

Probability theory I

Prof. Dr. Alexander Drewitz
Universität zu Köln

Preliminary version of July 16, 2019

If you spot any typos or mistakes, please drop an email to drewitz@math.uni-koeln.de.

Contents

1	Set functions	5
1.1	Systems of sets	5
1.1.1	Semirings, rings, and algebras	5
1.1.2	σ -algebras and Dynkin systems	11
1.2	Set functions	17
1.2.1	Properties of set functions	17
1.3	Carathéodory's extension theorem ('Maßerweiterungssatz')	25
1.3.1	Lebesgue measure	29
1.3.2	Lebesgue-Stieltjes measure	31
1.4	Measurable functions, random variables	32
1.5	Image measures, distributions	38
2	The Lebesgue integral	41
2.0.1	Integrals of simple functions	41
2.0.2	Lebesgue integral for measurable functions	43
2.0.3	Lebesgue vs. Riemann integral	46
2.1	Convergence theorems	47
2.1.1	Dominated and monotone convergence	47
2.2	Measures with densities, absolute continuity	50
2.2.1	Almost sure / almost everywhere properties	50
2.2.2	Hahn-Jordan decomposition	53
2.2.3	Lebesgue's decomposition theorem, Radon-Nikodym derivative	55
2.2.4	Integration with respect to image measures	55
2.3	Product spaces	56
2.4	Product measures	58
2.5	The theorems of Fubini and Tonelli	60
2.6	Fourier transform / characteristic functions	63
3	Classical and basic results in probability theory	65
3.1	Specific distributions	65
3.1.1	Discrete distributions	65
3.1.2	Distributions with densities	68
3.2	Independence	69
3.3	Covariance, variance	72
3.4	L^p spaces and some fundamental inequalities	77
3.5	Convergence of random variables	80
3.5.1	Almost sure convergence	80
3.5.2	Convergence in \mathcal{L}^p	81
3.5.3	Convergence in probability	82
3.5.4	Convergence in distribution	82
3.5.5	Some fundamental tools	85

3.5.6	Interdependence of types of convergence of random variables	86
3.6	Laws of large numbers	87
3.6.1	Weak law of large numbers	87
3.6.2	Strong law of large numbers	89
3.7	Convolution of measures	90
3.8	Central limit theorem	91
4	A primer on stochastic processes	103
4.1	Stochastic processes	103
4.2	Kolmogorov's existence and uniqueness theorem	105

Chapter 1

Set functions

In the introductory course ‘Introduction to probability and statistics’ (see [Dre18]),¹ a frequent motivation was the investigation of (finitely many) dice tosses or coin flips. We have seen that those experiments were suitably described by discrete probability spaces, see [Dre18, Section 1.2]. Also in the nice setting of probability distributions with a density we have learned how to deal with that setting by means of the Riemann integral to some extent. However, quite quickly we had reached the limit of that approach, since for example pointwise limits of Riemann integrable functions were not necessarily integrable anymore, see e.g. Section [Dre18, Section 1.8.3].

As promised in the introductory class, in this course we will provide a rigorous and self-contained introduction to the concept of Lebesgue integration, which will in particular comprise and generalize most of the content of [Dre18] regarding random variables and expectations.

The principal goals of the first two chapters are twofold:

- Investigate measures and in particular construct the Lebesgue measure on \mathbb{R}^d using a general extension theorem for elementary notions of volume;
- Introduce the Lebesgue integral for suitable functions defined on arbitrary measure spaces;

Recommended references to accompany this Chapter are [Els05], [Kle14], [Kal02] and [G08], as well as [Bau92]. Further sources on measure theory are [Bil95], [Coh13], [Doo94], [Rao04], [Hal50].

1.1 Systems of sets

1.1.1 Semirings, rings, and algebras

One of the main goals will be to measure subsets of some a priori abstract set Ω . In general, it will not be possible to do so in an appropriate manner for all subsets of Ω (the so-called ‘Maßproblem’ and ‘Inhaltsproblem’ (see [Els05]) as well as the Banach-Tarski paradox show the kind of problems that can arise when trying to do so; due to time constraints, we will not go into details here).

Maßproblem: We want to construct a function $\mu : 2^{\mathbb{R}^d} \rightarrow [0, \infty]$ (supposed to measure subsets of \mathbb{R}^d) with the following properties:

- (a) For any sequence $(A_n)_{n \in \mathbb{N}}$ of pairwise disjoint subsets $A_n \subset \mathbb{R}^d$, we have

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} \mu(A_n) \quad (\sigma\text{-additivity});$$

¹This course is not a prerequisite, however it might help intuition to have attended that course.

(b) For any isometry ('Bewegung') T of \mathbb{R}^d and any $A \subset \mathbb{R}^d$ one has

$$\mu(A) = \mu(T(A)) \quad (\text{invariance under isometries});$$

(c)

$$\mu([0, 1]^d) = 1 \quad (\text{normalization});$$

Theorem 1.1.1 (Satz v. Vitali (1905), Italian mathematician (1875–1932)). *Das Maßproblem ist unlösbar für $d = 1$.*

Theorem 1.1.2 (Satz v. Banach-Tarski (1924), Polish mathematicians (Stefan Banach, 1892 – 1945, Alfred Tarski, 1901 – 1983)). *For $d \geq 1$, let $A, B \subset \mathbb{R}^d$ be arbitrary sets with non-empty interior. Then there exists a sequence (C_n) of subsets $C_n \in \mathbb{R}^d$ and isometries (Bewegungen) $(T_n)_{n \in \mathbb{N}}$ such that*

$$A = \dot{\bigcup}_{n \in \mathbb{N}} C_n \quad \text{and} \quad B = \dot{\bigcup}_{n \in \mathbb{N}} T_n(C_n).$$

Therefore, a key role will be played by certain subsets of the power set $2^{\mathbb{R}^d}$ of \mathbb{R}^d which are nicely behaved. In fact, it will turn out that with little additional effort we will be able to develop a theory of measures (and subsequently integration) not only on suitable subsets of \mathbb{R}^d , but also of more general spaces Ω which will prove useful in many occasions.

In the following we introduce several such systems of subsets which play an integral part in the construction of those functions (so-called *measures*, introduced in Definition 1.2.3 below) which will be measuring the corresponding subsets (which will form so-called σ -algebras, see Definition 1.1.18).

The first definition is just a shorthand for systems of subsets of Ω which are closed under finite intersection.

Definition 1.1.3. *Let Ω be a non-empty set and let $\mathcal{S} \subset 2^\Omega$. \mathcal{S} is called a π -system (' π -System') if it is closed under (finite) intersections:*

$$A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}. \quad (1.1.1)$$

Remark 1.1.4. *Using induction ('vollständige Induktion') it is not hard to show that if \mathcal{S} is a π -system, $n \in \mathbb{N}$, and $A_1, \dots, A_n \in \mathcal{S}$, then*

$$\bigcap_{i=1}^n A_i \in \mathcal{S}.$$

(see exercise classes).

To us, this property will prove particularly valuable in combination with further properties, as will be seen in the set systems introduced below.

A standard way to construct functions which are supposed to measure many subsets of a set Ω is to first specify how to measure certain 'simple' subsets of Ω . Such simple subsets oftentimes form a semiring as introduced in the following definition.

Definition 1.1.5. *Let Ω be a non-empty set. A subset \mathcal{S} of 2^Ω is called a semiring ('Halbring', 'Semiring') over Ω if the following properties are fulfilled:*

(a)

$$\emptyset \in \mathcal{S}; \quad (1.1.2)$$

(b) \mathcal{S} is a π -system;

(c) for any $A, B \in \mathcal{S}$, there exist pairwise disjoint $C_1, \dots, C_n \in \mathcal{S}$ such that

$$A \setminus B = \dot{\bigcup}_{i=1}^n C_i.$$

One of the examples most relevant to us will be the following.

Example 1.1.6. Denote by

$$\mathcal{I} := \{(a, b] : a, b \in \mathbb{R}, a \leq b\}$$

the set of left-open right-closed intervals in \mathbb{R} . Then \mathcal{I} is a semiring over \mathbb{R} .

Proof. See exercise classes. □

The set of Cartesian products of elements of two semirings is a semiring again, as is stated in the following result.

Lemma 1.1.7. Let \mathcal{S}_1 be a semiring over Ω_1 and \mathcal{S}_2 be a semiring over Ω_2 . Then

$$\mathcal{S}_1 * \mathcal{S}_2 := \{A_1 \times A_2 : A_1 \in \mathcal{S}_1, A_2 \in \mathcal{S}_2\}$$

is a semiring over $\Omega_1 \times \Omega_2$.

Proof. We have to establish the properties of Definition 1.1.5.

Since $\emptyset \in \mathcal{S}_1$ and $\emptyset \in \mathcal{S}_2$, we get $\emptyset \in \mathcal{S}_1 * \mathcal{S}_2$.

Furthermore, to prove the second item let $A = A_1 \times A_2, B = B_1 \times B_2 \in \mathcal{S}_1 * \mathcal{S}_2$. Thus,

$$A \cap B = \underbrace{(A_1 \cap B_1)}_{\in \mathcal{S}_1} \times \underbrace{(A_2 \cap B_2)}_{\in \mathcal{S}_2} \in \mathcal{S}_1 * \mathcal{S}_2,$$

since \mathcal{S}_1 and \mathcal{S}_2 are stable under intersections.

To prove the last item, again let $A = A_1 \times A_2, B = B_1 \times B_2 \in \mathcal{S}_1 * \mathcal{S}_2$. Then $A \setminus B$ can be partitioned via

$$A \setminus B = ((A_1 \setminus B_1) \times A_2) \dot{\bigcup} ((A_1 \cap B_1) \times (A_2 \setminus B_2)). \quad (1.1.3)$$

Now by assumption there exist $C_1, \dots, C_n \in \mathcal{S}_1$ pairwise disjoint with

$$A_1 \setminus B_1 = \dot{\bigcup}_{i=1}^n C_i$$

and $D_1, \dots, D_m \in \mathcal{S}_2$ pairwise disjoint with

$$A_2 \setminus B_2 = \dot{\bigcup}_{i=1}^m D_i.$$

As a result, the right-hand side of (1.1.3) can be written as a pairwise disjoint union of elements of $\mathcal{S}_1 * \mathcal{S}_2$, which finishes the proof. □

We will use the standard convention that for $a, b \in \mathbb{R}^d$ we write $a \leq b$ if $a_i \leq b_i$ for all $i \in \{1, \dots, d\}$, and similarly for other types of (in)equalities. In addition, for $a, b \in \mathbb{R}^d$ with $a \leq b$ we will use the notation $(a, b) := \times_{i=1}^d (a_i, b_i) = \{x \in \mathbb{R}^d : a < x < b\}$ for the Cartesian product of (one-dimensional) intervals, and analogously for other types of intervals.

Corollary 1.1.8. For any $d \in \mathbb{N}$, the set

$$\mathcal{I}^d := \{(a, b] : a, b \in \mathbb{R}^d, a \leq b\} \quad (1.1.4)$$

of hyperrectangles (‘Hyperquader’) is a semiring over \mathbb{R}^d .

Proof. We proceed by induction (‘vollständige Induktion’) over the dimension d . The case $d = 1$ is Example 1.1.6.

Assume the statement holds for arbitrary $d \in \mathbb{N}$. Then $\mathcal{I}^{d+1} = \mathcal{I}^d * \mathcal{I}$, and hence \mathcal{I}^{d+1} is a semiring due to Lemma 1.1.7 and the induction hypothesis. \square

Definition 1.1.9. Let Ω be a non-empty set. A subset \mathcal{R} of 2^Ω is called a ring over Ω if the following properties are fulfilled:

$$(a) \quad \emptyset \in \mathcal{R}; \quad (1.1.5)$$

$$(b) \quad A, B \in \mathcal{R} \text{ implies } A \cup B \in \mathcal{R}; \quad (1.1.6)$$

$$(c) \quad \text{for any } A, B \in \mathcal{R}, \text{ one has } A \setminus B \in \mathcal{R}. \quad (1.1.7)$$

Example 1.1.10. Let Ω be an arbitrary non-empty set. Then the set \mathcal{R} of countable (we use countable in the sense of ‘at most’ countable, i.e., a set is countable if it is finite or has a bijection with \mathbb{N}) subsets of Ω is a ring.

Proof. Since \emptyset is countable we have $\emptyset \in \mathcal{R}$. Furthermore, if $A, B \in \mathcal{R}$, then A and B are both countable and hence so is $A \setminus B$. In this case, also $A \cup B$ is countable, and hence \mathcal{R} is a ring. \square

Example 1.1.11. Let \mathcal{S} be a semiring. Then the set

$$\mathcal{R} := \left\{ \bigcup_{i=1}^n S_i : n \in \mathbb{N}, S_i \in \mathcal{S} \forall i \in \{1, \dots, n\} \right\}$$

is a ring. It is also referred to as the ring generated by \mathcal{S} , and it is the smallest ring containing \mathcal{S} .

Proof. Since $\emptyset \in \mathcal{S}$, we immediately get $\emptyset \in \mathcal{R}$.

For $A, B \in \mathcal{R}$ we have $A = \bigcup_{i=1}^m A_i$, $B = \bigcup_{i=1}^n B_i$, with $A_i, B_i \in \mathcal{S}$, and we get at once that $A \cup B \in \mathcal{R}$.

The ‘hard’ part is to show the last property. For that purpose, assume again that $A, B \in \mathcal{R}$ and hence that $A = \bigcup_{i=1}^m A_i$, $B = \bigcup_{i=1}^n B_i$, with $A_i, B_i \in \mathcal{S}$. Then

$$A \setminus B = \bigcup_{i=1}^m A_i \setminus \bigcup_{j=1}^n B_j = \bigcup_{i=1}^m \bigcap_{j=1}^n A_i \setminus B_j.$$

Since \mathcal{S} is a semiring, we deduce that

$$A_i \setminus B_j = \bigcup_{k=1}^{n_{i,j}} C_k^{i,j},$$

some $n_{i,j} \in \mathbb{N}$ and $C_k^{i,j} \in \mathcal{S}$ for all $k \in \{1, \dots, n_{i,j}\}$. Hence,

$$A \setminus B = \bigcup_{i=1}^m \bigcap_{j=1}^n \bigcup_{k=1}^{n_{i,j}} C_k^{i,j} = \bigcup_{i=1}^m \bigcup_{f \in \times_{j=1}^n \{1, \dots, n_{i,j}\}} \underbrace{\bigcap_{j=1}^n C_{f(j)}^{i,j}}_{\in \mathcal{S}},$$

which shows that \mathcal{R} is a ring.

Furthermore, since rings are stable under finite unions, any ring containing \mathcal{S} must also contain \mathcal{R} . Thus, \mathcal{R} is the smallest ring containing \mathcal{S} . \square

Lemma 1.1.12. *For a non-empty set Ω , a subset $\mathcal{R} \subset 2^\Omega$ is a ring if and only if one of the following two conditions holds true:*

(a)

$$\emptyset \in \mathcal{R},$$

for any $A, B \in \mathcal{R}$ one has $A \Delta B \in \mathcal{R}$, and

for any $A, B \in \mathcal{R}$ one has $A \cup B \in \mathcal{R}$;

(b)

$$\emptyset \in \mathcal{R},$$

for any $A, B \in \mathcal{R}$ one has $A \Delta B \in \mathcal{R}$, and

for any $A, B \in \mathcal{R}$ one has $A \cap B \in \mathcal{R}$;

Proof. Let \mathcal{R} be a ring over Ω . Since

$$A \Delta B = (A \setminus B) \cup (B \setminus A),$$

we deduce that (a) holds true.

Now assume that \mathcal{R} fulfills (a). Then, since

$$A \cap B = (A \cup B) \Delta (A \Delta B)$$

we deduce that (b) holds true.

Now assume (b) to hold true. Then since

$$A \setminus B = A \Delta (A \cap B),$$

(1.1.7) follows. Furthermore, since $A \cup B = A \Delta (B \setminus A)$, we also deduce (1.1.6). Hence, \mathcal{R} is a ring. \square

Corollary 1.1.13. *If \mathcal{R} is a ring over a non-empty set Ω , then \mathcal{R} is also a semiring over Ω .*

Proof. This is an immediate consequence of Lemma 1.1.12 and the very definition of a semiring and a ring. \square

An even stronger (cf. Lemma 1.1.16) concept than that of a ring is given in the following definition.

Definition 1.1.14. *Let Ω be a non-empty set. A subset \mathcal{A} of 2^Ω is called an algebra² over Ω ('Algebra über Ω ') if the following properties are fulfilled:*

(a)

$$\Omega \in \mathcal{A}; \tag{1.1.8}$$

(b)

$$A \in \mathcal{A} \text{ implies } A^c \in \mathcal{A}; \tag{1.1.9}$$

²Some authors use the term *field* instead of algebra, see e.g. [Bil95].

(c)

$$A, B \in \mathcal{A} \text{ implies } A \cup B \in \mathcal{A}. \quad (1.1.10)$$

Lemma 1.1.15. *Let Ω be a non-empty set. A subset $\mathcal{A} \subset 2^\Omega$ is an algebra over Ω if and only if the following conditions are fulfilled:*

(a)

$$\Omega \in \mathcal{A}; \quad (1.1.11)$$

(b)

$$A \in \mathcal{A} \implies A^c \in \mathcal{A}; \quad (1.1.12)$$

(c)

$$A, B \in \mathcal{A} \implies A \cap B \in \mathcal{A}. \quad (1.1.13)$$

Proof. Let \mathcal{A} be an algebra. It only remains to show (1.1.13). But De Morgan's laws imply $A \cap B = (A^c \cup B^c)^c$, hence if $A, B \in \mathcal{A}$, then (1.1.10) and (1.1.9) yield that $A \cap B \in \mathcal{A}$.

Conversely assume that the system \mathcal{A} fulfills properties (1.1.11) to (1.1.13). It remains to show (1.1.10). But again by De Morgan's laws we obtain $A \cup B = (A^c \cap B^c)^c$, and hence (1.1.12) and (1.1.13) imply that $A \cup B \in \mathcal{A}$. Hence, \mathcal{A} is an algebra over Ω . \square

Lemma 1.1.16. *Let Ω be a non-empty set. A ring \mathcal{R} over Ω is an algebra over Ω if and only if $\Omega \in \mathcal{R}$.*

Proof. Let \mathcal{R} be a ring over Ω with $\Omega \in \mathcal{R}$. Then (1.1.8) immediately holds true. Choosing $A := \Omega \in \mathcal{R}$ in (1.1.7), we obtain that for arbitrary $B \in \mathcal{R}$ we have $B^c = \Omega \setminus B = \Omega \setminus B \in \mathcal{R}$, which implies that (1.1.9) holds true as well. In addition, (1.1.10) immediately follows from (1.1.6). Hence, \mathcal{R} is an algebra.

Conversely, if \mathcal{A} is an algebra, then $\emptyset = \Omega^c \in \mathcal{A}$ due to (1.1.9) and (1.1.8). Thus, (1.1.5) holds true. Furthermore, for $A, B \in \mathcal{A}$ we have $A \setminus B = A \cap B^c$, hence (1.1.7) is a consequence of (1.1.13) and (1.1.9). Lastly, (1.1.6) is implied by (1.1.10). \square

As an application, we have seen in Corollary 1.1.8 that \mathcal{I}^d is a semiring over \mathbb{R}^d , and we can generate a ring from it as in Example 1.1.11. Still \mathbb{R}^d is not contained in this ring, so due to the previous lemma this is still not an algebra.

We now go for a short detour to explain the motivation of the terms ‘ring’ and ‘algebra’ in the above context. For this purpose we recall that in (linear) algebra, a ring (‘Ring’) had been defined as a set R endowed with operations $+$: $R \times R \rightarrow R$ (addition) and \cdot : $R \times R \rightarrow R$ (multiplication) such that $(R, +)$ is an additive commutative group, and such that

(a)

$$\forall a, b, c \in R : \quad a \cdot (b \cdot c) = (a \cdot b) \cdot c \quad (\text{associativity})$$

(b)

$$\forall a, b, c \in R : \quad a \cdot (b + c) = a \cdot b + a \cdot c, \quad (a + b) \cdot c = a \cdot c + b \cdot c \quad (\text{distributivity})$$

Furthermore, in linear algebra, one way to introduce the concept of an ‘algebra’ is to demand it to be a ring R which in addition is a vector space over a field³ (‘Körper’) K such that in addition, for all $u, v \in R$ and $\alpha \in K$ one has

$$\alpha(uv) = (\alpha u)v = u(\alpha v).$$

³Adding insult to injury, there is again a possible overlap with terms here, since as we have seen before, some (although seemingly not too many) authors use the term ‘field’ for an algebra (in the set-theoretic sense). It seems that for notations in German there is slightly less confusion.

Lemma 1.1.17. *Let Ω be a non-empty set.*

- (a) *Endow the power set 2^Ω with Δ (symmetric difference) as addition and \cap (intersection) as multiplication. Then $(2^\Omega, \Delta, \cap)$ is a commutative ring with a zero ('Nullelement') \emptyset and a one ('Einselement') Ω .*
- (b) *A subset $\mathcal{R} \subset 2^\Omega$ is a ring in the sense of Definition 1.1.9 if and only if $(\mathcal{R}, \Delta, \cap)$ is a ring.*
- (c) *Let $\mathcal{A} \subset 2^\Omega$ be a ring or an algebra in the sense of Definition 1.1.14 or 1.1.9, respectively. Then $(\mathcal{A}, \Delta, \cap)$ is an algebra over the field $\{0, 1\}$.⁴ Here we define $0 \cdot A := \emptyset$ and $1 \cdot A := A$ for $A \in \mathcal{A}$.*

Proof. See exercise classes. □

1.1.2 σ -algebras and Dynkin systems

The systems of sets introduced above, i.e., semirings, rings, and algebras, only involved stability under finite operations (such as intersections and unions, for example). It turns out, however, that we want to be able to measure not only finite but countable intersections (or unions) of 'nice' sets.⁵ As a consequence, we will introduce systems of sets that are stable under such kinds of operations.

Definition 1.1.18. *Let Ω be a non-empty set. A subset \mathcal{F} of 2^Ω is called a σ -algebra over Ω ⁶ (' σ -Algebra über Ω ') if the following properties are fulfilled:*

$$(a) \quad \Omega \in \mathcal{F}; \tag{1.1.14}$$

$$(b) \quad A \in \mathcal{F} \text{ implies } A^c \in \mathcal{F}; \tag{1.1.15}$$

$$(c) \quad A_1, A_2, \dots \in \mathcal{F} \text{ implies } \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}. \tag{1.1.16}$$

Exercise 1.1.19. *If \mathcal{F} is a σ -algebra over Ω and $F \subset \Omega$, then*

$$\mathcal{F}_F := \mathcal{F} \cap F := \{F \cap G : G \in \mathcal{F}\}$$

is a σ -algebra over F (it is called the trace σ -algebra of F in \mathcal{F}).

Exercise 1.1.20. *Let Ω be a non-empty set.*

- (a) *Any σ -algebra \mathcal{F} also is an algebra, i.e., it is stable under finite unions.*
- (b) *$\mathcal{F} \subset 2^\Omega$ is a σ -algebra if and only if (1.1.14), (1.1.15), and*

$$A_1, A_2, \dots \in \mathcal{F} \text{ implies } \bigcap_{n \in \mathbb{N}} A_n \in \mathcal{F}. \tag{1.1.17}$$

⁴Where 0 is the neutral element of addition, and 1 the neutral element of multiplication. In particular, $0 + 1 = 1$, and $1 + 1 = 0$.

⁵In 'real life experiments' you might e.g. want to ask whether certain properties are fulfilled by an infinite sequence of coin tosses or dice rolls.

⁶Again, some authors use the term σ -field instead, see [Bil95].

While σ -algebras play an important role not only in probability theory but also e.g. analysis, the following concept of a Dynkin system has primarily been employed in probability theory.

Definition 1.1.21. Let Ω be a non-empty set. A subset \mathcal{D} of 2^Ω is called a Dynkin ('Dynkin-System')⁷ system (or also λ -system (' λ -System')) over Ω if the following properties are fulfilled:

$$(a) \quad \Omega \in \mathcal{D}; \quad (1.1.18)$$

$$(b) \quad A \in \mathcal{D} \text{ implies } A^c \in \mathcal{D}; \quad (1.1.19)$$

$$(c) \quad \text{If } A_1, A_2, \dots \in \mathcal{D} \text{ is a sequence of pairwise disjoint sets, then } \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}. \quad (1.1.20)$$

Exercise 1.1.22. Find an example of a Dynkin system that is not a semiring.

Lemma 1.1.23. Property (1.1.19) in Definition 1.1.21 can be substituted by the following: For any $A, B \in \mathcal{D}$ with $A \subset B$ one has $B \setminus A \in \mathcal{D}$.

Proof. Indeed, if this property holds, then for any $A \in \mathcal{D}$ we obtain, setting $B := \Omega \in \mathcal{F}$, that

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

Conversely, if \mathcal{D} is a Dynkin system and $A, B \in \mathcal{D}$ with $A \subset B$, then we have

$$B \setminus A = (A \dot{\cup} B^c)^c,$$

and the latter is contained in \mathcal{D} since Dynkin systems are closed under disjoint unions and complements. \square

We will investigate the relations between λ -systems and σ -algebras in more detail in Theorems 1.1.33 and 1.1.32 below.

Proposition 1.1.24. Let Λ be an arbitrary non-empty set, and let $(\mathcal{A}_\lambda)_{\lambda \in \Lambda}$ be a family of σ -algebras (or rings, or algebras, or λ -systems) over the same set Ω . Then

$$\bigcap_{\lambda \in \Lambda} \mathcal{A}_\lambda$$

is a σ -algebra (or ring, or algebra, or λ -system) over Ω again.

Proof. We only give the proof for σ -algebras, the remaining cases are proven in a similar way. Since $\Omega \in \mathcal{A}_\lambda$ for all $\lambda \in \Lambda$, we immediately get

$$\Omega \in \bigcap_{\lambda \in \Lambda} \mathcal{A}_\lambda.$$

Furthermore, assume $A \in \bigcap_{\lambda \in \Lambda} \mathcal{A}_\lambda$. Then $A \in \mathcal{A}_\lambda$ for all $\lambda \in \Lambda$, therefore $A^c \in \mathcal{A}_\lambda$ for all $\lambda \in \Lambda$, and hence

$$A^c \in \bigcap_{\lambda \in \Lambda} \mathcal{A}_\lambda.$$

Lastly, assume that $(A_n)_{n \in \mathbb{N}}$ is a sequence of sets such that $A_n \in \bigcap_{\lambda \in \Lambda} \mathcal{A}_\lambda$ for all $n \in \mathbb{N}$. Thus, for each $\lambda \in \Lambda$, we have $A_n \in \mathcal{A}_\lambda$ for all $n \in \mathbb{N}$, and hence $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}_\lambda$. As a consequence,

$$\bigcup_{n \in \mathbb{N}} A_n \in \bigcap_{\lambda \in \Lambda} \mathcal{A}_\lambda$$

for each $\lambda \in \Lambda$, too, which finishes the proof. \square

⁷In honor of Eugene Dynkin (1924–2014)

Exercise 1.1.25. Show that Proposition 1.1.24 does not hold true anymore if one replaces ‘ σ -algebra’ by ‘semiring’.

The following is a generalization of the very specific Example 1.1.11.

Definition 1.1.26. Let Ω be a non-empty set, and let $\mathcal{E} \subset 2^\Omega$. Proposition 1.1.24 implies that

$$\delta(\mathcal{E}) := \bigcap_{\substack{\mathcal{D} \text{ is a } \lambda\text{-system over } \Omega \\ \mathcal{D} \supset \mathcal{E}}} \mathcal{D}$$

is a λ -system again. It is called the λ -system generated by \mathcal{E} .

Similarly,

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathcal{F} \text{ is a } \sigma\text{-algebra over } \Omega \\ \mathcal{F} \supset \mathcal{E}}} \mathcal{F}$$

is a σ -algebra again. It is called the σ -algebra generated by \mathcal{E} .

In an analogous way, one can define rings and algebras generated by subsets of 2^Ω .

Note that it follows the previous definition that the σ -algebra $\sigma(\mathcal{E})$ is at the same time the smallest σ -algebra containing \mathcal{E} (and similarly for the remaining set systems).

Remark 1.1.27. For simpler set systems such as algebras and rings it is possible to explicitly represent their elements using by applying finitely many combinations of elementary set functions to elements of their generators (see Example 1.1.11, problem 1 on the first homework sheet, or also [Els05, Ch.1, §4]). For σ -algebras this is generally not possible anymore, see the Section ‘Constructing σ -fields’ in [Bil95] pp. 30. However, for many purposes it is actually sufficient to consider the generator of σ -algebras or Dynkin systems instead of the entire set system.

An prominent role will be played by the Borel- σ -algebra (French mathematician and politician Émile Borel (1871–1956)).

Definition 1.1.28. Let Ω be a non-empty set. A set $\tau \subset 2^\Omega$ is called a topology (‘Topologie’) if the following hold true:

(a)

$$\emptyset, \Omega \in \tau;$$

(b) if $(O_\lambda)_{\lambda \in \Lambda}$ is an arbitrary family of sets $O_\lambda \in \tau$, then

$$\bigcup_{\lambda \in \Lambda} O_\lambda \in \tau;$$

(c) if $O_1, O_2 \in \tau$, then

$$O_1 \cap O_2 \in \tau.$$

The pair (Ω, τ) is called a topological space. The elements of τ are called open sets, and the elements of

$$\{C \subset \Omega : C^c \in \tau\}$$

are called closed sets.

This definition can be motivated by having a closer look at metric spaces. (E.g., consider \mathbb{R}^d endowed with the Euclidean metric defined via $d(x, y) := \sqrt{\sum_{i=1}^d (x_i - y_i)^2}$, for $x, y \in \mathbb{R}^d$.) In fact, for a metric space (X, d) a subset $O \subset X$ had been defined to be *open* if for each $x \in O$ there exists $r \in (0, \infty)$ such that

$$B_r(x) := \{y \in X : d(x, y) < r\} \subset O.$$

It can be shown without too much effort that defining τ as the set of all open sets of the metric space (X, d) , it fulfills the properties of a topology as defined in Definition 1.1.28.

We will mostly be dealing with metric spaces (and, in fact, most of the times with \mathbb{R}^d endowed with the Euclidean metric), but the following definition does not come at any additional cost in its fully-fledged generality.

In the following, instead of τ for topology we write \mathcal{O} to denote the topology (inspired by the wording ‘open’ sets).

Definition 1.1.29. For an arbitrary topological space (X, \mathcal{O}) , the σ -algebra

$$\mathcal{B}(X) := \sigma(\mathcal{O}) = \bigcap_{\substack{\mathcal{F} \text{ is a } \sigma\text{-algebra over } \Omega \\ \mathcal{F} \supset \mathcal{O}}} \mathcal{F}$$

generated by the open subsets of X is called the Borel- σ -algebra (‘Borel- σ -Algebra’) on Ω . The sets $B \in \mathcal{B}(X)$ are called Borel sets or Borel-measurable sets.

Example 1.1.30. (a) Most often we will be interested in the case $X = \mathbb{R}^d$, $d \in \mathbb{N}$, and \mathcal{O} the set of open subsets of \mathbb{R}^d (in the topology induced by the Euclidean metric), i.e., in the σ -algebra $\mathcal{B}(\mathbb{R}^d)$.

(b) It will also turn out useful to consider the Borel- σ -algebra $\mathcal{B}(\overline{\mathbb{R}})$ of the two point compactification $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ of \mathbb{R} . The open sets of $\overline{\mathbb{R}}$ can be described as the largest topology (i.e., a system of open sets) on $\overline{\mathbb{R}}$ such that the map

$$\begin{aligned} \varphi : [-1, 1] &\rightarrow \overline{\mathbb{R}} \\ x &\mapsto \begin{cases} \pm\infty, & \text{if } x = \pm 1, \\ \tan(\pi x/2), & \text{otherwise,} \end{cases} \end{aligned}$$

is continuous, where we recall that the map φ is continuous if and only if all preimages $\varphi^{-1}(O)$, $O \subset \overline{\mathbb{R}}$ open, is open in $[-1, 1]$. Here, $[-1, 1]$ is endowed with the usual (trace-) topology of \mathbb{R} , i.e., the open subsets of $[-1, 1]$ are just the ones of the form $[-1, 1] \cap O$, for any $O \subset \mathbb{R}$ open in the topology induced by the Euclidean metric.

In other words, the open sets of $\overline{\mathbb{R}}$ are given by (unions of) the sets of the form

- $V \subset \mathbb{R}$ open;
- $[-\infty, a)$, $a \in \mathbb{R}$;
- $(a, \infty]$, $a \in \mathbb{R}$;

In particular, we see that $\{\infty\}$ and $\{-\infty\}$ are not open in this topology of $\overline{\mathbb{R}}$.

(c) in fact, it is not easy to construct sets in \mathbb{R}^d which are not in $\mathcal{B}(\mathbb{R}^d)$. An example of such sets are the so-called Vitali sets; however, apart from possibly peculiar counterexamples, all subsets of \mathbb{R}^d that we will encounter in this course are actually contained in $\mathcal{B}(\mathbb{R}^d)$.

Lemma 1.1.31. Each of the following subsets of $2^{\mathbb{R}^d}$ is a generator of the Borel- σ -algebra $\mathcal{B}(\mathbb{R}^d)$, and each of them is a π -system:

- (a) $\mathcal{E}_1 := \{(a, b) : a, b \in \mathbb{Q}^d \text{ with } a < b\} \cup \{\emptyset\}$;
- (b) $\mathcal{E}_2 := \{(a, b] : a, b \in \mathbb{Q}^d \text{ with } a < b\} \cup \{\emptyset\}$;
- (c) $\mathcal{E}_3 := \{[a, b) : a, b \in \mathbb{Q}^d \text{ with } a < b\} \cup \{\emptyset\}$;
- (d) $\mathcal{E}_4 := \{[a, b] : a, b \in \mathbb{Q}^d \text{ with } a \leq b\} \cup \{\emptyset\}$;
- (e) $\mathcal{E}_5 := \{(-\infty, a] : a \in \mathbb{Q}^d\}$;

- (f) $\mathcal{E}_6 := \{(-\infty, a) : a \in \mathbb{Q}^d\};$
- (g) $\mathcal{E}_7 := \{[a, \infty) : a \in \mathbb{Q}^d\};$
- (h) $\mathcal{E}_8 := \{(a, \infty) : a \in \mathbb{Q}^d\};$
- (i) $\mathcal{E}_9 := \{A \subset \mathbb{R}^d : A \text{ is compact}\};$
- (j) $\mathcal{E}_{10} := \{A \subset \mathbb{R}^d : A \text{ is closed}\}.$

We can replace any \mathbb{Q} in the above by \mathbb{R} and the statements still hold true.

In addition, in the case $d = 1$, and considered as subsets of $2^{\mathbb{R}}$, any of the above systems of sets is a generator of $\mathcal{B}(\mathbb{R})$ also (where in the cases (e) to (h) we have to include ∞ and $-\infty$, respectively, i.e., exchange the corresponding '(' or ')' by '[' or ']', respectively).

Proof. We prove only part of the result (the remaining cases can be proved in similar ways). Any open subset of \mathbb{R}^d can be written as a countable union of elements of \mathcal{E}_1 (or of \mathcal{E}_2 , \mathcal{E}_3 , or \mathcal{E}_4 for that matter). Hence, all open subsets of \mathbb{R}^n are contained in any σ -algebra containing any of the \mathcal{E}_j , $1 \leq j \leq 4$, and thus we deduce

$$\mathcal{B}(\mathbb{R}^n) \subset \sigma(\mathcal{E}_j), \quad \text{for all } 1 \leq j \leq 4. \quad (1.1.21)$$

Conversely, since any element of \mathcal{E}_1 is open, we immediately get $\sigma(\mathcal{E}_1) \subset \mathcal{B}(\mathbb{R}^n)$, which proves the desired equality $\sigma(\mathcal{E}_1) = \mathcal{B}(\mathbb{R}^n)$.

Furthermore, the complement of any element of \mathcal{E}_4 is open, $\mathcal{E}_4 \subset \mathcal{B}(\mathbb{R}^d)$, and hence $\sigma(\mathcal{E}_4) \subset \mathcal{B}(\mathbb{R}^d)$; thus, in combination with (1.1.21) we have

$$\sigma(\mathcal{E}_4) = \mathcal{B}(\mathbb{R}^d). \quad (1.1.22)$$

To continue, any element of \mathcal{E}_2 (or of \mathcal{E}_3) can be written as a countable union of elements of \mathcal{E}_4 ; indeed, we have for $a, b \in \mathbb{Q}^d$ with $a < b$ that

$$(a, b] = \bigcup_{\substack{r \in \mathbb{Q}^d \\ a < r < b}} [r, b],$$

and similarly in the case of \mathcal{E}_3 . As a consequence, $\mathcal{E}_2, \mathcal{E}_3 \subset \mathcal{E}_4$ and thus we obtain $\sigma(\mathcal{E}_2), \sigma(\mathcal{E}_3) \subset \sigma(\mathcal{E}_4)$, which in combination with (1.1.21) and (1.1.22) supplies us with

$$\sigma(\mathcal{E}_2) = \sigma(\mathcal{E}_3) = \mathcal{B}(\mathbb{R}^d).$$

Furthermore, $\mathcal{E}_9, \mathcal{E}_{10} \subset \mathcal{B}(\mathbb{R}^d)$, hence

$$\sigma(\mathcal{E}_9), \sigma(\mathcal{E}_{10}) \subset \mathcal{B}(\mathbb{R}^d). \quad (1.1.23)$$

Conversely all open subsets of \mathbb{R}^d are contained in $\sigma(\mathcal{E}_{10})$, hence

$$\sigma(\mathcal{E}_{10}) = \mathcal{B}(\mathbb{R}^d). \quad (1.1.24)$$

In addition, $\mathcal{E}_4 \subset \mathcal{E}_9$ and hence in combination with (1.1.23) and (1.1.22) we deduce $\sigma(\mathcal{E}_9) = \mathcal{B}(\mathbb{R}^d)$.

For the last point, alternatively we could argue by exhaustion as follows: For $C \in \mathcal{E}_{10}$ we have

$$C = \bigcup_{n \in \mathbb{N}} \underbrace{[n, n]^d}_{\text{compact}} \cap C,$$

and in combination with (1.1.24) we infer $\sigma(\mathcal{E}_9) = \mathcal{B}(\mathbb{R}^d)$.

The remaining parts are left as an exercise.

□

Theorem 1.1.32. *Let \mathcal{D} be a λ -system over a non-empty set Ω . Then \mathcal{D} is a π -system if and only if it is a σ -algebra.*

Proof. Let \mathcal{D} be a λ -system.

Any σ -algebra is obviously a π -system (see Exercise 1.1.20 (b), where all but finitely many A_n can be chosen to be \emptyset).

Conversely, assume that \mathcal{D} is a π -system. Obviously, (1.1.19) and (1.1.18) imply that (1.1.15) and (1.1.14) hold true, so it remains to show that (1.1.16) is fulfilled as well. For this purpose, we start with observing that \mathcal{D} is \setminus -closed. In fact, if $A, B \in \mathcal{D}$, then

$$A \setminus B = A \cap B^c \in \mathcal{D}.$$

Now let a sequence $(A_n)_{n \in \mathbb{N}}$ of sets with $A_n \in \mathcal{D}$ for all $n \in \mathbb{N}$ be given. Define $B_1 := A_1$, and for $n \geq 2$ set

$$B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i = A_n \setminus \bigcup_{i=1}^{n-1} B_i.$$

By definition the B_n form a sequence of pairwise disjoint sets, and inductively we obtain that $B_n \in \mathcal{D}$ for all $n \in \mathbb{N}$. As a consequence,

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n \in \mathcal{D}.$$

This implies (1.1.16). □

The following result is fundamental in probability theory, and in particular it will be helpful in order to show Theorem 1.2.17 below.

Theorem 1.1.33 (Dynkin's π - λ -Theorem). *Let Ω be a non-empty set and let $\mathcal{A} \subset 2^\Omega$ be a π -system. Then*

$$\delta(\mathcal{A}) = \sigma(\mathcal{A}). \quad (1.1.25)$$

Proof. Since any σ -algebra is a Dynkin system also, it is clear that ' \subset ' holds in (1.1.25). For the converse inclusion, we observe that it is sufficient to show that $\delta(\mathcal{A})$ is a π -system, which due to Theorem 1.1.32 would imply that $\delta(\mathcal{A})$ is a σ -algebra (containing \mathcal{A}), so ' \supset ' would follow. For this purpose we define for arbitrary $A \in \delta(\mathcal{A})$ the set

$$\delta_A(\mathcal{A}) := \{B \in \delta(\mathcal{A}) : A \cap B \in \delta(\mathcal{A})\}.$$

We claim that for any $A \in \delta(\mathcal{A})$, the set system $\delta_A(\mathcal{A})$ is a Dynkin system. Indeed, we have the following:

- $\Omega \in \delta_A(\mathcal{A})$, since $A \cap \Omega = A \in \delta(\mathcal{A})$ by choice of A ;
- if $B \in \delta_A(\mathcal{A})$, then also $B \cap A \in \delta(\mathcal{A})$; hence, Lemma 1.1.23 supplies us with $A \cap B^c = A \setminus (B \cap A) \in \delta(\mathcal{A})$, and thus $B^c \in \delta_A(\mathcal{A})$ also;
- if (B_n) is a sequence of pairwise disjoint sets such that $B_n \in \delta_A(\mathcal{A})$ for all $n \in \mathbb{N}$, then we have $B_n \cap A \in \delta(\mathcal{A})$ for all $n \in \mathbb{N}$, and furthermore those sets are pairwise disjoint. As a consequence,

$$\left(\bigcup_{n=1}^{\infty} B_n \right) \cap A = \bigcup_{n=1}^{\infty} (B_n \cap A) \in \delta(\mathcal{A}),$$

$$\text{i.e., } \bigcup_{n=1}^{\infty} B_n \in \delta_A(\mathcal{A}).$$

Hence, $\delta_A(\mathcal{A})$ forms a Dynkin system.

Next, we observe that since \mathcal{A} is a π -system, we have $\mathcal{A} \subset \delta_A(\mathcal{A})$ for all $A \in \mathcal{A}$. Since $\delta_A(\mathcal{A})$ is a Dynkin system for $A \in \mathcal{A}$ due to the above, we deduce that $\delta(\mathcal{A}) \subset \delta_A(\mathcal{A})$ for any $A \in \mathcal{A}$. This, however, implies that for all $B \in \delta(\mathcal{A})$ and $A \in \mathcal{A}$ we have $A \cap B \in \delta(\mathcal{A})$, hence $\mathcal{A} \subset \delta_A(\mathcal{A})$ for all $A \in \delta(\mathcal{A})$ now, and thus $\delta(\mathcal{A}) \subset \delta_A(\mathcal{A})$ for all $A \in \delta(\mathcal{A})$. In particular, this implies that $\delta(\mathcal{A})$ is a π -system. In combination with Theorem 1.1.32 this establishes the fact that $\delta(\mathcal{A})$ is a σ -algebra, hence $\delta(\mathcal{A}) \supset \sigma(\mathcal{A})$, which finishes the proof. \square

Corollary 1.1.34. *In the notation of Lemma 1.1.31 we have for each $j \in \{1, 2, \dots, 10\}$ that*

$$\mathcal{B}(\mathbb{R}^d) = \delta(\mathcal{E}_j).$$

Proof. This is a direct consequence of Lemma 1.1.31 in combination with Theorem 1.1.33. \square

1.2 Set functions

1.2.1 Properties of set functions

Definition 1.2.1. *For Ω a non-empty set and $\mathcal{E} \subset 2^\Omega$, a function $\mu : \mathcal{E} \rightarrow \overline{\mathbb{R}}$ is called a set function.*

The set function μ is called

- monotone if $\mu(A) \leq \mu(B)$ for all $A, B \in \mathcal{E}$ with $A \subset B$;
- additive if

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu(A_i)$$

for all pairwise disjoint $A_1, \dots, A_n \in \mathcal{E}$ with $\bigcup_{i=1}^n A_i \in \mathcal{E}$;

- subadditive if

$$\mu(A) \leq \sum_{i=1}^n \mu(A_i)$$

for all $A, A_1, \dots, A_n \in \mathcal{E}$ with $A \subset \bigcup_{i=1}^n A_i$;

- σ -additive if

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{1.2.1}$$

for any sequence of pairwise disjoint sets $A_1, A_2, \dots \in \mathcal{E}$ with $\bigcup_{i=1}^{\infty} A_i \in \mathcal{E}$ such that the right-hand side of the previous equation is well-defined;

- σ -subadditive if

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

for any $A, A_1, A_2, \dots \in \mathcal{E}$ with $A \subset \bigcup_{i=1}^{\infty} A_i$ such that the right-hand side of the previous equation is well-defined;

Remark 1.2.2. *In the context of Definition 1.2.1, as for ‘usual’ functions, we will commonly write $\mu \leq \nu$ for two set functions on \mathcal{E} if for all $A \in \mathcal{E}$ we have*

$$\mu(A) \leq \nu(A).$$

Apart from a short excursion (see Definition 1.2.11 below which will prove useful later on) we will only be interested in non-negative set functions. Some types of set functions will occur frequently, hence we introduce the following terminology.

