ist Stoppzeit bzgl. der kanonischen Filtration $\widetilde{\mathfrak{A}}$ zu $\widetilde{X}$, d.h. 'realisierbare Strategie'. Es gilt nämlich für $k = 0, \dots, N-1$

$$\{\tau = k\} = \bigcap_{i=0}^{k-1} \underbrace{\{X_i < a\}}_{\in \mathfrak{A}_i \subset \mathfrak{A}_{k-1}} \cap \underbrace{\{X_k \geq a\}}_{\in \mathfrak{A}_k} \in \mathfrak{A}_k$$

sowie

$$\{\tau = N\} = \bigcap_{i=0}^{N-1} \{X_i < a\} \in \mathfrak{A}_{N-1}.$$

Formal heißt (ii)

$$\tau = \inf\{i \in \{0, \dots, N\} : X_i = M\} \qquad \text{mit} \qquad M = \max_{i=0,\dots,N} X_i.$$

Dies ist i.a. keine Stoppzeit, d.h. eine 'nicht realisierbare Strategie'. Betrachte etwa das Cox-Ross-Rubinstein-Modell mit $d < 1 < u$. Für $N = 1$ gilt

$$\{\tau = 0\} = \{Y_1 = d\} \notin \{\emptyset, \Omega\} = \mathfrak{A}_0.$$

Lemma 2.

$$\sigma, \tau \text{ Stoppzeiten bzgl. } \widetilde{\mathfrak{A}} \quad \Rightarrow \quad \sigma + \tau,\ \min\{\sigma, \tau\},\ \max\{\sigma, \tau\} \text{ Stoppzeiten bzgl. } \widetilde{\mathfrak{A}}.$$

Proof. Verwende

$$\{\min(\sigma, \tau) \leq n\} = \{\sigma \leq n\} \cup \{\tau \leq n\},$$
$$\{\max(\sigma, \tau) \leq n\} = \{\sigma \leq n\} \cap \{\tau \leq n\}$$

und

$$\{\sigma + \tau \leq n\} = \bigcup_{k=0}^{n} \{\sigma = k\} \cap \{\tau = n - k\}.$$

□

Gegeben: Folge $\widetilde{X} = (X_n)_{n \in \mathbb{N}_0}$ von Zufallsvariablen auf $(\Omega, \mathfrak{A}, P)$, so daß für alle $n \in \mathbb{N}_0$ gilt:

(i) X_n $\mathfrak{A}_n$-meßbar,

(ii) $X_n \in \mathfrak{L}^1$.

Für eine Abbildung $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ definieren wir

$$X_\tau : \Omega \to \mathbb{R}$$

durch

$$X_\tau(\omega) = \begin{cases} X_{\tau(\omega)}(\omega) & \text{falls } \tau(\omega) < \infty \\ 0 & \text{sonst.} \end{cases}$$

Die folgenden beiden Sätze sind Varianten des *optional sampling theorem.*

Theorem 1.

$$\widetilde{X} \text{ Martingal bzgl. } \widetilde{\mathfrak{A}} \quad \Leftrightarrow \quad \forall\, \tau \text{ beschränkte Stoppzeit bzgl. } \widetilde{\mathfrak{A}} : \mathrm{E}(X_\tau) = \mathrm{E}(X_0).$$

Proof. '$\Rightarrow$' Sei τ eine Stoppzeit mit $\tau(\omega) \le N$ für alle $\omega \in \Omega$. Also

$$X_\tau = \sum_{n=0}^{N} 1_{\{\tau=n\}} \cdot X_n.$$

Also ist X_τ $\mathfrak{A}$-meßbar und $\mathrm{E}(|X_\tau|) \le \sum_{n=0}^{N} \mathrm{E}(|X_n|) < \infty$. Weiter

$$\begin{aligned} \mathrm{E}(X_\tau) &= \sum_{n=0}^{N} \mathrm{E}(1_{\{\tau=n\}} \cdot X_n) = \sum_{n=0}^{N} \mathrm{E}(1_{\{\tau=n\}} \cdot \mathrm{E}(X_N \,|\, \mathfrak{A}_n)) \\ &= \sum_{n=0}^{N} \mathrm{E}(\mathrm{E}(1_{\{\tau=n\}} \cdot X_N \,|\, \mathfrak{A}_n)) = \sum_{n=0}^{N} \mathrm{E}(1_{\{\tau=n\}} \cdot X_N) = \mathrm{E}(X_N) = \mathrm{E}(X_0). \end{aligned}$$

'$\Leftarrow$' Für $n < m$ und $A \in \mathfrak{A}_n$ ist zu zeigen

$$\int_A X_m \, dP = \int_A X_n \, dP.$$

Definiere

$$\tau = n \cdot 1_A + m \cdot 1_{\Omega \setminus A}.$$

Klar: τ ist beschränkte Stoppzeit. Also

$$\mathrm{E}(X_0) = \mathrm{E}(X_\tau) = \mathrm{E}(1_A \cdot X_n + 1_{\Omega\setminus A} \cdot X_m) = \mathrm{E}(X_m) - \mathrm{E}(1_A \cdot X_m) + \mathrm{E}(1_A \cdot X_n).$$

Beachte schließlich, daß n.V. insbesondere $\mathrm{E}(X_0) = \mathrm{E}(X_m)$ gilt. □

Theorem 2.1 und Theorem 1 beantworten die in Example 2.2 gestellten Fragen negativ, solange man eine obere Schranke für die Spieldauer akzeptiert.