Definition 1.2.3. Let \mathcal{S} be a semiring over Ω , and let $\mu : \mathcal{S} \rightarrow [0, \infty]$ be a set function with $\mu(\emptyset) = 0$. Then μ is called a

- (a) content ('Inhalt') if μ is additive;
- (b) pre-measure ('Prämaß') if μ is σ -additive;
- (c) measure ('Maß') if μ is a pre-measure and \mathcal{S} is a σ -algebra;
- (d) probability measure ('Wahrscheinlichkeitsmaß') if μ is a measure with $\mu(\Omega) = 1$.

Remark 1.2.4. Show that the items in Definition 1.2.3 become more and more restrictive. I.e., if (d) is satisfied, then (c) is also satisfied; if (c) is satisfied, then (b) is also satisfied, and so on. On the other hand, one can find examples of a content which is no pre-measure, of pre-measures which are no measures, and so on (exercise).

Definition 1.2.5. Let \mathcal{S} be a semiring over Ω and let $\mu : \mathcal{S} \rightarrow [0, \infty]$ be a content.

- (a) Then μ is called finite if $\mu(S) < \infty$ for all $S \in \mathcal{S}$. Furthermore, μ is called σ -finite if there exists a sequence (S_n) of sets $S_n \in \mathcal{S}$ with

•

$$\mu(S_n) < \infty \quad \forall n \in \mathbb{N},$$

and

•

$$\bigcup_{n=1}^{\infty} S_n = \Omega.$$

- (b) A set $N \in \mathcal{S}$ is called a (μ) -null set ('Nullmenge') if $\mu(N) = 0$.⁸

Definition 1.2.6. Let $\Omega \neq \emptyset$, \mathcal{F} a σ -algebra over Ω , and $\mu : \mathcal{F} \rightarrow [0, \infty]$ a measure, Then the triplet $(\Omega, \mathcal{F}, \mu)$ is called a measure space.

A measure space $(\Omega, \mathcal{F}, \mu)$ is called a σ -finite measure space if μ is σ -finite.

If μ is a probability measure, then a measure space $(\Omega, \mathcal{F}, \mu)$ is called a probability space.

Example 1.2.7 (Contents, pre-measures). (a) On the semiring \mathcal{I} of left-open right-closed intervals introduced in Example 1.1.6 we can define a σ -finite content as follows. Let $f : \mathbb{R} \rightarrow [0, \infty)$ be continuous and set

$$F(x) := \int_0^x f(r) \, dr, \quad x \in \mathbb{R},$$

and where the integral is to be interpreted in the Riemann sense. Then F is non-decreasing and hence

$$\mu(I) := F(b) - F(a),$$

for $I = (a, b] \in \mathcal{I}$ defines a content on \mathcal{I} . Indeed, we have $\mu(\emptyset) = F(a) - F(a)$ (for any $a \in \mathbb{R}$), and for $(a_i, b_i] \in \mathcal{I}$, $1 \leq i \leq n$, with

$$\bigcup_{i=1}^n (a_i, b_i] \in \mathcal{I},$$

⁸Some authors consider any $B \subset \Omega$ for which we find $N \in \mathcal{S}$ with $B \subset N$ and $\mu(N) = 0$ a null set.

we can, after possibly permuting the indices, write

$$(a, b] = \bigcup_{i=1}^n (a_i, b_i]$$

with $a = a_1 \leq b_1 = a_2 \leq b_2 = a_3 \leq \dots b_{n-1} = a_n \leq b_n = b$. Hence, since

$$\mu((a, b]) = F(b) - F(a) = \sum_{i=1}^n F(b_i) - F(a_i) = \sum_{i=1}^n \mu((a_i, b_i]),$$

we deduce that μ is additive and hence a content indeed.

For $S_n := (-n, n] \in \mathcal{S}$ we obtain that $\bigcup_{n \in \mathbb{N}} S_n = \mathbb{R}$ and $\mu(S_n) = F(n) - F(-n) < \infty$, hence μ is σ -finite.

Using problem 3 of Homework 3 (see also Example 1.1.11) we can uniquely extend μ to a content on not only the semiring \mathcal{S} , but also on the ring \mathcal{R} generated by \mathcal{S} .

- (b) For our purposes, the most important content / pre-measure arguably is the d -dimensional Lebesgue content / pre-measure λ^d . It is defined on the set of d -dimensional hyper-cuboids \mathcal{I}^d introduced in (1.1.4) as follows. For $(a, b] \in \mathcal{I}^d$ we have $a, b \in \mathbb{R}^d$ with $a \leq b$, and its d -dimensional Lebesgue pre-measure is defined as

$$\begin{aligned} \lambda^d : \mathcal{I}^d &\rightarrow [0, \infty), \\ (a, b] &\mapsto \prod_{i=1}^d (b_i - a_i). \end{aligned} \tag{1.2.2}$$

It is not too hard to verify that λ^d defines a content on \mathcal{I}^d ; indeed, for the case $d = 1$ this is a consequence of Part (a) of this example, where we choose $f \equiv 1$. For the case $d \geq 2$ we refer to the proof of Proposition 1.2.8 (alternatively, see [Bau92, Satz 4.3]).

and again choosing the sets $S_n := (-n, n]^d \in \mathcal{I}^d$ we get that $\mathbb{R}^d = \bigcup_{n \in \mathbb{N}} S_n$ and $\mu(S_n) = (2n)^d$, thus λ is σ -finite.

In fact, λ^d even defines a pre-measure:

Proposition 1.2.8. The function λ^d defined in (1.2.2) defines a pre-measure on \mathcal{I}^d .

Proof. The proof is slightly more technical than what we did in Part (a) of this example for the case $d = 1$; however, the key ideas are present here already. We refer to [Els05, Satz II.3.1] for a proof (in this source, several strategies for proving the result are given – the one following [Els05, Satz II.3.8 b)] is closest to Part (a) of this example). \square

Example 1.2.9 (Measures). (a) Let Ω be an arbitrary set and define the counting measure (‘Zählmaß’)

$$\begin{aligned} \mu : 2^\Omega &\rightarrow [0, \infty] \\ A &\mapsto |A|, \end{aligned}$$

where for $A \subset \Omega$ we denote by $|A|$ the number of elements of A if A is finite, and ∞ otherwise.

Exercise 1.2.10. Show that μ indeed defines a measure on 2^Ω . Is it σ -finite?

- (b) Let X be uncountable and define \mathcal{F} to be the σ -algebra over X that contains all sets $A \subset X$ for which either A or A^c is countable. (Check that this defines a σ -algebra indeed!) For $A \in \mathcal{F}$ define the measure μ via

$$\mu(A) := \begin{cases} 1, & \text{if } A \text{ uncountable,} \\ 0, & \text{if } A \text{ countable.} \end{cases} \quad (1.2.3)$$

Check that μ defines a measure on (X, \mathcal{F}) . We have $\mu(\emptyset) = 0$, and for a sequence (A_n) of pairwise disjoint sets with $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$, we have that $\bigcup_{n \in \mathbb{N}} A_n$ is countable if and only if A_n is countable for each $n \in \mathbb{N}$. Thus,

$$\mu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \begin{cases} 1, & \text{if some } A_n \text{ is uncountable,} \\ 0, & \text{if all } A_n \text{ are countable.} \end{cases}$$

In addition, if one A_{n^*} is uncountable, then, since the A_n are pairwise disjoint and in \mathcal{F} , we get that A_n is countable for all $n \in \mathbb{N}$ with $n \neq n^*$. Thus,

$$\sum_{n \in \mathbb{N}} \mu(A_n) = \begin{cases} 1, & \text{if some } A_n \text{ is uncountable,} \\ 0, & \text{if all } A_n \text{ are countable.} \end{cases}$$

This shows that μ is σ -additive and hence a measure.

Also, observe that there are many μ -null sets, not just \emptyset .

Furthermore, μ is σ -finite; indeed, μ is a finite (even a probability) measure since $\mu(X) < \infty$, and any finite measure is σ -finite as we can choose $S_n := X$ for all $n \in \mathbb{N}$.

On the other hand, if we replaced 1 by ∞ in (1.2.3), then μ would still be a measure (check!) but it would not be σ -finite anymore. Indeed, if we had a sequence (S_n) with $S_n \in \mathcal{F}$ and $\mu(S_n) < \infty$, then this would imply that S_n is countable for all $n \in \mathbb{N}$, and in particular we would infer that

$$\bigcup_{n \in \mathbb{N}} S_n$$

is countable again. Since we assumed X to be uncountable, a fortiori we would deduce $\bigcup_{n \in \mathbb{N}} S_n \neq X$, hence μ cannot be σ -finite.

The above were relatively simple examples of set functions, which could be defined explicitly for all sets we were interested in. If, however, we would try to directly define a measure on $\mathcal{B}(\mathbb{R}^d)$ which is consistent with the content of d -dimensional volume for the hyperrectangles $\mathcal{I}^d \subset \mathcal{B}(\mathbb{R}^d)$, we would run into troubles: The reason is just that we do not have an explicit hold on elements of $\mathcal{B}(\mathbb{R}^d)$ (as we had e.g. for generated algebras, cf. Homework 2 on sheet 1). Hence, one of our goals will be to give at least an abstract machinery of extending such simple notions of volume on basic sets to bigger systems of sets, see Section 1.3 below.

We now introduce the concept of a signed measure here for the sake of completeness. This is not very essential in probability theory, but on the one hand it will turn out that it does not really make things more complicated, and on the other hand it makes our short introduction to measure theory a bit more complete.

Definition 1.2.11. For a σ -algebra \mathcal{F} over Ω , we call a set function $\mu : \mathcal{F} \rightarrow [-\infty, \infty]$ with $\mu(\emptyset) = 0$ a signed measure if μ is σ -additive.⁹

Exercise 1.2.12. (a) Show that if μ is a signed measure, then its range is either a subset of $[-\infty, \infty)$ or of $(-\infty, \infty]$.

⁹In particular, we require all sums occurring on the right-hand side of (1.3.4) to be well-defined. Alternatively we could also demand the restriction of the ranges established in Exercise 1.2.12 in the definition already, if that makes you feel more comfortable.

- (b) If μ is a signed measure and $A, B \in \mathcal{F}$ with $A \subset B$ and $\mu(B) \in (-\infty, \infty)$, then also $\mu(A) \in (-\infty, \infty)$.

Although in probability theory we will mostly be interested in probability measures, it will turn out useful to be able to cover the case of ‘nicely’ behaved infinite contents (and measures) also.

Lemma 1.2.13. *Let \mathcal{S} be a semiring and let μ be a content on \mathcal{S} . Then:*

- (a) μ is monotone;
 (b) if \mathcal{S} is a ring, then for all $A, B \in \mathcal{S}$ we have

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B); \quad (1.2.4)$$

- (c) μ is subadditive, and if μ is σ -additive, then μ is also σ -subadditive;
 (d) if \mathcal{S} is a ring and $A_1, \dots, A_n \in \mathcal{S}$ with $\mu(\bigcup_{i=1}^n A_i) < \infty$, then

$$\mu\left(\bigcup_{k=1}^n A_k\right) = \sum_{k=1}^n (-1)^{k-1} \sum_{1 \leq i_1 < \dots < i_k \leq n} \mu(A_{i_1} \cap \dots \cap A_{i_k}) \quad (\text{inclusion-exclusion formula}). \quad (1.2.5)$$

Proof. (a) Let $A, B \in \mathcal{S}$ with $A \subset B$. Since \mathcal{S} is a semiring we can write $B \setminus A = \bigcup_{j=1}^n C_j$ some pairwise disjoint $C_j \in \mathcal{S}$, $1 \leq j \leq n$, some $n \in \mathbb{N}$. Thus, A, C_j , $1 \leq j \leq n$ is a family of pairwise disjoint elements of \mathcal{S} such that

$$B = A \dot{\cup} \bigcup_{j=1}^n C_j.$$

Hence, the additivity of the content μ gives

$$\mu(B) = \mu(A) + \sum_{j=1}^n \mu(C_j) \geq \mu(A),$$

which implies the monotonicity, since $\mu \geq 0$.

- (b) We have

$$A = \underbrace{(A \cap B)}_{\in \mathcal{S}} \dot{\cup} \underbrace{(A \setminus B)}_{\in \mathcal{S}}$$

and hence the additivity of μ supplies us with

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

where we also took advantage of $A \setminus B \dot{\cup} B = A \cup B$.

- (c) Let $A, A_1, \dots, A_n \in \mathcal{S}$ with $A \subset \bigcup_{i=1}^n A_i$. Similar to the proof of Theorem 1.1.32 we write $B_1 := A_1$ and for $k \geq 2$,

$$B_k := A_k \setminus \bigcup_{i=1}^{k-1} A_i = \bigcap_{i=1}^{k-1} A_k \setminus A_i \subset A_k,$$

so in particular

$$\bigcup_{k=1}^n A_k = \bigcup_{k=1}^n B_k. \quad (1.2.6)$$

By definition of a semiring, $A_k \setminus A_i$ is the finite disjoint union of elements of \mathcal{S} . Since \mathcal{S} is a π -system, it is not hard to show that the same applies to B_k ,

i.e.,

$$B_k = \dot{\bigcup}_{i=1}^{n_k} C_i^k, \quad \text{some pairwise disjoint } C_1^k, \dots, C_{n_k}^k \in \mathcal{S}.$$

Using this representation for B_k , in a similar fashion we get that

$$A_k \setminus B_k = \dot{\bigcup}_{i=1}^{m_k} D_i^k \quad \text{some pairwise disjoint } D_1^k, \dots, D_{m_k}^k \in \mathcal{S}.$$

In combination with (1.2.6) we therefore get

$$\begin{aligned} \mu(A) &= \mu\left(\dot{\bigcup}_{k=1}^n B_k \cap A\right) = \mu\left(\dot{\bigcup}_{k=1}^n \dot{\bigcup}_{i=1}^{n_k} C_i^k \cap A\right) \stackrel{\text{additivity}}{=} \sum_{k=1}^n \sum_{i=1}^{n_k} \mu(C_i^k \cap A) \\ &\leq \sum_{k=1}^n \left(\sum_{i=1}^{n_k} \mu(C_i^k) + \sum_{i=1}^{m_k} \mu(D_i^k) \right) \stackrel{\text{additivity}}{=} \sum_{k=1}^n \mu(A_k), \end{aligned}$$

which shows the subadditivity, and where we also took advantage of the monotonicity of μ in order to obtain the inequality.

If μ is σ -additive, then the σ -subadditivity follows in essentially the same way, replacing n in the above by ∞ .

- (d) Using Lemma 1.1.12, the proof proceeds in the same way as that of [Dre18, Lemma 1.3.10] (that proof was for μ a probability measure on a σ -algebra, but it proceeds in the same way for μ a content on a ring). □

The following definition will allow us to introduce the notion of continuity for functions defined on sets also.

Definition 1.2.14. *We write*

$$(A_n) \uparrow A \text{ as } n \rightarrow \infty,$$

if $(A_n)_{n \in \mathbb{N}}$ is a sequence of sets such that $A_n \subset A_{n+1}$ for all $n \in \mathbb{N}$, and $A = \bigcup_{n \in \mathbb{N}} A_n$. Similarly, we write

$$(A_n) \downarrow A \text{ as } n \rightarrow \infty,$$

if $(A_n)_{n \in \mathbb{N}}$ is a sequence of sets such that $A_{n+1} \subset A_n$ for all $n \in \mathbb{N}$, and $A = \bigcap_{n \in \mathbb{N}} A_n$.

Definition 1.2.15. *Let \mathcal{S} be a semiring and let μ be a content defined on \mathcal{S} .*

- (a) *We say that μ is continuous from below ('stetig von unten') if for any sequence of sets (A_n) with $A_n \uparrow A$ as well as $A_n, A \in \mathcal{S}$ for all $n \in \mathbb{N}$, one has*

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A).$$

- (b) *We say that μ is continuous from above ('stetig von oben') if for any sequence of sets (A_n) with $A_n \downarrow A$ as well as $A_n, A \in \mathcal{S}$ and $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$, one has*

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A).$$

- (c) *We say that μ is continuous in \emptyset ('stetig in \emptyset ') if for any sequence of sets (A_n) with $A_n \downarrow \emptyset$ as well as $A_n \in \mathcal{S}$ and $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$, one has*

$$\lim_{n \rightarrow \infty} \mu(A_n) = 0.$$

As the name of this section suggests, we will mostly be interested in measures, i.e., in particular the property of a set function being σ -additive will play a crucial role. Thus, it turns out useful to have other characterizations (and related properties) of σ -additivity available.

Proposition 1.2.16. *Let μ be a content on a ring \mathcal{R} . Consider the following properties:*

- (a) μ is σ -additive (i.e., it is a pre-measure);
- (b) μ is continuous from below;
- (c) μ is continuous from above;
- (d) μ is continuous in \emptyset ;

We have the following implications:

$$(a) \iff (b) \implies (c) \iff (d).$$

If in addition μ is finite, then in the previous display \implies can be replaced by \iff .

Proof. '(a) \implies (b)':

Let (A_n) and A be as in the assumptions of (b). As done several times before already, we define $B_1 := A_1$ as well as

$$B_n := A_n \setminus \underbrace{\bigcup_{i=1}^{n-1} A_i}_{\substack{\in \mathcal{R} \\ \in \mathcal{R}}}$$

for $n \geq 2$. Then the (B_n) form a sequence of pairwise disjoint sets with $B_n \in \mathcal{R}$, and such that $\bigcup_{i=1}^n B_i = \bigcup_{i=1}^n A_i$ for all $n \in \mathbb{N}$. As a consequence, and since μ is σ -additive by assumption, we get

$$\mu(A) = \mu\left(\bigcup_{i=1}^{\infty} B_i\right) \stackrel{\mu \sigma\text{-additive}}{=} \sum_{i=1}^{\infty} \mu(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(B_i) = \lim_{n \rightarrow \infty} \mu(A_n),$$

which shows that μ is continuous from below.

'(b) \implies (a)':

Let (A_n) be a sequence of pairwise disjoint sets with $A_n \in \mathcal{R}$ for all $n \in \mathbb{N}$ as well as $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{R}$. Defining $B_n := \bigcup_{i=1}^n A_i$, we have a sequence (B_n) with $B_n \in \mathcal{R}$ and such that $B_n \uparrow B := \bigcup_{i \in \mathbb{N}} B_i \in \mathcal{R}$ as $n \rightarrow \infty$. Henceforth, the continuity from below supplies us with

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu(B) \stackrel{\mu \text{ continuous from below}}{=} \lim_{n \rightarrow \infty} \mu(B_n) \stackrel{\text{additivity}}{=} \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(A_i) = \sum_{i=1}^{\infty} \mu(A_i),$$

which proves the σ -additivity of μ .

'(b) \implies (c)':

Let (A_n) and A be as in the assumptions of (c). Since $\mu(A_n) < \infty$ and $A_n \supset A_{n+1}$ for all $n \in \mathbb{N}$, the additivity of μ implies that

$$\mu(A_1 \setminus A_n) = \mu(A_1) - \mu(A_n). \quad (1.2.7)$$

Now because $A_n \downarrow A$ as $n \rightarrow \infty$, we get that $A_1 \setminus A_n \uparrow A_1 \setminus A$ as $n \rightarrow \infty$, and therefore the validity of (b) implies the existence of the left-hand side of

$$\mu(A_1 \setminus A) = \lim_{n \rightarrow \infty} \mu(A_1 \setminus A_n) \stackrel{(1.2.7)}{=} \lim_{n \rightarrow \infty} \mu(A_1) - \mu(A_n) = \mu(A_1) - \lim_{n \rightarrow \infty} \mu(A_n),$$

exists. Hence, using $\mu(A_1) < \infty$,

$$\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A_1) - \mu(A_1 \setminus A) = \mu(A),$$

where in the last equality we took advantage of the additivity of μ .

'(c) \implies (d)': This is obvious.

'(d) \implies (c)':

Let (A_n) and A be as in (c). Then $B_n := A_n \setminus A \downarrow \emptyset$ as $n \rightarrow \infty$. Furthermore, the monotonicity of μ in combination with $\mu(A_n) < \infty$ implies that $\mu(B_n) < \infty$ for all $n \in \mathbb{N}$. As a consequence, we deduce from the continuity of μ in \emptyset that

$$\mu(A_n) = \mu(B_n) + \mu(A) \rightarrow \mu(A) \quad \text{as } n \rightarrow \infty,$$

which implies the continuity of μ from above.

It remains to prove '(b) \iff (c)' in the case of μ being finite. For this purpose, let (A_n) and A be as in (b). Then $A \setminus A_n \downarrow \emptyset$ as $n \rightarrow \infty$ and $\mu(A_n) < \infty$ for all $n \in \mathbb{N}$. Hence, the continuity of μ in \emptyset in combination with the finiteness of μ supply us with

$$\mu(A) - \mu(A_n) = \mu(A \setminus A_n) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and hence $\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)$, which shows the continuity from below. \square

We now give a first application of Dynkin's π - λ -Theorem (Theorem 1.1.33) that will prove useful later on.

Theorem 1.2.17. *Let \mathcal{E} be a π -system over Ω . Assume there exists a sequence of sets (E_n) with $E_n \in \mathcal{E}$ for all $n \in \mathbb{N}$ and $\bigcup_{n \in \mathbb{N}} E_n = \Omega$. Furthermore, let μ_1, μ_2 denote two measures on $(\Omega, \sigma(\mathcal{E}))$ such that*

$$(a) \quad \mu_1(E) = \mu_2(E) \quad \forall E \in \mathcal{E}, \quad (1.2.8)$$

$$(b) \quad \mu_1(E_n) = \mu_2(E_n) < \infty \quad \forall n \in \mathbb{N}. \quad (1.2.9)$$

Then the measures μ_1 and μ_2 coincide and are σ -finite.

Proof. The σ -finiteness is a direct consequence of (1.2.9). Hence, it remains to show that μ_1 and μ_2 coincide. For $E \in \mathcal{E}$ with $\mu_1(E) < \infty$, define

$$\mathcal{D}_E := \{D \in \sigma(\mathcal{E}) : \mu_1(D \cap E) = \mu_2(D \cap E)\}.$$

Claim 1.2.18. \mathcal{D}_E is a Dynkin system.

Proof. We have $\Omega \in \mathcal{D}_E$ since the middle equality of $\mu_1(\Omega \cap E) = \mu_1(E) = \mu_2(E) = \mu_2(\Omega \cap E)$ follows from $E \in \mathcal{E}$.

Also, for $D \in \mathcal{D}_E$ we have

$$\mu_1(D^c \cap E) = \mu_1(E) - \mu_1(D \cap E) = \mu_2(E) - \mu_2(D \cap E) = \mu_2(D^c \cap E),$$

where the second equality follows from $D \in \mathcal{D}_E$ and the fact that $E \in \mathcal{E}$. Hence, we deduce $D^c \in \mathcal{D}_E$.

It remains to show that \mathcal{D}_E is stable under countable unions of pairwise disjoint sets. Thus, for a sequence (D_n) of pairwise disjoint sets with $D_n \in \mathcal{D}_E$ for all $n \in \mathbb{N}$, we deduce

$$\mu_1\left(\bigcup_{n \in \mathbb{N}} D_n \cap E\right) = \sum_{n \in \mathbb{N}} \mu_1(D_n \cap E) = \sum_{n \in \mathbb{N}} \mu_2(D_n \cap E) = \mu_2\left(\bigcup_{n \in \mathbb{N}} D_n \cap E\right),$$

where the first and last equality exploit that the $D_n \cap E$ are pairwise disjoint, and the middle equality takes advantage of the fact that $D_n \in \mathcal{D}_E$ for all $n \in \mathbb{N}$. \square

Due to the above claim, \mathcal{D}_E is a Dynkin system with $\mathcal{E} \subset \mathcal{D}_E$. Since \mathcal{E} is a π -system, Theorem 1.1.33 implies that $\sigma(\mathcal{E}) = \mathcal{D}_E$. In particular, this implies that for each $E \in \sigma(\mathcal{E})$ and $n \in \mathbb{N}$,

$$\mu_1(E \cap E_n) = \mu_2(E \cap E_n). \quad (1.2.10)$$

Setting $F_n := E_n \setminus \bigcup_{i=1}^{n-1} E_i$ we have the disjoint union

$$\dot{\bigcup}_{n \in \mathbb{N}} F_n = \Omega.$$

Thus, using that for $E \in \sigma(\mathcal{E})$ we have $F_n \cap E \in \sigma(\mathcal{E})$, we obtain in combination with (1.2.10) that

$$\mu_1(F_n \cap E) = \mu_1(E_n \cap (F_n \cap E)) = \mu_2(E_n \cap (F_n \cap E)) = \mu_2(F_n \cap E).$$

Summing these identities over $n \in \mathbb{N}$ we obtain $\mu_1(E) = \mu_2(E)$, which finishes the proof, since $E \in \sigma(\mathcal{E})$ had been chosen arbitrarily. \square

Corollary 1.2.19. *If μ_1 and μ_2 are measures on $(\Omega, \sigma(\mathcal{E}))$ with $\mu_1(\Omega) = \mu_2(\Omega) < \infty$ and such that (1.2.8) holds, then $\mu_1 = \mu_2$.*

Proof. We define $\tilde{\mathcal{E}} := \mathcal{E} \cup \{\Omega\}$. Then the assumptions of Theorem 1.2.17 are fulfilled with \mathcal{E} replaced by $\tilde{\mathcal{E}}$ and $E_n := \Omega$ for all $n \in \mathbb{N}$. Since $\sigma(\mathcal{E}) = \sigma(\tilde{\mathcal{E}})$, the result follows. \square

1.3 Carathéodory's extension theorem ('Maßerweiterungssatz')

In this section we will see how to extend contents to measures. As it turns out, in pursuing this endeavor it will be useful to make a detour via so-called 'outer measures' introduced below: Contents will give rise to 'nice' outer measures (see Theorem 1.3.9 below), which themselves (by restriction to 'well-behaved' sets) give rise to measures (see Theorem 1.3.5 below).

We start with investigating how to go from outer measures to measures, and start with the definition of the former.

Definition 1.3.1. *For Ω a non-empty set, a set function $\mu^* : 2^\Omega \rightarrow [0, \infty]$ is called an outer measure ('äußeres Maß'), if*

- (a) $\mu^*(\emptyset) = 0$;
- (b) μ^* is monotone;
- (c) μ^* is σ -subadditive.

Remark 1.3.2. *Since an outer measure μ^* is σ -subadditive by definition, we deduce that for $A_1, \dots, A_n \in 2^\Omega$, we have, setting $A_m := \emptyset \in 2^\Omega$ for $m > n$, that*

$$\mu^*\left(\bigcup_{i=1}^n A_i\right) = \mu^*\left(\bigcup_{i=1}^\infty A_i\right) \stackrel{\sigma\text{-subadditivity}}{\leq} \sum_{i=1}^\infty \mu^*(A_i) \stackrel{\text{Definition 1.3.1 (a)}}{=} \sum_{i=1}^n \mu^*(A_i).$$

Thus, an outer measure is also subadditive.

We will now introduce the concept of sets which are *measurable* with respect to an outer measure. Such sets will be the 'well-behaved' sets alluded to above; they form a σ -algebra on which (the restriction of) the outer measure induces a measure (see Theorem 1.3.5 below).

Definition 1.3.3. *Let Ω be a non-empty set and assume an outer measure $\mu^* : 2^\Omega \rightarrow [0, \infty]$ to be given. Then $A \in 2^\Omega$ is called μ^* -measurable (' μ^* -messbar'), if for all $B \in 2^\Omega$,*

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c). \quad (1.3.1)$$

Remark 1.3.4. Since an outer measure is subadditive due to Remark 1.3.2, we infer that

$$\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$$

holds true for all $A, B \in 2^\Omega$. Therefore, (1.3.1) is equivalent to

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

The next results is fundamental in our construction of measures and provides us with a recipe for how to obtain a measure from an outer measure.

Theorem 1.3.5 (Carathéodory). For an outer measure $\mu^* : 2^\Omega \rightarrow [0, \infty]$, the set

$$\mathcal{M}_{\mu^*} := \{A \subset \Omega : A \text{ is } \mu^*\text{-measurable}\}$$

is a σ -algebra over Ω , and the restriction

$$\mu^*|_{\mathcal{M}_{\mu^*}} \tag{1.3.2}$$

of μ^* to \mathcal{M}_{μ^*} is a measure.

Proof. We start with proving the following claim.

Claim 1.3.6. \mathcal{M}_{μ^*} is an algebra over Ω .

Proof. By Definition 1.3.1 (a) we have $\mu^*(\emptyset) = 0$, and hence (1.3.1) trivially holds true for $A = \Omega$.

Furthermore, since (1.3.1) is symmetric in A and A^c we immediately deduce that \mathcal{M}_{μ^*} is stable under complements.

It remains to show Property (c) of Definition 1.1.14. For this purpose, let $A_1, A_2 \in \mathcal{M}_{\mu^*}$, and let $B \subset \Omega$ be arbitrary. Then

$$\begin{aligned} \mu^*(B) &\stackrel{A_1 \in \mathcal{M}_{\mu^*}}{=} \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\ &\stackrel{A_2 \in \mathcal{M}_{\mu^*}}{=} \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\ &\stackrel{\mu^* \text{ subadditive}}{\geq} \mu^*((B \cap A_1) \cup (B \cap A_1^c \cap A_2)) + \mu^*(B \cap (A_1 \cup A_2)^c) \\ &= \mu^*(B \cap (A_1 \cup A_2)) + \mu^*(B \cap (A_1 \cup A_2)^c). \end{aligned}$$

This shows that $A \cup B \in \mathcal{M}_{\mu^*}$, and therefore \mathcal{M}_{μ^*} is an algebra. \square

In order to show that \mathcal{M}_{μ^*} is a σ -algebra, due to Theorem 1.1.32 it is sufficient to show that it is a λ -system;

Since \mathcal{M}_{μ^*} is an algebra due to Claim 1.3.6, we immediately get $\Omega \in \mathcal{M}_{\mu^*}$ and we also know that \mathcal{M}_{μ^*} is stable under complements. Thus, it remains to show the union of a countable family of pairwise disjoint elements of \mathcal{M}_{μ^*} is contained in \mathcal{M}_{μ^*} again. For this purpose, let (A_n) be a sequence of pairwise disjoint sets with $A_n \in \mathcal{M}_{\mu^*}$ for all $n \in \mathbb{N}$.

We start with inductively proving that for all $n \in \mathbb{N}$,

$$\mu^*\left(B \cap \bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mu^*(B \cap A_i) \quad \forall B \subset \Omega. \tag{1.3.3}$$

Indeed, for $n = 1$ this boils down to a tautology, so assume (1.3.3) holds for arbitrary $n \in \mathbb{N}$. Then, using that the A_i are pairwise disjoint and that $\bigcup_{i=1}^n A_i \in \mathcal{M}_{\mu^*}$,

$$\begin{aligned} \mu^*\left(B \cap \bigcup_{i=1}^{n+1} A_i\right) &= \mu^*\left(\left(B \cap \bigcup_{i=1}^{n+1} A_i\right) \cap \bigcup_{i=1}^n A_i\right) + \mu^*\left(\left(B \cap \bigcup_{i=1}^{n+1} A_i\right) \cap \left(\bigcup_{i=1}^n A_i\right)^c\right) \\ &= \sum_{i=1}^n \mu^*(B \cap A_i) + \mu^*(B \cap A_{n+1}), \end{aligned}$$

where we used the induction assumption in the last step. This finishes the induction step and therefore establishes (1.3.3).

We infer that

$$\mu^*(B) \geq \mu^*\left(B \cap \bigcup_{i=1}^n A_i\right) + \mu^*\left(B \cap \left(\bigcup_{i=1}^n A_i\right)^c\right) \geq \sum_{i=1}^n \mu^*(B \cap A_i) + \mu^*\left(B \cap \left(\bigcup_{i=1}^{\infty} A_i\right)^c\right),$$

where we used the fact that \mathcal{M}_{μ^*} is an algebra to get the inequality, and the inequality is a consequence of display (1.3.3) in combination with the monotonicity of μ^* . Taking $n \rightarrow \infty$ in the previous display and using that μ^* is σ -subadditive, we arrive at

$$\begin{aligned} \mu^*(B) &\geq \sum_{i=1}^{\infty} \mu^*(B \cap A_i) + \mu^*\left(B \cap \left(\bigcup_{i=1}^{\infty} A_i\right)^c\right) \\ &\geq \mu^*\left(B \cap \bigcup_{i=1}^{\infty} A_i\right) + \mu^*\left(B \cap \left(\bigcup_{i=1}^{\infty} A_i\right)^c\right), \end{aligned} \tag{1.3.4}$$

which shows that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{M}_{\mu^*}$. Therefore, \mathcal{M}_{μ^*} is a λ -system, which in combination with Claim 1.3.6 and Theorem 1.1.32 implies that \mathcal{M}_{μ^*} is a σ -algebra also.

Last but not least, we have to show (1.3.2). Since μ^* is an outer measure, we have $\mu^*(\emptyset) = 0$ by definition. Also, for a sequence (A_n) of pairwise disjoint sets with $A_n \in \mathcal{M}_{\mu^*}$ we have, choosing $B := \bigcup_{n=1}^{\infty} A_n$, that the first inequality in (1.3.4) supplies us with

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \geq \sum_{n=1}^{\infty} \mu^*(A_n).$$

In combination with the σ -subadditivity of μ^* this concludes the proof that $\mu^*|_{\mathcal{M}_{\mu^*}}$ is a measure. \square

Definition 1.3.7. A measure μ on a measurable space (Ω, \mathcal{F}) (or the measure space $(\Omega, \mathcal{F}, \mu)$ for that matter) is called *complete*, if for all $M \in \mathcal{F}$ with $\mu(M) = 0$, and all $N \subset M$, one has that $N \in \mathcal{F}$ as well.

Exercise 1.3.8. Show that the measure $\mu^*_{\mathcal{M}_{\mu^*}}$ on $(\Omega, \mathcal{M}_{\mu^*})$ from Theorem 1.3.5 is complete in the sense of Definition 1.3.7.

In what follows, we will assume the standard convention that

$$\inf \emptyset = \infty \quad \text{and} \quad \sup \emptyset = -\infty.$$

Theorem 1.3.9. Let $\mathcal{A} \subset 2^{\Omega}$ with $\emptyset \in \mathcal{A}$, and let $\mu : \mathcal{A} \rightarrow [0, \infty]$ be a set function with $\mu(\emptyset) = 0$. For $A \subset \Omega$ define the set function $\mu^* : 2^{\Omega} \rightarrow [0, \infty]$ via

$$\mu^*(A) := \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) : A_1, A_2, \dots \in \mathcal{A}, \text{ and } A \subset \bigcup_{i=1}^{\infty} A_i \right\}, \quad A \subset \Omega. \tag{1.3.5}$$

Then,

- (a) μ^* defines an outer measure;
- (b) if \mathcal{A} is a semiring and μ is a content, then $\mathcal{A} \subset \mathcal{M}_{\mu^*}$;
- (c) if \mathcal{A} is a semiring and if μ is not only a content but also σ -subadditive, then

$$\mu^*|_{\mathcal{A}} = \mu.$$

Proof. (a) Setting $A_n := \emptyset \in \mathcal{A}$, the definition in (1.3.5) of μ^* immediately yields $\mu^*(\emptyset) = 0$. The monotonicity of μ^* also is a direct consequence of (1.3.5).

It remains to show the σ -subadditivity. For this purpose, let (A_n) be a sequence with $A_n \subset \Omega$ for all $n \in \mathbb{N}$, and assume without loss of generality that $\mu^*(A_n) < \infty$ for all $n \in \mathbb{N}$. Then for $\varepsilon > 0$ arbitrary, choose for each $n \in \mathbb{N}$ a covering $(A_{n,i})_{i \in \mathbb{N}}$ of A_n with $A_{n,i} \in \mathcal{A}$ for all $i \in \mathbb{N}$ and such that

$$\sum_{i=1}^{\infty} \mu(A_{n,i}) \leq \mu^*(A_n) + 2^{-n} \varepsilon. \quad (1.3.6)$$

Then

$$\bigcup_{n \in \mathbb{N}} A_n \subset \bigcup_{i, n \in \mathbb{N}} A_{n,i},$$

and by the definition of μ^* in combination with (1.3.6), we deduce

$$\begin{aligned} \mu^*\left(\bigcup_{n \in \mathbb{N}} A_n\right) &\leq \sum_{n \in \mathbb{N}} \underbrace{\sum_{i \in \mathbb{N}} \mu(A_{n,i})}_{\stackrel{(1.3.6)}{\leq} \mu^*(A_n) + 2^{-n} \varepsilon} \leq \sum_{n \in \mathbb{N}} \mu^*(A_n) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ had been chosen arbitrarily, we deduce

$$\mu^*\left(\bigcup_{n \in \mathbb{N}} A_n\right) \leq \sum_{n \in \mathbb{N}} \mu^*(A_n),$$

which shows the σ -subadditivity of μ^* .

(b) For $A \in \mathcal{A}$ arbitrary we want to show that for all $E \subset \Omega$ we have

$$\mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E),$$

and without loss of generality we can assume $\mu^*(E) < \infty$. Then for $\varepsilon > 0$ arbitrary but fixed, we find a sequence (A_n) such that

- $A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$;
- $E \subset \bigcup_{n=1}^{\infty} A_n$;
-

$$\sum_{n=1}^{\infty} \mu(A_n) \leq \mu^*(E) + \varepsilon. \quad (1.3.7)$$

Now since \mathcal{A} is a semiring, for any $n \in \mathbb{N}$, we find $m_n \in \mathbb{N}$ and $A_{n,i}$, $1 \leq i \leq m_n$, such that

$$A_n \cap A^c = A_n \setminus A = \bigcup_{i=1}^{m_n} A_{n,i}.$$

As a consequence, we get

$$\begin{aligned} \mu^*(E \cap A) + \mu^*(E \cap A^c) &\leq \sum_{n=1}^{\infty} \mu(A_n \cap A) + \sum_{n=1}^{\infty} \mu^*\left(\bigcup_{i=1}^{m_n} A_{n,i}\right) \\ &\leq \sum_{n=1}^{\infty} \left(\mu(A_n \cap A) + \sum_{i=1}^{m_n} \mu(A_{n,i}) \right) = \sum_{n=1}^{\infty} \mu(A_n) \stackrel{(1.3.7)}{\leq} \mu^*(E) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ was chosen arbitrarily and using Remark 1.3.4, this finishes the proof.

- (c) Let $A \in \mathcal{A}$. On the one hand, by definition of μ^* we have $\mu^*(A) \leq \mu(A)$. On the other hand, for any covering $A_1, A_2, \dots \in \mathcal{A}$ with $A \subset \bigcup_{i=1}^{\infty} A_i$ we have

$$\mu(A) = \mu\left(\bigcup_{i=1}^{\infty} (A \cap A_i)\right) \leq \sum_{n=1}^{\infty} \mu(A_i \cap A), \quad (1.3.8)$$

where we used that μ is σ -subadditive in combination with the fact that $A_i \cap A \in \mathcal{A}$ for all $i \in \mathbb{N}$. Since display (1.3.8) holds for any suitable covering, this shows that $\mu(A) \leq \mu^*(A)$, which finishes the proof. \square

We can summarize Theorems 1.3.9 and 1.3.5 to obtain the following corollary.

Corollary 1.3.10. *Let \mathcal{S} be a semiring and let $\tilde{\mu}$ be a content on \mathcal{S} which is σ -subadditive. By $\tilde{\mu}^*$ we denote the corresponding outer measure as defined in (1.3.5).*

Then there exists a measure $\mu : \mathcal{M}_{\tilde{\mu}^} \rightarrow [0, \infty]$ such that $\tilde{\mu}$ and μ coincide on \mathcal{S} .*

Furthermore, if $\tilde{\mu}$ is σ -finite, then so is μ , and in this case the restriction of μ to $\sigma(\mathcal{S}) \subset \mathcal{M}_{\tilde{\mu}^}$ is the unique extension of $\tilde{\mu}$ to a measure on $\sigma(\mathcal{S})$.*

Proof. The existence of a measure μ on $\mathcal{M}_{\tilde{\mu}^*}$ extending $\tilde{\mu}$ follows from Theorems 1.3.9 and 1.3.5.

If $\tilde{\mu}$ is σ -finite, then there exists a sequence (S_n) such that $S_n \in \mathcal{S}$, $\bigcup_{n \in \mathbb{N}} S_n = \Omega$, and $\mu(S_n) < \infty$ for all $n \in \mathbb{N}$. In particular, this implies that μ is σ -finite.

Furthermore, using Theorem 1.2.17 we deduce that there exists at most one measure on $\sigma(\mathcal{S})$ which extends $\tilde{\mu}$, so the extension is unique. \square

1.3.1 Lebesgue measure

Our principal goal in this section is to extend the elementary content λ^d that we had defined on hyperrectangles in (1.2.2) to a measure on the σ -algebra generated by the hyperrectangles; due to Lemma 1.1.31, this σ -algebra coincides with $\mathcal{B}(\mathbb{R}^d)$.

Theorem 1.3.11 (d -dimensional Lebesgue measure). *There exists a uniquely determined σ -finite measure λ^d on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^d))$ such that*

$$\lambda^d((a, b]) = \prod_{i=1}^d (b_i - a_i) \quad \forall a, b \in \mathbb{R}^d \text{ with } a < b. \quad (1.3.9)$$

λ^d is called the d -dimensional (Borel-) Lebesgue measure.

Proof. In order to distinguish Lebesgue content and Lebesgue measure, let us write $\tilde{\lambda}^d$ for the Lebesgue content.

We will apply Corollary 1.3.10 to the case $\mathcal{S} = \mathcal{I}^d$ and $\tilde{\mu} = \tilde{\lambda}^d$. In combination with the fact that $\sigma(\mathcal{I}^d) = \mathcal{B}(\mathbb{R}^d)$ (cf. Lemma 1.1.31), the measure whose existence is implied by Corollary 1.3.10 will be the unique one defined on $\mathcal{B}(\mathbb{R}^d)$ and satisfying (1.3.9).

We have to show that

$$\tilde{\lambda}^d \text{ is } \sigma\text{-subadditive on } \mathcal{I}^d. \quad (1.3.10)$$

For this purpose let $a, b \in \mathbb{R}^d$ with $a \leq b$ be given, and let $(a(n), b(n)]$, where $a(n), b(n) \in \mathbb{R}^d$ with $a(n) \leq b(n)$ for all $n \in \mathbb{N}$, be a sequence of hypercubes (i.e., elements of \mathcal{I}^d) such that

$$(a, b] \subset \bigcup_{n=1}^{\infty} (a(n), b(n)].$$

It is sufficient to show that for the content $\tilde{\lambda}^d$ on \mathcal{I}^d we have

$$\tilde{\lambda}^d((a, b]) \leq \sum_{n=1}^{\infty} \tilde{\lambda}^d((a(n), b(n)]). \quad (1.3.11)$$

We will show this inequality by exploiting a continuity property of the content $\tilde{\lambda}^d$ in order to apply a compactness argument which then reduces the above to a finite setting. To be precise, for $\varepsilon > 0$ arbitrary choose for each $n \in \mathbb{N}$ an element $b^\varepsilon(n) \in \mathbb{R}^d$ such that $b^\varepsilon(n) > b(n)$ and such that

$$\tilde{\lambda}^d((a(n), b^\varepsilon(n)]) \leq \tilde{\lambda}^d((a(n), b(n)]) + 2^{-n}\varepsilon.$$

In addition, choose $a^\varepsilon \in \mathbb{R}^d$ such that $a^\varepsilon > a$ and such that

$$\tilde{\lambda}^d((a^\varepsilon, b]) \geq \tilde{\lambda}^d((a, b]) - \varepsilon. \quad (1.3.12)$$

Now

$$[a^\varepsilon, b] \subset (a, b] \subset \bigcup_{n=1}^{\infty} (a(n), b^\varepsilon(n)).$$

Since the left-hand side is a compact set, there exists $N_0 \in \mathbb{N}$ such that

$$[a^\varepsilon, b] \subset \bigcup_{n=1}^{N_0} (a(n), b^\varepsilon(n)).$$

Hence, using the fact that the content λ is (finitely) subadditive (cf. Lemma 1.2.13), we deduce

$$\begin{aligned} \tilde{\lambda}^d((a, b]) &\stackrel{(1.3.12)}{\leq} \tilde{\lambda}^d((a^\varepsilon, b]) + \varepsilon \\ &\leq \sum_{n=1}^{N_0} \tilde{\lambda}^d((a(n), b^\varepsilon(n)]) + \varepsilon \leq \sum_{n=1}^{N_0} \tilde{\lambda}^d((a(n), b(n)]) + 2\varepsilon. \end{aligned}$$

In particular, this implies

$$\tilde{\lambda}^d((a, b]) \leq \sum_{n=1}^{\infty} \tilde{\lambda}^d((a(n), b(n)]) + 2\varepsilon.$$

Since $\varepsilon > 0$ was chosen arbitrarily, this establishes (1.3.10) and hence (1.3.9).

The σ -finiteness follows immediately from the σ -finiteness of the Lebesgue content and the fact that Lebesgue content and Lebesgue measure coincide on \mathcal{I}^d . \square

Remark 1.3.12. For simplicity write λ^* for the outer measure induced by the d -dimensional Lebesgue content as defined in (1.3.5). Theorem 1.3.5 actually shows that the measure λ^d can be defined on the σ -algebra $\mathcal{L}(\mathbb{R}^d) := \mathcal{M}_{\lambda^*}$ of the so-called Lebesgue-measurable sets (in a unique way, and $\mathcal{L}(\mathbb{R}^d)$ is the completion of $\mathcal{B}(\mathbb{R}^d)$, i.e., the smallest complete σ -algebra containing $\mathcal{B}(\mathbb{R}^d)$; see [Els05, Corollary II.6.5]), where by λ^* we denote the outer measure induced by the content $\tilde{\lambda}$ and (1.3.5). From Theorem 1.3.9 we infer that

$$\mathcal{B}(\mathbb{R}^d) \subset \mathcal{L}(\mathbb{R}^d) \subset 2^\Omega,$$

and one can show that both of these inclusions are strict (see [Els05, Example II.4.6]).

1.3.2 Lebesgue-Stieltjes measure

Theorem 1.3.13 (Lebesgue-Stieltjes measures). *Let a non-decreasing function and right-continuous function $F : \mathbb{R} \rightarrow \mathbb{R}$ be given. Recall the semiring \mathcal{I} of left-open right-closed intervals in \mathbb{R} which had been introduced in Example 1.1.6. Then*

$$\begin{aligned}\tilde{\mu}_F : \mathcal{I} &\rightarrow [0, \infty), \\ (a, b] &\mapsto F(b) - F(a),\end{aligned}$$

defines a content, and there exists a unique σ -finite measure μ_F on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\mu_F(I) = \tilde{\mu}_F(I) \text{ for all } I \in \mathcal{I}.$$

μ_F is called the Lebesgue-Stieltjes measure of F .

Proof. We have $\tilde{\mu}_F(\emptyset) = 0$, and for pairwise disjoint intervals $(a(1), b(1)], \dots, (a(n), b(n)]$ with $(a, b] = \bigcup_{i=1}^n (a(i), b(i)]$ some $a, b \in \mathbb{R}$ with $a < b$ we have

$$\sum_{i=1}^n \tilde{\mu}_F((a(i), b(i)]) = F\left(\max_{1 \leq i \leq n} b(i)\right) - F\left(\min_{1 \leq i \leq n} a(i)\right) = \tilde{\mu}_F((a, b]),$$

since $a = \min_{1 \leq i \leq n} a(i)$ and $b = \max_{1 \leq i \leq n} b(i)$. This means that $\tilde{\mu}_F$ is additive also, and hence a content on \mathcal{I} .

Since \mathcal{I} is a semiring, according to Corollary 1.3.10, the only thing that is left to show is that $\tilde{\mu}_F$ is σ -subadditive and σ -finite. For the first, we proceed similarly to the proof of Theorem 1.3.11. Indeed, let $(a, b]$ as well as $((a(n), b(n)])_{n \in \mathbb{N}}$ with $a, b, a(n), b(n) \in \mathbb{R}$ be given such that $a \leq b$ and $a(n) \leq b(n)$ for all $n \in \mathbb{N}$, and such that

$$(a, b] \subset \bigcup_{n=1}^{\infty} (a(n), b(n)].$$

For $\varepsilon > 0$ given we take advantage of the right continuity of F in order to deduce the existence of $a_\varepsilon > a$ such that

$$\tilde{\mu}_F((a_\varepsilon, b]) \geq \tilde{\mu}_F((a, b]) - \varepsilon.$$

In addition, using the right continuity of F again, for each $n \in \mathbb{N}$ we find $b_\varepsilon(n) \in \mathbb{R}$ with $b_\varepsilon(n) > b(n)$ and

$$\tilde{\mu}_F((a(n), b(n)]) \geq \tilde{\mu}_F((a(n), b_\varepsilon(n)]) - \varepsilon 2^{-n}.$$

As in the proof of Theorem 1.3.11 we can then deduce that

$$\tilde{\mu}_F((a, b]) \leq \sum_{n \in \mathbb{N}} \tilde{\mu}_F((a(n), b(n)]) + 2\varepsilon.$$

Since $\varepsilon > 0$ had been chosen arbitrarily, we infer

$$\tilde{\mu}_F((a, b]) \leq \sum_{n \in \mathbb{N}} \tilde{\mu}_F((a(n), b(n)]),$$

which implies the desired σ -subadditivity.

Regarding the σ -finiteness, we observe that $\tilde{\mu}_F$ is σ -finite since $\tilde{\mu}_F((-n, n]) < \infty$ for all $n \in \mathbb{N}$, and hence so is μ_F . \square

Example 1.3.14. (a) Show that for $F : \mathbb{R} \rightarrow \mathbb{R}$ with $F(x) = x$ the Lebesgue-Stieltjes measure μ_F from Theorem 1.3.13 coincides the one-dimensional Lebesgue measure λ^1 .

(b) Let $f : \mathbb{R} \rightarrow [0, \infty)$ be continuous and set

$$F(x) := \int_0^x f(r) \, dr, \quad x \in \mathbb{R},$$

with the right-hand side interpreted as Riemann integral. Then F is non-decreasing and continuous (e.g. due to the Fundamental Theorem of Calculus), and hence we can use Theorem 1.3.13 to deduce the existence of a measure μ_F on $\sigma(\mathcal{I}) = \mathcal{B}(\mathbb{R})$.

Remark 1.3.15. If we have a closer look at the proof of Theorem 1.3.13 again, we discover that it did not hinge on the special nature of the Lebesgue content, but rather we only needed some continuity properties of the content. In particular, if F_1, \dots, F_n are non-decreasing and right-continuous functions from \mathbb{R} to \mathbb{R} just as in Theorem 1.3.13, then these give rise to measures μ_1, \dots, μ_n on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ just in the same vein as in Theorem 1.3.13. We can then perform the proof of Theorem 1.3.11 in essentially the same way as before and obtain that there exists a unique σ -finite measure μ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ (recall that $\sigma(\mathcal{I}^n) = \mathcal{B}(\mathbb{R}^n)$ due to Lemma 1.1.31) such that

$$\mu((a, b]) = \prod_{i=1}^n \mu_i((a_i, b_i]), \quad \forall a, b \in \mathbb{R}^n \text{ with } a < b.$$

We also write $\otimes_{i=1}^n \mu_i := \mu$ and call μ the product measure of μ_1, \dots, μ_n .

In Section 2.4 below we will see how to construct product measures not only as products of measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ but on arbitrary measurable spaces, and also how to construct infinite products of measures.

In what follows, most results and definitions concerning probability spaces can be generalized to finite measure spaces $(\Omega, \mathcal{F}, \mu)$, where μ is a finite measure. There is, however, oftentimes a more significant difference between the case of μ being finite or infinite, so we have to be a bit more careful when trying to transfer results we have for finite measure spaces to infinite measure spaces.

1.4 Measurable functions, random variables

The Riemann integral had been introduced by partitioning the domain of definition of the integrand (oftentimes intervals in \mathbb{R} or hypercubes in \mathbb{R}^d) into finer and finer pieces, and then consider the upper and lower Riemann sums. If they converged to the same limit as the partitions got finer and finer, this limit had been defined as the corresponding Riemann integral. The notion of integral we will be introducing later on,¹⁰ on the other hand, will be defined for real or complex functions defined on a measure space $(\Omega, \mathcal{F}, \mu)$. It can essentially be defined by partitioning the range of the integrand into finer and finer pieces. In particular, for an integrand f this will require that preimages of intervals, i.e., sets of the form $f^{-1}([a, b])$, can be measured by the measure underlying the domain of definition of the function. This means we want expressions of the type $\mu(f^{-1}([a, b]))$ to be well-defined, which is equivalent to $f^{-1}([a, b]) \in \mathcal{F}$. As a consequence, such functions will play a special role. (In the case of $(\Omega, \mathcal{F}, \mu)$ being a probability space, the definition of random variables in [Dre18, Def.1.7.1] had been general enough to serve our purposes.)

Definition 1.4.1. Let (Ω, \mathcal{F}) and (E, \mathcal{E}) be measurable spaces. A function $f : \Omega \rightarrow E$ is called a measurable function (‘messbare Funktion’) if for all $A \in \mathcal{E}$ its preimage under f is contained in \mathcal{F} , i.e., if

$$f^{-1}(A) := \{\omega \in \Omega : f(\omega) \in A\} \in \mathcal{F}, \quad \forall A \in \mathcal{E}.$$

¹⁰It is called the *Lebesgue integral*, although it not only refers to integrals with respect to the standard Lebesgue measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, but with respect to arbitrary measures.

In this case, f is also said to be $\mathcal{F} - \mathcal{E}$ -measurable ($\mathcal{F} - \mathcal{E}$ -messbar).

If not only a measurable space (Ω, \mathcal{F}) but a probability space $(\Omega, \mathcal{F}, \mu)$ is given, then f as above is called a random variable (*‘Zufallsvariable’*). If furthermore $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then f is referred to as a real random variable (*‘reelle Zufallsvariable’*, *‘Zufallsgröße’*), and if $(E, \mathcal{E}) = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$, we say that f is an extended real random variable (*‘erweiterte reelle Zufallsvariable’*).

Even if we have given only a measurable space (Ω, \mathcal{F}) , we denote the space of extended real functions by \mathcal{M} , and the subset of non-negative extended real-valued functions by \mathcal{M}^+ . For a random variable X the values $X(\omega)$, $\omega \in \Omega$, are also called realizations of the random variable X .

For the sake of simplicity, functions $f : \mathbb{R}^d \rightarrow \mathbb{R}^k$ which are $\mathcal{B}(\mathbb{R}^d) - \mathcal{B}(\mathbb{R}^k)$ -measurable are just called Borel-measurable.

Example 1.4.2. (a) Any constant function from (Ω, \mathcal{F}) to (E, \mathcal{E}) is $\mathcal{F} - \mathcal{E}$ -measurable.

(b) For $A \subset \Omega$ the indicator function of A is defined as

$$\begin{aligned} \mathbb{1}_A : \Omega &\rightarrow \{0, 1\} \\ \omega &\mapsto \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases} \end{aligned}$$

Show that if (Ω, \mathcal{F}) is a measurable space, then for $A \subset 2^\Omega$ the indicator function $\mathbb{1}_A$ is $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable if and only if $A \in \mathcal{F}$.

(c) Let Ω be a non-empty set and let \mathcal{F}, \mathcal{G} be two σ -algebras on Ω with $\mathcal{F} \subsetneq \mathcal{G}$. The identity function id_Ω from Ω to Ω , where $\text{id}_\Omega(\omega) := \omega$, is $\mathcal{G} - \mathcal{F}$ -measurable. However, it is not $\mathcal{F} - \mathcal{G}$ -measurable.

(d) If either $\mathcal{F} = 2^\Omega$ or $\mathcal{E} = \{E, \emptyset\}$, then any function from (Ω, \mathcal{F}) to (E, \mathcal{E}) is $\mathcal{F} - \mathcal{E}$ -measurable.

Remark 1.4.3. Let (Ω, \mathcal{F}) and (E, \mathcal{E}) be measurable spaces. Let furthermore $f : \Omega \rightarrow E$ be an $\mathcal{F} - \mathcal{E}$ -measurable function. Recall the definition of the trace- σ -algebra \mathcal{E}_G from Exercise 1.1.19 for $G \subset E$, and assume that $f(\Omega) \subset G$. Then, if $G \in \mathcal{E}$, the function f can also be interpreted as a $\mathcal{F} - \mathcal{E}_G$ -measurable function from Ω to G . This is not necessarily true if $G \notin \mathcal{E}$.

To us, random variables will be important to describe the outcomes of experiments that we consider to be random (prime examples being dice rolls or coin tosses). For more intuition on random variables we refer to [Dre18], in particular Section 1.7. Note, however, that in [Dre18] essentially all functions which occurred were measurable by definition anyway since we only investigated nice and simple settings. In general, the measurability has to be established when investigating arbitrary functions between measurable spaces. We will provide a useful tool for this in Theorem 1.4.7 below, but before we will establish some further general properties for measurable functions.

Theorem 1.4.4 (Compositions of measurable functions). Let $(\Omega_i, \mathcal{F}_i)$, $i \in \{1, 2, 3\}$, be measurable spaces, and let $f_i : \Omega_i \rightarrow \Omega_{i+1}$ be $\mathcal{F}_i - \mathcal{F}_{i+1}$ -measurable maps, for $i \in \{1, 2\}$.

Then the composition

$$\begin{aligned} f_2 \circ f_1 : \Omega_1 &\rightarrow \Omega_3 \\ \omega_1 &\mapsto f_2(f_1(\omega_1)) \end{aligned}$$

is a $\mathcal{F}_1 - \mathcal{F}_3$ -measurable map from Ω_1 to Ω_3 .

Proof. For arbitrary $F_3 \in \mathcal{F}_3$ we have

$$(f_2 \circ f_1)^{-1}(F_3) = f_1^{-1}(f_2^{-1}(F_3)) \in \mathcal{F}_1,$$

where the latter takes advantage of the fact that

$$f_2^{-1}(F_3) \in \mathcal{F}_2$$

since f_2 is $\mathcal{F}_2 - \mathcal{F}_3$ -measurable. This concludes the proof. \square

Since σ -algebras are often large, it is practically not feasible to check the measurability condition of Definition 1.4.1. As a remedy, Theorem 1.4.7 below states that it is sufficient to check it on a generator, which is often easier to achieve. We are now going to prepare its proof via a couple of auxiliary results.

Claim 1.4.5 ('Operationstreue'). *Given an arbitrary map f from a non-empty space X to a non-empty space Y , consider the preimage map*

$$\begin{aligned} f^{-1} : 2^Y &\rightarrow 2^X \\ A &\mapsto f^{-1}(A). \end{aligned}$$

Then the following properties hold:

(a) *For an arbitrary family (B_λ) , $\lambda \in \Lambda$, with $B_\lambda \in 2^Y$ for all $\lambda \in \Lambda$,*

$$f^{-1}\left(\bigcup_{\lambda \in \Lambda} B_\lambda\right) = \bigcup_{\lambda \in \Lambda} f^{-1}(B_\lambda),$$

and

$$f^{-1}\left(\bigcap_{\lambda \in \Lambda} B_\lambda\right) = \bigcap_{\lambda \in \Lambda} f^{-1}(B_\lambda);$$

(b) *for each $B \in 2^Y$,*

$$f^{-1}(B^c) = (f^{-1}(B))^c.$$

Proof. Exercise. \square

The previous claim helps in deriving the following lemma.

Lemma 1.4.6. *Let $f : \Omega \rightarrow E$ be an arbitrary mapping, and let \mathcal{H} be an arbitrary subset of 2^E . Then*

$$\sigma(f^{-1}(\mathcal{H})) = f^{-1}(\sigma(\mathcal{H})).$$

In particular, if \mathcal{H} is a σ -algebra over E , then $f^{-1}(\mathcal{H})$ is a σ -algebra over Ω .

Proof. We have

$$f^{-1}(\mathcal{H}) \subset f^{-1}(\sigma(\mathcal{H})),$$

and using Claim 1.4.5, we infer that the right-hand side of the previous display is a σ -algebra, whence

$$\sigma(f^{-1}(\mathcal{H})) \subset f^{-1}(\sigma(\mathcal{H})).$$

We now prove the converse inclusion using the good sets principle. Denote by \mathcal{G} those subsets of E the preimages of which under f which are contained in $\sigma(f^{-1}(\mathcal{H}))$:

$$\mathcal{G} := \{G \subset E : f^{-1}(G) \in \sigma(f^{-1}(\mathcal{H}))\}. \quad (1.4.1)$$

Then we have $E \in \mathcal{G}$, and furthermore, using Claim 1.4.5 we deduce that \mathcal{G} is stable under countable unions and complements. Hence, \mathcal{G} is a σ -algebra. In addition, as a consequence of (1.4.1) we get $\mathcal{H} \subset \mathcal{G}$. Thus, in particular we deduce that

$$f^{-1}(\sigma(\mathcal{H})) \subset f^{-1}(\mathcal{G}) \subset \sigma(f^{-1}(\mathcal{H})),$$

which finishes the proof. \square

We are now ready to prove the tool announced above, which shows that in order to prove measurability of a map, it is sufficient to consider the generator of the corresponding σ -algebra in the image space.

Theorem 1.4.7. *Let f be a mapping from the measurable space (Ω, \mathcal{F}) to the measurable space (E, \mathcal{E}) . Furthermore, let $\mathcal{G} \subset 2^E$ be any generator of \mathcal{E} . Then f is \mathcal{F} – \mathcal{E} -measurable if and only if $f^{-1}(\mathcal{G}) \subset \mathcal{F}$.*

Proof. If f is \mathcal{F} – \mathcal{E} -measurable, then since $\mathcal{G} \subset \mathcal{E}$ we deduce from the definition that $f^{-1}(\mathcal{G}) \subset \mathcal{F}$. To prove the converse inclusion, assume that $f^{-1}(\mathcal{G}) \subset \mathcal{F}$. Then

$$\sigma(f^{-1}(\mathcal{G})) \subset \mathcal{F},$$

and the left-hand side of this display coincides with $f^{-1}(\sigma(\mathcal{G})) = f^{-1}(\mathcal{E})$ due to Lemma 1.4.6, which finishes the proof. \square

We will now derive a corollary of the previous result for which we introduce the following notation.

Definition 1.4.8. *Let Λ be a non-empty set, and let X_λ , $\lambda \in \Lambda$ be mappings from Ω to sets E_λ . Furthermore, let \mathcal{E}_λ be σ -algebras on E_λ , $\lambda \in \Lambda$. We denote by $\sigma(X_\lambda : \lambda \in \Lambda)$ the smallest σ -algebra on Ω such that each X_λ is $\sigma(X_\lambda : \lambda \in \Lambda)$ – \mathcal{E}_λ -measurable. $\sigma(X_\lambda : \lambda \in \Lambda)$ is also called the σ -algebra generated by the X_λ , $\lambda \in \Lambda$.*

Corollary 1.4.9. *Let $\Lambda \neq \emptyset$ and measurable spaces (Ω, \mathcal{F}) , $(\tilde{\Omega}, \tilde{\mathcal{F}})$, as well as $(\Omega_\lambda, \mathcal{F}_\lambda)$, $\lambda \in \Lambda$, be given. Furthermore, assume maps $Y_\lambda : \tilde{\Omega} \rightarrow \Omega_\lambda$, $\lambda \in \Lambda$, to be given such that $\tilde{\mathcal{F}} = \sigma(Y_\lambda : \lambda \in \Lambda)$.*

Then a map $X : \Omega \rightarrow \tilde{\Omega}$ is \mathcal{F} – $\tilde{\mathcal{F}}$ -measurable if and only if the compositions $Y_\lambda \circ X : \Omega \rightarrow \Omega_\lambda$ are \mathcal{F} – \mathcal{F}_λ -measurable for all $\lambda \in \Lambda$.

Proof. If X is measurable, then all the compositions are measurable due to Theorem 1.4.4. If, on the other hand, all the $Y_\lambda \circ X$ are \mathcal{F} – \mathcal{F}_λ -measurable, then we start with observing that by definition of $\tilde{\mathcal{F}}$,

$$\tilde{\mathcal{G}} := \{Y_\lambda^{-1}(F) : \lambda \in \Lambda, F \in \mathcal{F}_\lambda\}$$

is a generator of $\tilde{\mathcal{F}}$. But since all Y_λ are measurable by assumption, we have $X^{-1}(\tilde{\mathcal{G}}) \subset \mathcal{F}$, and hence Theorem 1.4.7 implies that X has the desired measurability properties. \square

The following result is interesting in its own right, but it will also play an important role in proving the central Proposition 1.4.13 below. We recall the facts we had learned about topologies in Definition 1.1.28 and below. Furthermore, we remind ourselves that a function from a topological space (O_1, \mathcal{O}_1) to a topological space (O_2, \mathcal{O}_2) was defined to be continuous if and only if $f^{-1}(O) \in \mathcal{O}_1$ for all $O \in \mathcal{O}_2$ (this definition coincided with the definition for continuity in the case of metric spaces or \mathbb{R}^d).

Theorem 1.4.10. *Let (O_1, \mathcal{O}_1) and (O_2, \mathcal{O}_2) be topological spaces. Then any continuous map from (O_1, \mathcal{O}_1) to (O_2, \mathcal{O}_2) is $\mathcal{B}(O_1)$ – $\mathcal{B}(O_2)$ -measurable.*

Proof. Since f is continuous, by definition $f^{-1}(\mathcal{O}_2) \subset \mathcal{O}_1$, so Theorem 1.4.7 implies $f^{-1}(\mathcal{B}(\mathcal{O}_2)) \subset \mathcal{B}(\mathcal{O}_1)$, which finishes the proof. \square

Proposition 1.4.11. *Let $f : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. For $i \in \{1, \dots, d\}$ denote by*

$$\begin{aligned}\pi_i : \mathbb{R}^d &\rightarrow \mathbb{R}, \\ (x_1, \dots, x_d) &\mapsto x_i,\end{aligned}$$

the projection on the i -th coordinate.

Then the function the function f is $\mathcal{F} - \mathcal{B}(\mathbb{R}^d)$ -measurable if and only if $\pi_i \circ f$ is $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable for all admissible choices of i .

Proof. Once we show $\mathcal{B}(\mathbb{R}^d) = \sigma(\pi_i, 1 \leq i \leq d)$, the result follows from Corollary 1.4.9. Since the $\pi_i : \mathbb{R}^d \rightarrow \mathbb{R}$ are continuous, Theorem 1.4.10 implies their $\mathcal{B}(\mathbb{R}^d) - \mathcal{B}(\mathbb{R})$ -measurability, and we infer that $\sigma(\pi_i, 1 \leq i \leq d) \subset \mathcal{B}(\mathbb{R}^d)$. On the other hand, due to $[a, b] = \cap_{i=1}^d \pi_i^{-1}([a_i, b_i])$, $a \leq b \in \mathbb{R}^d$ and the fact that such hyperrectangles generate $\mathcal{B}(\mathbb{R}^d)$ due to Lemma 1.1.31, we infer $\mathcal{B}(\mathbb{R}^d) \subset \sigma(\pi_i, 1 \leq i \leq d)$, which finishes the proof. \square

Example 1.4.12. (a) *Let $f : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be a $\mathcal{F} - \mathcal{B}(\mathbb{R}^d)$ -measurable function. Then the function*

$$\begin{aligned}\|f\|_2 : (\Omega, \mathcal{F}) &\rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})), \\ \omega &\mapsto \|f(\omega)\|_2,\end{aligned}$$

is $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable, where for $x \in \mathbb{R}^n$ we write $\|x\|_2 := \sqrt{\sum_{i=1}^n x_i^2}$ for the so-called 2-norm on \mathbb{R}^d . In fact, as a composition of continuous functions, the function $\mathbb{R}^d \ni x \mapsto \|x\|_2 \in [0, \infty)$ is continuous again, and hence $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable due to Theorem and 1.4.10. Thus, due to Theorem 1.4.4 the function f is $\mathcal{F} - \mathcal{B}(\mathbb{R}^d)$ -measurable.

(b) *Let $f, g : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable functions. Then the functions*

$$f \vee g, \quad f \wedge g, \quad f^+, f^-, \quad \text{and } |f|$$

are $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable. Here, we use the standard notation that for $x, y \in \mathbb{R}$

$$x \vee y := \max(x, y), \quad x \wedge y := \min(x, y), \quad \text{as well as } x^+ := x \vee 0 \quad \text{and } x^- := -(x \wedge 0) \geq 0.$$

In the following result we will summarize a couple of important compositions of functions which supply us with measurable functions again.

Proposition 1.4.13. *Let f be a $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable function from Ω to \mathbb{R} , and let $g, h : \Omega \rightarrow \mathbb{R}^d$ be $\mathcal{F} - \mathcal{B}(\mathbb{R}^d)$ -measurable functions.*

Then $g + h, g - h, f \cdot g, g/f$ are also measurable functions.

Remark 1.4.14. *Here we must pay attention to how define g/f in the case that f (or f and g) are 0. For the setting of this result and the following proof, we set $x/0 := 0$ (which might seem awkward in the case $x \neq 0$, but more natural for $x = 0$; either way, scrutinizing the proof below we will see that any convention $x/0 := c$, some $c \in \mathbb{R}$, would work and still leave us with g/f measurable).*

Proof. For the case $g + h$, we observe that the function $\Omega \ni \omega \mapsto (g(\omega), h(\omega)) \in (\mathbb{R}^d \times \mathbb{R}^d)$ is a $\mathcal{F} - \mathcal{B}(\mathbb{R}^{2d})$ -measurable function due to Proposition 1.4.11. Furthermore, the function $\mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto x + y \in \mathbb{R}^d$ is continuous, so in combination with Theorem 1.4.4 we infer that the composition $g + h : \Omega \ni \omega \mapsto g(\omega) + h(\omega) \in \mathbb{R}^d$ is $\mathcal{F} - \mathcal{B}(\mathbb{R}^d)$ -measurable.

The remaining cases follows in a similar fashion, and the only point where we do have to pay some extra attention is g/f in the case that the denominator vanishes. Since we can write $g/f = g \cdot \frac{1}{f}$, and we know that f, g are measurable and that the product of two measurable functions is measurable again, due to Theorem 1.4.4 the only thing that remains to show is that $i : \mathbb{R} \ni x \mapsto 1/x \in \mathbb{R}$ is $\mathcal{B}(\mathbb{R}) - \mathcal{B}(\mathbb{R})$ -measurable. For this purpose, observe that for $U \subset \mathbb{R}$ open we have

$$i^{-1}(U) = \underbrace{i^{-1}\left(\underbrace{U \setminus \{0\}}_{\substack{\subset \mathbb{R} \setminus \{0\} \text{ open} \\ \subset \mathbb{R} \setminus \{0\} \text{ open, since } i \text{ continuous on } \mathbb{R} \setminus \{0\}}}\right)}_{\substack{\subset \mathbb{R} \setminus \{0\} \text{ open, since } i \text{ continuous on } \mathbb{R} \setminus \{0\}}} \cup (U \cap \{0\}) \in \mathcal{B}(\mathbb{R}).$$

Hence, the result follows in combination with Theorem 1.4.7. □

Proposition 1.4.15. *Let (f_n) be a sequence of functions in \mathcal{M} . Then the functions*

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

are all in \mathcal{M} again.

Proof. For any $a \in \mathbb{R}$ we have

$$\left\{ \sup_{n \in \mathbb{N}} f_n \in (a, \infty] \right\} = \bigcup_{n \in \mathbb{N}} f_n^{-1}((a, \infty])$$

Thus, using Lemma 1.1.31 and Theorem 1.4.7 yields that $\sup_{n \in \mathbb{N}} f_n$ has the desired measurability property. The case of $\inf_{n \in \mathbb{N}} f_n$ can be shown similarly or otherwise by using the identity $\inf_{n \in \mathbb{N}} f_n = -\sup_{n \in \mathbb{N}} (-f_n)$ and invoking Proposition 1.4.13 twice.

With regard to $\limsup_{n \rightarrow \infty} f_n$, we observe that

$$\limsup_{n \rightarrow \infty} f_n = \inf_{n \in \mathbb{N}} \sup_{m \geq n} f_m,$$

and use the first part of this Proposition to first conclude that $\sup_{m \geq n} f_m$ has the desired measurability properties, and then to conclude that the same holds true for $\inf_{n \in \mathbb{N}} \sup_{m \geq n} f_m$, which finishes this part.

Again, for the case $\liminf_{n \rightarrow \infty} f_n$ we can follow one of the two alternative routes outlined in the proof of the first part. □

Corollary 1.4.16. *Let (f_n) be a sequence of functions in \mathcal{M} such that*

$$f(\omega) := \lim_{n \rightarrow \infty} f_n(\omega)$$

exists for all $\omega \in \Omega$.

Then $f \in \mathcal{M}$.

Another result that comes in handy is the following.

Lemma 1.4.17 (factorization lemma ('Maßtheoretischer Dreisatz')). *Let (Ω, \mathcal{F}) and (Ω', \mathcal{F}') be two measurable spaces and let $T : \Omega \rightarrow \Omega'$ and $f : \Omega \rightarrow \mathbb{R}$ be two mappings. Then f is $T^{-1}(\mathcal{F}') - \mathcal{B}$ -measurable if and only if there exists a $\mathcal{F}' - \mathcal{B}$ -measurable mapping $\varphi : \Omega' \rightarrow \mathbb{R}$ such that $f = \varphi \circ T$.*

Proof. Exercise. □

1.5 Image measures, distributions

As introduced in [Dre18, Definition 1.8.12], for real-valued random variables the concept of its (cumulative) distribution function plays a prominent role.

Definition 1.5.1. Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and mapping to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then the function

$$F_X : \mathbb{R}^d \rightarrow [0, 1],$$

$$t \mapsto \mathbb{P}(X \leq t) = \mathbb{P}\left(X \in \bigtimes_{i=1}^d (-\infty, t_i]\right),$$

is called the (cumulative) distribution function (or cdf) of X ('Verteilungsfunktion von X '). Similarly, if μ is a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, then the function

$$F_\mu : \mathbb{R}^d \rightarrow [0, 1],$$

$$t \mapsto \mu\left(\bigtimes_{i=1}^d (-\infty, t_i]\right),$$

is called the distribution function of μ ('Verteilungsfunktion von μ ').

Of particular importance to us will be the case $d = 1$, and in [Dre18, Thm. 1.8.16] we had found the following properties of distribution functions.

Theorem 1.5.2. If F is the distribution function of a real random variable or of a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then

(a) F is non-decreasing;

(b)

$$\lim_{t \rightarrow -\infty} F(t) = 0, \quad \lim_{t \rightarrow \infty} F(t) = 1;$$

(c) F is right-continuous (i.e., for all $t_0 \in \mathbb{R}$ one has $F(t_0) = \lim_{t \downarrow t_0} F(t)$);

We had also introduced the following definition and theorem as [Dre18, Def. 1.8.17, Thm. 1.8.18].

Definition 1.5.3. Any function $F : \mathbb{R} \rightarrow [0, 1]$ that satisfies the three properties given in Theorem 1.5.2 is called a distribution function ('Verteilungsfunktion').

The following result complements Theorem 1.5.2, and combined they establish that there is a correspondence between random variables and distribution functions.

Theorem 1.5.4. If F is any distribution function, then there exists a unique probability measure μ_F on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $F_{\mu_F} = F$.

However, at that point we were not able to prove this result. Since we are now in the position to do so, we give the proof.

Proof of Thm. 1.5.4. See homework problems. □

Observing that the distribution function F_X of a real random variable X depends only on the probability measure $\mathbb{P}(X \in \cdot)$, and combining Theorems 1.5.2 and 1.5.4, we deduce that there exists a one-to-one correspondence between probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and distribution functions $F : \mathbb{R} \rightarrow [0, 1]$.

Corollary 1.5.5. *The mapping induced by Definition 1.5.1, which maps a probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to its distribution function F_μ defines a bijection from the space of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to the space of distributions functions on \mathbb{R} .*

Theorem 1.5.6. *Let $(\Omega, \mathcal{F}, \nu)$ be a measure space and let (E, \mathcal{E}) be a measurable space. Furthermore, assume a $\mathcal{F} - \mathcal{E}$ -measurable mapping f to be given. Then the set function*

$$\begin{aligned}\nu \circ f^{-1} : \mathcal{E} &\rightarrow [0, \infty], \\ F &\mapsto \nu(f^{-1}(F)),\end{aligned}$$

defines a measure on (E, \mathcal{E}) . We will also sometimes denote it by ν_f for simplicity of notation. If ν is a probability measure (so f is a random variable), ν_f is also called the distribution of f .

Proof. The proof proceeds in the same way as the one of [Dre18, Theorem 1.7.6], which is for probability measures. □

Definition 1.5.7. *The measure $\nu \circ f^{-1}$ introduced in Theorem 1.5.6 is called the image measure (or pushforward) of ν under f . In the case of f being a random variable (i.e., ν is a probability measure), the image measure $\nu \circ f^{-1}$ is called the distribution (or law) of f .*

As we will see, in probability theory, oftentimes a random variable X itself will not be of too much importance to us; rather, what we will be interested in usually is its law.

Example 1.5.8. *Arguably the most prominent example of a distribution is the so-called Normal distribution or Gaussian distribution ('Normalverteilung', 'Gaußverteilung'). We say that for $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$, a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is $\mathcal{N}(\mu, \sigma^2)$ distributed if*

$$\mathbb{P}(X \leq s) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^s e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx, \quad \forall s \in \mathbb{R}.$$

This really defines a probability measure, i.e.,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1. \quad (1.5.1)$$

Indeed, we have

$$\left(\int_{-\infty}^{\infty} e^{-x^2/2} dx \right)^2 = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{-x^2/2} e^{-y^2/2} dx \right) dy = \int_0^{\infty} 2\pi r e^{-r^2/2} dr = -2\pi e^{-r^2/2} \Big|_{r=0}^{\infty} = 2\pi,$$

where we took advantage of Polar coordinates in standard Riemann integration in the second equality. In particular, (1.5.1) follows.

In higher dimensions $d \geq 2$ we can still define the normal distribution and say that $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is $\mathcal{N}(\mu, \Sigma)$ -distributed, where $\mu \in \mathbb{R}^d$ and Σ is a symmetric positive definite matrix in $\mathbb{R}^{d \times d}$, if

$$\mathbb{P}(X \leq s) = \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|\det(\Sigma)|}} \int_{-\infty}^{s_1} \dots \int_{-\infty}^{s_d} e^{-\frac{1}{2}(x-\mu)^t \cdot (\Sigma^{-1}(x-\mu))} dx_1 \dots dx_d, \quad \forall s \in \mathbb{R}^d,$$

i.e., its density with respect to λ^d is given by

$$\frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{|\det(\Sigma)|}} e^{-\frac{1}{2}(x-\mu)^t \cdot (\Sigma^{-1}(x-\mu))}.$$

Chapter 2

The Lebesgue integral

We recall here that the Riemann integral had been constructed by directly partitioning the domain of the integrand into finer and finer partitions (see Section 2.0.3 below for a short reminder). As already outlined at the beginning of Section 1.4, the Lebesgue integral will essentially be constructed by first partitioning the image of the integrand (into finer and finer partitions), and then use these partitions of the image in order to obtain a partition of the domain and hence define the integral.

This procedure is most easily performed for simple functions as follows.

2.0.1 Integrals of simple functions

Definition 2.0.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f : (\Omega, \mathcal{F}, \mu) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable function such that $f(\Omega)$ is a finite set. Then f is called a simple function ('einfache Funktion', 'Treppenfunktion'). The set (in fact a vector space, see Lemma 2.0.4 below) of all simple functions will be denoted by \mathcal{T} , and by \mathcal{T}^+ we denote the set of all non-negative simple functions.

Lemma 2.0.2. A function $f : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is simple if and only if there exist pairwise disjoint sets $F_1, \dots, F_n \in \mathcal{F}$ and numbers $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ such that

$$f = \sum_{i=1}^n \alpha_i \mathbb{1}_{F_i}. \quad (2.0.1)$$

Remark 2.0.3. As in the literature, we will call a representation of f as in Lemma 2.0.2 a normal representation 'Normaldarstellung' of f ; note that some authors also demand that $\bigcup_{i=1}^n F_i = \Omega$ for (2.0.1) to be called a normal representation. In fact, if $\bigcup_{i=1}^n F_i \subsetneq \Omega$, then $\sum_{i=1}^{n+1} \alpha_i \mathbb{1}_{F_i}$ with $\alpha_{n+1} = 0$ and $F_{n+1} = \Omega \setminus \bigcup_{i=1}^n F_i$ is a normal representation with $\bigcup_{i=1}^{n+1} F_i = \Omega$.

Proof. If f is of the form (2.0.1), then obviously $f \in \mathcal{T}$. On the other hand, if $f \in \mathcal{T}$, then $f(\Omega) = \{\alpha_1, \dots, \alpha_n\} \subset \mathbb{R}$, and we have

$$f = \sum_{i=1}^n \alpha_i \mathbb{1}_{f^{-1}(\{\alpha_i\})},$$

so f is of the form (2.0.1) with pairwise disjoint F_i , since $f^{-1}(\{\alpha_i\}) \in \mathcal{F}$ due to the $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurability of f . \square

Lemma 2.0.4. The product of two simple functions is simple again, and so is the linear combination of finitely many simple functions. In particular, \mathcal{T} is a vector space.

Proof. Exercise. \square

From now on onwards we will use the convention that $0 \cdot \infty = 0$.

Lemma 2.0.5. *For a measure space $(\Omega, \mathcal{F}, \mu)$, let $f \in \mathcal{T}^+$ be a simple function with normal representations*

$$f = \sum_{i=1}^n \alpha_i \mathbb{1}_{F_i} = \sum_{j=1}^m \beta_j \mathbb{1}_{G_j},$$

for $m, n \in \mathbb{N}$, $\alpha_i, \beta_j \in \mathbb{R}$, and $F_i, G_j \in \mathcal{F}$. Then

$$\sum_{i=1}^n \alpha_i \mu(F_i) = \sum_{j=1}^m \beta_j \mu(G_j).$$

Proof. For F_i, G_j with $F_i \cap G_j \neq \emptyset$ we get that $f(\omega) = \alpha_i = \beta_j$ for $\omega \in F_i \cap G_j$, and as a consequence

$$\sum_{i=1}^n \alpha_i \mu(F_i) = \sum_{i=1}^n \sum_{j=1}^m \alpha_i \mu(F_i \cap G_j) = \sum_{j=1}^m \sum_{i=1}^n \beta_j \mu(F_i \cap G_j) = \sum_{j=1}^m \beta_j \mu(G_j),$$

where in the first and third equality we took advantage of the facts that if $\alpha_i \neq 0$, then $F_i \subset \bigcup_{j=1}^m G_j$, and if $\beta_j \neq 0$, then $G_j \subset \bigcup_{i=1}^n F_i$. \square

Definition 2.0.6. *Let $f \in \mathcal{T}^+$ with normal representation $f = \sum_{i=1}^n \alpha_i \mathbb{1}_{F_i}$ with $\alpha_i \geq 0$ and $F_i \in \mathcal{F}$ for all $i \in \{1, \dots, n\}$. Then the (μ) -integral of f is defined as*

$$\int_{\Omega} f \, d\mu := \int_{\Omega} f(\omega) \, \mu(d\omega) := \sum_{i=1}^n \alpha_i \mu(F_i) \in [0, \infty].$$

Remark 2.0.7. *Note that due to Lemma 2.0.5, the previous Definition 2.0.6 is well-defined.*

We collect some important properties of integrals of non-negative simple functions

Lemma 2.0.8. (a) *Let $F \in \mathcal{F}$. Then*

$$\int_{\Omega} \mathbb{1}_F \, d\mu = \mu(F).$$

(b) *For $f, g \in \mathcal{T}^+$ and $c \geq 0$,*

$$\int_{\Omega} (cf + g) \, d\mu = c \int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu \quad (\text{linearity}). \quad (2.0.2)$$

(c) *For $f, g \in \mathcal{T}^+$ with $f \leq g$*

we have

$$\int_{\Omega} f \, d\mu \leq \int_{\Omega} g \, d\mu.$$

Proof. (a) A normal representation of f is given by $1 \cdot \mathbb{1}_F$, the integral of which is $\mu(F)$ by definition.

(b) We start with noting that due to Lemma 2.0.4, $cf + g$ is a simple function again, and since it is non-negative, we have $cf + g \in \mathcal{T}^+$ and can therefore consider its integral.

Furthermore, if $\sum_{i=1}^n \alpha_i \mathbb{1}_{F_i}$ and $\sum_{j=1}^m \beta_j \mathbb{1}_{G_j}$ are normal representations of f and g , respectively, then $\sum_{k=1}^{m \cdot n} \gamma_k \mathbb{1}_{H_k}$ is a normal representation for $cf + g$, where

$$\gamma_k = c\alpha_i + \beta_j$$

and

$$H_k = F_i \cap G_j$$

with $k = (i-1)m + j \in \{1, \dots, mn\}$, $i \in \{1, \dots, n\}$, $j \in \{1, \dots, m\}$. (2.0.2) then follows immediately from the definition of the integral.

- (c) This follows from the linearity and non-negativity of the integral in combination with the fact that $g, g - f \in \mathcal{T}^+$. □

Exercise 2.0.9. Show that if $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{F_i} \in \mathcal{T}^+$ with $F_i \in \mathcal{F}$ and $\alpha_i \geq 0$ for all $1 \leq i \leq n$ is not necessarily a normal representation, one still has

$$\int_{\Omega} f \, d\mu = \sum_{i=1}^n \alpha_i \mu(F_i).$$

2.0.2 Lebesgue integral for measurable functions

As is often done in mathematics, we are now to introduce the integral of suitable measurable functions by reducing it to something simpler, i.e., to the integral of simple functions. In order to pull through this procedure, the following approximation result is fundamental.

Lemma 2.0.10. Let (Ω, \mathcal{F}) be a measurable space. Then $f : (\Omega, \mathcal{F}) \rightarrow ([0, \infty], \mathcal{B}([0, \infty]))$ is $\mathcal{F} - \mathcal{B}([0, \infty])$ -measurable if and only if there exists a non-decreasing sequence (f_n) of functions $f_n \in \mathcal{T}^+$ such that

$$f_n \rightarrow f \quad \text{as } n \rightarrow \infty.$$

Proof. Proposition 1.4.15 immediately supplies us with the fact that if (f_n) is a sequence of functions as in the assumptions, $\lim_{n \rightarrow \infty} f_n = \sup_{n \in \mathbb{N}} f_n \in \mathcal{M}^+$.

On the other hand, assume $f \in \mathcal{M}^+$ to be given. For $n \in \mathbb{N}$ and $i \in \{0, 1, \dots, n2^n - 1\}$ we define the function

$$f_n := \sum_{i=0}^{n2^n} i2^{-n} \mathbf{1}_{A_{n,i}},$$

where for $n \in \mathbb{N}$ and $i \in \{0, 1, \dots, n2^n - 1\}$ we set $A_{n,i} := f^{-1}([i2^{-n}, (i+1)2^{-n})) \in \mathcal{F}$, as well as $A_{n,n2^n} := f^{-1}([n, \infty])$. It is apparent from the definition that $f_n \in \mathcal{T}^+$, that f_n is non-decreasing, and that $f_n \rightarrow f$ pointwise, which finishes the proof. □

The obvious procedure would now be to define the integral of a non-negative measurable (extended) real-valued function f as the monotone limit of a sequence of integrals of simple functions (f_n) approximating it monotonically. I.e.,

$$\int_{\Omega} f \, d\mu := \lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu,$$

where one can invoke Lemma 2.0.8 to ensure that the integrals on the right-hand side are non-decreasing in n , and hence the limit exists. However, we have to make sure that the limit of those integrals does not depend on the very choice of the approximating sequence of simple functions. For this purpose, we prove the following lemma.

Lemma 2.0.11. Let $(f_n), (g_n)$ be two non-decreasing sequences of functions in \mathcal{T}^+ such that $\lim_{n \rightarrow \infty} f_n = \lim_{n \rightarrow \infty} g_n$. Then

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} g_n \, d\mu. \quad (2.0.3)$$

For the proof of Lemma 2.0.11 we take advantage of the following claim.