Theorem 2. Sei $\widetilde{X}$ Martingal und τ Stoppzeit mit

$$P(\{\tau < \infty\}) = 1 \quad \wedge \quad \mathrm{E}(|X_\tau|) < \infty \quad \wedge \quad \lim_{n\to\infty} \int_{\{\tau > n\}} |X_n| \, dP = 0. \tag{1}$$

Dann

$$\mathrm{E}(X_\tau) = \mathrm{E}(X_0).$$

Proof. Für $\tau_N = \min\{\tau, N\}$ gilt

$$|\operatorname{E}(X_\tau) - \operatorname{E}(X_{\tau_N})| \leq \int_{\{\tau > N\}} |X_\tau| \, dP + \int_{\{\tau > N\}} |X_N| \, dP$$

und somit

$$\lim_{N \to \infty} \operatorname{E}(X_{\tau_N}) = \operatorname{E}(X_\tau).$$

Theorem 1 und Lemma 2 liefern $\operatorname{E}(X_0) = \operatorname{E}(X_{\tau_N})$. □

Example 2. In Example 2.2 gelte: $(Y_i)_{i \in \mathbb{N}}$ i.i.d. mit $P_{Y_1} = 1/2 \cdot (\varepsilon_1 + \varepsilon_{-1})$. Einsatz $H_n = 2^n$ in $(n+1)$-ten Spiel (Verdopplungsstrategie). Nach Theorem 1 (einfacher: Example 2.2) definiert $Z_0 = 0$ und

$$Z_n = \sum_{i=0}^{n-1} 2^i \cdot Y_{i+1}, \qquad n \in \mathbb{N},$$

ein Martingal. Für die Stoppzeit

$$\tau = \inf\{i \in \mathbb{N} : Y_i = 1\}$$

ergibt sich

(i) $\tau = \inf\{n \in \mathbb{N}_0 : Z_n > 0\}$,

(ii) $Z_\tau = 1$,

(iii) $P(\{\tau = n\}) = 2^{-n}$, also τ f.s. endlich und $\operatorname{E}(\tau) = 2$.

Jedoch ist $\tau > n$ äquivalent zu $Z_n = -1 - \cdots - 2^{n-1} = -(2^n - 1)$, so daß

$$\int_{\{\tau > n\}} |Z_n| \, dP = (2^n - 1) \cdot \sum_{m=n+1}^{\infty} 2^{-m} = 1 - 2^{-n}.$$

Example 3 (Das Ruin-Problem)**.** Betrachte das Glücksspiel aus Example 2.2 mit

$$P_{Y_i} = p \cdot \varepsilon_1 + (1-p) \cdot \varepsilon_{-1}$$

für festes $p \in \,]0,1[$. Startkapital $C > 0$. Ziel: Gewinn $G > 0$. Spiele bis G erreicht oder C verspielt. Also

$$\tau = \inf\left\{n \in \mathbb{N}_0 : X_n = G \vee X_n = -C\right\}.$$

Bestimme die Ruin-Wahrscheinlichkeit

$$p_{-C} = P(\{X_\tau = -C\})$$

sowie den Erwartungswert der Spieldauer τ.

Dazu zeigt man vorab

$$\exists\, \alpha > 0 \;\exists\, \gamma \in \,]0,1[\;\; \forall\, j \in \mathbb{N}_0 : \quad P(\{\tau > j\}) \leq \alpha \cdot \gamma^j, \tag{2}$$

siehe Einführung in die Stochastik, Bsp. II.1.3.
Mit (2) folgt
$$P(\{\tau = \infty\}) \leq \liminf_{j\to\infty} P(\{\tau > j\}) = 0$$
und weiter
$$\mathrm{E}(\tau) = \sum_{j=1}^{\infty} P(\{\tau \geq j\}) < \infty.$$
Also
$$1 = P(\{\tau < \infty\}) = p_G + p_{-C} \tag{3}$$
mit der Gewinn-Wahrscheinlichkeit
$$p_G = P(\{X_\tau = G\}).$$