Claim 2.0.12. *Let $f \in \mathcal{T}^+$ and let (f_n) be an increasing sequence with $f_n \in \mathcal{T}^+$ for all $n \in \mathbb{N}$ such that*

$$f \leq \lim_{n \rightarrow \infty} f_n. \quad (2.0.4)$$

Then

$$\int_{\Omega} f \, d\mu \leq \lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu.$$

Proof. Write $f = \sum_{i=1}^m \alpha_i \mathbb{1}_{F_i}$ for a normal representation of f . Then for $\varepsilon \in (0, 1)$ we consider the set

$$M_n := \{f_n \geq (1 - \varepsilon)f\} \in \mathcal{F}.$$

From (2.0.4) we deduce that $M_n \uparrow \Omega$, hence the continuity of the measure μ from below implies

$$\begin{aligned} \int_{\Omega} f \, d\mu &= \sum_{i=1}^m \alpha_i \mu(F_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^m \alpha_i \mu(F_i \cap M_n) = \lim_{n \rightarrow \infty} \int_{\Omega} f \cdot \mathbb{1}_{M_n} \, d\mu \\ &\leq \lim_{n \rightarrow \infty} \int_{\Omega} \frac{1}{1 - \varepsilon} f_n \, d\mu = \lim_{n \rightarrow \infty} \frac{1}{1 - \varepsilon} \int_{\Omega} f_n \, d\mu. \end{aligned}$$

Since $\varepsilon \in (0, 1)$ was chosen arbitrarily, taking $\varepsilon \downarrow 0$ it follows that $\int_{\Omega} f \, d\mu \leq \lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu$, which finishes the proof. \square

Proof of Lemma 2.0.11. From Claim 2.0.12 we deduce that for all $m \in \mathbb{N}$,

$$\int_{\Omega} g_m \, d\mu \leq \lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu.$$

Hence, taking $m \rightarrow \infty$ we infer

$$\lim_{m \rightarrow \infty} \int_{\Omega} g_m \, d\mu \leq \lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu.$$

Exchanging the roles of f_n and g_m we therefore obtain (2.0.3). \square

We can now introduce the integral for measurable extended real-valued functions, which by Lemma 2.0.11 is well-defined.

Definition 2.0.13. *Let $f \in \mathcal{M}^+$ and let (f_n) be any sequence as in Lemma 2.0.10. Then*

$$\int_{\Omega} f \, d\mu := \int_{\Omega} f(\omega) \, \mu(d\omega) := \lim_{n \rightarrow \infty} \int_{\Omega} f_n \, d\mu$$

is called the (μ) -integral of f .

As in the case of integrals of non-negative simple functions we derive the following basic properties.

Lemma 2.0.14. (a) *For $f, g \in \mathcal{M}^+$ and $c \in [0, \infty]$,*

$$\int_{\Omega} (cf + g) \, d\mu = c \int_{\Omega} f \, d\mu + \int_{\Omega} g \, d\mu \quad (\text{linearity}).$$

(b) *For $f, g \in \mathcal{M}^+$ with $f \leq g$*

we have

$$\int_{\Omega} f \, d\mu \leq \int_{\Omega} g \, d\mu.$$

The lemma can be proven taking advantage of its validity in the case of integrands in \mathcal{T}^+ (see Lemma 2.0.8) and then decomposing into positive and negative parts as well as taking limits. We omit the details.

Corollary 2.0.15. *An alternative way (which sometimes comes handy) to reduce the integral of f as in Definition 2.0.13 to integrals of non-negative simple functions is via*

$$\int_{\Omega} f \, d\mu = \sup_{\substack{g \in \mathcal{T}^+ \\ g \leq f}} \int_{\Omega} g \, d\mu.$$

Proof. Exercise. □

Having introduced the integral for non-negative measurable functions, we would like to extend it to a suitable class of measurable functions that can take positive and negative values at once. It will turn out, however, that we do not only need the notion of Lebesgue integral for extended real-valued functions, but also for complex valued functions. Therefore, for a complex number $z = x + yi \in \mathbb{C}$ with $x, y \in \mathbb{R}$, we denote its real part x by $\operatorname{Re}(z)$ and its imaginary part y by $\operatorname{Im}(z)$, so $z = \operatorname{Re}(z) + \operatorname{Im}(z) \cdot i$.

Exercise 2.0.16. *Show that a function $f : (\Omega, \mathcal{F}, \mu) \rightarrow (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ is $\mathcal{F} - \mathcal{B}(\mathbb{C})$ -measurable if and only if the functions $\operatorname{Re} f$ and $\operatorname{Im} f$ are $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable.*

Proof. We note that the functions $\mathbb{C} \ni z \mapsto \operatorname{Re}(z) \in \mathbb{R}$ and $\mathbb{C} \ni z \mapsto \operatorname{Im}(z) \in \mathbb{R}$ are continuous. Therefore, if f is measurable, so are the functions $\operatorname{Re} f$ and $\operatorname{Im} f$ due to Theorems 1.4.10 and 1.4.4.

If, on the other hand, the functions $\operatorname{Re} f$ and $\operatorname{Im} f$ are $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable, then so is the function $i \cdot \operatorname{Im} f$, and hence $f = \operatorname{Re} f + i \cdot \operatorname{Im} f$ is measurable as well due to Proposition 1.4.13 (where we identify \mathbb{C} with \mathbb{R}^2). □

Definition 2.0.17. *Let $f : (\Omega, \mathcal{F}, \mu) \rightarrow (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ be an $\mathcal{F} - \mathcal{B}(\mathbb{C})$ -measurable function. Then f is called Lebesgue-integrable or (μ) -integrable if the integrals*

$$\int_{\Omega} (\operatorname{Re} f)^{\pm} \, d\mu, \quad \int_{\Omega} (\operatorname{Im} f)^{\pm} \, d\mu$$

are all finite. If this is the case, the quantity

$$\int_{\Omega} f \, d\mu := \int_{\Omega} (\operatorname{Re} f)^+ \, d\mu - \int_{\Omega} (\operatorname{Re} f)^- \, d\mu + i \int_{\Omega} (\operatorname{Im} f)^+ \, d\mu - i \int_{\Omega} (\operatorname{Im} f)^- \, d\mu$$

is called the Lebesgue integral of f (or also μ -integral of f).

Furthermore, for $A \in \mathcal{F}$ we introduce the notation

$$\int_A f \, d\mu := \int_{\Omega} \mathbb{1}_A \cdot f \, d\mu, \tag{2.0.5}$$

if the function $\mathbb{1}_A \cdot f$ is Lebesgue-integrable.

Above all, in the case of real-valued integrands it will turn out useful to be slightly less demanding in the above definition.

Definition 2.0.18. *Let $f \in \mathcal{M}$. We define*

$$\int_{\Omega} f \, d\mu := \int_{\Omega} f^+ \, d\mu - \int_{\Omega} f^- \, d\mu$$

as long as one of the two terms on the right-hand side is finite.¹ In this case, we call f quasi-integrable. We write \mathcal{L}^1 or $\mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ for the set $\{f \in \mathcal{M} : \int_{\Omega} f \, d\mu \in \mathbb{R}\}$, the (vector) space of integrable functions.

Similarly, for a random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ we define its expectation

$$\mathbb{E}[X] := \int_{\Omega} X \, d\mathbb{P},$$

whenever X is quasi-integrable with respect to \mathbb{P} .

The Reader may convince herself that the notion of integral introduced in Definition 2.0.18 coincides with that of Definition 2.0.17 for measurable f that takes values in \mathbb{R} only.

Also, the property of linearity as given for non-negative functions in Lemma 2.0.14 directly transfers to complex-valued or $\overline{\mathbb{R}}$ -valued functions if the resulting sum is well-defined in $\overline{\mathbb{R}}$ (i.e., we don't have expressions like $\infty - \infty$ appearing).

2.0.3 Lebesgue vs. Riemann integral

Recall that the Riemann integral has been defined for real-valued functions which are defined on (subsets of) \mathbb{R}^d instead of on more general sets (as is the case for the Lebesgue integral, which essentially can be applied to real-valued functions defined on arbitrary measure spaces). Generalizing Definition 2.0.18 to the case of functions that are defined only on a part of the underlying space, for $f : A \rightarrow \mathbb{R}$ which is $\mathcal{L}(A) - \mathcal{B}(\mathbb{R})$ -measurable, where $A \in \mathcal{B}(\mathbb{R}^d)$ and $\mathcal{L}(A) := \mathcal{L}(\mathbb{R}^d)_A$ the trace σ -algebra, we say that f is *Lebesgue integrable* if

$$\int_A f \, d\lambda^d := \int_{\mathbb{R}^d} \tilde{f} \, d\lambda^d$$

(where $\tilde{f} : \mathbb{R}^d \rightarrow \mathbb{R}$ is defined to coincide with f on A , and as 0 on A^c) is well-defined and finite. We will have a closer look at the case of A being an interval in \mathbb{R} and recall that the Riemann integral had been defined as follows. $f : [a, b] \rightarrow \mathbb{R}$ was called *Riemann integrable* if for any sequence (\mathcal{I}_n) with

$$\mathcal{I}_n : a = t_0^{(n)} < t_1^{(n)} < \dots < t_{m_n}^{(n)} = b$$

such that

$$\max_{1 \leq i \leq m_n} t_i^{(n)} - t_{i-1}^{(n)} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and for any sequence of points

$$\xi_i^{(n)} \in [t_{i-1}^{(n)}, t_i^{(n)}], \quad n \in \mathbb{N}, 1 \leq i \leq m_n,$$

we have that the limit

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{m_n} f(\xi_i^{(n)}) (t_i^{(n)} - t_{i-1}^{(n)})$$

exists, is finite, and is independent of the very choice of (\mathcal{I}_n) and $(\xi_i^{(n)})$. In this case we write $\int_a^b f(x) \, dx$ for the corresponding limit.

Exercise 2.0.19. (a) Show that f as above is Lebesgue integrable if and only if $\int |f| \, d\lambda^d < \infty$.

(b) Show that the function

$$f : [0, 1] \rightarrow \{0, 1\}$$

$$x \mapsto \begin{cases} 1, & \text{if } x \in \mathbb{Q}, \\ 0, & \text{if } x \notin \mathbb{Q}, \end{cases}$$

is Lebesgue integrable but not Riemann integrable.

¹You might sometimes also encounter the notation $\mu(f)$ for the Lebesgue integral $\int_{\Omega} f \, d\mu$.

(c) Show that the function

$$g : [1, \infty) \rightarrow \mathbb{R}$$

$$x \mapsto \frac{\sin(x)}{x}$$

is (improperly ('uneigentlich')) Riemann integrable (i.e., the limit $\lim_{n \rightarrow \infty} \int_1^n g(x) dx$ exists) but not Lebesgue integrable.

Theorem 2.0.20. *If $f : [a, b] \rightarrow \mathbb{R}$ is Riemann integrable, then f is bounded and $(\mathcal{L}([a, b]), \mathcal{B}(\mathbb{R}))$ -measurable. In particular, f is Lebesgue integrable and the Riemann integral of f coincides with the Lebesgue integral of f .*

Since this result is not central to our further exposition, we refer to the exercise classes for a proof.

As a consequence of the previous result, we will henceforth also write dx instead of $\lambda(dx)$ if there is not danger of confusion.

2.1 Convergence theorems

In contrast to the Riemann integral, the Lebesgue integral is pretty robust when it comes to exchanging limits and integration. The following subsection collects the convergence theorems that are most relevant to us in what is to come.

2.1.1 Dominated and monotone convergence

Theorem 2.1.1 (Monotone convergence theorem (MCT) (B. Levi (1875 – 1961, Italian polymath))). *Let (f_n) be a non-decreasing sequence of measurable functions $f_n \in \mathcal{M}^+$. Then*

$$\int_{\Omega} \lim_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu.$$

Proof. We first of all notice that $\lim_{n \rightarrow \infty} f_n \in \mathcal{M}^+$ due to Proposition 1.4.15, and hence its integral is well-defined (recall Def. 2.0.13).

The monotonicity of the integral (Lemma 2.0.14) immediately implies that

$$\int_{\Omega} \lim_{n \rightarrow \infty} f_n d\mu \geq \int_{\Omega} f_m d\mu$$

for each $m \in \mathbb{N}$, and that the right-hand side is monotone in m , whence we conclude that

$$\int_{\Omega} \lim_{n \rightarrow \infty} f_n d\mu \geq \lim_{m \rightarrow \infty} \int_{\Omega} f_m d\mu.$$

To show the reverse inequality, we will take advantage of Corollary 2.0.15. For this purpose, consider arbitrary $g \in \mathcal{T}^+$ with $g \leq \lim_{n \rightarrow \infty} f_n$. Then for $\varepsilon > 0$ arbitrary we have that

$$\mathbb{1}_{\{f_n \geq (1-\varepsilon)g\}} \cdot g \uparrow g,$$

and as a consequence

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu \geq (1 - \varepsilon) \lim_{n \rightarrow \infty} \int_{\Omega} \underbrace{\mathbb{1}_{\{f_n \geq (1-\varepsilon)g\}} \cdot g}_{\in \mathcal{T}^+} d\mu \geq (1 - \varepsilon) \int_{\Omega} g d\mu,$$

where the limits exist due to monotonicity and Lemma 2.0.14, and the second inequality follows from Claim 2.0.12. Taking $\varepsilon > 0$ to 0 we obtain

$$\lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu \geq \int_{\Omega} g d\mu,$$

for any $g \in \mathcal{T}^+$ with $g \leq f$. Thus, the desired inequality is a consequence of Corollary 2.0.15. \square

Having this result at our disposal, we can immediately prove the following result which is of importance on its own.

Lemma 2.1.2 (Lemma of Fatou (1878 – 1929, French mathematician)). *Let (f_n) be a sequence with $f_n \in \mathcal{M}^+$. Then*

$$\liminf_{n \rightarrow \infty} \int f_n \, d\mu \geq \int \liminf_{n \rightarrow \infty} f_n \, d\mu.$$

Proof. Since for each $n \in \mathbb{N}$ we have that for $m \geq n$,

$$f_m \geq \inf_{k \geq n} f_k,$$

the monotonicity of the integral supplies us with

$$\inf_{k \geq n} \int f_k \, d\mu \geq \int \inf_{k \geq n} f_k \, d\mu.$$

Taking limits on both sides we get

$$\liminf_{n \rightarrow \infty} \int f_n \, d\mu \geq \lim_{n \rightarrow \infty} \int \inf_{k \geq n} f_k \, d\mu \stackrel{\text{Thm. 2.1.1}}{=} \int \liminf_{n \rightarrow \infty} f_n \, d\mu.$$

□

Exercise 2.1.3. *Show that the conclusion of Lemma 2.1.2 does not hold in general if we dispose of the assumption $f_n \geq 0$.*

Definition 2.1.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and assume given a sequence (A_n) of subsets $A_n \subset \Omega$. Then the ‘limes superior’ of the sequence (A_n) is defined as*

$$\limsup_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

The ‘limes inferior’ of the sequence (A_n) is defined as

$$\liminf_{n \rightarrow \infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$$

Note that if $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$ in the previous definition, then also $\limsup_{n \rightarrow \infty} A_n \in \mathcal{F}$ and $\liminf_{n \rightarrow \infty} A_n \in \mathcal{F}$.

Exercise 2.1.5. *Show the following identities:*

•

$$\limsup_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\};$$

•

$$\liminf_{n \rightarrow \infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ such that } \exists n_0 \in \mathbb{N} \text{ with } \omega \in A_n \, \forall n \geq n_0\};$$

Corollary 2.1.6. *Let a measure space $(\Omega, \mathcal{F}, \mu)$ and a sequence (A_n) with $A_n \in \mathcal{F}$ be given. Then*

$$\mu(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} \mu(A_n), \tag{2.1.1}$$

and if μ is finite, then also

$$\mu(\limsup_{n \rightarrow \infty} A_n) \geq \limsup_{n \rightarrow \infty} \mu(A_n), \tag{2.1.2}$$

Proof. (2.1.1) is a direct consequence of Fatou's lemma with $f_n := \mathbb{1}_{A_n}$, since

$$\mu(\liminf A_n) = \int \liminf_{n \rightarrow \infty} \mathbb{1}_{A_n} d\mu \leq \liminf_{n \rightarrow \infty} \int \mathbb{1}_{A_n} d\mu = \liminf_{n \rightarrow \infty} \mu(A_n).$$

To obtain (2.1.2), observe that (2.1.1) with A_n replaced by A_n^c reads

$$\mu(\liminf A_n^c) \leq \liminf_{n \rightarrow \infty} \mu(A_n^c).$$

Using this in combination with the identity $(\limsup_{n \rightarrow \infty} A_n)^c = \liminf_{n \rightarrow \infty} A_n^c$, the finiteness of μ , and the fact that $\mu(\Omega) - \liminf_{n \rightarrow \infty} \mu(A_n^c) = \limsup_{n \rightarrow \infty} \mu(A_n)$, we therefore deduce

$$\mu(\limsup_{n \rightarrow \infty} A_n) \geq \limsup_{n \rightarrow \infty} \mu(A_n).$$

□

Another benefit of Fatou's lemma is that it serves in proving the dominated convergence theorem below, which (besides the monotone convergence theorem) is one of the principal results allowing the interchange of integration and limits.

Theorem 2.1.7 (Lebesgue's dominated convergence theorem (DCT)). *Let (f_n) be a sequence in $\mathcal{M}(\Omega, \mathcal{F}, \mu)$, and assume there exists $g \in \mathcal{M}^+(\Omega, \mathcal{F}, \mu)$ with $|f_n| \leq g$ for all $n \in \mathbb{N}$, as well as $\int g d\mu < \infty$. Furthermore, assume that f_n converges μ -almost surely to some $f \in \mathcal{M}$ (see Definition 2.2.1 below for almost sure convergence).*

Then $\int |f| d\mu < \infty$, and

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

Proof. Using Fatou's lemma we obtain

$$\int g d\mu + \liminf_{n \rightarrow \infty} \left(\pm \int f_n d\mu \right) = \liminf_{n \rightarrow \infty} \int (g \pm f_n) d\mu \geq \int (g \pm f) d\mu = \int g d\mu \pm \int f d\mu. \quad (2.1.3)$$

Subtracting $\int g d\mu \in [0, \infty)$ on both sides supplies us with

$$\int f d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu \leq \limsup_{n \rightarrow \infty} \int f_n d\mu = -\liminf_{n \rightarrow \infty} \int -f_n d\mu \leq \int f d\mu,$$

and hence finishes the proof. □

Example 2.1.8. *Let $f \in \mathcal{M}^+$. Then*

$$\int f d\mu = \int_{[0, \infty)} \mu(f > t) \lambda(dt).$$

Indeed, let (f_n) be any sequence as in Lemma 2.0.10. Then, using MCT,

$$\int_{\Omega} f d\mu = \lim_{n \rightarrow \infty} \int_{\Omega} f_n d\mu = \lim_{n \rightarrow \infty} \int_{[0, \infty)} \nu(f_n > t) \lambda(dt) = \int_{[0, \infty)} \nu(f > t) \lambda(dt).$$

where the first equality is the definition of the integral for measurable non-negative functions (or else by MCT), the second equality is easy to check for simple functions in normal form, and the third equality takes advantage of the MCT as well as the convergence $\{f_n > t\} \uparrow \{f > t\}$ in combination with the continuity from below of the measure μ .

2.2 Measures with densities, absolute continuity

2.2.1 Almost sure / almost everywhere properties

Definition 2.2.1. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and let P be a property such that for each $\omega \in \Omega$ it can be decided whether the property P holds true or not. We then say that the property P holds (μ) -almost everywhere / (μ) -a.e. ((μ) -fast überall / (μ) -f.ü.)) if there exists a (μ) -null set $N \in \mathcal{F}$ such that P holds for all $\omega \in N^c$.

If μ is a probability measure, we also say that the property P holds for (μ) -almost all / (μ) -a.a. (μ -fast alle / μ -f.a.) ω instead.

Example 2.2.2. Consider the Cantor function $F : [0, 1] \rightarrow [0, 1]$. Then F is differentiable $\lambda|_{[0,1]}$ -almost everywhere with $F'(x) = 0$ for all such x (see problem 2 e) on homework sheet 7). Indeed, in the homework we have seen that $F'(x) = 0$ for all $x \in [0, 1] \setminus C$, with C denoting the Cantor set. You have furthermore shown that $\lambda(C) = 0$, so in particular this means that F is differentiable $\lambda|_{[0,1]}$ -a.e.

Lemma 2.2.3. Let $f \in \mathcal{M}^+(\Omega, \mathcal{F}, \mu)$. Then

$$\int f \, d\mu = 0 \quad \text{if and only if} \quad f = 0 \, \mu\text{-a.e.}$$

Proof. The statement is obvious for simple functions. For general $f \in \mathcal{M}^+$, choose a monotone approximating sequence of non-negative simple functions f_n with $f_n \uparrow f$. Then $f = 0$ μ -a.e. if and only if for all $n \in \mathbb{N}$, we have $f_n = 0$ μ -a.e. By the observed validity of the statement for simple functions, the latter is equivalent to $\int f_n \, d\mu = 0$ for all $n \in \mathbb{N}$, which due to the definition of the integral via

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

is equivalent to $\int f \, d\mu = 0$. This proves the result. \square

Lemma 2.2.3 is interesting in its own right, but it also proves useful in deriving the following result.

Proposition 2.2.4. Let

- (a) $f, g \in \mathcal{M}^+(\Omega, \mathcal{F}, \mu)$, or let
- (b) $f, g \in \mathcal{M}(\Omega, \mathcal{F}, \mu)$ and let f or g be μ -integrable.

Then, if $f \leq g$ μ -a.e., we have

$$\int_{\Omega} f \, d\mu \leq \int_{\Omega} g \, d\mu.$$

Proof. Exercise. \square

The MCT Theorem 2.1.1 gives rise to a class of measures that is of particular importance in probability theory, as is explained in the following corollary to it.

Corollary 2.2.5. Let $f \in \mathcal{M}^+$ be defined on a measure space $(\Omega, \mathcal{F}, \mu)$. Then the set function

$$\begin{aligned} \nu : \mathcal{F} &\rightarrow [0, \infty], \\ A &\mapsto \int_A f \, d\mu, \end{aligned} \tag{2.2.1}$$

defines a measure on (Ω, \mathcal{F}) .

Proof. Since $f \geq 0$ we have $\nu \geq 0$, and furthermore $\nu(\emptyset) = 0$. To show the σ -additivity, let (A_n) be a sequence of pairwise disjoint sets with $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$. Then

$$\nu\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \int_{\bigcup_{n \in \mathbb{N}} A_n} f \, d\mu = \int_{\Omega} \left(\sum_{n \in \mathbb{N}} \mathbb{1}_{A_n}\right) \cdot f \, d\mu \stackrel{\text{Thm. 2.1.1}}{=} \sum_{n \in \mathbb{N}} \int_{\Omega} f \cdot \mathbb{1}_{A_n} \, d\mu = \sum_{n \in \mathbb{N}} \nu(A_n).$$

□

Remark 2.2.6. In fact, if instead of $f \in \mathcal{M}^+$ we only assume that $f \in \mathcal{M}$ is μ -quasi-integrable, we still get that (2.2.1) defines a signed measure.

For those who have attended ‘Introduction to stochastics’, we add here the remark that the result of Corollary 2.2.5 is very good news for us. Indeed, recall from [Dre18, Section 1.8.3] that we did run into severe troubles trying to define measures as in (2.2.1) using the Riemann integral. It turns out that by use of the Lebesgue integral everything we’re after works out smoothly.

Definition 2.2.7. In the context of Corollary 2.2.5, we say that f is a density of ν with respect to μ . We write $\nu = f \cdot \mu$, or

$$f = \frac{d\nu}{d\mu}.$$

Example 2.2.8. Recall the Normal distribution on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ introduced in Example 1.5.8. We observe that a $\mathcal{N}(\mu, \Sigma)$ distributed random variable ($\mu \in \mathbb{R}^d$, Σ a symmetric positive definite matrix) has a distribution (recall Def. 1.5.7) which has density

$$\frac{1}{\sqrt{(2\pi)^d |\det(\Sigma)|}} e^{-\frac{1}{2}(x-\mu) \cdot (\Sigma^{-1}(x-\mu))}, \quad x \in \mathbb{R}^d,$$

with respect to λ^d .

The following result explains us how to integrate with respect to measures that have a density.

Proposition 2.2.9. Let $g \in \mathcal{M}$ such that $g \geq 0$ μ -a.s. or $\int |g| \, d\mu < \infty$. Furthermore, assume $\nu = f \cdot \mu$ with $f \in \mathcal{M}^+$. Then

$$\int_{\Omega} |g| \, d\nu < \infty \quad \text{if and only if} \quad \int_{\Omega} |g| \cdot f \, d\mu < \infty,$$

and if the integrals are finite, then

$$\int_{\Omega} g \, d\nu = \int_{\Omega} g \cdot f \, d\mu \in \mathbb{R}.$$

Proof. Assume $g \in \mathcal{M}^+$. Then, due to Lemma 2.0.10, there exists a non-decreasing sequence (g_n) of functions in \mathcal{T}^+ such that $\lim_{n \rightarrow \infty} g_n = g$. Writing $g_n = \sum_{i=1}^{m_n} \alpha_i^n \mathbb{1}_{A_i^n}$, for a normal representation of g_n with $A_i^n \in \mathcal{F}$ and $\alpha_i^n \geq 0$ we have

$$\int_{\Omega} g_n \, d\nu = \sum_{i=1}^{m_n} \alpha_i^n \nu(A_i^n) = \sum_{i=1}^{m_n} \alpha_i^n \int_{\Omega} f \cdot \mathbb{1}_{A_i^n} \, d\mu = \int_{\Omega} g_n \cdot f \, d\mu,$$

where in the second equality we took advantage of the fact that $\nu(A_i^n) = \int_{A_i^n} f \, d\mu$ by definition of the measure ν . Using the MCT, the result follows.

If, on the other hand, $g \in \mathcal{M}$, then we decompose $g = g^+ - g^-$, and proceed similarly to the above for both, g^+ and g^- . □

Proposition 2.2.10. *Let ν and μ be measures on (Ω, \mathcal{F}) , and assume that ν is σ -finite. Furthermore, assume that f and g are densities of ν with respect to μ . Then we have*

$$f = g \quad \mu - \text{a.e.}$$

Proof. Since ν is σ -finite we find $S_n \in \mathcal{F}$ such that $\nu(S_n) < \infty$ and $\Omega = \bigcup_{n \in \mathbb{N}} S_n$. Define the set $L_n := \{f > g\} \cap S_n$ of those points in S_n where f takes values larger than g . We deduce

$$0 = \nu(L_n) - \nu(L_n) = \int_{L_n} f \, d\mu - \int_{L_n} g \, d\mu = \int_{\Omega} (f - g) \mathbb{1}_{\{f > g\} \cap S_n} \, d\mu.$$

Since $(f - g) \mathbb{1}_{\{f > g\}} > 0$, this implies $\mu(L_n) = 0$. Therefore,

$$\mu(f > g) = \mu\left(\bigcup_{n \in \mathbb{N}} L_n\right) \leq \sum_{n \in \mathbb{N}} \mu(L_n) = 0.$$

In a similar manner we obtain

$$\mu(f < g) = 0,$$

so $\mu(f \neq g) = 0$ and therefore $f = g$ μ -a.e. □

More generally, we introduce the following concepts relating two measures on the same measurable space.

Definition 2.2.11. *Let μ, ν be two measures on a measurable space (Ω, \mathcal{F}) . We say that*

- (a) ν is absolutely continuous ('absolutstetig') with respect to μ (and write $\nu \ll \mu$), if for each $F \in \mathcal{F}$ with $\mu(F) = 0$ we also have $\nu(F) = 0$;
- (b) μ and ν are equivalent if $\mu \ll \nu$ and $\nu \ll \mu$;
- (c) ν is singular ('singulär') with respect to μ (and write $\nu \perp \mu$), if there exists $F \in \mathcal{F}$ such that $\mu(F) = 0$ and $\nu(F^c) = 0$.

In the case of finite measures, there is an obvious justification for Part (a) of this terminology given in the following lemma.

Lemma 2.2.12. *Let μ and ν be measures on a measurable space (Ω, \mathcal{F}) such that ν is finite. Then ν is absolutely continuous with respect to μ if and only if for each $\varepsilon > 0$ there exists $\delta > 0$ such that for any $F \in \mathcal{F}$,*

$$\mu(F) \leq \delta \quad \text{implies} \quad \nu(F) \leq \varepsilon.$$

Proof. If the ε - δ -condition holds, then for any set $F \in \mathcal{F}$ with $\mu(F) = 0$ we have $\nu(F) = 0$, so ν is absolutely continuous with respect to μ .

On the other hand, assume that the condition does not hold true. Then we find $\varepsilon > 0$ and a sequence (F_n) with $F_n \in \mathcal{F}$ such that $\mu(F_n) \leq 2^{-n}$ and $\nu(F_n) > \varepsilon$. Setting

$$F := \limsup_{n \rightarrow \infty} F_n = \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} F_m$$

we deduce for each $n \in \mathbb{N}$ that

$$\mu(F) \leq \mu\left(\bigcup_{m \geq n} F_m\right) \leq \sum_{m \geq n} \mu(F_m) \leq 2^{-n+1},$$

so $\mu(F) = 0$. On the other hand, using Corollary 2.1.6 and the fact that μ is finite,

$$\nu(F) \geq \limsup_{n \rightarrow \infty} \nu(F_n) \geq \varepsilon > 0.$$

Thus, ν is not absolutely continuous with respect to μ . □

2.2.2 Hahn-Jordan decomposition

The following result on the Radon-Nikodym derivative (Theorem 2.2.16) is not only of importance in the context of probability theory, but is also of significance to functional analysis. Before, however, we introduce a key result for proving it, which is interesting in its own right. We begin with giving an intuitive definition.

Definition 2.2.13. Let μ be a signed measure on (Ω, \mathcal{F}) . A set $A \in \mathcal{F}$ is called positive (negative), if for every $F \in \mathcal{F}$ we have that $\mu(A \cap F) \geq 0$ ($\mu(A \cap F) \leq 0$).

Theorem 2.2.14 (Hahn-Jordan decomposition (Hans Hahn (1879–1934), Camille Jordan (1838–1922))). Let μ be a signed measure on (Ω, \mathcal{F}) . Then there exist $\Omega^+, \Omega^- \in \mathcal{F}$ such that $\Omega = \Omega^+ \dot{\cup} \Omega^-$ and the following hold:

(a) Setting $\mu^+ := \mu(\cdot \cap \Omega^+)$ as well as $\mu^- := -\mu(\cdot \cap \Omega^-)$, both μ^+ and μ^- define non-negative measures (note that in particular we have $\mu^-(\Omega^+) = 0$ as well as $\mu^+(\Omega^-) = 0$). In this case, we say that Ω^+ and Ω^- form a so-called Hahn decomposition of Ω with respect to μ .

(b)

$$\mu = \mu^+ - \mu^- \quad (\text{Jordan decomposition}).$$

In addition, the Hahn decomposition of Ω is unique up to null sets with respect to the measure $\mu^+ + \mu^-$.

Proof. According to Exercise 1.2.12 we can assume w.l.o.g. (without loss of generality, ‘o.B.d.A.’, ‘ohne Beschränkung der Allgemeinheit’) that $\mu(F) \in (-\infty, \infty]$ for all $F \in \mathcal{F}$.

We define

$$c := \inf_{\substack{F \in \mathcal{F} \\ F \text{ negative}}} \mu(F). \quad (2.2.2)$$

We can find a sequence (F_n) of negative subsets of \mathcal{F} such that $\lim_{n \rightarrow \infty} \mu(F_n) = c$, the countable union of negative sets is negative again, we infer that

$$\Omega^- := \bigcup_{n=1}^{\infty} F_n$$

is negative and that $\mu(\Omega^-) = c$. Indeed, since $-\mu|_{\Omega^-}$ is a ‘common’ non-negative measure, the latter equality follows from

$$c = \lim_{n \rightarrow \infty} \mu(F_n) \geq \mu(\Omega^-) \geq \inf_{\substack{F \in \mathcal{F} \\ F \text{ negative}}} \mu(F) \geq c,$$

where the second limit exists since the sequence is monotone.

We claim that $\Omega^+ := \Omega \setminus \Omega^-$ is a positive set. Indeed, assume it was not. Then there would exist $G_0 \in \mathcal{F}$ with $G_0 \subset \Omega^+$ such that $\mu(G_0 \cap \Omega^+) < 0$. Now G_0 cannot be a negative set since in that case $G_0 \dot{\cup} \Omega^-$ would be a negative set with $\mu(G_0 \cup \Omega^-) < c$, a contradiction to (2.2.2). Thus, there exists a minimal $k_1 \in \mathbb{N}$ such that G_0 contains a set $G_1 \in \mathcal{F}$ with $\mu(G_1) \geq \frac{1}{k_1}$. Then

$$\mu(G_0 \setminus G_1) = \mu(G_0) - \mu(G_1) \leq -\frac{1}{k_1};$$

since the RHS is negative, the same reasoning applied to G_0 before can now be applied to $G_0 \setminus G_1$ in order to infer that $G_0 \setminus G_1$ contains a set $G_2 \in \mathcal{F}$ with $\mu(G_2) \geq \frac{1}{k_2}$, and such that k_2 is

minimized among all such admissible subsets. Due to $\mu(G_0) \in (-\infty, \infty)$, Exercise 1.2.12 would in particular yield that

$$\lim_{n \rightarrow \infty} \frac{1}{k_n} = 0,$$

and as a consequence, for every $G \in \mathcal{F}$ subset of

$$G^* := G_0 \setminus \bigcup_{n=1}^{\infty} G_n$$

we would have $\mu(G) \leq 0$. In particular, $G^* \in \mathcal{F}$ would be negative. But then

$$\mu(G^*) = \mu(G_0) - \sum_{n=1}^{\infty} \mu(G_n) \leq \mu(G_0) < 0,$$

which as before contradicts the assumption that $\mu(\Omega^-)$ is minimal among all negative sets. Therefore, we must have that Ω^+ is a positive set and we have proved the claim and can now conclude the proof:

(a) It immediately follows that

$$\mu^-(\Omega^+) = \mu^+(\Omega^-) = 0. \quad (2.2.3)$$

Similarly, $\mu^+ \geq 0$ since Ω^+ is a positive set and $\mu^- \geq 0$ since Ω^- is a negative set.

(b) This is a consequence of the additivity of signed measures.

Uniqueness: Exercise. □

Lemma 2.2.15. *For a signed measure μ on (Ω, \mathcal{F}) , we have with the notation of Theorem 2.2.14:*

(a)

$$\mu^+(A) = \sup_{\substack{F \in \mathcal{F} \\ F \subset A}} \mu(F), \quad \forall A \in \mathcal{F},$$

and μ^+ is also called the positive variation of μ .

(b)

$$\mu^-(A) = - \inf_{\substack{F \in \mathcal{F} \\ F \subset A}} \mu(F), \quad \forall A \in \mathcal{F},$$

and μ^- is also called the negative variation of μ .

Furthermore, we call the measure

$$|\mu| := \mu^+ + \mu^-$$

the total variation of μ .

Proof. (a) It follows from Theorem 2.2.14 that $\mu(\cdot \cap \Omega^-)$ is non-positive and $\mu(\cdot \cap \Omega^+)$ is non-negative, so

$$\sup_{\substack{F \in \mathcal{F} \\ F \subset A}} \mu(F) = \sup_{\substack{F \in \mathcal{F} \\ F \subset A \cap \Omega^+}} \mu(F) = \mu^+(A).$$

The remaining part follows in a similar manner. □

2.2.3 Lebesgue's decomposition theorem, Radon-Nikodym derivative

Theorem 2.2.16 (Lebesgue's decomposition theorem, Radon-Nikodym theorem). *Let μ and ν be σ -finite measures on a measurable space (Ω, \mathcal{F}) .*

Then there exist unique measures ν_{ac} and ν_s on (Ω, \mathcal{F}) such that the following hold:

$$\nu = \nu_{ac} + \nu_s, \quad \text{and} \quad \nu_{ac} \ll \mu, \quad \nu_s \perp \mu.$$

Furthermore, there exists $f \in \mathcal{M}^+$ with

$$\nu_{ac} = f \cdot \mu$$

and f is μ -a.s. uniquely determined.

Corollary 2.2.17. *Under the same assumptions as in Theorem 2.2.16,*

$$\nu \ll \mu \quad \text{if and only if } \nu \text{ has a density with respect to } \mu.$$

You are asked to prove most of Theorem 2.2.16 in homework problem 2 on sheet 8. Here, we will only address the uniqueness part. For this purpose take another decomposition of μ with ν'_s and ν'_{ac} denoting the corresponding singular and absolutely continuous parts. Let $A \in \mathcal{F}$ such that $\mu(A) = 0$ and $\nu_s(A^c) = 0$. Then $\nu'_{ac}(A) = 0$ since $\mu(A) = 0$, and hence

$$\nu_s(F) = \nu(F \cap A) = \nu'_s(F \cap A) \leq \nu'_s(F) \quad \forall F \in \mathcal{F}.$$

In particular, we infer $\nu_s \leq \nu'_s$ and thus $\nu'_{ac} \leq \nu_{ac}$. But then $\nu'_s - \nu_s = \nu_{ac} - \nu'_{ac}$ is a measure which at the same times is singular and absolutely continuous w.r.t. μ so it must vanish. Hence, $\nu_s = \nu'_s$ and $\nu'_{ac} = \nu_{ac}$, which proves the uniqueness of the Lebesgue decomposition.

Now that we know that ν_{ac} is unique, the μ -a.e. uniqueness of f follows from the fact that (see homework sheet) $\nu_{ac} := f \cdot \mu$ for some $f \in \mathcal{M}^+$ and Proposition 2.2.10.

Remark 2.2.18. (a) *Using Theorem 2.2.14, Theorem 2.2.16 can be extended to signed measures (which we won't do here).*

(b) *Above we had introduced all the machinery we needed in order to give a self-contained proof of Theorem 2.2.16. If you do have a basic knowledge of functional analysis, you might also want to have a look at another proof of the cited results that takes advantage of Riesz' representation theorem (see the Proof of [Kle14, Theorem 7.33], for example).*

2.2.4 Integration with respect to image measures

The following is a generalization of [Dre18, Proposition 1.9.10].

Theorem 2.2.19 (Change of variable formula ('Transformationssatz')). *Assume a measure space $(\Omega, \mathcal{F}, \mu)$ as well as a measurable space $(\tilde{\Omega}, \tilde{\mathcal{F}})$ and a $\mathcal{F} - \tilde{\mathcal{F}}$ -measurable map $\varphi : \Omega \rightarrow \tilde{\Omega}$ be given. Denote the image measure $\mu \circ \varphi^{-1}$ by $\tilde{\mu}$. Then for $f \in \mathcal{M}(\tilde{\Omega}, \tilde{\mathcal{F}})$ the integral*

$$\int_{\Omega} f \circ \varphi \, d\mu$$

exists if and only if the integral

$$\int_{\tilde{\Omega}} f \, d\tilde{\mu}$$

exists in $\overline{\mathbb{R}}$ (i.e., in the sense of quasi-integrability). In this case both integrals coincide.

The proof is contained in Homework 8.1.

2.3 Product spaces

As outlined in [Dre18, Section 1.14.3] of [Dre18] already, it will be crucial for us to be able to construct infinite sequences of random variables defined on the same probability space. One way to ensure that the underlying probability space is ‘rich enough’ to accommodate for this wealth and structure of random variables will be to consider infinite product spaces, which are products of measure spaces and play a crucial role in probability theory.²

Definition 2.3.1. Let Λ be an arbitrary non-empty index set, and let Ω_λ , $\lambda \in \Lambda$, be a family of non-empty sets. Then we define the product space (‘Produktraum’) (or also the product of the Ω_λ , $\lambda \in \Lambda$)

$$\bigtimes_{\lambda \in \Lambda} \Omega_\lambda$$

to be the set of all maps $f : \Lambda \rightarrow \bigcup_{\lambda \in \Lambda} \Omega_\lambda$ such that $f(\lambda) \in \Omega_\lambda$ for all $\lambda \in \Lambda$.

In the case that the Ω_λ are equal to some set Ω for all $\lambda \in \Lambda$, we also write Ω^Λ for the corresponding product space.

Example 2.3.2. Without going into detail, it might be suggesting itself that if we want to model an infinite sequence of coin tosses, the space $\{0, 1\}^\mathbb{N}$ might be a candidate for outcomes of such an experiment to lie in, where 0 can be identified with tails and 1 with heads. Then $\omega \in \{0, 1\}^\mathbb{N}$ is nothing else than an infinite sequence $(\omega(n))_{n \in \mathbb{N}}$ of elements in $\omega(n) \in \{0, 1\}$ for all $n \in \mathbb{N}$.

Definition 2.3.3. For $I \subset J \subset \Lambda$ we introduce the projections

$$\begin{aligned} \pi_I^J : \bigtimes_{\lambda \in J} \Omega_\lambda &\rightarrow \bigtimes_{\lambda \in I} \Omega_\lambda \\ \omega &\mapsto \omega|_I. \end{aligned}$$

In particular, if $J = \Lambda$ we will write π_I , and if $I = \{\lambda\}$, we will use the notation π_λ^J . For $\pi_{\{\lambda\}}^I$ we will also just write $\pi_{\{\lambda\}}$.

We are now going to introduce the notion of a product- σ -algebra. In case you have seen this before, it will be very much in the spirit of the definition of the product topology, where instead of measurability one asks for continuity of the projection maps. We recall the definition here for completeness.

Definition 2.3.4. Let a family $(\Omega_\lambda, \tau_\lambda)$, $\lambda \in \Lambda$, of topological spaces be given. The corresponding product topology τ is defined as the smallest topology on $\bigtimes_{\lambda \in \Lambda} \Omega_\lambda$ such that for each $\lambda' \in \Lambda$, the coordinate maps $\pi_{\lambda'} : \bigtimes_{\lambda \in \Lambda} \Omega_\lambda \rightarrow \Omega_{\lambda'}$ are continuous with respect to the topologies τ and $\tau_{\lambda'}$.

With this in mind, we can now proceed to the definition of the product- σ -algebra.

Definition 2.3.5. Let a family $(\Omega_\lambda, \mathcal{F}_\lambda)$, $\lambda \in \Lambda$, of measurable spaces be given. The corresponding product- σ -algebra $\bigotimes_{\lambda \in \Lambda} \mathcal{F}_\lambda$ is defined as the smallest σ -algebra on $\bigtimes_{\lambda \in \Lambda} \Omega_\lambda$ such that for each $\lambda' \in \Lambda$, the coordinate maps $\pi_{\lambda'} : \bigtimes_{\lambda \in \Lambda} \Omega_\lambda \rightarrow \Omega_{\lambda'}$ are $\bigotimes_{\lambda \in \Lambda} \mathcal{F}_\lambda - \mathcal{F}_{\lambda'}$ -measurable. As before, if $\mathcal{F}_0 = \mathcal{F}_\lambda$ for all $\lambda \in \Lambda$, then we also abbreviate $\bigotimes_{\lambda \in \Lambda} \mathcal{F}_\lambda$ by $(\mathcal{F}_0)^{\otimes \Lambda}$.

Lemma 2.3.6. In the setting of Definition 2.3.5, for any $I \subset J \subset \Lambda$, the mapping π_I^J is $\bigotimes_{\lambda \in J} \mathcal{F}_\lambda - \bigotimes_{\lambda \in I} \mathcal{F}_\lambda$ -measurable.

Proof. This is a direct consequence of Corollary 1.4.9. □

²In spirit, this will be very much related to the concept of products of topological spaces in case you have seen this before

Definition 2.3.7. Let a family $(\Omega_\lambda, \mathcal{F}_\lambda)$, $\lambda \in \Lambda$, of measurable spaces be given. Then any set of the form

$$(\pi_J)^{-1}(F) \in \bigotimes_{\lambda \in \Lambda} \mathcal{F}_\lambda, \quad F \in \bigotimes_{\lambda \in J} \mathcal{F}_\lambda, \quad J \subset \Lambda \text{ finite}, \quad (2.3.1)$$

is called a cylinder set.

Lemma 2.3.6 implies that cylinder sets are $\mathcal{F} = \bigotimes_{\lambda \in \Lambda} \mathcal{F}_\lambda$ -measurable.