Nun Anwendung des optional sampling theorem. Klar: τ ist unbeschränkt, deshalb verwenden wir Theorem 2.
Definiere $M_0 = 0$ und
$$M_n = \sum_{i=1}^{n} (Y_i - \mathrm{E}(Y_i)) = X_n - na,$$
wobei $a = 2p - 1$. Dann ist $\widetilde{M}$ ein Martingal, siehe Example 2.2. Wir verifizieren die weiteren Voraussetzungen von Theorem 2.
Es gilt
$$|M_\tau| \leq |X_\tau| + \tau \cdot |a| \leq \max\{G, C\} + |a| \cdot \tau,$$
und somit
$$\mathrm{E}(|M_\tau|) \leq \max\{G, C\} + |a| \cdot \mathrm{E}(\tau) < \infty.$$
Ferner
$$\begin{aligned}\int_{\{\tau>n\}} |M_n| \, dP &\leq \int_{\{\tau>n\}} (|X_n| + |a| \cdot n) \, dP \\ &\leq \max\{G, C\} \cdot P(\{\tau > n\}) + |a| \cdot n \cdot P(\{\tau > n\}),\end{aligned}$$
und somit sichert (2)
$$\lim_{n\to\infty} \int_{\{\tau>n\}} |M_n| \, dP = 0.$$
Theorem 2 liefert
$$0 = \mathrm{E}(M_0) = \mathrm{E}(M_\tau) = \mathrm{E}(X_\tau) - \mathrm{E}(\tau) \cdot a = G \cdot p_G - C \cdot p_{-C} - \mathrm{E}(\tau) \cdot a. \tag{4}$$

1. Fall: Faires Spiel, d.h.
$$p = \frac{1}{2}.$$
Dann $a = 0$, und (3) sowie (4) sichern
$$p_G = \frac{C}{C+G}, \qquad p_{-C} = \frac{G}{C+G}.$$

Weiterhin ist $(X_n^2 - n)_{n\in\mathbb{N}_0}$ ein Martingal, und die Voraussetzungen von Theorem 2 sind erfüllt. Also

$$0 = \mathrm{E}(X_0^2 - 0) = \mathrm{E}(X_\tau^2 - \tau) = \mathrm{E}(X_\tau^2) - \mathrm{E}(\tau),$$

so daß

$$\mathrm{E}(\tau) = \mathrm{E}(X_\tau^2) = G^2 \cdot p_G + C^2 \cdot p_{-C} = C \cdot G.$$

2. Fall Unfaires Spiel, d.h.

$$p \neq \frac{1}{2}.$$

Mit der Wahl von

$$r = \frac{1-p}{p} \neq 1$$

wird durch $Z_n = r^{X_n}$ ein Martingal definiert, siehe Example 2.3, und die Voraussetzungen von Theorem 2 sind erfüllt. Man erhält

$$1 = \mathrm{E}(Z_\tau) = r^G \cdot p_G + r^{-C} \cdot p_{-C},$$

und zusammen mit (3) folgt

$$p_G = \frac{1 - r^{-C}}{r^G - r^{-C}}, \qquad p_{-C} = \frac{r^G - 1}{r^G - r^{-C}}.$$

Schließlich zeigt (4)

$$\mathrm{E}(\tau) = \frac{G \cdot p_G - C \cdot p_{-C}}{2p-1}.$$

Numerische Berechnungen zeigen: kleine Abweichungen von $p = 1/2$ führen zu drastischen Änderungen der Ruin-Wahrscheinlichkeit.

Abschließend betrachten wir im Fall $p > 1/2$ die einseitigen Gewinn- und Ruin-Wahrscheinlichkeiten $P(\bigcup_{n=1}^\infty \{X_n = G\})$ bzw. $P(\bigcup_{n=1}^\infty \{X_n = -C\})$. Es gilt

$$P\left(\bigcup_{n=1}^\infty \{X_n = -C\}\right) = \lim_{G\to\infty} p_{-C} = r^C,$$

und Satz IV.2.4 sichert

$$P\left(\bigcup_{n=1}^\infty \{X_n = G\}\right) = 1.$$

4 Branching Processes

...

5 Ausblick

...

Literature

H. Bauer, *Probability Theory*, de Gruyter, Berlin, 1996.

P. Billingsley, *Probability and Measure*, Wiley, New York, first edition 1979, third edition 1995.

Y. S. Chow, H. Teicher, *Probability Theory*, Springer, New York, first editon 1978, third edition 1997.

R. M. Dudley, *Real Analysis and Probability*, Cambridge University Press, Cambridge, 2002.

J. Elstrodt, *Maß- und Integrationstheorie*, Springer, Berlin, first edition 1996, fifth edition, 2007.

K. Floret, *Maß- und Integrationstheorie*, Teubner, Stuttgart, 1981.

P. Gänssler, W. Stute, *Wahrscheinlichkeitstheorie*, Springer, Berlin, 1977.

E. Hewitt, K. Stromberg, *Real and Abstract Analysis*, Springer, Berlin, 1965.

A. Klenke, *Wahrscheinlichkeitstheorie*, Springer, Berlin, first edition 2006, second edition 2008.

K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, New York, 1967.

A. N. Širjaev, *Wahrscheinlichkeit*, Deutscher Verlag der Wissenschaften, Berlin, 1988.
A. N. Shiryayev, *Probability*, Springer, New York, 1984.

J. Yeh, *Martingales and Stochastic Analysis*, World Scientific, Singapore, 1995.

Index