The principal reason that cylinder sets will play an important role in what follows is that it is generally easier to assign probabilities to them than to arbitrary measurable subsets of an infinite space. Indeed, imagine the setting of infinitely many coin tosses again. As long as you only want to understand events involving the outcome of finitely many of these coin tosses, you are easily able to assign them a probability under the assumption that the coin tosses are fair and independent. This becomes much more intricate in case you consider events that depend on the outcome of infinitely many coin tosses.

Crucially, as will turn out below, specifying a probability measure on cylinder sets already uniquely characterizes the measure (see Theorem 4.2.1 below), so there is no need to specify probabilities for an even bigger subclass of measurable sets.

Exercise 2.3.8. Assume the setting of Definition 2.3.7. Show that cylinder sets form an algebra, but not a σ -algebra.

We have seen before that the Borel- σ -algebra plays a prominent role in our studies; the following result sheds some light on its behaviour under taking products.

Theorem 2.3.9. Let Λ be an at most countable set and assume that for a family $(\Omega_\lambda, \tau_\lambda)$, $\lambda \in \Lambda$, each $(\Omega_\lambda, \tau_\lambda)$ is a Polish space³ (complete separable metric spaces, if you prefer). Then, setting $\Omega := \bigtimes_{\lambda \in \Lambda} \Omega_\lambda$ and denoting by τ the product topology of Ω , we get that (Ω, τ) is a Polish space (or a complete separable metric spaces for that matter) again, and

$$\sigma(\tau) = \bigotimes_{\lambda \in \Lambda} \mathcal{B}(\tau_\lambda). \quad (2.3.2)$$

Proof. For simplicity of notation we assume w.l.o.g. that $\Lambda = \mathbb{N}$ (or a finite subset of \mathbb{N} with the respective modifications in the notation below) for the first part of the proof. Denote by d_n a metric on Ω_n that induces the topology τ_n , and with respect to which Ω_n is complete. We define on Ω a new metric

$$d(\omega, \omega') := \sum_{n \in \mathbb{N}} 2^{-n} \frac{d_n(\omega(n), \omega'(n))}{1 + d_n(\omega(n), \omega'(n))},$$

and it is left as an exercise to check that d induces the product topology on Ω , and that Ω is complete with respect to d . Furthermore, Ω is separable, as will follow from an argument below. We now prove (2.3.2). By definition, for each $\lambda \in \Lambda$, the projections $\pi_\lambda : \Omega \rightarrow \Omega_\lambda$, $\lambda \in \Lambda$, are continuous maps from the topological space (Ω, τ) to the topological space $(\Omega_\lambda, \tau_\lambda)$. As a consequence of Theorem 1.4.7, the σ -algebra $\bigotimes_{\lambda \in \Lambda} \mathcal{B}(\tau_\lambda)$ is generated by all sets of the form $\pi_\lambda^{-1}(O)$, $O \in \tau_\lambda$, $\lambda \in \Lambda$, and in combination with the aforementioned continuity of the projections we get that $\pi_\lambda^{-1}(O) \in \tau$, whence $\bigotimes_{\lambda \in \Lambda} \mathcal{B}(\tau_\lambda) \subset \sigma(\tau)$.

To prove the converse inclusion, we first of all observe that the (at most) countable product of separable metric spaces is separable again. Indeed, for $\lambda \in \Lambda$ we denote by D_λ a countable dense subset of Ω_λ . For each $\lambda \in \Lambda$ choose and fix $\tilde{\omega}_\lambda \in D_\lambda$ arbitrarily and set

$$D := \left\{ \omega \in \bigtimes_{\lambda \in \Lambda} D_\lambda : \omega_{\lambda'} \neq \tilde{\omega}_{\lambda'} \text{ for finitely many } \lambda' \in \Lambda \right\}.$$

³Recall that a Polish space was defined as a separable topological space, for which there exists a complete metric that induces its topology.

Then D is countable and dense in Ω , so Ω is separable.

Now let

$$\mathcal{O}_\lambda := \left\{ B_{\frac{1}{n}}(\omega) : \omega \in D_\lambda, n \in \mathbb{N} \right\},$$

with $B_{\frac{1}{n}}(\omega) = \{\tilde{\omega} \in \Omega_\lambda : d_\lambda(\omega, \tilde{\omega}) < \frac{1}{n}\}$, then \mathcal{O}_λ is a countable basis of the topology of Ω_λ (i.e., any open set in Ω_λ can be written as a (a priori possibly uncountable) union of elements of \mathcal{O}_λ). Thus,

$$\bigcup_{n=1}^{\infty} \left\{ \bigcap_{i=1}^n \pi_{\lambda_i}^{-1}(B_{\lambda_i}) : B_{\lambda_i} \in \mathcal{O}_{\lambda_i}, \{\lambda_1, \dots, \lambda_n\} \subset \Lambda \right\}. \quad (2.3.3)$$

is a basis for the topology τ .

Now since Ω is separable, for an arbitrary basis \mathcal{B} of the topology τ , any open set in τ can be written as the countable union of elements in \mathcal{B} . Thus, since the elements of (2.3.3) are contained in $\otimes_{\lambda \in \Lambda} \mathcal{B}(\tau_\lambda)$, this finishes the proof of $\tau \subset \otimes_{\lambda \in \Lambda} \mathcal{B}(\tau_\lambda)$ and thus also $\sigma(\tau) \subset \otimes_{\lambda \in \Lambda} \mathcal{B}(\tau_\lambda)$. \square

We immediately obtain the following important corollary.

Corollary 2.3.10. *For each $d \in \mathbb{N}$, we have*

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R})^{\otimes d}.$$

2.4 Product measures

Oftentimes, such as e.g. in the case of \mathbb{R}^d , we do know how to integrate with respect to the ‘one-dimensional measures’ (such as with respect to the Lebesgue measure, where the fundamental theorem of calculus provides us with a powerful tool to actually compute integrals), but integration with respect to the product measure seems to be harder when it comes to actual computations (recall Theorem 1.3.11 as well as Remark 1.3.15). The Theorems 2.5.1 and 2.5.2 below provide a useful technique to reduce the integral with respect to the product measure to integrals with respect to the marginals.

In order to be able to rigorously formulate them, we first have to introduce the concept of a product measure in a more general setting than that of Remark 1.3.15.

Definition 2.4.1. *Assume measurable spaces $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ and (E, \mathcal{E}) be given and write $\Omega := \Omega_1 \times \Omega_2$. For arbitrary $A \subset \Omega$ as well as $\tilde{\omega}_j \in \Omega_j$, $1 \leq j \leq 2$, we call*

$$A_{\tilde{\omega}_1} := \{\omega_2 \in \Omega_2 : (\tilde{\omega}_1, \omega_2) \in A\}$$

the $\tilde{\omega}_1$ -section of A (‘ $\tilde{\omega}_1$ -Schnitt von A ’), and similarly we call

$$A^{\tilde{\omega}_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \tilde{\omega}_2) \in A\}$$

the $\tilde{\omega}_2$ -section of A (‘ $\tilde{\omega}_2$ -Schnitt von A ’).

If $f : \Omega \rightarrow E$, then we call

$$f_{\tilde{\omega}_1} : \Omega_2 \rightarrow E, \quad \omega_2 \mapsto f(\tilde{\omega}_1, \omega_2)$$

the $\tilde{\omega}_1$ -section of f (‘ $\tilde{\omega}_1$ -Schnitt von f ’), and similarly

$$f^{\tilde{\omega}_2} : \Omega_1 \rightarrow E, \quad \omega_1 \mapsto f(\omega_1, \tilde{\omega}_2)$$

the $\tilde{\omega}_2$ -section of f (‘ $\tilde{\omega}_2$ -Schnitt von f ’).

Lemma 2.4.2. *Assume the setting of Definition 2.4.1.*

(a) If $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$, then

$$A_{\tilde{\omega}_1} \in \mathcal{F}_2, \quad A^{\tilde{\omega}_2} \in \mathcal{F}_1.$$

(b) If f is $\mathcal{F}_1 \otimes \mathcal{F}_2 - \mathcal{E}$ -measurable, then $f_{\tilde{\omega}_1}$ is $\mathcal{F}_2 - \mathcal{E}$ -measurable, and similarly $f^{\tilde{\omega}_2}$ is $\mathcal{F}_1 - \mathcal{E}$ -measurable.

Proof. (a) Fix $\tilde{\omega}_1 \in \Omega_1$ and consider the system of sets

$$\mathcal{A} := \{A \in \mathcal{F}_1 \otimes \mathcal{F}_2 : A_{\tilde{\omega}_1} \in \mathcal{F}_2\}.$$

We claim that \mathcal{A} is a σ -algebra. Indeed, $\Omega \in \mathcal{A}$ since $\Omega_{\tilde{\omega}_1} = \Omega_2 \in \mathcal{F}_2$. Furthermore, since $(A^c)_{\tilde{\omega}_1} = \Omega_2 \setminus (A_{\tilde{\omega}_1})$, we get that $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$. Lastly, for (A_n) with $A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$, we deduce

$$\left(\bigcup_{n \in \mathbb{N}} A_n \right)_{\tilde{\omega}_1} = \bigcup_{n \in \mathbb{N}} (A_n)_{\tilde{\omega}_1},$$

and hence we deduce $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$, too, and \mathcal{A} is a σ -algebra. Furthermore, for $F_1 \in \mathcal{F}_1$, $F_2 \in \mathcal{F}_2$ we have $F_1 \times F_2 \in \mathcal{F}_1 \otimes \mathcal{F}_2$, since

$$(F_1 \times F_2)_{\tilde{\omega}_1} = \begin{cases} F_2, & \text{if } \tilde{\omega}_1 \in F_1, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Since $\mathcal{F}_1 * \mathcal{F}_2$ (recall the notation of Lemma 1.1.7) generates the σ -algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$, we deduce that $\mathcal{A} = \mathcal{F}_1 \otimes \mathcal{F}_2$, which finishes the proof.

Analogously, one can show $A^{\tilde{\omega}_2} \in \mathcal{F}_1$.

(b) Fix $\tilde{\omega}_1 \in \Omega_1$. For $F \in \mathcal{E}$ we have

$$(f_{\tilde{\omega}_1})^{-1}(F) = (f^{-1}(F))_{\tilde{\omega}_1},$$

and the claim follows from the fact that $f^{-1}(F) \in \mathcal{F}_1 \otimes \mathcal{F}_2$ in combination with (a).

Similarly for $\tilde{\omega}_2 \in \Omega_2$ and $f^{\tilde{\omega}_2}$.

□

Proposition 2.4.3. For $j \in \{1, 2\}$, let μ_j be a measure on a measurable space $(\Omega_j, \mathcal{F}_j)$.

(a) If μ_2 is σ -finite, then for any $F \in \mathcal{F}_1 \otimes \mathcal{F}_2$, the function

$$\Omega_1 \ni \omega_1 \mapsto \mu_2(F_{\omega_1}) \tag{2.4.1}$$

is $\mathcal{F}_1 - \mathcal{B}(\overline{\mathbb{R}})$ -measurable, and the function

$$\mu : \mathcal{F}_1 \otimes \mathcal{F}_2 \mapsto \int_{\Omega_1} \mu_2(F_{\omega_1}) \mu_1(d\omega_1) \tag{2.4.2}$$

defines a measure such that for any $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$,

$$\mu(A \times B) = \mu_1(A) \cdot \mu_2(B). \tag{2.4.3}$$

(b) If both, μ_1 and μ_2 are σ -finite, then there is exactly one measure $\mu_1 \otimes \mu_2$ on $\mathcal{F}_1 \otimes \mathcal{F}_2$ such that (2.4.3) holds with μ replaced by $\mu_1 \otimes \mu_2$. In this case,

$$\mu_1 \otimes \mu_2(F) = \int_{\Omega_1} \mu_2(F_{\omega_1}) \mu_1(d\omega_1) = \int_{\Omega_2} \mu_1(F^{\omega_2}) \mu_2(d\omega_2) \quad \forall F \in \mathcal{F}_1 \otimes \mathcal{F}_2. \tag{2.4.4}$$

$\mu_1 \otimes \mu_2$ is called the product measure of μ_1 and μ_2 , and it is σ -finite.

Proof. (a) As before, by the usual exhaustion procedure we can assume that μ_2 is finite. Then desired measurability of the function in (2.4.1) follows again by a good sets principle and we will omit the details here.

We can rewrite the function in (2.4.2) as

$$\mu(F) = \int_{\Omega_1} \int_{\Omega_2} \mathbb{1}_F(\omega_1, \omega_2) \mu_2(d\omega_2) \mu_1(d\omega_1),$$

from which one can observe by applying the MCT that this expression defines a measure on $\mathcal{F}_1 \otimes \mathcal{F}_2$. By setting $F := A \times B$ we immediately obtain (2.4.3).

(b) By (a) and symmetry, the middle and right-hand side expressions of (2.4.4) both define measures satisfying (2.4.3). Since both, μ_1 and μ_2 are σ -finite, we can use Theorem 1.2.17 to deduce the desired uniqueness, so the two measures coincide and $\mu_1 \otimes \mu_2$ is well-defined and uniquely determined by (2.4.4). \square

Remark 2.4.4. σ -finiteness of μ_2 is needed Part (a) of Proposition 2.4.3, since otherwise the function $\omega_1 \mapsto \mu_2(F_{\omega_1})$ is not necessarily measurable anymore, see [Beh87, p.96].

The above can immediately be generalized to the product of finitely many measures, which in particular is a generalization of the observation of Remark 1.3.15.

Theorem 2.4.5. Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, $1 \leq i \leq n$, be σ -finite measure spaces. Then there exists a unique σ -finite measure on the product- σ -algebra $\mathcal{F} := \otimes_{1 \leq i \leq n} \mathcal{F}_i$ such that

$$\mu(F_1 \times \dots \times F_n) = \prod_{i=1}^n \mu_i(F_i),$$

for all $F_i \in \mathcal{F}_i$, $1 \leq i \leq n$.

$\otimes_{i=1}^n \mu_i := \mu_1 \otimes \dots \otimes \mu_n := \mu$ is called the product measure of the μ_i , $1 \leq i \leq n$, and in the case that all $(\Omega_i, \mathcal{F}_i, \mu_i)$ are equal, we write $\mu_1^{\otimes n}$ for the product measure.

Proof. We will not give the proof here since we will prove a more general result in Theorem 4.2.1 below. See [Els05, Satz V.1.12] for a proof. \square

2.5 The theorems of Fubini and Tonelli

Theorem 2.5.1 (Tonelli's theorem (Italian mathematician (1885–1946))). Let $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$ be σ -finite measure spaces, and let $f \in \mathcal{M}^+(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. Then the function

$$\omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \quad \text{is in } \mathcal{M}^+(\Omega_1, \mathcal{F}_1), \quad (2.5.1)$$

the function

$$\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \quad \text{is in } \mathcal{M}^+(\Omega_2, \mathcal{F}_2), \quad (2.5.2)$$

and the equality

$$\int_{\Omega_1 \times \Omega_2} f d\mu_1 \otimes \mu_2 = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) = \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1) \right) \mu_2(d\omega_2) \quad (2.5.3)$$

holds true.

Proof. We see that (2.5.1) to (2.5.3) hold true for $f = \mathbb{1}_F$, $F \in \mathcal{F}_1 \otimes \mathcal{F}_2$, due to Proposition 2.4.3. Using the linearity of the integral for non-negative functions as in Lemma 2.0.14, we obtain that the validity extends to simple non-negative functions which are $\mathcal{F}_1 \otimes \mathcal{F}_2 - \mathcal{B}(\mathbb{R})$ -measurable. Choosing a non-decreasing sequence (f_n) of non-negative simple functions as in Lemma 2.0.13 with $\lim_{n \rightarrow \infty} f_n = f$, we use Proposition 1.4.15 and the MCT to deduce the measurability stated in (2.5.1) and (2.5.2) for $f \in \mathcal{M}^+$ from their validity for simple functions:

$$\int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) = \int_{\Omega_2} \lim_{n \rightarrow \infty} f_n(\omega_1, \omega_2) d\mu_2(\omega_2) = \lim_{n \rightarrow \infty} \int_{\Omega_2} f_n(\omega_1, \omega_2) d\mu_2(\omega_2).$$

Similarly for the roles of ω_1 and ω_2 exchanged. The equalities in (2.5.3) then follow by the MCT and their validity for simple functions:

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f d\mu_1 \otimes \mu_2 &= \int_{\Omega_1 \times \Omega_2} \lim_{n \rightarrow \infty} f_n d\mu_1 \otimes \mu_2 = \lim_{n \rightarrow \infty} \int_{\Omega_1 \times \Omega_2} f_n d\mu_1 \otimes \mu_2 \\ &= \lim_{n \rightarrow \infty} \int_{\Omega_1} \left(\int_{\Omega_2} f_n(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1), \end{aligned}$$

and similarly for the second equality in (2.5.3). \square

It turns out that we are still allowed to integrate coordinatewise even if f is not necessarily non-negative. However, we have to demand integrability with respect to the product measure to replace non-negativity. This is the content of Fubini's theorem.

Theorem 2.5.2 (Fubini's theorem (Italian mathematician (1879–1943))). *Assume the setting of Theorem 2.5.1, except that instead of $f \in \mathcal{M}^+(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ we only require $f = g + ih$, with $g, h \in \mathcal{M}(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$. Then, if f is $\mu_1 \otimes \mu_2$ -integrable,*

$$(a) \quad A_1^c := \{\omega_1 \in \Omega_1 : f(\omega_1, \cdot) \text{ is not } \mu_2\text{-integrable}\} \in \mathcal{F}_1 \quad (2.5.4)$$

is a μ_1 -null set;

$$(b) \quad A_2^c := \{\omega_2 \in \Omega_2 : f(\cdot, \omega_2) \text{ is not } \mu_1\text{-integrable}\} \in \mathcal{F}_2 \quad (2.5.5)$$

is a μ_2 -null set;

$$(c) \quad \omega_1 \mapsto \int_{\Omega_2} f(\omega_1, \omega_2) d\mu_2(\omega_2) \quad \text{is in } \mathcal{M}(A_1, \mathcal{F}_1|_{A_1}), \quad (2.5.6)$$

and

$$\omega_2 \mapsto \int_{\Omega_1} f(\omega_1, \omega_2) d\mu_1(\omega_1) \quad \text{is in } \mathcal{M}(A_2, \mathcal{F}_2|_{A_2}), \quad (2.5.7)$$

and

$$\begin{aligned} \int_{\Omega_1 \times \Omega_2} f d\mu_1 \otimes \mu_2 &= \int_{A_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\ &= \int_{A_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2) \mu_1(d\omega_1) \right) \mu_2(d\omega_2). \end{aligned} \quad (2.5.8)$$

Proof. From the $\mu_1 \otimes \mu_2$ -integrability of f we deduce that $|f|$ is integrable with respect to $\mu_1 \otimes \mu_2$ as well. Thus, Tonelli's theorem implies

$$\int_{\Omega_1} \left(\int_{\Omega_2} |f(\omega_1, \omega_2)| \mu_2(d\omega_2) \right) \mu_1(d\omega_1) = \int_{\Omega_1 \times \Omega_2} |f| d\mu_1 \otimes \mu_2 < \infty. \quad (2.5.9)$$

The inner integral on the left-hand side is a measurable function due to (2.5.1) and so the finiteness of the outer integral on the left-hand side implies

$$\int_{\Omega_2} |f(\omega_1, \omega_2)| \mu_2(d\omega_2) < \infty \quad \text{for } \mu_1 - a.a. \omega_1 \in \Omega_1,$$

i.e.,

$$A_1^c = \left\{ \omega_1 \in \Omega_1 : \int_{\Omega_2} |f(\omega_1, \omega_2)| \mu_2(d\omega_2) = \infty \right\}$$

is in \mathcal{F}_1 due to (2.5.1) and a μ_1 -null set, which establishes (2.5.4). In particular,

$$\text{for all } \omega_1 \in A_1, \text{ the function } |f(\omega_1, \cdot)| \text{ is } \mu_2\text{-integrable.} \quad (2.5.10)$$

Thus we deduce from the linearity of the integral that for $\omega_1 \in A_1$ we have

$$\begin{aligned} \int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) &= \int_{\Omega_2} (\operatorname{Re} f)^+(\omega_1, \omega_2) \mu_2(d\omega_2) \\ &\quad - \int_{\Omega_2} (\operatorname{Re} f)^-(\omega_1, \omega_2) \mu_2(d\omega_2) \\ &\quad + i \int_{\Omega_2} (\operatorname{Im} f)^+(\omega_1, \omega_2) \mu_2(d\omega_2) \\ &\quad - i \int_{\Omega_2} (\operatorname{Im} f)^-(\omega_1, \omega_2) \mu_2(d\omega_2), \end{aligned} \quad (2.5.11)$$

where all integrals on the right-hand side exist in $[0, \infty)$ due to (2.5.10) and

$$(\operatorname{Re} f)^+, (\operatorname{Re} f)^-, (\operatorname{Im} f)^+, (\operatorname{Im} f)^- \in [0, |f|]. \quad (2.5.12)$$

Thus, (2.5.6) follows since restricted to A_1 , all integrals on the right-hand side of (2.5.11) are finite and in $\mathcal{M}^+(A_1, \mathcal{F}_1|_{A_1})$. The bounds (2.5.12) in combination with (2.5.9) imply that linearity of the integral supplies us with

$$\begin{aligned} \int_{A_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) &= \int_{A_1} \left(\int_{\Omega_2} (\operatorname{Re} f)^+(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\ &\quad - \int_{A_1} \left(\int_{\Omega_2} (\operatorname{Re} f)^-(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\ &\quad + i \int_{A_1} \left(\int_{\Omega_2} (\operatorname{Im} f)^+(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1) \\ &\quad - i \int_{A_1} \left(\int_{\Omega_2} (\operatorname{Im} f)^-(\omega_1, \omega_2) \mu_2(d\omega_2) \right) \mu_1(d\omega_1), \end{aligned}$$

where the last equality follows from the fact that all integrands of the outer integrals are non-negative.

We may now apply Tonelli's theorem to each summand on the right-hand side of the last display to deduce that it equals

$$\begin{aligned} &\int_{\Omega_1 \times \Omega_2} (\operatorname{Re} f)^+ d\mu_1 \otimes \mu_2 \\ &\quad - \int_{\Omega_1 \times \Omega_2} (\operatorname{Re} f)^- d\mu_1 \otimes \mu_2 \\ &\quad + \int_{\Omega_1 \times \Omega_2} (\operatorname{Im} f)^+ d\mu_1 \otimes \mu_2 \\ &\quad - \int_{\Omega_1 \times \Omega_2} (\operatorname{Im} f)^- d\mu_1 \otimes \mu_2 \\ &= \int_{\Omega_1 \times \Omega_2} f d\mu_1 \otimes \mu_2, \end{aligned}$$

where the last equality follows from the finiteness of the integrals on the left-hand side and the linearity of the integral. This implies the first equality in (2.5.8).

The remaining statements are obtained by exchanging the roles of $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$. \square

2.6 Fourier transform / characteristic functions

Recall the notation for image measures introduced in Theorem 1.5.6 and Definition 1.5.7.

Definition 2.6.1. Let μ be a finite measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Its characteristic function is defined via the Fourier transform

$$\varphi_\mu(t) := \int_{\mathbb{R}^d} e^{it \cdot x} \mu(dx) = \int_{\mathbb{R}^d} \cos(t \cdot x) \mu(dx) + i \int_{\mathbb{R}^d} \sin(t \cdot x) \mu(dx).$$

The characteristic function of a random variable X taking values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is defined as the characteristic function of its distribution:

$$\varphi_X(t) := \varphi_{\mathbb{P} \circ X^{-1}}(t) = \int_{\mathbb{R}^d} e^{it \cdot x} \mathbb{P}_X(dx) \stackrel{\text{Thm. 2.2.19}}{=} \mathbb{E}[e^{it \cdot X}]$$

For the next example we need the following result which is interesting and useful in its own right.

Proposition 2.6.2 (Interchange of integration and differentiation). Let I be an interval containing more than one point, $t_0 \in I$, let $(\Omega, \mathcal{F}, \mu)$ be a measure space, and let $f : I \times X \rightarrow \mathbb{C}$ with the following properties:

- (a) for all $t \in I$, the function $f(t, \cdot)$ is integrable;
- (b) $\frac{\partial f}{\partial t}(t_0, \omega)$ exists for all $\omega \in \Omega$;
- (c) there exists a neighborhood U of t_0 as well as $g \in \mathcal{M}^+(\Omega, \mathcal{F})$ integrable such that for all $t \in U \cap I$ with $t \neq t_0$, one has for μ -a.a. $\omega \in \Omega$ that

$$\left| \frac{f(t, \omega) - f(t_0, \omega)}{t - t_0} \right| \leq g(\omega).$$

Then the function $F : I \ni t \mapsto \int_{\Omega} f(t, \omega) \mu(d\omega)$ is (at least one-sidedly) differentiable in t_0 , the function $\frac{\partial f}{\partial t}(t_0, \cdot)$ is integrable, and

$$F'(t_0) = \int_{\Omega} \frac{\partial f}{\partial t}(t_0, \omega) \mu(d\omega).$$

Proof. See exercise 3 on homework sheet 10. \square

Remark 2.6.3. Under the appropriate assumptions, important computational tools that we got to know for real-valued functions remain valid for complex-valued functions also. This applies to the Fundamental Theorem of Calculus, integration by parts, etc., where we can essentially prove the respective results by decomposing a complex-valued function into its real and imaginary parts, and then perform the proof for each of these parts.

Exemplifying we go through the example of the Fundamental Theorem of Calculus here: Let $f : [a, b] \rightarrow \mathbb{C}$ a continuous complex-valued function with $f = g + ih$ and g, h real-valued functions, and $F = G + iH$ an antiderivative to F with G and H antiderivatives to g and h . Then $\int_{[a, b]} f d\lambda = \int_{[a, b]} g d\lambda + i \int_{[a, b]} h d\lambda = G(b) - G(a) + i(H(b) - H(a)) = F(b) - F(a)$, where the penultimate equality follows from the Fundamental Theorem of Calculus.

Example 2.6.4. Let $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$, and let $X \sim \mathcal{N}(\mu, \sigma^2)$. In order to compute φ_X we first of all note that using the previous result on integration by measures with densities (see Proposition 2.2.9) and using substitution,

$$\varphi_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{itx} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{it(\sigma x + \mu)} dx = e^{it\mu} \varphi_Y(\sigma t) \quad (2.6.1)$$

where $Y \sim \mathcal{N}(0, 1)$. Using Proposition 2.6.2 we can then differentiate $\varphi_Y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} e^{itx} dx$ to get that

$$\varphi_Y'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \underbrace{e^{-x^2/2}(-x)}_{=:u'(x)} \underbrace{(-i)e^{itx}}_{=:v(x)} dx.$$

We can continue using integration by parts to obtain

$$\varphi_Y'(t) = \frac{1}{\sqrt{2\pi}} \left(\underbrace{e^{-x^2/2}(-i)e^{itx}}_{=0} \Big|_{x=-\infty}^{\infty} - \int_{-\infty}^{\infty} e^{-x^2/2} t e^{itx} dx \right) = -t \varphi_Y(t) \quad (2.6.2)$$

and note that

$$\varphi_Y(0) = \mathbb{E}[e^0] = 1. \quad (2.6.3)$$

From the theory of ODEs (i.e., Analysis II in your case, cf. Existence and Uniqueness Theorem of Picard-Lindelöf, [Wal93, Theorem II.6.I]) we know that the initial value problem given by (2.6.2) and (2.6.3) has a unique solution which is given by

$$\varphi_Y(t) = e^{-t^2/2}.$$

Plugging this into (2.6.1), we obtain

$$\varphi_X(t) = e^{i\mu t - (\sigma t)^2/2}.$$

The above considerations generalize to X a d -dimensional (μ, Σ) -distributed random variable ($\mu \in \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d \times d}$ positive definite) to obtain

$$\varphi_X(t) = e^{i\mu \cdot t - t \cdot (\Sigma t)/2}, \quad t \in \mathbb{R}^d.$$

Theorem 2.6.5. Any finite measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is uniquely characterized by its characteristic function.

The proof of this result will be given later on (see page 96), when we have a better probabilistic understanding of its tools.

Chapter 3

Classical and basic results in probability theory

A large chunk of this chapter will be based on the lecture notes [Dre18] accompanying the course ‘Introduction to Stochastics’ which can be found [here](#). As a consequence, we will not repeat proofs of results that are proven in the same way as in [Dre18] but rather refer to that source instead. In particular, you might want to have a look at [Dre18, Section 1.2] for motivating the concept of a probability.

As regards to other sources, [Kle14], [Kal02] and [Dur10] make particularly good reads for foundations of probability theory. All three sources cover significantly more than what we can hope for in this course.

3.1 Specific distributions

When putting our previous setting of measure theory and integration into a probabilistic context, we will usually consider some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to be given. Random experiments will then be described via random variables $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E, \mathcal{E})$ as defined in Definition 1.4.1. Furthermore, elements of the form $\{X \in A\} := \{\omega \in \Omega : X(\omega) \in A\} (\in \mathcal{F}), A \in \mathcal{E}$, will be called *events*, and they will constitute those outcomes of (random) experiments that we will be able to assign a probability to. See also [Dre18, Example 1.2.1]. In fact, this example (as well as Remark 3.1.1 below) also exemplifies that we will oftentimes and without loss of generality choose $(\Omega, \mathcal{F}) := (E, \mathcal{E})$; the latter is usually naturally given by the model of the experiment, whereas the former may as well be some abstract space lurking in the background. What is important to us, however, is the image measures $\mathbb{P} \circ X^{-1}$.

The following observation will prove useful in the next sections when we introduce various different distributions. This is largely taken from the corresponding [section](#) [Dre18, Section 1.8]. We refer to that source for further examples and motivation.

Remark 3.1.1. *Given any distribution μ (i.e., a probability measure on (E, \mathcal{E})) one can construct a random variable X with law μ as follows. Take (E, \mathcal{E}, μ) as the underlying probability space and choose $X : E \ni \omega \mapsto \omega \in E$ to be the identity on E . Then X defines a random variable from (E, \mathcal{E}, μ) to (E, \mathcal{E}) with law μ .*

In particular, as a consequence of this remark, if we want to describe an random experiment whose outcome is a value in E , then we can choose (E, \mathcal{E}, μ) as the underlying probability space for suitable \mathcal{E} and μ .

3.1.1 Discrete distributions

We recall the notion of a distribution introduced in Definition 1.5.7, and we also repeat the definition [Dre18, Definition 1.3.8] of the δ - or Dirac-measure.

Definition 3.1.2. Let (E, \mathcal{E}) be a measurable space. For $x \in E$, the Dirac measure / Dirac distribution / delta measure in x is defined via

$$\begin{aligned}\delta_x : \mathcal{E} &\rightarrow [0, 1]. \\ F &\mapsto \mathbf{1}_F(x).\end{aligned}$$

We will call any distribution on a measurable space (E, \mathcal{E}) which is of the form

$$\sum_{n \in \mathbb{N}} \alpha_n \delta_{x_n},$$

where $x_n \in E$ and $\alpha_n \geq 0$ with $\sum_{n \in \mathbb{N}} \alpha_n = 1$ a *discrete distribution*. Similarly, we call any random variable which has a discrete distribution a *discrete random variable*.

In the remaining part of this section, if not mentioned otherwise explicitly, we will always assume the underlying probability space to be $(\Omega, \mathcal{F}, \mathbb{P})$, and random variables map to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Example 3.1.3. A random variable X is called Bernoulli distributed with parameter $p \in [0, 1]$ (named after the Swiss mathematician Jacob Bernoulli (1655–1705)) if

$$\mathbb{P}(X = 1) = p, \quad \text{and} \quad \mathbb{P}(X = 0) = 1 - p.$$

In this case one writes $X \sim \text{Ber}_p$ and the law / distribution $\mathbb{P} \circ X^{-1}$ is referred to as the Bernoulli distribution Ber_p which, using Definition 3.1.2, can be written as

$$\text{Ber}_p = p\delta_1 + (1 - p)\delta_0.$$

A random variable that is Bernoulli distributed describes a coin flip (biased if $p \neq 1/2$), for example. Assume w.l.o.g. that the coin shows heads with probability p and tails with probability $1 - p$.

Example 3.1.4. A random variable X is called Binomially distributed with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$, if

$$\text{for each } k \in \{0, 1, \dots, n\} \text{ one has } \mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}. \quad (3.1.1)$$

In this case, one writes $X \sim \text{Bin}_{n,p}$ and its distribution is referred to as the Binomial distribution $\text{Bin}_{n,p}$, which can be written as

$$\text{Bin}_{n,p} = \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} \delta_k.$$

Example 3.1.5. A random variable X is called geometrically distributed with success parameter $p \in (0, 1)$, if

$$\text{for all } k \in \mathbb{N} \text{ one has } \mathbb{P}(X = k) = p(1 - p)^{k-1}. \quad (3.1.2)$$

In this case we write $X \sim \text{Geo}_p$, and its distribution is referred to as the Geometric distribution Geo_p , which can be written as

$$\text{Geo}_p = \sum_{k=1}^{\infty} p(1 - p)^{k-1} \delta_k.$$

Remark 3.1.6. Some authors call X geometrically distributed if instead of (3.1.2),

$$\text{for all } k \in \mathbb{N}_0 \text{ one has } \mathbb{P}(X = k) = p(1 - p)^k.$$

Example 3.1.7. A random variable X is called Poisson distributed with parameter $\nu > 0$ if $X : \Omega \rightarrow \mathbb{N}_0$ and

$$\mathbb{P}(X = k) = e^{-\nu} \frac{\nu^k}{k!} \quad \forall k \in \mathbb{N}_0.$$

In this case we write $X \sim \text{Poi}_\nu$, and its distribution is referred to as the Poisson distribution Poi_ν (named after the French mathematician Siméon Denis Poisson (1781 – 1840)), which can be written as

$$\text{Poi}_\nu = e^{-\nu} \sum_{k=0}^{\infty} \frac{\nu^k}{k!} \delta_k.$$

Poisson distributed random variables are e.g. used to describe the number of customers that have called a customer service center in a certain time interval. The reason for such a description being feasible is given by Theorem 3.1.8 below.

Theorem 3.1.8 (Poisson limit theorem). Let (p_n) be a sequence of numbers from $[0, 1]$ such that the limit $\nu := \lim_{n \rightarrow \infty} np_n$ exists in $(0, \infty)$. Then for each $k \in \mathbb{N}_0$,

$$\lim_{n \rightarrow \infty} \text{Bin}_{n,p_n}(k) = \text{Poi}_\nu(k).$$

Proof. For $k \in \mathbb{N}_0$ fixed we have

$$\text{Bin}_{n,p_n}(k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n!}{k!(n-k)!} \frac{(p_n n)^k}{n^k} \left(1 - \frac{p_n n}{n}\right)^{n-k} \xrightarrow{n \rightarrow \infty} \frac{\nu^k}{k!} e^{-\nu} = \text{Poi}_\nu(k).$$

□

This result explains the fact that the Poisson distribution is used for modeling e.g. the number of customers that contact a call center during a certain time interval: We partition the time interval into n subintervals of equal width, and as we take n to infinity, it is reasonable to assume that in any of the subintervals either zero or one customers are calling. Due to symmetry, it furthermore seems reasonable to assume that the probability of a customer calling in a subinterval has a probability decaying like p/n some $p \in (0, \infty)$, and that the fact that a customer has called during one subinterval does not influence the probabilities that a customer is calling during another time interval.¹ Thus, the probability of k customers calling during the original time interval should be approximated by $\text{Bin}_{n,p/n}(k)$ if n is large. The above Theorem 3.1.8 now shows that the Binomial distribution is the right candidate for this.

Example 3.1.9. Let $N \in \mathbb{N}$, and $K, n \in \{0, 1, \dots, N\}$. A random variable X is called hypergeometrically distributed with parameters K, N, n if $X : \{0, 1, \dots, N\} \rightarrow \{0, 1, \dots, N\}$ with

$$\mathbb{P}(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}} \quad \text{for } k \in \{0 \vee n + K - N, \dots, n \wedge K\}, \quad (3.1.3)$$

and $\mathbb{P}(X = k) = 0$ otherwise.

In this case we write $X \sim \text{Hyp}(N, K, n)$, and its distribution is referred to as the Hypergeometric distribution $\text{Hyp}_{N,K,n}$ with parameters N, K , and n .

Example 3.1.10. Let $X \sim \text{Geo}_p$. Then the distribution function of X is given by

$$F_X(t) = \begin{cases} 0, & \text{if } t < 1, \\ \sum_{j=1}^{\lfloor t \rfloor} p(1-p)^{j-1} = p \frac{1-(1-p)^{\lfloor t \rfloor}}{1-(1-p)} = 1 - (1-p)^{\lfloor t \rfloor}, & \text{if } t \geq 1. \end{cases}$$

Exercise 3.1.11. If X is a discrete real random variable, then F_X has jumps exactly at the points in $\{x \in X(\Omega) : \mathbb{P}(X^{-1}\{x\}) > 0\}$ and is constant otherwise.

¹These are slightly delicate issues; in fact, if the customer center in question is e.g. that of an energy retailer and there is a power outage during some part of the time interval we consider, then these assumptions will generally not be met. However, they seem reasonable to assume during normal operation.

3.1.2 Distributions with densities

Definition 3.1.12. A function $f \in \mathcal{M}^+(\Omega, \mathcal{F}, \mu)$ with the property that

$$\int_{\mathbb{R}} f \, d\mu = 1$$

is called a probability density (‘Wahrscheinlichkeitsdichte’).

For the time being we will mostly be interested in the case $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$.

We will call any distribution on (Ω, \mathcal{F}) which is absolutely continuous with respect to λ^d a *continuous distribution*. Similarly, we call any random variable which has a continuous distribution a *continuous random variable*.

Also, we remark in passing that the distinction between discrete and continuous distributions is not as essential anymore as it used to be in the introductory lecture. This is because we now have one comprising framework for discrete and continuous distributions, since both of them can be considered as probability measures on (Ω, \mathcal{F}) now.

Example 3.1.13. (a) For $a, b \in \mathbb{R}$ with $a < b$ the uniform distribution (‘Gleichverteilung’) on the interval $[a, b]$ has the density

$$\mathbb{R} \ni x \mapsto \frac{1}{b-a} \mathbb{1}_{[a,b]}(x).$$

We write $\text{Uni}([a, b])$ for the uniform distribution on the interval $[a, b]$, and the corresponding distribution function is given by

$$F(t) = \begin{cases} 0, & \text{if } t \leq a, \\ \frac{t-a}{b-a}, & \text{if } t \in (a, b), \\ 1, & \text{if } t \geq b. \end{cases}$$

(b) Let $\kappa \in (0, \infty)$. The exponential distribution (‘Exponentialverteilung’) with parameter κ has density

$$\mathbb{R} \ni x \mapsto \begin{cases} \kappa e^{-\kappa x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

We write $X \sim \text{Exp}(\kappa)$ if X is a random variable that is exponentially distributed with parameter $\kappa > 0$.

(c) The normal or Gaussian distribution (‘Normalverteilung’ or ‘Gaußverteilung’, named after the German mathematician Carl Friedrich Gauss (1777–1855)) with parameters $\mu \in \mathbb{R}$ and $\sigma^2 \in (0, \infty)$ (seen in Example 1.5.8 already) has the density

$$\mathbb{R} \ni x \mapsto \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

and we had agreed to write $X \sim \mathcal{N}(\mu, \sigma^2)$ if X is a random variable that is normally distributed with parameters μ and σ^2 .

It should also be noted here that the cumulative distribution function of the standard Normal distribution $\mathcal{N}(0, 1)$ is usually denoted by

$$\Phi(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} \, dx, \quad (3.1.4)$$

and that there is no closed expression for general values of t for the right-hand side. There are, however, tables to look up those values for a variety of different values for t .

We will get back to those distributions after having introduced the concept of expectation.

3.2 Independence

A key concept in probability theory is the notion of ‘independence’, which in the setting of events has been introduced as follows in [Dre18, Def. 1.6.1], see [here](#). In fact, there are even people saying that the concept of independence is the principal distinction of probability from measure theory.

The motivation for the definition of independence has been the following definition of the conditional probability.

Definition 3.2.1. *Let $F, G \in \mathcal{F}$ be such that $\mathbb{P}(G) > 0$. Then we define the conditional probability of F given G as*

$$\mathbb{P}(F | G) := \frac{\mathbb{P}(F \cap G)}{\mathbb{P}(G)}. \quad (3.2.1)$$

In terms of the interpretation of relative frequencies given in [Dre18, Section 1.2], this means that if $\mathbb{P}(F | G) = \mathbb{P}(F)$ (i.e., if $\mathbb{P}(F \cap G) = \mathbb{P}(F)\mathbb{P}(G)$), then the (limiting) relative frequency of F is not changed if we restrict to those experiments for which G occurs.

This gave rise to the following definition.

Definition 3.2.2. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, events $A, B \in \mathcal{F}$ are called independent if*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

As it turns out, we will need a more general concept of independence as introduced in the following definition.

Definition 3.2.3. *A family (\mathcal{E}_λ) , $\lambda \in \Lambda$, with $\mathcal{E}_\lambda \subset \mathcal{F}$ is called independent if for any $J \subset \Lambda$ finite and any choice of $F_j \in \mathcal{E}_j$ for $j \in J$, one has*

$$\mathbb{P}\left(\bigcap_{j \in J} F_j\right) = \prod_{j \in J} \mathbb{P}(F_j). \quad (3.2.2)$$

An important special case is when $\mathcal{E}_\lambda = \{F_\lambda\}$ for all $\lambda \in \Lambda$, and with $F_\lambda \in \mathcal{F}$. In this case we say that the family of events (F_λ) , $\lambda \in \Lambda$, is independent.

Remark 3.2.4. *The family (\mathcal{E}_λ) , $\lambda \in \Lambda$, is independent if and only if for any $J \subset \Lambda$ finite, the family (\mathcal{E}_λ) , $\lambda \in J$, is independent.*

Proposition 3.2.5. *Let (\mathcal{E}_λ) , $\lambda \in \Lambda$ be an independent family with $\mathcal{E}_\lambda \subset \mathcal{F}$. Then the family $(\delta(\mathcal{E}_\lambda))$, $\lambda \in \Lambda$, of Dynkin systems is also independent.*

Proof. From Remark 3.2.4 we deduce that we can assume Λ to be finite.

For $\lambda' \in \Lambda$ arbitrary but fixed define $\mathcal{D}_{\lambda'}$ to be the set of all $F \in \mathcal{F}$ such that if we replace $\mathcal{E}_{\lambda'}$ by $\{F\}$, then the resulting family (\mathcal{E}_λ) , $\lambda \in \Lambda$, is still independent. Then $\mathcal{D}_{\lambda'}$ is a Dynkin system (exercise).

Now we have $\mathcal{E}_{\lambda'} \subset \mathcal{D}_{\lambda'}$, and hence also $\delta(\mathcal{E}_{\lambda'}) \subset \mathcal{D}_{\lambda'}$. Thus, by definition of the independence property, we deduce that the family we obtain when replacing $\mathcal{E}_{\lambda'}$ by $\delta(\mathcal{E}_{\lambda'})$ is still independent. Repeating this step for each remaining $\lambda \in \Lambda \setminus \{\lambda'\}$, the result follows. \square

Corollary 3.2.6. *If (\mathcal{E}_λ) , $\lambda \in \Lambda$ is an independent family such that each \mathcal{E}_λ is a π -system, then $(\sigma(\mathcal{E}_\lambda))$, $\lambda \in \Lambda$ also is an independent family.*

Proof. This follows from Proposition 3.2.5 in combination with the π - λ -Theorem 1.1.32. \square

Proposition 3.2.7. *Let (\mathcal{E}_λ) , $\lambda \in \Lambda$, be an independent family such that each \mathcal{E}_λ is a π -system. Consider a partition of Λ into subsets J_i , $i \in I$, and denote*

$$\mathcal{F}_i := \sigma\left(\bigcup_{j \in J_i} \mathcal{E}_j\right).$$

Then (\mathcal{F}_i) , $i \in I$, is an independent family.

Proof. For $i \in I$ denote

$$\tilde{\mathcal{E}}_i := \left\{ E_{j_1} \cap \dots \cap E_{j_n} : n \in \mathbb{N}, \{j_1, \dots, j_n\} \subset J_i, \text{ and } E_{j_k} \in \mathcal{E}_{j_k} \forall k \in \{1, \dots, n\} \right\}.$$

Since each \mathcal{E}_λ is a π -system, so is each $\tilde{\mathcal{E}}_i$ and hence $\sigma(\tilde{\mathcal{E}}_i) = \delta(\tilde{\mathcal{E}}_i)$. Since the $(\tilde{\mathcal{E}}_i)$, $i \in I$, still form an independent family, the result follows with Corollary 3.2.6. \square

Definition 3.2.8. *A family of random variables (X_λ) , $\lambda \in \Lambda$, is called independent if the family of σ -algebras $\sigma(X_\lambda)$, $\lambda \in \Lambda$, is independent.*

We refer to [Dre18] on more background for the concept of independence, in particular see [Dre18, Remark 1.7.15].

In [Dre18, Claim 1.8.4] the following claim had been derived, which you might want to try your hands at (without looking it up) if you haven't seen it before.

Claim 3.2.9. *The sum*

$$S_n := \sum_{j=1}^n X_j$$

of independent random variables X_1, \dots, X_n , each distributed according to Ber_p , is distributed according to $\text{Bin}_{n,p}$.

The concept of a family of random variables that are independent and all have the same distribution is so important that it has its own name.

Definition 3.2.10. *A family of random variables (X_λ) , $\lambda \in \Lambda$, is called independent identically distributed (i.i.d.) ('unabhängig identisch verteilt' (u.i.v.)), if*

- (a) *the family (X_λ) , $\lambda \in \Lambda$, is an independent family of random variables, and*
- (b) *if the X_λ , $\lambda \in \Lambda$, all have the same distribution.*

The Borel-Cantelli lemmas

In order to prove this theorem we need some further results, which are important and of interest on their own.

Lemma 3.2.11 (Borel-Cantelli lemma (Italian mathematician Francesco Paolo Cantelli (1875–1966))). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and assume given a sequence (A_n) of events $A_n \in \mathcal{F}$.*

(a) *If*

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) < \infty, \tag{3.2.3}$$

we have

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$$

(b) If

$$\sum_{n \in \mathbb{N}} \mathbb{P}(A_n) = \infty,$$

and if in addition the (A_n) are independent, then

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.$$

The [proof](#) is that of [[Dre18](#), Lemma 1.12.6].

Remark 3.2.12. It is important to note here that the independence assumption in part (b) of Lemma 3.2.11 cannot be dropped. To see this, consider for example a single fair coin toss modeled on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and denote for all $n \in \mathbb{N}$ by A_n the event that the coin shows tails. Then $\mathbb{P}(A_n) = \frac{1}{2}$ for all $n \in \mathbb{N}$, so $\sum_{n \in \mathbb{N}_0} \mathbb{P}(A_n) = \infty$, but $\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = \mathbb{P}(A_n) = \frac{1}{2} \neq 1$.

Example 3.2.13. (a) A popular application is the so-called ‘infinite monkey theorem’. It states that a monkey which is randomly hitting keys (in an i.i.d. fashion, and such that any key, lower and upper case, has a positive probability of being hit) of a computer keyboard will almost surely type any given text, such as e.g. Tolstoy’s ‘War and Peace’. It is left to the reader to make this statement more precise. k

(b) Consider a sequence (X_n) of independent random variables such that

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2} \frac{1}{n \ln(n+1)}$$

and

$$\mathbb{P}(X_n = 0) = 1 - \frac{1}{2} \frac{1}{n \ln(n+1)}.$$

Then, setting $A_n := \{\frac{|X_n|}{n} \geq 1\}$, we obtain

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \sum_{n=1}^{\infty} \frac{1}{n \ln(n+1)} = \infty,$$

where the latter equality can be shown by Cauchy’s condensation test (‘Cauchy’sches Verdichtungskriterium’). Therefore, the sequence (A_n) fulfills the condition of the second part of the Borel-Cantelli lemma. We will come back to this example in the context of the law of large numbers.

Kolmogorov’s 0 – 1-law

Definition 3.2.14. Let (\mathcal{F}_n) be a sequence of σ -algebras with $\mathcal{F}_n \subset \mathcal{F}$. We define the corresponding tail- σ -algebra (‘terminale σ -Algebra’) as

$$\mathcal{T} := \mathcal{T}((\mathcal{F}_n)) := \bigcap_{n=1}^{\infty} \sigma\left(\bigcup_{m \geq n} \mathcal{F}_m\right).$$

The intuition is the following: An event $A \in \mathcal{F}$ is contained in the tail- σ -algebra if in order to decide whether or not it occurs (i.e., whether or not $\omega \in A$) we can discard the ‘information’ from finitely many of the \mathcal{F}_n . It becomes more clear in the context of an example.

Example 3.2.15. Let (X_n) be a sequence of random variables and consider the corresponding tail- σ -algebra

$$\mathcal{T} := \mathcal{T}((\sigma(X_n))).$$

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

Then we do for example have that (check!)

•

$$\left\{ \lim_{n \rightarrow \infty} S_n \text{ exists} \right\} \in \mathcal{T} \quad \text{and}$$

•

$$\left\{ \lim_{n \rightarrow \infty} \frac{S_n}{n} = 0 \right\} \in \mathcal{T}.$$

Theorem 3.2.16 (Kolmogorov's 0 – 1-law). *Let (\mathcal{F}_n) be a sequence which is an independent family of σ -algebras with $\mathcal{F}_n \subset \mathcal{F}$. Then the tail- σ -algebra \mathcal{T} is \mathbb{P} -trivial, i.e.,*

$$\mathbb{P}(A) \in \{0, 1\} \quad \forall A \in \mathcal{T}.$$

Proof. We are going to show that for all $A \in \mathcal{T}$,

$$\mathbb{P}(A) = \mathbb{P}(A)^2, \tag{3.2.4}$$

which will imply the result.

For this purpose, for arbitrary fixed $A \in \mathcal{T}$ define

$$\mathcal{D} := \{D \in \mathcal{F} : \mathbb{P}(A \cap D) = \mathbb{P}(A)\mathbb{P}(D)\}.$$

Our strategy is to show that \mathcal{D} is a Dynkin system with

$$\mathcal{T} \subset \mathcal{D}, \tag{3.2.5}$$

which in particular will imply (3.2.4).

The fact that \mathcal{D} is a Dynkin system is shown along the by now standard lines and we will not go into further detail. In order to establish (3.2.5), we observe that due to Proposition 3.2.7 and the fact that $A \in \mathcal{T}$ we have

$$\sigma_n := \sigma\left(\bigcup_{k=1}^n \mathcal{F}_k\right) \subset \mathcal{D},$$

so we also obtain

$$\sigma_\infty := \bigcup_{n \in \mathbb{N}} \sigma_n \subset \mathcal{D},$$

and consequently

$$\delta(\sigma_\infty) \subset \mathcal{D}. \tag{3.2.6}$$

But since the σ_n are non-decreasing in n , we deduce that σ_∞ is a π -system, so by the π - λ -Theorem we infer

$$\delta(\sigma_\infty) = \sigma(\sigma_\infty). \tag{3.2.7}$$

On the other hand, $\cup_{n \in \mathbb{N}} \mathcal{F}_n \subset \sigma(\sigma_\infty)$, and therefore in particular also $\mathcal{T} \subset \sigma(\sigma_\infty)$. Combining this with (3.2.7) and (3.2.6), we infer (3.2.5) which finishes the proof. \square

3.3 Covariance, variance

Definition 3.3.1. For $X \in \mathcal{L}^1$, the expression on the right-hand side of

$$\text{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] \in [0, \infty]$$

is called the variance of X .

From the expression on the right-hand side it is clear that the variance is always non-negative, since the random variable in the expectation on the right-hand side is non-negative. Furthermore, this expression shows that the variance gauges the expected quadratic deviation of X from its expectation $\mathbb{E}[X]$. It is a simple measure for how strongly the random variable X fluctuates around its mean.

Using the linearity of expectation, we can rewrite the variance as

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2,$$

which holds true in the case $\mathbb{E}[X^2] = \infty$ as well. Thus, we immediately obtain the following corollary.

Corollary 3.3.2. *For $X \in \mathcal{L}^1$, we have $\text{Var}(X) < \infty$ if and only if $\mathbb{E}[X^2] < \infty$.*

Definition 3.3.3. *The covariance of two random variables X and Y is defined as*

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - 2\mathbb{E}[X]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \quad (3.3.1)$$

if the right-hand side is well-defined in $[-\infty, \infty]$.

The two random variables are called uncorrelated if $\text{Cov}(X, Y) = 0$.

Again we note that variance and covariance only depend on the random variables involved through their corresponding distributions.

In some sense the covariance $\text{Cov}(X, Y)$ tells us how strongly X and Y are correlated, i.e., how strongly they *tend* to ‘change together’. If both X and Y tend to take values above their expectation on the same subset of Ω , and also tend to take values below their expectations on similar sets, then according to (3.3.1) this should imply that their covariance is positive; on the other hand, if X tends to take values above its expectation on subsets of Ω where Y tends to take values below its expectation, and vice versa, then this would suggest that their covariance is negative. Therefore, if X and Y are independent one might possibly guess that $\text{Cov}(X, Y)$ vanishes. This is indeed the case as Theorem 3.3.6 below shows. Note, however, that the converse is not generally true as will be asked to show in Exercise 3.3.8.

We now collect some properties of covariances and variances in the following result.

Proposition 3.3.4. *Let X and Y be random variables with $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$, and let $a, b, c, d \in \mathbb{R}$. Then*

(a)

$$\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y);$$

in particular,

$$\text{Var}(a(X + b)) = a^2 \text{Var}(X); \quad (3.3.2)$$

(b)

$$|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}(X) \text{Var}(Y)};$$

Proof. • Using the linearity of expectation we get

$$\begin{aligned} \text{Cov}(aX + b, cY + d) &= \mathbb{E}[(aX + b - \mathbb{E}[aX + b])(cY + d - \mathbb{E}[cY + d])] \\ &= ac \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = ac \text{Cov}(X, Y). \end{aligned}$$

•

$$\begin{aligned} |\text{Cov}(X, Y)| &\leq \mathbb{E}[|X - \mathbb{E}[X]| \cdot |Y - \mathbb{E}[Y]|] \leq \mathbb{E}[(X - \mathbb{E}[X])^2]^{\frac{1}{2}} \mathbb{E}[(Y - \mathbb{E}[Y])^2]^{\frac{1}{2}} \\ &= \sqrt{\text{Var}(X) \text{Var}(Y)}, \end{aligned}$$

where the inequality is a consequence of the Cauchy-Schwarz inequality.²

□

Before continuing, we bring a small result which is easy to prove but nevertheless oftentimes important and helpful.

Claim 3.3.5. *Let $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E_1, \mathcal{E}_1)$ and $Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (E_2, \mathcal{E}_2)$ be two independent random variables. Then*

$$\mathbb{P}_{(X,Y)} = \mathbb{P}_X \otimes \mathbb{P}_Y. \quad (3.3.3)$$

Proof. Due to the independence assumption on X and Y , on the \cap -stable generator of rectangles of $\mathcal{E}_1 \otimes \mathcal{E}_2$ the two probability measures $\mathbb{P}_{(X,Y)}$ and $\mathbb{P}_X \otimes \mathbb{P}_Y$ coincide. Thus, Corollary 1.2.19 yields (3.3.3). □

Theorem 3.3.6. *Let $X, Y \in \mathcal{L}^1$ be independent random variables. Then $XY \in \mathcal{L}^1$ and*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]. \quad (3.3.4)$$

In particular, independent random variables are uncorrelated.

Proof. We recall the statement of Claim 3.3.5. Therefore, the change of variable formula Theorem 2.2.19 in combination with Tonelli's theorem implies that

$$\begin{aligned} \mathbb{E}[|XY|] &= \int_{\mathbb{R}} z \mathbb{P}_{|XY|}(dz) = \int_{\mathbb{R}^2} |xy| \mathbb{P}_{(X,Y)}(d(x,y)) = \int_{\mathbb{R}^2} |xy| \mathbb{P}_X \otimes \mathbb{P}_Y(d(x,y)) \\ &= \int_{\mathbb{R}^2} |xy| \mathbb{P}_X(dx) \mathbb{P}_Y(dy) = \int_{\mathbb{R}} |x| \mathbb{P}_X(dx) \int_{\mathbb{R}} |y| \mathbb{P}_Y(dy) = \mathbb{E}[|X|] \mathbb{E}[|Y|]. \end{aligned}$$

Now if $X, Y \in \mathcal{L}^1$, then the right-hand side (and therefore all expressions appearing) are finite. In particular, in this case we can remove the absolute value signs and obtain (reading the previous display from right to left, and replacing Tonelli by Fubini) that $XY \in \mathcal{L}^1$ as well as (3.3.4). □

Remark 3.3.7. *Iterating the above we obtain the following generalization of Theorem 3.3.6: Let X_1, \dots, X_n be a family of independent random variables which are either all in \mathcal{L}^1 or all non-negative. Then*

$$\mathbb{E}\left[\prod_{j=1}^n X_j\right] = \prod_{j=1}^n \mathbb{E}[X_j].$$

Exercise 3.3.8. *Find an example of real random variables X, Y which are uncorrelated but not independent.*

We now compute some variances of distributions we got to know earlier in this course.

Example 3.3.9. (a) *Let $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$ and $\sigma^2 \in (0, \infty)$. It is not hard to compute $\mathbb{E}[X] = \mu$ (see e.g. [Dre18, Example 1.9.6]). Then we get using Proposition 2.2.9 that*

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &\stackrel{x \mapsto \sigma x + \mu}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma x)^2 e^{-\frac{x^2}{2}} dx = \frac{\sigma^2}{\sqrt{2\pi}} \underbrace{\left(-xe^{-\frac{x^2}{2}} \right) \Big|_{x=-\infty}^{\infty}}_{=0} + \underbrace{\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx}_{=\sqrt{2\pi}} = \sigma^2, \end{aligned}$$

²Here, the Cauchy-Schwarz inequality is applied to the symmetric bilinear form defined via $\mathcal{L}^2 \times \mathcal{L}^2 \ni (f, g) \mapsto \int f \cdot g \, d\mu \in \mathbb{R}$ – which is not necessarily an inner product since we can have $(f, f) = 0$ even if $f \neq 0$ (and only $\mu(f \neq 0) = 0$) – however, the (standard) proof of the Cauchy-Schwarz for inner products does not depend on the missing implication $(f, f) = 0 \implies f = 0$.

where we used integration by parts for the penultimate equality. Hence, we observe that the second parameter in $\mathcal{N}(\mu, \sigma^2)$ denotes the variance of the random variable. In particular, this means that the normal distribution is completely distributed by its expectation and its variance.

Furthermore, we deduce that the standard normal distribution from Example 3.1.13 (c) has mean 0 and variance 1.

- (b) Let $X \sim \text{Geo}(p)$ for $p \in (0, 1)$. We first compute $\mathbb{E}[X]$ and for this purpose we take advantage of the following useful trick. For $q \in (-1, 1)$, the formula for the geometric series supplies us with

$$\sum_{j=1}^{\infty} q^j = \frac{q}{1-q}.$$

Since the left-hand side defines a power series that is absolutely convergent on $(-1, 1)$, we know from Analysis I that its derivative can be computed term by term. Thus, differentiating both sides of the equation gives

$$\sum_{j=1}^{\infty} j q^{j-1} = \frac{(1-q) - q(-1)}{(1-q)^2} = \frac{1}{(1-q)^2}. \quad (3.3.5)$$

Using this identity for $q = 1 - p$ we can compute

$$\mathbb{E}[X] = \sum_{j=1}^{\infty} j \mathbb{P}(X = j) = \sum_{j=1}^{\infty} j p (1-p)^{j-1} = \frac{p}{p^2} = \frac{1}{p}. \quad (3.3.6)$$

We now have to compute $\mathbb{E}[X^2]$. For this purpose we differentiate (3.3.5) once again (and again, the left-hand side can be differentiated term by term on $(-1, 1)$ due to its absolute convergence) to obtain

$$\sum_{j=1}^{\infty} j(j-1) q^{j-2} = \frac{2}{(1-q)^3}. \quad (3.3.7)$$

Thus, we get using the change of variable formula that

$$\begin{aligned} \mathbb{E}[X^2] &= \sum_{j=1}^{\infty} j^2 \mathbb{P}(X = j) = \sum_{j=1}^{\infty} j^2 p (1-p)^{j-1} \\ &= p(1-p) \sum_{j=1}^{\infty} j(j-1) (1-p)^{j-2} + p \sum_{j=1}^{\infty} j (1-p)^{j-1} = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}, \end{aligned}$$

where we took advantage of (3.3.6) and (3.3.7) to get the third equality. Thus, we can compute

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

If we want to compute the variance of the sum of random variables, the following result turns out to be useful by decomposing it into a sum of variances and corresponding covariances.

Proposition 3.3.10. *Let X_1, \dots, X_n be random variables in \mathcal{L}^2 . Then*

$$\text{Var}\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \text{Var}(X_j) + \sum_{1 \leq i, j \leq n, i \neq j} \text{Cov}(X_i, X_j).$$

Proof. Due to Proposition 3.3.4 (a), without loss of generality, we can assume $\mathbb{E}[X_i] = 0$ for all $1 \leq i \leq n$. Using the linearity of expectation we get

$$\begin{aligned} \text{Var} \left(\sum_{j=1}^n X_j \right) &= \mathbb{E} \left[\left(\sum_{j=1}^n X_j \right)^2 \right] - \underbrace{\left(\mathbb{E} \left[\sum_{j=1}^n X_j \right] \right)^2}_{=0 \text{ by assumption}} = \sum_{i,j=1}^n \mathbb{E}[X_i X_j] \\ &= \sum_{i=1}^n \text{Var}(X_i) + \sum_{1 \leq i,j \leq n, i \neq j} \text{Cov}(X_i, X_j). \end{aligned}$$

Note that

$$\mathbb{E}[|X_i X_j|] \leq (\mathbb{E}[X_i^2])^{\frac{1}{2}} (\mathbb{E}[X_j^2])^{\frac{1}{2}} < \infty$$

due to Cauchy-Schwarz' inequality, hence all expectations in the above equations are well-defined, and so are all the sums. \square

If the random variables in the above result turn out to be uncorrelated, all covariances in the above result vanish and the computation of the variance becomes significantly simpler. The corresponding result is used so often that it deserves its own name.

Corollary 3.3.11 (Bienaymé formula (Irénée-Jules Bienaymé (1796–1878), French probabilist and statistician)). *Let X_1, \dots, X_n be (pairwise) uncorrelated random variables in \mathcal{L}^2 . Then*

$$\text{Var} \left(\sum_{j=1}^n X_j \right) = \sum_{j=1}^n \text{Var}(X_j).$$

Example 3.3.12. (a) *Let $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be independent random variables. Then $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.*

Now you may have a look at [Dre18, Example 3.3.12] to convince yourself that it took a little bit of not so nice calculus to prove this.

Using Theorem 2.6.5 and Example 2.6.4 we are in the position to derive this result in a significantly neater way. Indeed, using the independence of X and Y in combination with Example 2.6.4 we compute

$$\begin{aligned} \varphi_{X+Y}(t) &= \mathbb{E}[e^{it(X+Y)}] \stackrel{\text{Thm. 3.3.6}}{=} \mathbb{E}[e^{itX}] \mathbb{E}[e^{itY}] = e^{i\mu_1 t - (\sigma_1^2 t^2)/2} e^{i\mu_2 t - (\sigma_2^2 t^2)/2} \\ &= e^{i(\mu_1 + \mu_2)t - ((\sigma_1^2 + \sigma_2^2)t^2)/2}. \end{aligned}$$

In combination with Theorem 2.6.5 and Example 2.6.4 it therefore follows that $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

(b) *Let $X \sim \text{Bin}_{n,p}$ for some $n \in \mathbb{N}$ and $p \in [0, 1]$. In Claim 3.2.9 we had seen that X has the same distribution as $\sum_{j=1}^n Y_j$, where the Y_j are independent random variables distributed according to Ber_p . Now $\text{Var}(Y_j)$ is easy to compute since $\mathbb{E}[Y_j] = p$ and $\mathbb{E}[Y_j^2] = p$. Thus, $\text{Var}(Y_j) = p(1 - p)$. Now since $\text{Var}(X)$ depends on X only through its distribution, we get the first equality of*

$$\text{Var}(X) = \text{Var} \left(\sum_{j=1}^n Y_j \right) = \sum_{j=1}^n \text{Var}(Y_j) = np(1 - p),$$

where in the second equality we used Corollary 3.3.11.

The following lemma is interesting in its own right, but a generalization of it will play an important role when we introduce the concept of conditional expectations (which heuristically will amount to averaging over partial information of \mathcal{F} only) in Section ?? below. It can be interpreted in the sense that the best approximation to a random variable X by a constant c is its expectation $c = \mathbb{E}[X]$ (if distance is measured in terms of the second moment of $X - c$).

Lemma 3.3.13. *Let $X \in \mathcal{L}^2$ be a random variable. Then the function*

$$\mathbb{R} \ni s \mapsto \mathbb{E}[(X - s)^2]$$

is minimized at $s = \mathbb{E}[X]$. In particular, we have $\mathbb{E}[(X - s)^2] \geq \text{Var}(X)$ for all $s \in \mathbb{R}$.

Proof. We compute using the linearity of expectation that

$$\mathbb{E}[(X - s)^2] = \mathbb{E}[X^2] - 2s\mathbb{E}[X] + s^2 = (\mathbb{E}[X^2] - \mathbb{E}[X]^2) + (\mathbb{E}[X] - s)^2.$$

From this it is obvious that the function attains its minimum for $s = \mathbb{E}[X]$, in which case it equals $\text{Var}(X)$. This finishes the proof. \square

3.4 L^p spaces and some fundamental inequalities

Definition 3.4.1. *Let $f \in \mathcal{M}$. We define its essential supremum as*

$$\text{ess sup } f := \inf\{M \in \mathbb{R} : \mu(f \geq M) = 0\},$$

with the standard convention $\inf \emptyset = \infty$.

Similarly, its essential infimum is defined as

$$\text{ess inf } f := \sup\{m \in \mathbb{R} : \mu(f \leq m) = 0\},$$

with the standard convention $\sup \emptyset = -\infty$.

Exercise 3.4.2. *Show that the essential supremum could be equivalently defined as*

$$\text{ess sup } f := \inf\{M \in \mathbb{R} : \mu(f > M) = 0\},$$

and similarly that

$$\text{ess inf } f = \sup\{m \in \mathbb{R} : \mu(f < m) = 0\},$$

Definition 3.4.3. *Let $p \in (0, \infty)$. For $f \in \mathcal{M}$ we define*

$$\|f\|_p := \left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}} \in [0, \infty].$$

In addition, set

$$\|f\|_{\infty} := \text{ess sup } |f| \in [0, \infty].$$

For $p \in (0, \infty]$ we then set

$$\mathcal{L}^p := \mathcal{L}^p(\Omega, \mathcal{F}, \mu) := \{f \in \mathcal{M} : \|f\|_p < \infty\},$$

which is consistent with the notation from Definition 2.0.18. Motivated by this definition one also uses the notation

$$\|f\|_{\mathcal{L}^p(\Omega, \mathcal{F}, \mu)} := \|f\|_p.$$

By $\|\cdot\|_p : \mathcal{L}^p \ni f \mapsto \|f\|_p$ we denote the mapping that maps functions to their respective norms.

Proposition 3.4.4. *For $p \in [1, \infty]$ the mapping $\|\cdot\|_p$ introduced in Definition 3.4.3 is a semi-norm on \mathcal{L}^p .*

For the proof of the triangle inequality we need another result which is important on its own.

Theorem 3.4.5 (Minkowski's inequality). *Let $f, g \in \mathcal{M}$ such that $f + g$ is well-defined (in the sense that ' $\infty - \infty$ ' does not occur). Then for any $p \in [1, \infty]$,*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \quad (3.4.1)$$

The proof of Theorem 3.4.5 will take advantage of yet another inequality that we have not studied so far.

Proof of Proposition 3.4.4. $\|\cdot\|_p$ maps from \mathcal{L}^p to $[0, \infty)$, so we only have to show that it is absolutely homogeneous and fulfills the triangle inequality. Absolute homogeneity ($\|cf\|_p = |c|\|f\|_p$ for all $c \in \mathbb{R}$ and $f \in \mathcal{L}^p$) follows from the linearity of the integral. The validity of the triangle inequality is a consequence of Minkowski's inequality. \square

While for $p \in [1, \infty]$ the above result in combination with the fact that \mathcal{L}^p is a vector space provides us with the fact that \mathcal{L}^p is actually a semi-normed vector space, it is not hard to observe that $\|\cdot\|_p$ does *not* in general define a norm on \mathcal{L}^p . Indeed, for any $f \in \mathcal{L}^p$ we can choose some $g \in \mathcal{L}^p$ such that $f \neq g$ and $\mu(f \neq g) = 0$ and get $\|f - g\|_p = 0$.

An elegant way out of this quandary is to consider an appropriate quotient space. To be precise,

$$\text{let } \mathcal{N} \text{ denote the set of all } f \in \mathcal{L}^p \text{ such that } \mu(f \neq 0) = 0. \quad (3.4.2)$$

Applying Lemma 2.2.3 we can deduce that

$$\mathcal{N} = \{f \in \mathcal{L}^p : \|f\|_p = 0\}. \quad (3.4.3)$$

It is not hard to show that \mathcal{N} forms a subspace of \mathcal{L}^p , and thus we can define the quotient space

$$L^p := \mathcal{L}^p / \mathcal{N} = \{\tilde{f} := f + \mathcal{N} : f \in \mathcal{L}^p\}.$$

Hence, elements of L^p are equivalence classes of functions in \mathcal{L}^p , and $f, g \in \mathcal{L}^p$ are in the same equivalence class (usually written $f \sim g$) if and only if $f - g \in \mathcal{N}$, i.e., according to (3.4.3),

$$\int |f - g|^p d\mu = 0.$$

Thus, in combination with Proposition 3.4.4

$$\|f\|_p \leq \|f - g\|_p + \|g\|_p = \|g\|_p,$$

and similarly we get

$$\|g\|_p \leq \|f\|_p,$$

so $\|f\|_p = \|g\|_p$. As a consequence, we obtain the following result.

Corollary 3.4.6. *For $p \in [1, \infty]$, the space L^p is a normed vector space with norm $\|\cdot\|_p$.*

Proof. Using Proposition 3.4.4 in combination with the fact that \mathcal{N} as introduced in (3.4.2) is a subspace of \mathcal{L}^p , we obtain that L^p is a semi-normed vector space. Since for $f \in \mathcal{L}^p$ we have $\|f\|_p = 0$ if and only if $f \in \mathcal{N}$, we deduce that $\|\cdot\|_p$ is definite on L^p , and hence the latter endowed with $\|\cdot\|_p$ is a normed vector space. \square

In a slight abuse of nomenclature, one usually also refers to elements of L^p as *functions*, although, strictly speaking, they are equivalence classes of functions. One reason for this is that in probability theory (and also functional analysis, where L^p spaces play an important role) people are most often mainly interested in the almost sure behaviour (recall Section 2.2.1), i.e., in properties that do not change if the random variable is modified on a set of measure zero; in particular, this implies that for $f \in L^p$, any representative $\tilde{f} \in \mathcal{L}^p$ of the equivalence class of f would have the same (almost everywhere) properties. An example that you have gotten to know already is the distribution of a random variable (recall Definition 1.5.1), which did not change if we modified a random variable on a set of measure zero.

In fact, one even has that L^p endowed with $\|\cdot\|_p$ is complete.

Theorem 3.4.7. *For $p \in [0, \infty]$, the vector space L^p endowed with $\|\cdot\|_p$ is a Banach space.*

Since this result is not central to this class, we refer to the proof of [Els05, Korollar 2.6] for a proof.

We now give some further properties of the above spaces, most of which have been derived in [Dre18] already. Since for $0 < p \leq q$ we have $|x|^p \leq 1 + |x|^q$ for all $x \in \mathbb{R}$ we immediately obtain the inclusion

$$\mathcal{L}^q(\mu) \subset \mathcal{L}^p(\mu), \quad (3.4.4)$$

if μ is a finite measure.

Example 3.4.8. *Consider the measurable function*

$$f(x) = \mathbb{1}_{[1, \infty)}(x) \frac{1}{x}, \quad x \in \mathbb{R}.$$

Then Theorem 2.1.1 (MCT) implies that

$$\int_{\mathbb{R}} |f(x)|^p \lambda(dx) = \lim_{n \rightarrow \infty} \int_{[1, n]} |f(x)|^p \lambda(dx).$$

Then we can use Theorem 2.0.20 to deduce

$$\int_{\mathbb{R}} |f(x)|^p \lambda(dx) = \lim_{n \rightarrow \infty} \frac{1}{p-1} \left(-n^{-p+1} + 1 \right)$$

Thus, we see that if $p > 1$, then

$$\int_{\mathbb{R}} |f(x)|^p \lambda(dx) = \frac{1}{p-1} < \infty,$$

whereas for $p \in (0, 1)$,

$$\int_{\mathbb{R}} |f(x)|^p \lambda(dx) = \infty$$

(and the same applies for $p = 1$).

In particular, the right-hand side is infinite for $p \in (0, 1]$ and finite for $p \in (1, \infty)$. Thus, $f \in \mathcal{L}^p$ for $p \in (1, \infty)$ but $f \notin \mathcal{L}^p$ for $p \in (0, 1]$.

In order to prove the fundamental Hölder inequality below we will need the following auxiliary result.

Lemma 3.4.9 (Young's inequality (English mathematician William Henry Young (1863–1942))). *Let $a, b \in [0, \infty)$ and $p, q \in (1, \infty)$ such that*

$$\frac{1}{p} + \frac{1}{q} = 1. \quad (3.4.5)$$

Then

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \quad (3.4.6)$$

Proof. See the proof of [Dre18, Lemma 1.9.14]. □

Theorem 3.4.10 (Hölder inequality (German mathematician Otto Ludwig Hölder (1859–1937))). *Let $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then, for $f, g \in \mathcal{M}(\Omega, \mathcal{F}, \mu)$ one has*

$$\int_{\Omega} |f \cdot g| d\mu \leq \left(\int_{\Omega} |f|^p d\mu \right)^{\frac{1}{p}} \left(\int_{\Omega} |g|^q d\mu \right)^{\frac{1}{q}}. \quad (3.4.7)$$

Proof. This is the proof of [Dre18, Thm. 1.9.16], so we omit it here. \square

Remark 3.4.11. (a) In particular, if $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^q$, then $fg \in \mathcal{L}^1$.

(b) The special case of $p = q = 1/2$ gives a special case of the Cauchy-Schwarz (Augustin-Louis Cauchy (1789–1857), Hermann Schwarz (1843–1921) inequality you might know from linear algebra (or might get to know in functional analysis) for inner products.

(c) Hölder’s inequality not only holds for expectations (which will be interpreted as integration against probability measures in ‘Probability Theory I’) but also for more general integrals in.

We now have all the tools to prove Theorem 3.4.5, which will be part of the last homework sheet.

3.5 Convergence of random variables

Since this section introduces some core notions of probability theory, and in less generality this has been treated in the corresponding part [Dre18, Section 1.11], which can be found [here](#).

As in analysis, asymptotic investigation play a fundamental role in probability theory, in particular when it comes to the fundamental limit theorems that we will be investigating below. As a first step to build a theoretical base for this we will introduce the fundamental types of convergence that we will encounter in probability theory and give their dependencies.

In what follows, if not mentioned otherwise (S, d) is a separable metric space.

3.5.1 Almost sure convergence

This is one of the strongest types of convergence that we will consider, and we will introduce it for random variables taking values in separable metric spaces.

We will need an auxiliary result before giving the precise definition.

Lemma 3.5.1. *Let (S, d) be a metric space. If $X, Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (S, d)$ are two random variables, then the mapping $\Omega \ni \omega \mapsto d(X(\omega), Y(\omega))$ defines an $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable real valued random variable.*

Proof. We first of all note that the mapping $\varphi_1 : \Omega \ni \omega \mapsto (X(\omega), Y(\omega)) \in S \times S$ is $\mathcal{F} - (\mathcal{B}(S) \otimes \mathcal{B}(S))$ -measurable by definition of the product- σ -algebra (and the assumption that the X and Y are random variables). Furthermore, since d is a metric, the function

$$\varphi_2 : S \times S \ni (x, y) \mapsto d(x, y) \in [0, \infty) \quad (3.5.1)$$

is continuous. Therefore, as a consequence of Theorem 1.4.10, φ_2 is $\mathcal{B}(S \times S) - \mathcal{B}(\mathbb{R})$ -measurable. Since (S, d) is separable, Theorem 2.3.9 implies that $\mathcal{B}(S \times S) = \mathcal{B}(S) \otimes \mathcal{B}(S)$ and hence φ_2 is $\mathcal{B}(S \times S) - \mathcal{B}(\mathbb{R})$ -measurable also. Therefore, due Theorem 1.4.4, it follows that the composition $\varphi_2 \circ \varphi_1$, which equals $d(X, Y)$, is measurable and hence a random variable. \square

Definition 3.5.2. *Let (X_n) be a sequence of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and mapping into a metric space (S, d) , and let X be another such random variable. We say that X_n converges (\mathbb{P}) -almost surely (or a.s.) (‘fast sicher’ (or else ‘f.s.’) to X , and we write*

$$X_n \xrightarrow{\text{a.s.}} X \quad \text{as } n \rightarrow \infty,$$

or

$$\lim_{n \rightarrow \infty} X_n = X \quad \mathbb{P}\text{-a.s.},$$

if

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} d(X_n, X) = 0\right) = \mathbb{P}\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\right\}\right) = 1. \quad (3.5.2)$$

Remark 3.5.3. (a) Note that from Lemma 3.5.1 in combination with Proposition 1.4.15 we infer that the probabilities in (3.5.2) are well-defined.

(b) In particular, note that if X_n converges to X pointwise, then we have almost sure convergence as well. The reason that pointwise convergence is not so important to us is that modifications that only effect null sets cannot be noticed from a point of view of the probability measure.

(c) Property (3.5.2) can be rephrased as

$$\mathbb{P}(\limsup_{n \rightarrow \infty} d(X_n, X) > 0) = 0.$$

In the setting of a general measure space $(\Omega, \mathcal{F}, \mu)$, where μ does not necessarily have mass 1, if for functions (f_n) and f one has $\mu(\limsup_{n \rightarrow \infty} d(f_n, f) > 0) = 0$, or equivalently $\mu(\{\lim_{n \rightarrow \infty} d(f_n, f) = 0\}^c) = 0$, then (f_n) is said to ‘converge μ -almost everywhere (or μ -a.e.) to f ’.

3.5.2 Convergence in \mathcal{L}^p

This is yet another fairly strong type of convergence which in a slightly more general form plays an important role in (functional) analysis, too. Here, we will focus on the case of real-valued random variables.

Definition 3.5.4. Let $p > 0$, let (X_n) be a sequence of (equivalence classes of) random variables in $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$, and let $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ as well. Then we say that X_n converges to X in $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$, and write

$$X_n \xrightarrow{\mathcal{L}^p} X$$

if

$$\|X_n - X\|_p \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

As long as we do not impose any further assumptions (which we don’t do for the time being), none of the above two types of convergence is actually stronger than the other.

Example 3.5.5. Let \mathbb{P} denote the uniform distribution on $[0, 1)$

(a) Consider for $n \geq 1$ and $k \in \{0, 1, \dots, 2^n - 1\}$ the random variables

$$X_{n,k} := \mathbb{1}_{[k2^{-n}, (k+1)2^{-n})}$$

and define $Y_1 := X_{1,0}$, $Y_2 := X_{1,1}$, $Y_3 := X_{2,0}$, $Y_4 := X_{2,1}$, \dots (this is the ‘lexicographic ordering’). Then $\limsup_{n \rightarrow \infty} Y_n = 1$ and $\liminf_{n \rightarrow \infty} Y_n = 0$, and in particular Y_n does not converge almost surely. On the other hand, for $p > 0$, any $n \in \mathbb{N}$, and $k \in \{0, \dots, 2^n - 1\}$ we have

$$\mathbb{E}[|X_{n,k} - 0|^p] = \mathbb{P}([0, 2^{-n})) = 2^{-n},$$

and the right-hand side converges to 0 as $n \rightarrow \infty$. Therefore, $Y_n \xrightarrow{\mathcal{L}^p} 0$ as $n \rightarrow \infty$.

This example shows that convergence in \mathcal{L}^p does not imply almost sure convergence.

(b) Fix $p > 0$ and consider the random variables $X_n := n^{\frac{1}{p}} \mathbb{1}_{[0, 1/n]}$. Then for any $\omega \in (0, 1)$ fixed we have

$$X_n(\omega) = n^{\frac{1}{p}} \mathbb{1}_{[0, 1/n]}(\omega),$$

and the right-hand side converges to 0 as $n \rightarrow \infty$. Therefore,

$$\left\{ \lim_{n \rightarrow \infty} X_n = 0 \right\} = (0, 1),$$

and since $\mathbb{P}((0,1)) = 1$ this implies that $\lim_{n \rightarrow \infty} X_n = 0$ almost surely.

On the other hand, a moment's thought reveals that since $X_n \rightarrow X$ holds \mathbb{P} -a.s. as $n \rightarrow \infty$, the only possible limit in \mathcal{L}^p would be an almost surely constant random variable $X = 0$. Now for all $n \in \mathbb{N}$ one has

$$\mathbb{E}[|X_n - 0|^p] = 1,$$

and therefore X_n does not converge to 0 in \mathcal{L}^p .

This example shows that almost sure convergence does not imply convergence in \mathcal{L}^p .

3.5.3 Convergence in probability

Definition 3.5.6. Let (X_n) be a sequence of random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$ mapping into a separable metric space (S, d) , and let X be another such random variable. We say that X_n converges in probability ('konvergiert in Wahrscheinlichkeit' oder 'konvergiert stochastisch') to X if for all $\varepsilon > 0$,

$$\mathbb{P}(d(X_n, X) \geq \varepsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (3.5.3)$$

In this case we write

$$X_n \xrightarrow{\mathbb{P}} X \quad \text{as } n \rightarrow \infty.$$

Again, as a consequence of Lemma 3.5.1, the probability appearing in (3.5.3) is well-defined.

3.5.4 Convergence in distribution

In a slight abuse of notation, we will say that μ is a measure on a metric space (S, d) if, in fact, it is a measure on the measurable space $(S, \mathcal{B}(S))$, where as before the Borel- σ -algebra on S is defined as the σ -algebra generated by the open sets of S (which again are induced by the metric d).

Definition 3.5.7. Let (μ_n) be a sequence of finite measures on a separable metric space (S, d) and let μ be yet another finite measure on (S, d) . We say that (μ_n) converges in weakly ('konvergiert schwach') to μ if for all continuous bounded functions $f \in C_b(S)$ from S to \mathbb{R} we have

$$\int_S f \, d\mu_n \rightarrow \int_S f \, d\mu \quad \text{as } n \rightarrow \infty.$$

In this case we write

$$\mu_n \xrightarrow{w} \mu \quad \text{as } n \rightarrow \infty,$$

where w stands for 'weakly'.

In addition, given (S, d) -valued random variables X_n and X defined on possibly different probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ and $(\Omega, \mathcal{F}, \mathbb{P})$, we say that X_n converges to X in distribution as $n \rightarrow \infty$, if

$$\mathbb{P}_n \circ X_n^{-1} \xrightarrow{w} \mathbb{P} \circ X^{-1} \quad \text{as } n \rightarrow \infty.$$

In this case we write

$$X_n \xrightarrow{\mathcal{L}} X \quad \text{as } n \rightarrow \infty,$$

or also

$$X_n \xrightarrow{\mathcal{D}} X \quad \text{as } n \rightarrow \infty.$$

Here, \mathcal{L} and \mathcal{D} stand for 'law' and 'distribution', respectively. Yet another very common notation is

$$X_n \Longrightarrow X \quad \text{as } n \rightarrow \infty.$$

Theorem 3.5.8. *If (S, d) is a metric space and μ, ν are two finite measures on (S, d) with*

$$\int f d\mu = \int f d\nu \quad \forall f \in C_b(S) \text{ with } f \geq 0,$$

then $\mu = \nu$.

Proof. According to Theorem 1.2.17 it is sufficient to show that μ and ν coincide on the π -system of open sets (which is generating $\mathcal{B}(S)$).

For this purpose, let $U \subset S$ be open and for $x \in S$ define $d(x, U^c) := \inf_{y \in U^c} d(x, y)$. Then for any $n \in \mathbb{N}$, the function $f_n(x) := 1 \wedge nd(x, U^c)$ is in $C_b(S)$ and we have $f_n \uparrow \mathbf{1}_U$. Therefore, by assumption and the MCT we infer that

$$\mu(U) = \lim_{n \rightarrow \infty} \int f_n d\mu = \lim_{n \rightarrow \infty} \int f_n d\nu = \nu(U).$$

□

The following result gives a powerful characterization of weak convergence. As we will see in the proof of Corollary 3.5.10 already, it will turn out very useful to have different characterizations of weak convergence available.

Theorem 3.5.9 (Portmanteau theorem). *For a sequence (μ_n) of probability measures on the metric space (S, d) and μ another probability measure, the following conditions are equivalent:*

(a)

$$\mu_n \rightarrow \mu \quad \text{weakly};$$

(b)

$$\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\mu$$

for all $f \in C_b(S)$ which are uniformly continuous;

(c)

$$\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$$

for all $F \subset S$ closed;

(d)

$$\liminf_{n \rightarrow \infty} \mu_n(O) \geq \mu(O)$$

for all $O \subset S$ open;

(e)

$$\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$$

for all $A \in \mathcal{B}(S)$ with $\mu(\partial A) = 0$ (such a set A is also called a μ -continuity set).

Proof. ‘(a) \implies (b)’: This is immediate from the definition.

‘(b) \implies (c)’: Similarly to the proof of Theorem 3.5.8, setting

$$f_m(x) := (1 - md(x, F))^+,$$

we get, since each f_m is bounded and uniformly continuous, and since $\mathbf{1}_F \leq f_m \leq \mathbf{1}_{F_m}$, that

$$\limsup_{n \rightarrow \infty} \mu_n(F) \leq \limsup_{n \rightarrow \infty} \int f_m d\mu_n = \int f_m d\mu \leq \mu(F_m),$$

where for $\varepsilon > 0$ we define $F^\varepsilon := \{x \in S : d(x, F) < \varepsilon\}$.

If F is closed, then by the continuity of μ from above we obtain taking $m \rightarrow \infty$ that (c) holds.

‘(c) \implies (d)’: (d) follows from (c) by taking complements.

‘(c)&(d) \implies (e)’: We obtain

$$\mu(\overline{A}) \stackrel{(c)}{\geq} \limsup \mu_n(\overline{A}) \geq \limsup \mu_n(A) \geq \liminf_{n \rightarrow \infty} \mu_n(A) \geq \liminf_{n \rightarrow \infty} \mu_n(A^\circ) \stackrel{(d)}{\geq} \mu(A^\circ).$$

For A a μ -continuity set the left-hand side and the right-hand side of the previous display coincide, which yields (e).

‘(e) \implies (a)’: we choose $f \in C_b(S)$ arbitrary, but by linearity of the integral, we assume without loss of generality that $f(S) \subset [0, 1]$. Then using Example 2.1.8 we obtain that

$$\int_S f d\mu = \int_0^\infty \mu(f > t) dt = \int_0^1 \mu(f > t) dt.$$

Now since f is continuous, we deduce that $\partial\{f > t\} \subset \{f = t\}$. But we know that we can have $\mu(f = t) > 0$ for at most countably many $t \in [0, 1]$, so (e) implies that for λ -almost all $t \in [0, 1]$, we have that $\mu_n(f > t) \rightarrow \mu(f > t)$ as $n \rightarrow \infty$, which in combination with the DCT implies

$$\int f d\mu_n = \int_0^1 \mu_n(f > t) dt \rightarrow \int_0^1 \mu(f > t) dt = \int f d\mu,$$

which implies (a). □

Oftentimes we will be dealing with real random variables, and the following equivalent criterion for convergence in distribution of real random variables will come handy (which we had proven separately in the introductory lecture).

Corollary 3.5.10. *Let (X_n) be a sequence of real random variables and let X also be a real random variable. Denote the corresponding distribution functions by F_n and F , respectively. Then the following are equivalent:*

(a)

$$X_n \implies X;$$

(b) *For all points t of continuity of F , one has*

$$F_n(t) \rightarrow F(t) \quad \text{as } n \rightarrow \infty. \tag{3.5.4}$$

Proof. We only prove ‘(a) \implies (b)’ here. The key point is to observe that t is a point of continuity of F if and only if $(-\infty, t]$ is a $\mathbb{P} \circ X^{-1}$ -continuity set. Indeed, F is right-continuous due to Theorem 1.5.2, so we have that F is continuous at t if and only if

$$0 = F(t) - \underbrace{\lim_{h \downarrow 0} F(t-h)}_{=\mathbb{P} \circ X^{-1}((-\infty, t)) \text{ due to Prop. 1.2.16}},$$

and the right-hand side of the last display equals $\mathbb{P} \circ X^{-1}(\{t\})$, i.e., $\mathbb{P} \circ X^{-1}(\partial(-\infty, t])$, which establishes the claim.

The last equivalence of the Portmanteau theorem now immediately supplies us with (3.5.4). □

3.5.5 Some fundamental tools

Markov's and Chebyshev's inequalities

We will introduce some fundamental inequalities. These play a central role in probability and are some of the standard tools one has to feel comfortable to apply.

Proposition 3.5.11 (Markov's inequality (Andrey Andreyevich Markov (1856–1922))). *Let X be a real random variable and let $\varepsilon > 0$. Then, for any increasing function $\varphi : [0, \infty) \rightarrow [0, \infty)$ with $\varphi(\varepsilon) > 0$ one has*

$$\mathbb{P}(|X| \geq \varepsilon) \leq \frac{\mathbb{E}[\varphi(|X|)]}{\varphi(\varepsilon)}. \quad (3.5.5)$$

The proof is contained in [Dre18], but since it is short we reproduce it here.

Proof. Since φ is monotone increasing we have the inequality

$$\varphi(|X|) \geq \mathbf{1}_{|X| \geq \varepsilon} \varphi(\varepsilon),$$

and taking expectations on both sides supplies us with

$$\mathbb{E}[\varphi(|X|)] \geq \mathbb{P}(|X| \geq \varepsilon) \varphi(\varepsilon),$$

which implies (3.5.5). \square

Corollary 3.5.12 (Chebyshev's inequality (Pafnuty Chebyshev (1821–1894))). *Let X be in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Then*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{\varepsilon^2} = \frac{\text{Var}(X)}{\varepsilon^2}. \quad (3.5.6)$$

Proof. This follows from Proposition 3.5.11 by choosing the random variable in (3.5.5) as $X - \mathbb{E}[X]$ and $\varphi(x) := x^2$. \square

Remark 3.5.13. *Inequalities of the type (3.5.6) which bound the probability that X deviates from a certain quantity, such as its expectation, are also referred to as ‘concentration inequalities’.*

Theorem 3.5.14 (Jensen's inequality (Danish mathematician Johan Jensen (1859 – 1925))). *Let X be a real random variable in \mathcal{L}^1 and let $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ be a convex function (if X is a non-negative random variable, then it is sufficient for φ to be a convex function defined on $[0, \infty)$). Then*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)] \in (-\infty, \infty]. \quad (3.5.7)$$

The proof is that of [Dre18, Thm. 1.12.9].

Remark 3.5.15. (a) *Using Theorem 1.4.4 in combination with the fact that convex functions from \mathbb{R} (or $[0, \infty)$) to \mathbb{R} are $(\mathbb{R}, \mathcal{B}(\mathbb{R})) - (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ measurable (either exercise, or: for affine functions this is clear, and otherwise it follows from the proof of Theorem ?? below) we deduce that $\varphi \circ X$ is a random variable again, and hence at least we do have the measurability assumptions to speak of the expectation of $\varphi \circ X$.*

(b) *If $\tilde{\varphi}$ is a concave function on \mathbb{R} , then $-\tilde{\varphi}$ is a convex function, hence Theorem 3.5.14 yields*

$$\tilde{\varphi}(\mathbb{E}[X]) \geq \mathbb{E}[\tilde{\varphi}(X)]$$

for $X \in \mathcal{L}^1$.

- (c) This immediately supplies us with another proof for the inclusion $\mathcal{L}^q \subset \mathcal{L}^p$ for $q, p \in (0, \infty)$ with $q > p$ which we had derived in (3.4.4). Indeed, since the function $\tilde{\varphi}(x) := x^{\frac{p}{q}}$ is concave on $[0, \infty)$ and since $|X|$ is non-negative, we get for $X \in \mathcal{L}^q$ that

$$\infty > \tilde{\varphi}(\mathbb{E}[|X|^q]) \geq \mathbb{E}[\tilde{\varphi}(|X|^q)] = \mathbb{E}[|X|^p].$$

Thus, $\mathbb{E}[|X|^p] < \infty$ which implies $X \in \mathcal{L}^p$.

Example 3.5.16. (a) The absolute value function $\varphi(x) := |x|$ yields

$$|\mathbb{E}[X]| \leq \mathbb{E}[|X|].$$

- (b) Choosing the convex function $\varphi(x) := x^2$, Jensen's inequality supplies us with

$$\mathbb{E}[|X|]^2 \leq \mathbb{E}[X^2].$$

3.5.6 Interdependence of types of convergence of random variables

Having introduced all the above types of convergence, it is natural to try to order them in terms of strength. As we have seen in Example 3.5.5, there is no general implications between convergence in $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ and \mathbb{P} -almost sure convergence. However, for the remaining ones we do have the following hierarchy.

Theorem 3.5.17. Let X_n, X be real random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and let $p > 0$.

- (a) If either $\lim_{n \rightarrow \infty} X_n = X$ almost surely, or if $X, X_n \in \mathcal{L}^p$ and $X_n \xrightarrow{\mathcal{L}^p} X$, then

$$X_n \xrightarrow{\mathbb{P}} X.$$

- (b) If $X_n \xrightarrow{\mathbb{P}} X$, then

$$X_n \Longrightarrow X. \quad (3.5.8)$$

- (c) If $0 < p < q < \infty$ and if (X_n) and X are in \mathcal{L}^q such that $X_n \xrightarrow{\mathcal{L}^q} X$, then $X_n \xrightarrow{\mathcal{L}^p} X$ as well.

- (d) If

$$\text{for all } \varepsilon > 0 \text{ one has } \sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) < \infty, \quad (3.5.9)$$

then $\lim_{n \rightarrow \infty} X_n = X$ \mathbb{P} -a.s.³

In particular, if $X_n \xrightarrow{\mathbb{P}} X$, then there exists a subsequence (X_{n_k}) of (X_n) such that

$$X_{n_k} \longrightarrow X \quad \mathbb{P} - \text{a.s.}$$

The proof is exactly that of [Dre18, Thm. 1.13.1], so we omit it here.

Remark 3.5.18. (a) Show that the converses of the convergence implications given in Theorem 3.5.17 (a) to (c) do not hold true in general.

- (b) Also note that a substantial part of the above implications might break down if instead of \mathbb{P} we consider an infinite measure on (Ω, \mathcal{F}) .

Theorem 3.5.19 (Egorov's theorem). Let $X_n, X \in \mathcal{M}(\Omega, \mathcal{F}, \mathbb{P})$ be real random variables such that \mathbb{P} -a.s., $X_n \rightarrow X$.

Then for every $\varepsilon > 0$ there exists $A \in \mathcal{F}$ such that $\mathbb{P}(A) < \varepsilon$ and such that X_n converges to X uniformly on A^c .

Exercise 3.5.20. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a discrete probability space. Show that in this setting, if $X_n \xrightarrow{\mathbb{P}} X$ already implies that $\lim_{n \rightarrow \infty} X_n = X$ holds \mathbb{P} -almost surely.

³If (3.5.9) holds true one says that X_n converges fast or almost completely to X .

3.6 Laws of large numbers

One central topic in probability theory is the asymptotic analysis of random systems and one of the simplest and more or less realistic situations to imagine is arguably a very long (or, possibly slightly less realistic, an infinite) sequence of independent coin tosses or dice rolls. For the sake of simplicity let's have a look at the situation of independent fair coin tosses, and define for $n \in \mathbb{N}$ a random variable X_n on $(\Omega, \mathcal{F}, \mathbb{P})$ that takes the value 1 if the coin of the n -th toss shows heads, whereas it takes the value -1 if the coin shows tails.⁴ Now we know that $\mathbb{E}[X_n] = 0$, and also for the sum

$$S_n := \sum_{j=1}^n X_j \quad (3.6.1)$$

we have $\mathbb{E}[S_n] = 0$ by the linearity of expectation.

Definition 3.6.1. *The sequence S_n as defined in (3.6.1) is also called simple random walk (SRW) ('einfache Irrfahrt').*

For $x \in \mathbb{Z}$ we will sometimes write $\mathbb{P}_x(S_n \in \cdot) := \mathbb{P}(S_n + x \in \cdot)$ to denote the law of simple random walk started in x .

If you have attended the introductory class, it might be worthwhile to notice that simple random walk is a very basic example of a Markov chain.

Oftentimes, instead of investigating the expectation, one is interested e.g. in realizationwise statements, or statements concerning probabilities of certain events. In our current setting for example, one might want to ask what values $S_n(\omega)$ 'typically' takes. Now, although $\mathbb{E}[S_n] = 0$ for all $n \in \mathbb{N}$, it is obvious that $S_n(\omega) = 0$ can only hold true if n is even. In fact, even when n is even, 0 is not *the* typical value for S_n to take, in the sense that it is realised with a high probability or at least with a probability that is bounded away from 0 for $n \rightarrow \infty$. Indeed, for $n = 2k$ even we get with Stirling's formula that

$$\mathbb{P}(S_n = 0) = \binom{2k}{k} \left(\frac{1}{2}\right)^n \sim \frac{(2k/e)^{2k} \sqrt{2\pi \cdot 2k}}{((k/e)^k \sqrt{2\pi k})^2} 2^{-2k} = \frac{1}{\sqrt{k\pi}}, \quad (3.6.2)$$

where for sequences (a_n) and (b_n) of positive real numbers we write $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} a_n/b_n = 1$.

Exercise 3.6.2. *Using an explicit computation as in (3.6.2), show that although $\mathbb{P}(S_n = 0) \rightarrow 0$ due to (3.6.2), for $n = 2k$ the function $\mathbb{Z} \ni m \mapsto \mathbb{P}(S_n = m)$ is maximised for $m = 0$.*

Thus (3.6.2) tells us that $\mathbb{P}(S_n = 0)$ goes to zero at the order of $n^{-\frac{1}{2}}$. One might therefore be tempted to guess that if instead of just considering 0, we were replacing it by intervals of the type $[-c\sqrt{n}, c\sqrt{n}]$, then we would obtain a non-trivial limiting probability for S_n to take values in such intervals. This is indeed the case (and not only if the X_n describe coin tosses, but for far more general distributions of X) as will be established in the central limit theorem (see Theorem 3.8.1 below). For the time being, however, we start with having a look at a simpler result at cruder scales.

3.6.1 Weak law of large numbers

We will start with investigating the so-called empirical mean.

Definition 3.6.3. *Given a realization $X_1(\omega), \dots, X_n(\omega)$ of \mathbb{R}^d -valued random variables, its empirical mean is defined as*

$$\frac{1}{n} S_n(\omega) = \frac{1}{n} \sum_{j=1}^n X_j(\omega). \quad (3.6.3)$$

⁴The corresponding distribution $\mathbb{P} \circ X_n$ is also called *Rademacher distribution*, named after the German-American mathematician Hans Rademacher.

In order to be able to prove something meaningful about the empirical mean, we will take advantage of Chebyshev's inequality introduced in Corollary 3.5.12 above.

As suggested by (3.6.2) and the heuristics developed subsequently, we might guess that the empirical mean defined in (3.6.3) will converge to 0 under suitable assumptions on the sequence (X_n) .

In order to be able to treat the d -dimensional case at once, we generalize our definition of expectation to random variable $X : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ we define its *expectation* as

$$\begin{pmatrix} \mathbb{E}[\pi_1(X)] \\ \vdots \\ \mathbb{E}[\pi_d(X)] \end{pmatrix} \quad (3.6.4)$$

where we recall that the measurability of the coordinate functions $\pi_i(X)$ comes as a consequence of Proposition 1.4.11. It will then be left as an easy exercise to generalize the laws of large numbers below to \mathbb{R}^d -valued random variables whose coordinate functions fulfil the assumptions of those results.

Definition 3.6.4. A sequence (X_n) of elements of $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ satisfies a weak law of large numbers if

$$\frac{1}{n} \left(\sum_{j=1}^n X_j - \mathbb{E}[X_j] \right) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty. \quad (3.6.5)$$

Historically, a weak law of large numbers had first been rigorously derived by Jakob Bernoulli in [Ber13]. Nevertheless, the intuition for such a statement must have been around at that time already since in a correspondence Jakob Bernoulli writes to Gottfried Wilhelm Leibniz in October 1703 [vdWB75, pp. 509–513]: ‘Obwohl aber seltsamerweise durch einen sonderbaren Naturinstinkt auch jeder Dummste ohne irgend eine vorherige Unterweisung weiss, dass je mehr Beobachtungen gemacht werden, umso weniger die Gefahr besteht, dass man das Ziel verfehlt, ist es doch ganz und gar nicht Sache einer Laienuntersuchung, dieses genau und geometrisch zu beweisen.’

Theorem 3.6.5 (Weak law of large numbers). Let (X_n) be a sequence of pairwise uncorrelated random variables in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ and let (α_n) be a sequence of real numbers such that

$$\frac{\sum_{j=1}^n \text{Var}(X_j)}{\alpha_n^2} \rightarrow 0. \quad (3.6.6)$$

Then for all $\varepsilon > 0$,

$$\mathbb{P} \left(\left| \frac{\sum_{j=1}^n (X_j - \mathbb{E}[X_j])}{\alpha_n} \right| \geq \varepsilon \right) \leq \frac{\sum_{j=1}^n \text{Var}(X_j)}{\alpha_n^2 \varepsilon^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (3.6.7)$$

In particular, if the sequence (X_n) is even i.i.d., then it satisfies a weak law of large numbers.

The proof is a consequence of Chebyshev's inequality (Corollary 3.5.12) and Bienaymé's formula (Corollary 3.3.11). We omit the details and refer to [Dre18, Thm. 1.14.6] for a proof.

Example 3.6.6. Let a sequence (X_n) as in Definition 3.6.1 of simple random walk be given. Then the sequence (X_n) satisfies a weak law of large numbers.

Indeed, by assumption the (X_j) are independent and hence in particular pairwise uncorrelated. In addition, we have

$$\text{Var}(X_j) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 1 - 0 = 1.$$

Thus, in particular $X_j \in \mathcal{L}^2$, and the assumption of Theorem 3.6.5 are satisfied for any sequence (α_n) of positive reals with $\alpha_n/\sqrt{n} \rightarrow \infty$ as $n \rightarrow \infty$, which supplies us with

$$\frac{1}{\alpha_n} \sum_{j=1}^n X_j \xrightarrow{\mathbb{P}} 0$$

and in particular

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{\mathbb{P}} 0.$$

In particular the sequence (X_n) satisfies a weak law of large numbers.

It occurs quite frequently in probability theory that triangular arrays $(X_{n,k})$, $1 \leq k \leq n$, of random variables play an important role. In this setting we get the following generalization of Theorem 3.6.5.

Theorem 3.6.7. *Let $(X_{n,k})$, $1 \leq k \leq n$, $n \in \mathbb{N}$ be a triangular array of random variables in $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ such that for each $n \in \mathbb{N}$, the random variables $X_{n,1}, \dots, X_{n,n}$ are pairwise uncorrelated. Furthermore, let (α_n) be a sequence of real numbers such that setting*

$$S_n := \sum_{j=1}^n X_{n,j},$$

we have that

$$\frac{\text{Var}(S_n)}{\alpha_n^2} \rightarrow 0. \quad (3.6.8)$$

Then

$$\frac{S_n - \mathbb{E}[S_n]}{\alpha_n} \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty.$$

The proof is exactly the same as that of [Dre18, Theorem 1.14.9].

3.6.2 Strong law of large numbers

Definition 3.6.8. *A sequence (X_n) of elements of $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ satisfies the strong law of large numbers if*

$$\mathbb{P}\left(\limsup_{n \rightarrow \infty} \left| \frac{1}{n} \sum_{j=1}^n (X_j - \mathbb{E}[X_j]) \right| = 0\right) = 1,$$

which is the same as saying that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n (X_j - \mathbb{E}[X_j]) = 0 \quad \mathbb{P} - a.s.$$

Theorem 3.6.9 (Strong law of large numbers). *Let (X_n) be a sequence of independent identically distributed random variables in $\mathcal{L}^4(\Omega, \mathcal{F}, \mathbb{P})$. Then (X_n) satisfies a strong law of large numbers.*

Proof. Possibly replacing X_i by $X_i - \mathbb{E}[X_i]$ we can assume without loss of generality that $\mathbb{E}[X_i] = 0$. Setting $S_n := \sum_{i=1}^n X_i$, according to Theorem 3.5.17 (d) it is sufficient to show that for all $\varepsilon > 0$ we have

$$\sum_{n=0}^{\infty} \mathbb{P}(|n^{-1}S_n| \geq \varepsilon) < \infty. \quad (3.6.9)$$

For this purpose, we apply Markov's inequality with the function $\varphi(x) = x^4$, which entails

$$\mathbb{P}(|n^{-1}S_n| \geq \varepsilon) \leq \frac{\mathbb{E}[n^{-4}S_n^4]}{\varepsilon^4}. \quad (3.6.10)$$

Now

$$\mathbb{E}[S_n^4] = \sum_{1 \leq i, j, k, l \leq n} \mathbb{E}[X_i X_j X_k X_l].$$

Using that the (X_n) are independent we deduce that $\mathbb{E}[X_i X_j X_k X_l]$ can be non-zero only if each of the indices i, j, k, l appears at least twice among i, j, k, l . We can therefore continue the above equality to get

$$\mathbb{E}[S_n^4] \leq \sum_{i=1}^n \mathbb{E}[X_i^4] + C \sum_{\substack{i,j=1 \\ i \neq j}}^n \mathbb{E}[X_i^2 X_j^2] \leq n\mathbb{E}[X_1^4] + Cn^2\mathbb{E}[X_1^2]^2,$$

with C a finite constant. Plugging this into (3.6.10) we get

$$\mathbb{P}(|n^{-1}S_n| \geq \varepsilon) \leq \frac{n\mathbb{E}[X_1^4] + Cn^2\mathbb{E}[X_1^2]^2}{n^4\varepsilon^4},$$

which is summable over $n \in \mathbb{N}$ since $\mathbb{E}[X_1^2], \mathbb{E}[X_1^4] < \infty$. Therefore, (3.6.9) follows which finishes the proof. \square

Remark 3.6.10. (a) The implications of Theorem 3.6.9 also hold if we replace the condition $X \in \mathcal{L}^4(\Omega, \mathcal{F}, \mathbb{P})$ by $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. This has been proven by Etemadi [Ete81]; the proof is elementary and you should feel encouraged to read it (the article is available online through the university network at <http://link.springer.com/article/10.1007%2FBF01013465>)

(b) As the name suggests, if (X_n) satisfies a strong law of large numbers it also satisfies a weak law of large numbers. This is a direct consequence of Theorem 3.5.17 (a) applied to the sequence $(n^{-1} \sum_{i=1}^n X_i)$ of random variables and where the limiting random variable in Theorem 3.5.17 (a) is given by the constant 0.

3.7 Convolution of measures

As outlined above, the scaling (i.e., division by n) in the law of large numbers does not look like the most accurate information one might be able to obtain on a sequence of i.i.d. variables under nice assumptions. In order to prepare for the Central Limit Theorem, we will therefore introduce some tools that will prove helpful in its derivation.

Definition 3.7.1. Let μ, ν be two finite (possibly signed) measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then their convolution is defined as

$$(\mu * \nu)(B) := \int_{\mathbb{R}^d} \nu(B - x) \mu(dx), \quad B \in \mathcal{B}(\mathbb{R}^d), \quad (3.7.1)$$

where $B - x := \{y \in \mathbb{R}^d : y + x \in B\}$.

Alternatively, if $f, g \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$, then their convolution is defined as the function

$$(f * g)(y) := \int_{\mathbb{R}^d} f(y - x)g(x) dx \stackrel{x \mapsto y-x}{=} \int_{\mathbb{R}^d} g(y - x)f(x) dx = (g * f)(y).$$

Note that due to

$$\nu(B - x) = \int_{\mathbb{R}^d} \mathbf{1}_B(x + y) \nu(dy), \quad (3.7.2)$$

the right-hand side of (3.7.1) is well-defined. Furthermore, plugging (3.7.2) into (3.7.1) and applying Tonelli's theorem we also infer that

$$\mu * \nu = \nu * \mu \quad (\text{commutativity of convolution}).$$

Also, using Tonelli's theorem it can be shown that $f * g$ is well-defined and in \mathcal{L}^1 once $f, g \in \mathcal{L}^1$ (exercise).

The following result is the main reason the convolution plays an important role in probability theory.

Theorem 3.7.2. *Let $X, Y : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ be independent random variables. Then*

$$\mathbb{P}_{X+Y} = \mathbb{P}_X * \mathbb{P}_Y.$$

Proof. Writing $\sigma : \mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto x + y \in \mathbb{R}^d$ and noting that $\mathbb{P}_{X+Y} = (\mathbb{P}_X \otimes \mathbb{P}_Y) \circ \sigma^{-1}$ (c.f. (3.3.3)), we obtain for $B \in \mathcal{B}(\mathbb{R}^d)$ that

$$\begin{aligned} \mathbb{P}_{X+Y}(B) &\stackrel{\text{Thm. 2.2.19}}{=} \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbf{1}_B(x + y) (\mathbb{P}_X \otimes \mathbb{P}_Y)(d(x, y)) = \int_{\mathbb{R}^d} \mathbb{P}_Y(B - x) \mathbb{P}_X(dx) \\ &= (\mathbb{P}_X * \mathbb{P}_Y)(B), \end{aligned}$$

where in the penultimate equality we took advantage of Tonelli's theorem. \square

Lemma 3.7.3. *Let X_1, \dots, X_n be independent real random variables whose distributions have densities $\varphi_1, \dots, \varphi_n$ with respect to the Lebesgue measure λ . Then the distribution of the random variable $\sum_{i=1}^n X_i$ is absolutely continuous with respect to λ with density*

$$\varphi_1 * \varphi_2 * \dots * \varphi_n,$$

which is well-defined due to the associativity of convolution.

Proof. Exercise. \square

3.8 Central limit theorem

As the name suggests, the central limit theorem is one of the main result in probability theory. On the one hand, it gives us a somewhat more precise result of the fluctuations of the sum of well-behaved independent identically distributed random variables than the results we know from the laws of large numbers. On the other hand, it plays an important role in statistics since it justifies using the normal distribution in many models.

To motivate the central limit theorem, let us get back to (3.6.2) where we had shown that for simple random walk S_n ,

$$\mathbb{P}(S_{2k} = 0) \sim \frac{1}{\sqrt{2k\pi}}.$$

In fact, in this setting it is not hard to show that not only the probability of finding simple random walk in 0 at time $2k$ has a square root decay in k , but also the probabilities of finding simple random walk at a distance of order \sqrt{k} at time $2k$ (we restrict ourselves to even times for simplicity), see [Gre18, Section 1.15] for further details.

As a consequence, if we look for a rescaling of S_n by some scale function $\varphi(n)$ such that $S_n/\varphi(n)$ converges in distribution to a non-trivial limiting distribution, then the above suggests that \sqrt{n} is the only possible order of $\varphi(n)$ – and, as it will turn out below, the desired convergence does indeed take place.

Yet another motivation for the central limit theorem can be derived from the laws of large numbers: From those we know that under suitable assumptions on a sequence of i.i.d. random variables we have

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} S_n - \mathbb{E}[X_1] \right) = 0.$$

To obtain information on a finer scale than in the central limit theorem we can now ask if there exists an exponent $\beta \in (0, \infty)$ such that the sequence $n^\beta (\frac{1}{n} S_n - \mathbb{E}[X_1])$ might hopefully converge to a non-trivial limiting random variable instead of 0. The first motivational thread via the investigation of simple random walk then suggests that $\beta = 1/2$. Indeed, this always has to be the case as long as the X_n are assumed to have finite variance since due to Bienaymé's formula we have

$$\text{Var} \left(n^\beta (n^{-1} S_n - \mathbb{E}[X_1]) \right) = n^{2\beta} \frac{1}{n^2} n = n^{2\beta-1},$$

which can only converge to a non-trivial limit if $\beta = \frac{1}{2}$.

While the central limit theorem will not give us any information on probabilities of finding e.g. simple random walk at single points, it does indeed imply that the right scale for rescaling is \sqrt{n} ; and not only does it do so for simple random walk, but for a very general class of distributions.

Theorem 3.8.1 (Central limit theorem). *Let a sequence (X_n) of independent identically distributed random variables with $X_n : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ such that $\mathbb{E}[X_1] = \mu \in \mathbb{R}^d$ and $\mathbb{E}[|X_1 - \mu|^2] \in (0, \infty)$ be given. Then the sequence of random variables defined via*

$$Y_n := \frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{n}}, \quad n \in \mathbb{N}, \quad (3.8.1)$$

converges in distribution to a $\mathcal{N}(0, \Sigma)$ distributed random variable where $\Sigma_{i,j} := \text{Cov}(\pi_i(X_1), \pi_j(X_1))$, $1 \leq i, j \leq d$, is the (positive semi-definite) covariance matrix.

Remark 3.8.2. (a) *The Y_n are shifted in such a way that $\mathbb{E}[Y_n] = 0$ and $\text{Cov}(\pi_i(Y_n), \pi_j(Y_n)) = \Sigma_{i,j}$ (the latter being a consequence of Bienaymé's formula, see Cor. 3.3.11), so expectation and covariance structure already coincide with those of a $\mathcal{N}(0, \Sigma)$ -distributed variable.*

(b) *It is surprising that, as long as the X_n have finite second moments the limiting distribution is the normal distribution, independent of the specific distribution of the X_i s. This phenomenon is also called universality (of the normal distribution).*

The fact that the normal distribution appears in this context is due to the fact that if the X_n are i.i.d. $\mathcal{N}(0, \Sigma)$ distributed, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \sim \mathcal{N}(0, \Sigma), \quad (3.8.2)$$

i.e., the Y_n as defined in (3.8.1) are again $\mathcal{N}(0, \Sigma)$ distributed for all $n \in \mathbb{N}$.

(c) *There is a plethora of other, more general conditions which imply the validity (3.8.1). In particular, similarly to the case of the weak law of large numbers Theorem 3.6.7, there is a version of the central limit theorem for triangular arrays as well.*

(d) *The finiteness of the second moment is in fact essential in Theorem 3.8.1. If it is not assumed, however, then one can still obtain other types of convergence results to non-trivial distributions (so-called α -stable distributions) for different rescalings than the division by \sqrt{n} in (3.8.1).*

(e) *One can ask whether the sequence (Y_n) might even converge in probability to some random variable Z . In fact, in this case we would have that $Y_{2n} - Y_n$ would converge to 0 in probability due to*

$$\mathbb{P}(|Y_{2n} - Y_n| \geq \varepsilon) \leq \mathbb{P}(|Y_{2n} - Z| \geq \varepsilon/2) + \mathbb{P}(|Y_n - Z| \geq \varepsilon/2) \rightarrow 0, \quad \text{as } n \rightarrow \infty,$$

and using Theorem 3.5.17 we would deduce that

$$Y_{2n} - Y_n \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty. \quad (3.8.3)$$

However, assuming $d = 1$, $\mu = 0$ and $\sigma^2 = 1$ for simplicity of notation, we rewrite

$$Y_{2n} - Y_n = \frac{1}{\sqrt{2}} \frac{\sum_{i=n+1}^{2n} X_i}{\sqrt{n}} - \left(1 - \frac{1}{\sqrt{2}}\right) \frac{\sum_{i=1}^n X_i}{\sqrt{n}},$$

and observe that due to the CLT, both $\frac{\sum_{i=n+1}^{2n} X_i}{\sqrt{n}}$ and $\frac{\sum_{i=1}^n X_i}{\sqrt{n}}$ converge in distribution to a $\mathcal{N}(0, 1)$ -variable, both of which are independent. Using Part (a) of Example 3.3.12 we can therefore infer that $Y_{2n} - Y_n$ must converge to a $\mathcal{N}(0, (\frac{1}{\sqrt{2}})^2 + (1 - \frac{1}{\sqrt{2}})^2)$ -distributed random variable. In particular, this contradicts (3.8.3), hence $Y_{2n} - Y_n$ cannot converge to 0 in probability, so there cannot exist a random variable Z as postulated above.

There are at least two essentially different strategies to prove the central limit theorem. The first one works well in the case $d = 1$ and is a more or less self-contained and direct proof along the lines of the proof of [Geo09, Theorem 5.28]. The second one uses the technique of characteristic functions. It has the disadvantage that it is less self-contained; it is, however, more robust under variations of the very setting given in Theorem 3.8.1 and can be extended without too much effort to more general situations, such as higher dimensions or dependencies between the random variables X_n . We will follow the second approach and need a couple of general and auxiliary result which will also prove to be beneficial later on and in the lecture Probability II when establishing so-called ‘functional Central limit theorems’.

There are a couple of important properties of characteristic functions which are not hard to prove.

Lemma 3.8.3. *Let X and Y be random variables mapping to $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, and let $a \in \mathbb{R}$ as well as $b, t \in \mathbb{R}^d$ be arbitrary. Then*

(a)

$$\varphi_X(0) = 1;$$

(b)

$$\varphi_{aX+b}(t) = e^{ib \cdot t} \varphi_X(at);$$

(c)

$$|\varphi_X(t)| \leq 1 \quad \forall t \in \mathbb{R}^d; \quad (3.8.4)$$

(d) If X and Y are independent, then

$$\varphi_{X+Y}(t) = \varphi_X(t) \varphi_Y(t);$$

(e) The function $\mathbb{R}^d \ni t \mapsto \varphi_X(t)$ is uniformly continuous.

Proof. (a) Obvious, since $e^0 = 1$.

(b) We have

$$\varphi_{aX+b}(t) = \mathbb{E}[e^{it \cdot (aX+b)}] = e^{ib \cdot t} \varphi_X(at).$$

(c) If $e^{it \cdot X}$ was a real random variable, the statement would follow immediately with Jensen's inequality applied to the convex function $\mathbb{R} \ni x \mapsto |x|$. Since $e^{it \cdot X}$ is not real, however, we can approximate it by simple $(\mathbb{C}, \mathcal{B}(\mathbb{C}))$ -valued random variables X_n (as we did with real-valued functions) such that $|X_n| \leq 1$ and $X_n \rightarrow e^{it \cdot X}$. For simple random variables, (3.8.4) is a simple consequence of the fact that the unit ball around 0 in \mathbb{C} is a convex set. Taking the limit and using Fubini's theorem then implies the result.

(d) If X and Y are independent random variables, then so are $e^{it \cdot X}$ and $e^{it \cdot Y}$, for each $t \in \mathbb{R}^d$. Therefore,

$$\varphi_{X+Y}(t) = \mathbb{E}[e^{it \cdot (X+Y)}] = \mathbb{E}[e^{it \cdot X}] \mathbb{E}[e^{it \cdot Y}] = \varphi_X(t) \varphi_Y(t).$$

(e) Exercise.

□

We recall the following result that we had mentioned before already, and we will actually prove it here.

Theorem (Theorem 2.6.5). *Any finite measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is uniquely characterized by its characteristic function.*

For μ a finite measure as above, denote for $\sigma > 0$ by

$$\mu^{(\sigma)} := \mu * \mathcal{N}(0, \sigma^2 \text{Id}), \quad (3.8.5)$$

i.e., the convolution of μ with a d -dimensional Normal distribution with mean 0 and covariance matrix $\sigma^2 \text{Id}$, where Id is the identity matrix in $\mathbb{R}^{d \times d}$ (recall Example 1.5.8).

The following is a common paradigm in Fourier analysis: If we put in an arbitrarily 'rough' measure μ , then convoluting it with something 'smooth' (in our case a measure which is absolutely continuous with respect to λ) supplies us with something smooth as well. Here and in the lemma below, the 'smooth' measure $\mathcal{N}(0, \sigma^2 \text{Id})$ plays the role of a 'mollifier', just in case you've seen this concept in functional analysis.

Lemma 3.8.4. *Let μ be a finite measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Then the convolution $\mu^{(\sigma)}$ has a density $f^{(\sigma)}$ with respect to λ^d which is given by*

$$f^{(\sigma)}(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_\mu(t) \exp \left\{ -ix \cdot t - \frac{\sigma^2(t \cdot t)}{2} \right\} dt, \quad (3.8.6)$$

where we recall that φ_μ denotes the characteristic function of μ .

Proof. We write

$$h^{(\sigma)}(x) := \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} e^{-\frac{x \cdot x}{2\sigma^2}}, \quad x \in \mathbb{R}^d, \quad (3.8.7)$$

for the density of an \mathbb{R}^d -valued $\mathcal{N}(0, \sigma^2 \text{Id})$ -distributed random variable Z .

We start with observing that

$$\begin{aligned} \mu^{(\sigma)}(B) &= \int_{\mathbb{R}^d} \mathbb{P}_Z(B - x) \mu(dx) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_B(x + y) e^{-\frac{y \cdot y}{2\sigma^2}} \lambda^d(dy) \mu(dx) \\ &\stackrel{y \mapsto y-x}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbb{1}_B(y) h^{(\sigma)}(y - x) \lambda^d(dy) \mu(dx) = \int_B \left(\int_{\mathbb{R}^d} h^{(\sigma)}(y - x) \mu(dy) \right) \lambda^d(dy), \end{aligned}$$

where the penultimate equality follows from the change of variable formula Theorem 2.2.19 and the last equality is due to Tonelli's theorem. Thus, $\mu^{(\sigma)}$ has density

$$f^{(\sigma)}(x) = \int_{\mathbb{R}^d} h^{(\sigma)}(y - x) \mu(dy) \quad (3.8.8)$$

with respect to λ^d .

Then, generalizing Example 2.6.4 to the d -dimensional case, we obtain that

$$\varphi_{\mathcal{N}(0, \sigma^2 \text{Id})}(t) = e^{-\frac{\sigma^2 t \cdot t}{2}}.$$

and hence the identity

$$\frac{1}{(2\pi/\sigma^2)^{\frac{d}{2}}} \int_{\mathbb{R}^d} e^{it \cdot x} e^{-\frac{\sigma^2 x \cdot x}{2}} \stackrel{\text{def}}{=} \varphi_{\mathcal{N}(0, \frac{1}{\sigma^2} \text{Id})}(t) = e^{-\frac{t \cdot t}{2\sigma^2}} = (2\pi\sigma^2)^{\frac{d}{2}} h^{(\sigma)}(t).$$

Plugging this into the right-hand side of (3.8.8), we deduce that

$$f^{(\sigma)}(x) = \int_{\mathbb{R}^d} h^{(\sigma)}(y - x) \mu(dy) = \int_{\mathbb{R}^d} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp \left\{ i(y - x) \cdot t - \frac{\sigma^2 t \cdot t}{2} \right\} dt \mu(dy).$$

Using Fubini's theorem the latter equals

$$\begin{aligned} & \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} e^{iy \cdot t} \mu(dy) \right) \exp \left\{ -ix \cdot t - \frac{\sigma^2 t \cdot t}{2} \right\} dt \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_\mu(t) \exp \left\{ -ix \cdot t - \frac{\sigma^2 t \cdot t}{2} \right\} dt, \end{aligned}$$

which establishes (3.8.6) and hence finishes the proof. \square

Lemma 3.8.5. *For μ and $\mu^{(\sigma)}$ as above we have that*

$$\mu^{(\sigma)} \xrightarrow{w} \mu, \quad \text{as } \sigma \downarrow 0. \quad (3.8.9)$$

(So far we've only been concerned with convergence of sequences, not families of measures. We can either retreat to considering $\mu^{(1/n)}$ instead of $\mu^{(\sigma)}$ and then take $n \rightarrow \infty$, or otherwise the real meaning of the convergence in (3.8.9) is that the stated convergence takes place along any subsequence $(\sigma_n)_n$ with $\sigma_n > 0$ and $\lim_{n \rightarrow \infty} \sigma_n = 0$.)

Proof. μ and $\mu^{(\sigma)}$ have the same mass $\mu(\mathbb{R}^d) \in (0, \infty)$, so w.l.o.g. we can assume that it equals one. Then choose independent \mathbb{R}^d -valued random variables X and Y on the same probability space such that $\mathbb{P} \circ X^{-1} = \mu$ and $\mathbb{P} \circ Y^{-1} = \mathcal{N}(0, \text{Id})$. Then $X + \sigma Y$ has law $\mu^{(\sigma)}$, and \mathbb{P} -a.s.

$$X + \sigma Y \rightarrow X.$$

As a consequence, Theorem 3.5.17 implies that $\mu^{(\sigma)}$ converges weakly to μ , which finishes the proof. \square

The following result is not explicitly needed for proving Theorem 2.6.5, but we nevertheless give it here since it is important and interesting on its own.

Corollary 3.8.6 (Fourier inversion formula). *Let $\mu = f \cdot \lambda^d$ for some probability density f defined on \mathbb{R}^d . Then, if $\varphi_\mu \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$, we have for λ^d -almost all $x \in \mathbb{R}^d$ that*

$$f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_\mu(t) e^{-ix \cdot t} dt.$$

Proof. We have that the integrand on the right-hand side of (3.8.6) converges to $\varphi_\mu(t) e^{-ix \cdot t}$ as $\sigma \downarrow 0$, and furthermore, for each $\sigma > 0$, its absolute value is upper bounded by $|\varphi_\mu|$, which by assumption is in \mathcal{L}^1 . Therefore, using the DCT and taking $\sigma \downarrow 0$ in (3.8.6),

$$\lim_{\sigma \downarrow 0} f^{(\sigma)}(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_\mu(t) e^{-ix \cdot t} dt =: h(x).$$

In combination with the fact that $\mu^{(\sigma)} = f^{(\sigma)} \cdot \lambda^d$ according to Lemma 3.8.4 and using our result on integration with respect to measures with densities (Theorem 2.2.19), this implies (due to the DCT and the fact that

$$\sup_{x \in \mathbb{R}^d} f^{(\sigma)}(x) \leq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\varphi_\mu(t)| dt =: C \in (0, \infty),$$

that for any continuous function $v \geq 0$ with compact support we have that

$$C \sup_{x \in \mathbb{R}^d} v(x) \geq |v f^{(\sigma)}(y)| \quad \forall y \in \mathbb{R}^d,$$

and hence the left-hand side of the previous display is an integrable majorising function, so the DCT gives

$$\int v d\mu^{(\sigma)} = \int v f^{(\sigma)} d\lambda^d \rightarrow \int v h d\lambda^d, \quad \text{as } \sigma \downarrow 0.$$

On the other hand, Lemma 3.8.5 implies that

$$\int v d\mu^{(\sigma)} \rightarrow \int v d\mu = \int v f d\lambda^d, \quad \text{as } \sigma \downarrow 0,$$

so the right-hand sides of the last two displays coincide for all such v , i.e., we have

$$\int v h d\lambda^d = \int v f d\lambda^d \tag{3.8.10}$$

for all $v \geq 0$ in $C_b(\mathbb{R}^d)$ with compact support.

Now if only $v \in C_b(\mathbb{R}^d)$ instead of v having compact support as well, then we can choose a sequence a monotone increasing sequence (v_n) of functions with

- $v_n \geq 0$, $v_n \in C_b(\mathbb{R}^d)$ with compact support;
- $v_n(x) \in [0, \max_{x \in \mathbb{R}^d} |v(x)|] \subset [0, \infty)$;
- For all $x \in \mathbb{R}^d$ we have $v_n(x) \rightarrow v(x)$ as $n \rightarrow \infty$.

As a consequence (3.8.10) implies

$$\int v h d\lambda^d = \lim_{n \rightarrow \infty} \int v_n h d\lambda^d \stackrel{(3.8.10)}{=} \lim_{n \rightarrow \infty} \int v_n f d\lambda^d = \int v f d\lambda^d,$$

where the first and third equality are due to the MCT.

Therefore, Theorem 3.5.8 implies that

$$f \cdot \lambda^d = h \cdot \lambda^d,$$

so $f = h$ holds λ^d -almost everywhere according to Proposition 2.2.10. □

Proof of Theorem 2.6.5. Assume that $\varphi_\mu = \varphi_\nu$. Then, by Lemma 3.8.4 we get that

$$\mu^{(\sigma)} = \nu^{(\sigma)},$$

and due to Lemma 3.8.5 the left-hand side converges weakly to μ , whereas the right-hand side converges weakly to ν . I.e.,

$$\int_{\mathbb{R}^d} v d\mu = \lim_{\sigma \downarrow 0} \int_{\mathbb{R}^d} v d\mu^{(\sigma)} = \lim_{\sigma \downarrow 0} \int_{\mathbb{R}^d} v d\nu^{(\sigma)} = \int_{\mathbb{R}^d} v d\nu$$

for all $v \in C_b(\mathbb{R}^d)$. Therefore, by Theorem 3.5.8, $\mu = \nu$, which finishes the proof. □

We will give a couple of important implications of Theorem 2.6.5. We start with generalizing Example 3.3.12 (a) and giving a small hit parade of characteristic functions.

Proposition 3.8.7. *The following distributions have the given characteristic functions:*

(a) δ_x with $x \in \mathbb{R}$:

$$\varphi(t) = e^{itx}.$$

(b) Ber_p with $p \in (0, 1)$:

$$\varphi(t) = pe^{it} + (1 - p).$$

(c) $\text{Bin}_{n,p}$ with $p \in (0, 1)$, $n \in \mathbb{N}$:

$$\varphi(t) = (pe^{it} + (1-p))^n.$$

(d) Poi_ν with $\nu \in (0, \infty)$:

$$\varphi(t) = e^{\nu(e^{it}-1)}$$

(e) $\text{U}_{[0,a]}$ with $a \in (0, \infty)$:

$$\varphi(t) = \frac{e^{iat} - 1}{iat}$$

(f) Exp_κ with $\kappa \in (0, \infty)$:

$$\varphi(t) = \frac{\kappa}{\kappa - it}.$$

(g) $\mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}$, $\sigma^2 \in (0, \infty)$:

$$\varphi(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}.$$

Proof. For the normal distribution this has been shown in Exercise 2.6.4, the remaining parts are left as an exercise. \square

Combining Proposition 3.8.7 and Theorem 2.6.5, we directly obtain the following important result which (see Theorem 3.7.2) tells us how some sums of two independent random variables with the same type of distributions is distributed.

Corollary 3.8.8. (a) For $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 \in (0, \infty)$, we have

$$\mathcal{N}_{\mu_1, \sigma_1^2} * \mathcal{N}_{\mu_2, \sigma_2^2} = \mathcal{N}_{\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2};$$

(b) For $\nu_1, \nu_2 \in (0, \infty)$, we have

$$\text{Poi}_{\nu_1} * \text{Poi}_{\nu_2} = \text{Poi}_{\nu_1 + \nu_2};$$

(c) For $p \in (0, 1)$ and $m, n \in \mathbb{N}$ we have

$$\text{Bin}_{n,p} * \text{Bin}_{m,p} = \text{Bin}_{n+m,p}.$$

We now proceed with our preparations to proving the Central Limit theorem and start with the following auxiliary result, which we use in the proof of Theorem 3.8.10, but it is also of general interest.

Lemma 3.8.9. Let μ, ν be probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, and assume that $\mu = f \cdot \lambda^d$ for some density f . Then $\mu * \nu$ has density

$$g(x) := \int_{\mathbb{R}^d} f(x-y) \nu(dy)$$

with respect to λ^d .

Proof. For $B \in \mathcal{B}(\mathbb{R}^d)$ arbitrary we get

$$\begin{aligned} (\mu * \nu)(B) &= \int_{\mathbb{R}^d} \mu(B-y) \nu(dy) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_B(x+y) \mu(dx) \nu(dy) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \mathbf{1}_B(x+y) f(x) dx \nu(dy) = \int_B \int_{\mathbb{R}^d} f(x-y) \nu(dy) dx, \end{aligned}$$

which proves the result, and where to obtain the last equality we substituted $x \mapsto x-y$ and used Tonelli's theorem. \square

The following is actually a weak version of Lévy's continuity theorem. It is, however, significantly easier to prove than the general version and still sufficient for our purposes for the time being.

Theorem 3.8.10. *Assume probability measures μ, μ_1, μ_2, \dots on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ to be given, and denote the corresponding characteristic functions by $\varphi, \varphi_1, \varphi_2, \dots$.*

If

$$\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi(t) \quad \forall t \in \mathbb{R}^d,$$

then

$$\mu_n \xrightarrow{w} \mu \quad \text{as } n \rightarrow \infty. \quad (3.8.11)$$

Proof. Let $g \in C_b(S)$ have compact support. In particular, this implies that g is uniformly continuous. Then using the notation from (3.8.5) for $\mu^{(\sigma)}$ for any $\sigma > 0$, the triangle inequality implies

$$\begin{aligned} \left| \int g \, d\mu_n - \int g \, d\mu \right| &\leq \left| \int g \, d\mu_n - \int g \, d\mu_n^{(\sigma)} \right| \\ &\quad + \left| \int g \, d\mu_n^{(\sigma)} - \int g \, d\mu^{(\sigma)} \right| + \left| \int g \, d\mu^{(\sigma)} - \int g \, d\mu \right| \end{aligned} \quad (3.8.12)$$

Recalling that $\mu^{(\sigma)}$ was defined as the convolution $\mu * \mathcal{N}(0, \sigma^2 \text{Id})$, using the notation $h^{(\sigma)}$ from (3.8.7) for denoting the mollifier, and using Lemma 3.8.9, we can write that

$$\int g \, d\mu_n^{(\sigma)} = \int \left(\int g(x) h^{(\sigma)}(x-y) \mu_n(dy) \right) \lambda^d(dx) = \int g * h^{(\sigma)} \, d\mu_n,$$

where in the last equality we used Fubini's theorem (since the function $(x, y) \mapsto g(x) h^{(\sigma)}(x-y)$ is $\mu \otimes \lambda^d$ -integrable due to

$$\begin{aligned} \int |g(x) h^{(\sigma)}(x-y)| \mu_n \otimes \lambda^d(dx, dy) &= \int \left(\int |g(x)| h^{(\sigma)}(x-y) \mu_n(dy) \right) \lambda^d(dx) \\ &= \int |g| \, d\mu_n^{(\sigma)} < \infty, \end{aligned}$$

where in the first equality we used Tonelli's theorem, in the second we used Lemma 3.8.9, and the inequality comes from the fact that $\mu_n^{(\sigma)}$ is a finite measure and g is bounded.) Therefore, and taking advantage of Lemma 3.8.9 and Tonelli's theorem, we can upper bound

$$\left| \int g \, d\mu_n - \int g \, d\mu_n^{(\sigma)} \right| \leq \int |g(x) - (g * h^{(\sigma)})(x)| \mu_n(dx),$$

and due to the uniform continuity and the boundedness of g we get that $g * h^{(\sigma)}$ converges uniformly to g , i.e., for any $\varepsilon > 0$, for all $\sigma > 0$ small enough we have

$$\sup_{x \in \mathbb{R}^d} |g(x) - (g * h^{(\sigma)})(x)| < \varepsilon.$$

Thus, for any $\varepsilon > 0$ we have that for all $\sigma > 0$ small enough and all $n \in \mathbb{N}$,

$$\left| \int g \, d\mu_n - \int g \, d\mu_n^{(\sigma)} \right| \leq \varepsilon. \quad (3.8.13)$$

The exact same argument works for the last term on the RHS to yield

$$\left| \int g \, d\mu^{(\sigma)} - \int g \, d\mu \right| \leq \varepsilon. \quad (3.8.14)$$

Now in order to digest the second summand on the RHS of (3.8.12), note that Lemma 3.8.4 implies that $\mu^{(\sigma)}$ and $\mu_n^{(\sigma)}$ have densities

$$x \mapsto \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_\mu(t) \exp \left\{ -ix \cdot t - \frac{\sigma^2(t \cdot t)}{2} \right\} dt$$

and

$$x \mapsto \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi_{\mu_n}(t) \exp \left\{ -ix \cdot t - \frac{\sigma^2(t \cdot t)}{2} \right\} dt$$

with respect to λ^d . Thus, we get

$$\left| \int g d\mu_n^{(\sigma)} - \int g d\mu^{(\sigma)} \right| \leq \iint |g(x)| |\varphi_\mu(t) - \varphi_{\mu_n}(t)| \exp \left\{ -\frac{\sigma^2(t \cdot t)}{2} \right\} dt dx \quad (3.8.15)$$

and the integrand on the right-hand side can be upper bounded by

$$\mathbb{R}^d \times \mathbb{R}^d \ni (x, y) \mapsto 2|g(x)| \exp \left\{ -\frac{\sigma^2(t \cdot t)}{2} \right\}$$

which is integrable with respect to $\lambda^d \otimes \lambda^d$ since g is continuous with compact support. Therefore, the assumptions of the DCT are fulfilled and hence for any $\sigma > 0$ the RHS of (3.8.15) converges to 0 as $n \rightarrow \infty$.

In combination with (3.8.13), (3.8.14) and (3.8.12), this proves that

$$\left| \int g d\mu_n - \int g d\mu \right| \rightarrow 0 \quad (3.8.16)$$

for g as above. The Portmanteau Theorem (Theorem 3.5.9), however, demands the convergence of bounded continuous functions with non-compact support also. In order to derive the required convergence, let $g \in C_b(\mathbb{R}^d)$ not necessarily with bounded support, and choose a sequence of continuous functions (h_m) with $h_m \in [0, 1]$ and compact support (in particular, h_m is uniformly continuous) such that $h_m \uparrow 1$. In particular, the reasoning above implies

$$\int h_m d\mu_n \rightarrow \int h_m d\mu \quad \text{as } n \rightarrow \infty, \quad (3.8.17)$$

and

$$\int gh_m d\mu_n \rightarrow \int gh_m d\mu \quad \text{as } n \rightarrow \infty,$$

since h_m and gh_m are bounded and continuous with compact support. Furthermore,

$$\left| \int gh_m d\mu_n - \int g d\mu_n \right| \leq \sup |g(x)| \int (1 - h_m) d\mu_n \quad (3.8.18)$$

and

$$\left| \int gh_m d\mu - \int g d\mu \right| \leq \sup |g(x)| \int (1 - h_m) d\mu, \quad (3.8.19)$$

From (3.8.17) and the fact that the μ_n and μ are probability measures we infer that

$$\int (1 - h_m) d\mu_n \rightarrow \int (1 - h_m) d\mu \quad \text{as } n \rightarrow \infty.$$

In particular, for $\varepsilon > 0$ we can choose $m \in \mathbb{N}$ such that for all n large enough, the right-hand sides of (3.8.18) and (3.8.19) are upper bounded by ε . All in all, putting things together we obtain for $g \in C_b(S)$ and h_m as above that for any $\varepsilon > 0$ there exist $m, n_0 \in \mathbb{N}$ such that

$$\left| \int g d\mu_n - \int g d\mu \right| \leq \underbrace{\left| \int g d\mu_n - \int gh_m d\mu_n \right|}_{\substack{(3.8.18) \\ \leq \varepsilon}} + \underbrace{\left| \int gh_m d\mu_n - \int gh_m d\mu \right|}_{\substack{(3.8.16) \\ \leq \varepsilon}} + \underbrace{\left| \int gh_m d\mu - \int g d\mu \right|}_{\substack{(3.8.19) \\ \leq \varepsilon}}$$

for all $n \geq n_0$. This shows the desired weak convergence (3.8.11) and hence finishes the proof. \square

Theorem 3.8.11. *Let X be an $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ -valued random variable with $\mathbb{E}[\|X\|_2^m] < \infty$ for some $m \in \mathbb{N}$. Then for $t \in \mathbb{R}^d$ fixed we obtain for $h \in \mathbb{R}$ with $h \neq 0$ that*

$$\varphi_X(ht) = \sum_{k=0}^m \frac{i^k \mathbb{E}[(ht \cdot X)^k]}{k!} + o(|h|^m), \quad \text{as } h \rightarrow 0. \quad (3.8.20)$$

Proof. For $h \in \mathbb{R}$ set

$$\psi(h) := \varphi_X(ht).$$

For $h \in \mathbb{R}$ as well as $k \in \{0, \dots, m\}$ we obtain

$$\psi^{(k)}(h) = \mathbb{E}[(it \cdot X)^k e^{iht \cdot X}]. \quad (3.8.21)$$

Indeed, we use induction. The equality is clear for $k = 0$, and for $k \in \{1, \dots, m-1\}$ we obtain

$$\begin{aligned} \psi^{(k)}(h) &= \lim_{\alpha \rightarrow 0} \frac{\psi^{(k-1)}(h + \alpha) - \psi^{(k-1)}(h)}{\alpha} = \lim_{\alpha \rightarrow 0} \frac{\mathbb{E}[(it \cdot X)^{k-1} e^{i(h+\alpha)t \cdot X}] - \mathbb{E}[(it \cdot X)^{k-1} e^{iht \cdot X}]}{\alpha} \\ &= \lim_{\alpha \rightarrow 0} \mathbb{E}\left[\frac{(it \cdot X)^{k-1} e^{iht \cdot X} (e^{i\alpha t \cdot X} - 1)}{\alpha}\right], \end{aligned}$$

where in the second equality we took advantage of the induction assumption. Due to the fact that $|e^{ih} - 1| \leq |h|$ for $h \in \mathbb{R}$ arbitrary, we obtain that

$$\left| \frac{(it \cdot X)^{k-1} e^{iht \cdot X} (e^{i\alpha t \cdot X} - 1)}{\alpha} \right| \leq |t \cdot X|^k, \quad \forall \alpha \in \mathbb{R}, \alpha \neq 0.$$

According to our assumptions, the right-hand side is in $\mathcal{L}^k(\Omega, \mathcal{F}, \mathbb{P})$, and furthermore we have

$$\lim_{\alpha \rightarrow 0} \frac{(it \cdot X)^{k-1} e^{iht \cdot X} (e^{i\alpha t \cdot X} - 1)}{\alpha} = (it \cdot X)^k e^{iht \cdot X}.$$

Therefore, the dominated convergence theorem (see Theorem 2.1.7) implies that (3.8.21) holds. Therefore, (3.8.20) follows in combination with Taylor's formula. \square

Proof of Theorem 3.8.1. Without loss of generality we can assume that the X_n are centered for all $n \in \mathbb{N}$. Denote the characteristic function of X_n by φ_X and let φ_n denote the characteristic function of $Y_n = \frac{\sum_{i=1}^n X_i}{\sqrt{n}}$. Then for any $t \in \mathbb{R}^d$ fixed, using that $\mathbb{E}[t \cdot X_n] = \sum_{i=1}^d t_i \mathbb{E}[\pi_i(X_n)] = 0$ as well as that

$$\mathbb{E}[(t \cdot X_n)^2] = \sum_{1 \leq i, j \leq n} t_i t_j \underbrace{\mathbb{E}[\pi_i(X_n) \pi_j(X_n)]}_{\Sigma_{i,j}} = t^T \Sigma t,$$

we obtain

$$\varphi_n(t) \stackrel{\text{Lemma 3.8.3}}{=} \left(\varphi_X(t/\sqrt{n}) \right)^n \stackrel{\text{Thm. 3.8.11 for } m=2}{=} \left(1 - \frac{t^T \Sigma t}{2n} + o(n^{-1}) \right)^n \rightarrow e^{-\frac{t^T \Sigma t}{2}}, \quad \text{as } n \rightarrow \infty. \quad (3.8.22)$$

The result now follows using Theorem 3.8.10 in combination with the fact that the right-hand side of (3.8.22) is the characteristic function of a $\mathcal{N}(0, \Sigma)$ -distributed random variable. \square

Exercise 3.8.12. *For a sequence of random variables (X_n) as in the assumption of Theorem 3.8.1, the central limit theorem implies the validity of a weak law of large numbers for (X_n) . Indeed, since the distribution function Φ of the standard normal distribution (see (3.1.4)) is continuous, Theorem 3.8.1 implies that for arbitrary $M > 0$ we have*

$$\mathbb{P}\left(\underbrace{\frac{\sum_{i=1}^n (X_i - \mu)}{\sqrt{\sigma^2 n}}}_{A_{n,M}} \notin (-M, M]\right) \rightarrow \Phi(-M) + (1 - \Phi(M)). \quad (3.8.23)$$

Now for any $M \in (0, \infty)$ and $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that for all $n \geq N$ one has

$$B_{n,\varepsilon} := \left\{ \left| \frac{\sum_{i=1}^n (X_i - \mu)}{n} \right| > \varepsilon \right\} \subset A_{n,M}.$$

As a consequence, we obtain for any such M and ε , in combination with (3.8.23), that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(B_{n,\varepsilon}) \leq \Phi(-M) + (1 - \Phi(M)).$$

Since M was arbitrary and $\lim_{M \rightarrow \infty} \Phi(-M) + (1 - \Phi(M)) = 0$, this implies

$$\lim_{n \rightarrow \infty} \mathbb{P}(B_{n,\varepsilon}) = 0.$$

As in addition $\varepsilon > 0$ was arbitrary, this implies the desired weak law of large numbers for (X_n) .

Example 3.8.13. (a) Using the strong law of large numbers, for a random walk with drift (i.e., $S_n = \sum_{j=1}^n X_j$ where the X_j are i.i.d. with $\mathbb{P}(X_1 = 1) = p$, $\mathbb{P}(X_1 = -1) = 1 - p$, and $p \in (1/2, 1)$) one has that for all $\varepsilon > 0$,

$$\mathbb{P}(|S_n - n(2p - 1)| \geq n\varepsilon) \rightarrow 0.$$

Therefore, the first order (i.e. linear in n) term of the position of S_n at time n will asymptotically be given by $2p - 1$. In order to obtain a better understanding, it is of course tempting to ask for the lower order corrections. For this purpose we apply the central limit theorem; using that the variance of X_n is given by

$$\text{Var}(X_n) = \mathbb{E}[X_n^2] - \mathbb{E}[X_n]^2 = 1 - (2p - 1)^2 = 1 - 4p^2 + 4p - 1 = 4p(1 - p) := \sigma^2$$

we obtain

$$\frac{S_n - n(2p - 1)}{\sqrt{\sigma^2 n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

In particular, this implies that the ‘typical’ fluctuations of S_n around its expected value $n(2p - 1)$ are of the order \sqrt{n} .

Chapter 4

A primer on stochastic processes

4.1 Stochastic processes

Definition 4.1.1. A stochastic process is a family (X_t) , $t \in T$, of random variables mapping from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into a measurable space (E, \mathcal{E}) . Here, T is an arbitrary non-empty set.

Example 4.1.2. In the setting of the law of large numbers and the central limit theorem, the sequences

$$(X_n), \quad n \in \mathbb{N},$$

$$\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{n}, \quad n \in \mathbb{N},$$

and

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}}, \quad n \in \mathbb{N},$$

are all stochastic processes (with $T = \mathbb{N}$ or $T = \mathbb{N}_0$, respectively, and $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$).

We will primarily consider the setting that the X_t are real random variables, and in this case we also refer to (X_t) as a *real stochastic process*.¹

Most of the times we will actually interpret t as ‘time’, and hence natural choices are $T = \mathbb{N}_0$ or also $T = [0, \infty)$. The above definition, however, is more general.

We have seen in the theory of random variables that the distribution of random variables has played a very important role. In fact, the very structure of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ underlying a random variable X was often irrelevant, and what was more crucial to us was the law \mathbb{P}_X of the random variable.

In a similar way, in the theory of stochastic processes, a key role is played by the distribution of a process. By definition, the distribution of a real process (X_t) , $t \in T$, would be a probability law on \mathbb{R}^T endowed with a suitable σ -algebra \mathcal{T} on \mathbb{R}^T which makes the mapping

$$(\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^T, \mathcal{T})$$

$$\Omega \ni \omega \mapsto (X_t(\omega))_{t \in T}$$

itself a $\mathcal{F} - \mathcal{T}$ -measurable random variable. Since by definition of a stochastic process, the only ‘regularity’ assumption we made was that the X_t were random variables (i.e., they are $\mathcal{F} - \mathcal{B}(\mathbb{R})$ -measurable functions from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$), the natural σ -algebra to choose for \mathbb{R}^T is the product- σ -algebra (recall Definition 2.3.5). We recall that the product σ -algebra had been generated by the coordinate projections π_t , $t \in T$, and equivalently the product σ -algebra is generated by the cylinder sets (recall Definition 2.3.7). Thus, those subsets of \mathbb{R}^T for which finitely many coordinates are contained in certain measurable subsets of \mathbb{R} will play an important role; this then leads to the following concept.

¹In fact, if not mentioned otherwise, we will assume all the stochastic processes to be real in the following.

Definition 4.1.3. Let a stochastic process (X_t) , $t \in T$, be given. The finite dimensional distributions of the process are given by the probability measures

$$\mu_S(B) := \mathbb{P}((X_s)_{s \in S} \in B), \quad B \in \mathcal{B}(\mathbb{R}^S),$$

on $(\mathbb{R}^S, \mathcal{B}(\mathbb{R}^S))$, where $S \subset T$ finite.

In particular in the case of say continuous time (as opposed to discrete such as $T = \mathbb{N}_0$), i.e. if e.g. $T = \mathbb{R}$, then the finite dimensional distributions do not contain all the mathematically interesting information. However, this issue will only play a role and be addressed in more advanced classes.

For the time being, it is worthwhile to notice that for any given stochastic process (X_t) , $t \in T$, the family of finite dimensional distributions satisfies the following consistency condition:

$$\mathbb{P}((X_t)_{t \in J} \in (\pi_I^J)^{-1}(B)) = \mathbb{P}((X_t)_{t \in I} \in B) \quad \forall I \subset J \subset T, \quad J \text{ finite}, \quad B \in \mathcal{B}(\mathbb{R}^I), \quad (4.1.1)$$

or, which is the same,

$$\mathbb{P}_{((X_t)_{t \in J})} \circ (\pi_I^J)^{-1} = \mathbb{P}_{((X_t)_{t \in I})} \quad \forall I \subset J \subset T, \quad J \text{ finite}, \quad (4.1.2)$$

where the projection

$$\pi_I^J : \mathbb{R}^J \ni (x_j)_{j \in J} \mapsto (x_i)_{i \in I} \in \mathbb{R}^I$$

had been introduced in Definition 2.3.3.

Remark 4.1.4. In fact, some authors index the family of finite dimensional distributions by ordered tuples (t_1, \dots, t_n) with $n \in \mathbb{N}$, $t_i \in T$ for all $i \in \{1, \dots, n\}$. This, however, turns out to be more complicated since one has to impose a condition on how permuting acts on the elements of the family, i.e., how μ_{t_1, \dots, t_n} and $\mu_{t_{\pi(1)}, \dots, t_{\pi(n)}}$ are related to each other for an arbitrary permutation π of $\{1, \dots, n\}$ (i.e., $\pi \in \mathcal{S}_n$ with \mathcal{S}_n denoting the symmetric group). Indeed, if the family of finite dimensional distributions is to be generated by a stochastic process in the sense that

$$\mu_{t_1, \dots, t_n}(B_1 \times \dots \times B_n) = \mathbb{P}((X_{t_1}, \dots, X_{t_n}) \in B_1 \times \dots \times B_n),$$

then one obviously must have the condition that

$$\mu_{t_1, \dots, t_n}(B_1 \times \dots \times B_n) = \mu_{t_{\pi(1)}, \dots, t_{\pi(n)}}(B_{\pi(1)} \times \dots \times B_{\pi(n)}). \quad (4.1.3)$$

Therefore, once one has specified $\mu_{t_{\pi(1)}, \dots, t_{\pi(n)}}$ for some $t_1, \dots, t_n \in T$ and an arbitrary permutation $\pi \in \mathcal{S}_n$, then (4.1.3) already characterizes $\mu_{t_{\tilde{\pi}(1)}, \dots, t_{\tilde{\pi}(n)}}$ for any $\tilde{\pi} \in \mathcal{S}_n$. Therefore, it is sufficient and more convenient to index the finite dimensional distributions just by finite subsets of T (i.e., unordered tuples) instead of by ordered tuples t_1, \dots, t_n with $t_i \in T$.

Motivated the observation in (4.1.2), we introduce the notion of a consistent family of probability measures. For this purpose, recall the definition of the projection operators given in Definition 2.3.3.

Definition 4.1.5. Let T be an arbitrary non-empty set. If P_I , $I \subset T$ finite, is a family of probability measures such that P_I is a probability measure on $(\mathbb{R}^I, \mathcal{B}(\mathbb{R})^{\otimes I})$, then we call the family consistent, if we have

$$P_J \circ (\pi_I^J)^{-1} = P_I \quad \forall I \subset J \subset T, \quad J \text{ finite}. \quad (4.1.4)$$

4.2 Kolmogorov's existence and uniqueness theorem

The following result has first appeared (in slightly weaker form) in [Kol33, III § IV]

Theorem 4.2.1 (Kolmogorov's existence and uniqueness theorem). *Let (E_λ) , $\lambda \in \Lambda$, be an arbitrary family of Polish spaces, and let (\mathcal{B}_λ) , $\lambda \in \Lambda$, be the corresponding family of Borel σ -algebras.*

Furthermore, let P_I , $I \subset \Lambda$ finite, be a consistent family of probability measures P_I on $(\times_{\lambda \in I} E_\lambda, \otimes_{\lambda \in I} \mathcal{B}_\lambda)$.

Then there exists a unique probability measure P on $(\times_{\lambda \in \Lambda} E_\lambda, \otimes_{\lambda \in \Lambda} \mathcal{B}_\lambda)$ such that

$$P \circ (\pi_J)^{-1} = P_J \quad \forall J \subset \Lambda \text{ finite}, \quad (4.2.1)$$

where

$$\pi_J : \times_{\lambda \in \Lambda} E_\lambda \ni (\omega_\lambda)_{\lambda \in \Lambda} \mapsto (\omega_\lambda)_{\lambda \in J} \in \times_{\lambda \in J} E_\lambda$$

is again the projection.

P is also called the projective limit ('projektiver Limes') of the family (P_J) , $J \subset \Lambda$ finite.

For the proof, we will need the following notation and auxiliary result.

Definition 4.2.2. *Let E be a Hausdorff topological space or else a metric space, and let μ be a measure defined on the Borel σ -algebra $\mathcal{B}(E)$.*

(a) μ is called a Borel measure ('Borelmaß'), if

$$\mu(K) < \infty \quad \forall K \subset E \text{ compact}.$$

(b) μ is called inner regular ('regulär von innen') if

$$\mu(B) = \sup \{ \mu(K) : K \subset B \text{ compact} \}.$$

(c) μ is called outer regular ('regulär von außen') if

$$\mu(B) = \inf \{ \mu(O) : O \supset B \text{ open} \}.$$

(d) μ is called regular ('regulär') if it is inner regular and outer regular.

It should be noted that there is a variety of different definitions of the term 'Borel measure'. We will stick to the one above.

Lemma 4.2.3. *Let μ be either a finite Borel measure on a Polish space E , or let μ be a measure on $(\mathbb{R}^d, \mathcal{B}^d)$ such that $\mu(A) < \infty$ for any bounded $A \in \mathcal{B}(E)$.*

Then μ is regular.

Remark 4.2.4. *In the Polish space setting the above result is sometimes referred to as Ulam's theorem.*

Proof. Since our main emphasis is on $E = \mathbb{R}^d$, we will restrict ourselves to giving the proof in this simpler case (as a treat, we may on the other hand discard with the finiteness of μ and only require $\mu(A) < \infty$ for any $A \in \mathcal{B}(\mathbb{R}^d)$ bounded – in particular, we include the Lebesgue measure this way). For a proof of the general version, see [Bau92, Lemma 26.2] for instance.

We start with showing the following claim.

Claim 4.2.5. For all $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\mu(A) = \inf \{ \mu(O) : O \supset A \text{ open} \} = \sup \{ \mu(F) : F \subset A \text{ closed} \}. \quad (4.2.2)$$

Proof. We start with proving the first equality. Recalling the notation \mathcal{I}^d for the semi-ring of rectangles in \mathbb{R}^d , according to Theorem 1.3.9 we can extend $\mu|_{\mathcal{I}^d}$ to an outer measure μ^* on $2^{\mathbb{R}^d}$. Now Theorem 1.3.5 in combination with Theorem 1.3.9 imply that $\mu^*|_{\mathcal{B}(\mathbb{R}^d)}$ is a measure, and due to Theorem 1.2.17, we obtain that

$$\mu^*|_{\mathcal{B}(\mathbb{R}^d)} = \mu.$$

Using (1.3.5) we therefore get

$$\mu(A) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) : A_1, A_2, \dots \in \mathcal{I}^d, \text{ and } A \subset \bigcup_{i=1}^{\infty} A_i \right\}, \quad A \in \mathcal{B}(\mathbb{R}^d). \quad (4.2.3)$$

Using furthermore that μ is continuous from above, for every $\varepsilon > 0$ and any sequence of (A_i) on the right-hand side of (4.2.3) we deduce the existence of a sequence (O_i) of open rectangles with

- $A_i \subset O_i$, and
- $\mu(O_i) < \mu(A_i) + 2^{-i}\varepsilon$

for all $i \in \mathbb{N}$. In combination with (4.2.3) this yields the outer regularity of μ .

Passing to complements we deduce

$$\mu(A) = \sup \{ \mu(F) : F \subset A \text{ closed} \}$$

and thus recover the second equality in (4.2.2). \square

Since measures are continuous from below (see Prop. 1.2.16), for any $F \subset A$ closed we get that $\mu(F) = \sup_{n \in \mathbb{N}} \mu(F_n)$, where $F_n := F \cap B(0, n)$ is compact, with $B(0, n)$ denoting the closed unit ball of radius n in \mathbb{R}^d . In combination with (4.2.2) this implies the desired regularity. \square

Proof of Thm. 4.2.1. We know from Exercise 2.3.8 that the cylinder sets form an algebra over $\times_{\lambda \in \Lambda} E_\lambda$, which we will denote by \mathcal{A} for simplicity.

From Definition 2.3.5 we infer that \mathcal{A} is a generator of $\bigotimes_{\lambda \in \Lambda} \mathcal{B}_\lambda$, and being an algebra, we deduce that \mathcal{A} is a π -system. Hence, we may readily check that the assumptions of the newly added Corollary 1.2.19 are fulfilled and we deduce uniqueness, i.e., there is at most one probability measure on $(\times_{\lambda \in \Lambda} E_\lambda, \bigotimes_{\lambda \in \Lambda} \mathcal{B}_\lambda)$ satisfying (4.2.1).

It remains to show the existence of such a probability measure,² and we will take advantage of Carathéodory's existence theorem, or rather its Corollary 1.3.10. For that purpose, we have to show that

$$\text{the right-hand side of (4.2.1) defines a } \sigma\text{-subadditive content } P \text{ on the semiring } \mathcal{A} \quad (4.2.4)$$

via

$$P(Z) := P_I(B), \quad \text{all } Z \in \mathcal{A}, \text{ i.e., } Z = (\pi_I)^{-1}(B), \text{ where } B \in \bigotimes_{\lambda \in I} \mathcal{B}_\lambda, I \subset \Lambda \text{ finite.} \quad (4.2.5)$$

We start with showing that P as in (4.2.5) is actually well-defined. For this purpose assume that there are $B_1 \in \bigotimes_{\lambda \in I} \mathcal{B}_\lambda$ and $B_2 \in \bigotimes_{\lambda \in J} \mathcal{B}_\lambda$, where $I, J \subset \Lambda$ finite such that $Z = (\pi_I)^{-1}(B_1) = (\pi_J)^{-1}(B_2)$. Then there exists $B_3 \in \bigotimes_{\lambda \in I \cup J} \mathcal{B}_\lambda$ such that

$$Z = (\pi_{I \cup J})^{-1}(B_3).$$

²In fact, if any such measure exists, it must be a probability measure since $\times_{\lambda \in \Lambda} E_\lambda \in \mathcal{A}$, and we know that e.g. $P(\times_{\lambda \in \Lambda} E_\lambda) = P_\lambda(E_\lambda) = 1$ for arbitrary $\lambda \in \Lambda$, since the P_λ are probability measures.

From the consistency condition (4.2.1) and the fact that $B_3 = \{x \in \times_{\lambda \in I \cup J} E_\lambda : \pi_I^{I \cup J}(x) \in B_1\}$, we deduce that

$$P_{I \cup J}(B_3) = P_I(B_1),$$

and similarly we obtain

$$P_{I \cup J}(B_3) = P_J(B_2),$$

so $P(Z)$ as introduced in (4.2.5) does not depend on the specific representation of Z , and hence P is well-defined.

Next, we have to show (4.2.4). It is obvious that $P(\emptyset) = 0$ and that $P \geq 0$. To show additivity, let $Z_1, Z_2 \in \mathcal{A}$ be two disjoint sets. Then there exist $I \subset \Lambda$ finite as well as $B_1, B_2 \in \bigotimes_{\lambda \in I} \mathcal{B}_\lambda$ disjoint such that

$$Z_1 = \pi_I^{-1}(B_1) \quad \text{and} \quad Z_2 = \pi_I^{-1}(B_2). \quad (4.2.6)$$

Since Z_1 and Z_2 are disjoint, we get that B_1 and B_2 must be disjoint as well, and thus we deduce, using (4.2.6), the fact that

$$\pi_I^{-1}(B_1 \cup B_2) = \pi_I^{-1}(B_1) \cup \pi_I^{-1}(B_2),$$

as well as that P_I is a measure on $\bigotimes_{\lambda \in I} \mathcal{B}_\lambda$, that

$$P(Z_1 \dot{\cup} Z_2) = P_I(B_1 \dot{\cup} B_2) = P_I(B_1) + P_I(B_2) = P(Z_1) + P(Z_2),$$

hence P is additive and a content on \mathcal{A} .

In order to show that P is σ -subadditive, due to Proposition 1.2.16 it is sufficient to check that P is continuous in \emptyset .

We will prove this by contradiction. Indeed, assume otherwise that there is a decreasing sequence (Z_n) , $n \in \mathbb{N}$, of sets $Z_n \in \mathcal{A}$ with $\lim_{n \rightarrow \infty} Z_n = \emptyset$ and such that $\limsup_{n \rightarrow \infty} P(Z_n) > 0$. By possibly passing to a subsequence, w.l.o.g. assume that $P(Z_n) \geq \delta$ for some $\delta > 0$ and all $n \in \mathbb{N}$. Furthermore, we can assume $Z_n = \pi_{I_n}^{-1}(A_n)$ some $A_n \in \bigotimes_{\lambda \in I_n} \mathcal{B}_\lambda$ for all $n \in \mathbb{N}$, where $I_n \subset \Lambda$ finite (w.l.o.g., the I_n can and will be chosen to be increasing sets in $n \in \mathbb{N}$). Now recall that $\times_{\lambda \in I_n} E_\lambda$ is Polish (due to Theorem 2.3.9). Using Lemma 4.2.3, we therefore deduce that there exists a sequence (K_n) , $n \in \mathbb{N}$, of compact sets $K_n \in \bigotimes_{\lambda \in I_n} \mathcal{B}_\lambda$, and such that $K_n \subset A_n$ for all $n \in \mathbb{N}$, and

$$P_{I_n}(K_n) \geq \delta(1 - 2^{-(n+1)}) \quad \text{for all } n \in \mathbb{N}. \quad (4.2.7)$$

We set $Y_n := (\pi_{I_n}^{-1})(K_n)$ and furthermore

$$\tilde{Y}_n := \bigcap_{i=1}^n Y_i \in \mathcal{A}.$$

Then (\tilde{Y}_n) is a non-increasing sequence of elements of $\bigotimes_{\lambda \in \Lambda} E_\lambda$ with

$$\tilde{Y}_n \neq \emptyset, \quad \forall n \in \mathbb{N}. \quad (4.2.8)$$

Indeed, since $\tilde{Y}_n \subset Z_n$, we have

$$\begin{aligned} P(Z_n) - P(\tilde{Y}_n) &= P(Z_n \setminus \tilde{Y}_n) \leq P\left(\bigcup_{i=1}^n (Z_i \setminus Y_i)\right) \leq \sum_{i=1}^n P(Z_i \setminus Y_i) \\ &= \sum_{i=1}^n P_{I_i}(A_i \setminus K_i) \stackrel{(4.2.7)}{\leq} \delta \sum_{i=1}^n 2^{-(i+1)} \leq \frac{\delta}{2}, \end{aligned}$$

where the first inequality we took advantage of the fact that

$$Z_n \setminus \tilde{Y}_n = \bigcup_{i=1}^n Z_n \setminus Y_i$$

as well as the fact that the Z_n are monotone decreasing. Therefore $P(\tilde{Y}_n) > 0$ since $P(Z_n) \geq \delta$, so in particular $\tilde{Y}_n \neq \emptyset$.

Our goal now is to construct from this an element $z \in \lim_{n \rightarrow \infty} Z_n$, which would finish our proof by contradiction. For this purpose, due to (4.2.8), we can and do choose a sequence (y_n) , $n \in \mathbb{N}$, such that $y_n \in \tilde{Y}_n$, and the monotonicity of (\tilde{Y}_n) implies that

$$y_n \in \tilde{Y}_k \quad \forall n \geq k.$$

Hence, choosing t in some I_n , $n \in \mathbb{N}$, and projecting both sides of the previous display on the t -th coordinate, we deduce that

$$\pi_t(y_m) \in \tilde{Y}_n \subset \pi_t^{I_n}(K_n) \quad \forall m \geq n. \quad (4.2.9)$$

Since $\bigcup_{n \in \mathbb{N}} I_n = \lim_{n \rightarrow \infty} I_n$ is a countable subset of Λ , we can order its elements as t_1, t_2, \dots . We now apply a diagonal argument in combination with (4.2.9) to deduce that there exists a subsequence (y_{n_k}) , $k \in \mathbb{N}$, such that for each $j \in \mathbb{N}$ the sequence $(\pi_{t_j}(y_{n_k}))$ converges in each of the $\pi_{t_j}^{I_m}(K_m)$, $m \in \mathbb{N}$, as $k \rightarrow \infty$ (mind that the $\pi_{t_j}^{I_n}(K_n)$ are compact subsets of E_{t_j} since the projections $\pi_{t_j}^{I_n}$ are continuous and the K_n are compact in $\times_{\lambda \in I_n} E_\lambda$.) Hence, for each $t \in \bigcup_{n \in \mathbb{N}} I_n$, the limit

$$\lim_{k \rightarrow \infty} \pi_t(y_{n_k}) =: \tilde{y}(t) \in \pi_t^{I_m}(K_m) \subset \pi_t(Z_m) \quad (4.2.10)$$

(for all $m \in \mathbb{N}$) exists.

For some $x \in \times_{\lambda \in \Lambda} E_\lambda$ we now define $z \in \bigcap_{n \in \mathbb{N}} \tilde{Y}_n$ via

$$y(t) := \begin{cases} \tilde{y}(t), & \text{if } t \in \bigcup_{n \in \mathbb{N}} I_n, \\ x(t), & \text{otherwise.} \end{cases}$$

Then $y \in \bigcap_{n \in \mathbb{N}} Z_n$ due to (4.2.10), so the intersection is non-empty which was all that remained to finish the proof. \square

Remark 4.2.6. One can also prove Kolmogorov's extension theorem by first establishing it for the case that T is countable (using the so-called Ionescu-Tulcea theorem) and then generalize it to uncountable T ; See [Kle14, Section 14.3] for this approach.

In particular, if one is only interested in the result for $T = \mathbb{N}_0$, then there are easier proofs available than the one we gave (and the consistency condition is also easier to formulate).

Example 4.2.7. (a) We are now in the position to show that i.i.d. sequences of random variables, such as e.g. postulated in the strong law of large numbers, actually do exist! For this purpose, all we have to show is the existence of an i.i.d. sequence of random variables (X_n) on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}_{X_n} = \mu$ for all $n \in \mathbb{N}$, where μ is some given probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

We now apply Kolmogorov's extension theorem to the case $\Lambda := \mathbb{N}$ and $E_n := \mathbb{R}^d$, $\mathcal{E}_n := \mathcal{B}(\mathbb{R}^d)$, for all $n \in \Lambda$, and we define the measures $\mathbb{P}_n := \mu$ on (E_n, \mathcal{E}_n) . For $I \subset \mathbb{N}$ finite we consider the finite product measure

$$\mathbb{P}_I = \bigotimes_{n \in I} \mathbb{P}_n = \mu^{\bigotimes_{n \in I}}$$

on $(\times_{n \in I} E_n, \bigotimes_{n \in I} \mathcal{E}_n)$, as defined in Theorem 2.4.5.

It is not hard to show that the \mathbb{P}_I , $I \subset \mathbb{N}$ finite, form a consistent family of probability measures. Indeed, for $I \subset J \subset \mathbb{N}$ with J finite, we have for

$$B = \bigtimes_{n \in I} B_n \in \bigotimes_{n \in I} \mathcal{E}_n$$

(i.e., $B_n \in \mathcal{E}_n$ for each $n \in I$), that

$$(\pi_I^J)^{-1}(B) = \left(\bigtimes_{n \in I} B_n \right) \times ((\mathbb{R}^d)^{J \setminus I}) \subset \times_{n \in J} E_n.$$

So, according to the definition of the product measure,

$$\begin{aligned} \mathbb{P}_J((\pi_I^J)^{-1}(B)) &= \mathbb{P}_J\left(\left(\bigtimes_{n \in I} B_n\right) \times ((\mathbb{R}^d)^{J \setminus I})\right) = \prod_{n \in I} \mathbb{P}_n(B_n) \prod_{n \in J \setminus I} \mathbb{P}_n(E_n) \\ &= \prod_{n \in I} \mathbb{P}_n(B_n) = \mathbb{P}_I(B). \end{aligned}$$

Using Theorem 1.2.17 we thus deduce that

$$\mathbb{P}_J \circ (\pi_I^J)^{-1} = \mathbb{P}_I,$$

and as a consequence, Theorem 4.2.1 implies that there exists a probability measure \mathbb{P} on

$$((\mathbb{R}^d)^\mathbb{N}, \mathcal{B}(\mathbb{R}^d)^{\otimes \mathbb{N}})$$

such that for any $I \subset \mathbb{N}$ finite, $\mathbb{P} \circ \pi_I^{-1} = \mathbb{P}_I$. In particular, defining we can define a sequence of random variables (X_n) on the probability space $((\mathbb{R}^d)^\mathbb{N}, \mathcal{B}(\mathbb{R}^d)^{\otimes \mathbb{N}}, \mathbb{P})$ via

$$X_n : (\mathbb{R}^d)^\mathbb{N} \ni \omega \mapsto \omega(n) \in \mathbb{R}^d$$

with the desired properties as required above.

(b) Bernoulli percolation on \mathbb{Z}^d

Fix $p \in [0, 1]$. For $T \subset \mathbb{Z}^d$ finite define the probability measure \mathbb{P}_T^p on $(\{0, 1\}^T, \mathcal{B}(\mathbb{R}^T)) = (\{0, 1\}^T, 2^{\{0, 1\}^T})$ by setting

$$\mathbb{P}_T^p(\{f\}) := \prod_{x \in T} p^{f(x)} (1-p)^{1-f(x)} \quad (4.2.11)$$

(with the convention that $0^0 := 1$). Then the family of probability measures \mathbb{P}_T , $T \subset \mathbb{Z}^d$ finite, is consistent. Indeed, for $S \subset T \subset \mathbb{Z}^d$ finite we have for $f \in 2^S$ that

$$\begin{aligned} \mathbb{P}_T^p((\pi_S^T)^{-1}(\{f\})) &= \sum_{g \in (\pi_S^T)^{-1}(\{f\})} \prod_{x \in T} p^{g(x)} (1-p)^{1-g(x)} \\ &= \sum_{g \in (\pi_S^T)^{-1}(\{f\})} \prod_{x \in S} p^{g(x)} (1-p)^{1-g(x)} \prod_{x \in T \setminus S} p^{g(x)} (1-p)^{1-g(x)} \\ &= \prod_{x \in S} p^{f(x)} (1-p)^{1-f(x)} = \mathbb{P}_S^p(\{f\}); \end{aligned}$$

Since $\{0, 1\}^T$ is a discrete space, this shows $\mathbb{P}_T^p \circ (\pi_S^T)^{-1} = \mathbb{P}_S^p$ and hence the consistency.

Therefore, Theorem 4.2.1 supplies us with the existence of a probability measure \mathbb{P}^p on $(\{0, 1\}^{\mathbb{Z}^d}, \mathcal{B}(\{0, 1\}^{\mathbb{Z}^d})) = (\{0, 1\}^{\mathbb{Z}^d}, \mathcal{B}(\{0, 1\}^{\mathbb{Z}^d}))$ with the property that its projections / pushforwards on $(\{0, 1\}^T, 2^{\{0, 1\}^T})$ are given by the expression in (4.2.11).

In fact, the existence of this measure can also be derived ‘by foot’, showing that the corresponding content is σ -additive, see [Kle14, Thm. 14.36].

Remark 4.2.8. This will not be of utmost importance to us in this class, but we should note that in some sense from a certain point of view the σ -algebra \mathcal{B}^T is not very suitable. Indeed, in

the case $[0, \infty)$ it is natural to ask whether or not a stochastic process is a continuous functions, or the probability of this being the case. I.e., one might be interested in probabilities of the type

$$\mathbb{P}(\text{the function } [0, \infty) \ni t \mapsto X_t \text{ is continuous}).$$

In particular, for this to make sense we would need that the set

$$\{\omega \in \Omega : [0, \infty) \ni t \mapsto X_t(\omega) \text{ is continuous}\}$$

is contained in $\mathcal{B}(\mathbb{R})^{\otimes T}$. This, however, is generally not the case. This will be investigated in detail in more advanced probability classes.

Acknowledgment: I would like to thank Cedric Neumann for pointing out various typos.

Bibliography

- [AT07] Robert J. Adler and Jonathan E. Taylor. *Random fields and geometry*. Springer Monographs in Mathematics. Springer, New York, 2007.
- [Bau92] Heinz Bauer. *Maß- und Integrationstheorie*. de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, second edition, 1992.
- [Bau02] Heinz Bauer. *Wahrscheinlichkeitstheorie*. de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, fifth edition, 2002.
- [Beh87] E. Behrends. *Maß- und Integrationstheorie*. Hochschultext. [University Textbooks]. Springer-Verlag, Berlin, 1987.
- [Ber13] J. Bernoulli. *Ars conjectandi*. Landmarks of science. Impensis Thurnisiorum, fratrum, 1713.
- [Bil95] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. A Wiley-Interscience Publication.
- [Boc33] S. Bochner. Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse. *Math. Ann.*, 108(1):378–410, 1933.
- [Coh13] Donald L. Cohn. *Measure theory*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser/Springer, New York, second edition, 2013.
- [Dir68] P. G. Lejeune Dirichlet. *Vorlesungen über Zahlentheorie*. Herausgegeben und mit Zusätzen versehen von R. Dedekind. Vierte, umgearbeitete und vermehrte Auflage. Chelsea Publishing Co., New York, 1968.
- [Doo94] J. L. Doob. *Measure theory*, volume 143 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994.
- [Dre17] Alexander Drewitz. Probability theory II (lecture notes), 2017.
- [Dre18] Alexander Drewitz. Introduction to probability and statistics (lecture notes), 2018.
- [Dud02] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.
- [Dur10] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.
- [Els05] Jürgen Elstrodt. *Maß- und Integrationstheorie*. Springer-Lehrbuch. [Springer Textbook]. Springer-Verlag, Berlin, fourth edition, 2005. Grundwissen Mathematik. [Basic Knowledge in Mathematics].

- [Ete81] N. Etemadi. An elementary proof of the strong law of large numbers. *Z. Wahrsch. Verw. Gebiete*, 55(1):119–122, 1981.
- [Gö8] J. Gärtner. *Skript Maß- und Integrationstheorie*. 2008.
- [Geo09] Hans-Otto Georgii. *Stochastik*. de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, expanded edition, 2009. Einführung in die Wahrscheinlichkeitstheorie und Statistik. [Introduction to probability and statistics].
- [Hal50] Paul R. Halmos. *Measure Theory*. D. Van Nostrand Company, Inc., New York, N. Y., 1950.
- [Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.
- [Kle14] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. A comprehensive course.
- [Kol33] A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933.
- [LL10] Gregory F. Lawler and Vlada Limic. *Random walk: a modern introduction*, volume 123 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
- [Rao04] M. M. Rao. *Measure theory and integration*, volume 265 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker, Inc., New York, second edition, 2004.
- [Rud99] Walter Rudin. *Reelle und komplexe Analysis*. R. Oldenbourg Verlag, Munich, 1999. Translated from the third English (1987) edition by Uwe Krieg.
- [Sie75] Waclaw Sierpiński. *Oeuvres choisies*. PWN–Éditions Scientifiques de Pologne, Warsaw, 1975.
- [vdWB75] B.L. van der Waerden and J. Bernoulli. *Die Werke von Jakob Bernoulli: Bd. 3: Wahrscheinlichkeitsrechnung*. Die Werke von Jakob Bernoulli. Birkhäuser Basel, 1975.
- [Wal93] Wolfgang Walter. *Gewöhnliche Differentialgleichungen*. Springer-Lehrbuch. [Springer Textbook]. Springer-Verlag, Berlin, fifth edition, 1993. Eine Einführung. [An introduction].