Measure theory for probabilists

Peter Pfaffelhuber

Version: October 11, 2024

Prelude

Modern probability theory (including statistics) is based on measure theory. This manuscript is part of a course with the aim to introduce measure theory for students with a solid backround in mathematics, which aim to dive deeper into probability theory.

In various parts of mathematics, we aim to assign a set some value, and describe it as its content, volume, etc. In probability, this value is the probability of the set. Since this concept of assigning values to sets has the same features in many areas (e.g. if two sets are disjoint, the volume of their union is the sum of the volumes), several areas are dealing with measure theory.

We approach measure theory in several steps. First, in Chapter 1, we have to deal with set systems (i.e. sets of sets), since it turns out that it leads to contradictions if we assign volumes to all sets. Here, we will learn about semi-rings and σ-fields as specific set systems. Second, in Chapter 2, we construct measures on these set systems. We will do so by constructing outer measures (defined on all sets) and restricting them to a σ-field. Third, in Chapter 3, we will be dealing with measurable functions and integrals with respect to measures. In probabilistic terms, these are random variables, and their expectations. Fourth, in Chapter 4, we will study certain subsets of measurable functions (or random variables), known as $\mathcal{L}^p$-spaces. Last, in Chapter 5, we will be dealing with product spaces, which are important for the theory of stochastic processes. Since various notions (Borel sets, compact systems) are from set-theoretic topology, we give a repetition of the relevant terms in Appendix A.
There are various textbooks in measure theory with a focus on probability. The following have guided me as references for the purpose of this manuscript.

- Bogatchev, Vladimir I. Measure Theory. Springer, 2007
- Billingsley, Patrick. Probability and Measure. Wiley, third edition, 1995
- Kallenberg, Olaf. Foundations of Modern Probability Theory. Springer, third edition, 2021
- Klenke, Achim. Probability theory. A comprehensive course. Springer, 2014

The present english version of this manuscript was written based on the German version with the help of DeepL.

Contents

1 Set systems

Probability theory formalises the colloquially used word *probable*. This is (in the broadest sense) a property of a possible outcome of an experiment. Fundamental to probability theory is the concept of an *event*, which is intended to describe everything that can happen in the experiment. Events are represented by subsets of an abstract basic space, which is always called Ω. The aim of this section is to assign a probability to as many subsets of Ω as possible. This leads to the concept of a σ-algebra, because these contain exactly the subsets of the base space to which probabilities are then assigned in the next section. In other words, elements of σ-algebras will be events in the above sense. The other set systems introduced in this section will be used to define suitable σ-algebras.

1.1 Semi-rings, rings and σ-fields

The notions in this section are connected as follows: For $\mathcal{C} \subseteq 2^\Omega$, we have

$$\mathcal{C}\ \sigma\text{-field} \quad \Longrightarrow \quad \mathcal{C}\ \text{ring} \quad \Longrightarrow \quad \mathcal{C}\ \text{semi-ring}.$$

Some more properties of the three notions are given in table 1.

Definition 1.1 (Semi-ring, ring, $\boldsymbol{\sigma}$-field)**.** *Let Ω be a set and $\emptyset \neq \mathcal{H}, \mathcal{R}, \mathcal{F} \subseteq 2^\Omega$.*

1. *$\mathcal{H}$ is* $\cap$-stable *(or closed under $\cap$, or a π-system), if ($A, B \in \mathcal{H} \Rightarrow A \cap B \in \mathcal{H}$). It is called* σ-$\cap$-stable *(or closed under σ-$\cap$) if ($A_1, A_2, ... \in \mathcal{H} \Rightarrow \bigcap_{i=n}^\infty A_n \in \mathcal{H}$).*
   It is called $\cup$-stable *(or closed under $\cup$), if ($A, B \in \mathcal{H} \Rightarrow A \cup B \in \mathcal{H}$). It is called* $\sigma - \cup$-stable *(or closed under σ-$\cup$) if ($A_1, A_2, ... \in \mathcal{H} \Rightarrow \bigcup_{i=n}^\infty A_n \in \mathcal{H}$).*
   It is complement-stable *(or closed under complements), if $A \in \mathcal{H} \Rightarrow A^c \in \mathcal{H}$. It is* set-difference-stable *(or closed under set-differences), if ($A, B \in \mathcal{H} \Rightarrow B \setminus A \in \mathcal{H}$).*

2. *$\mathcal{H}$ is a* semi-ring*, if it is (i) closed under $\cap$ and (ii) $\forall A, B \in \mathcal{H} \exists C_1, \dots, C_n \in \mathcal{H}$ with* [1] *$B \setminus A = \biguplus_{i=1}^n C_i$.*

3. *$\mathcal{R}$ is a* ring*, if it is closed under $\cup$ and set-differences.*

4. *$\mathcal{F}$ is a* σ-field *(or σ-algebra), if $\Omega \in \mathcal{F}$, it is closed under complements and closed under σ-$\cup$. Then, $(\Omega, \mathcal{F})$ is called* measurable space.

Remark 1.2 (Relationships between the collections of sets)**.**

1. Every ring $\mathcal{R}$ is a semi-ring: *For closedness under $\cap$, we write for $A, B \in \mathcal{R}$*

   $$A \cap B = A \setminus (A \setminus B) \in \mathcal{R}.$$

   The second property is trivial.

2. Every σ-field $\mathcal{F}$ is a ring: *We need to understand that $\mathcal{F}$ is closed under set-differences. For this, we write for $A, B \in \mathcal{F}$*

   $$B \setminus A = B \cap A^c = (B^c \cup A)^c.$$

[1] We write $A \uplus B$ for $A \cup B$ if $A \cap B = \emptyset$.

	$\mathcal{C}$ semi-ring	$\mathcal{C}$ ring	$\mathcal{C}$ σ-field
$\mathcal{C}$ is $\cap$-stable	$\bullet$	$\circ$	$\circ$
$\mathcal{C}$ is σ-$\cap$-stable			$\circ$
$\mathcal{C}$ is $\cup$-stable		$\bullet$	$\circ$
$\mathcal{C}$ is σ-$\cup$-stable			$\bullet$
$\mathcal{C}$ is set-difference-stable		$\bullet$	$\circ$
$\mathcal{C}$ is complement-stable			$\bullet$
$B \setminus A = \biguplus_{i=1}^n C_i$	$\bullet$	$\circ$	$\circ$
$\Omega \in \mathcal{C}$			$\bullet$

Table 1: For $\mathcal{C} \subseteq 2^\Omega$, we list all properties for semi-rings, rings and σ-fields. $\bullet$ means that the respective property is a hypothesis in the definition, whereas $\circ$ means that the respective property is a result following from the definition of the collection of subsets.

Example 1.3 (Semi-rings, σ-fields)**.**

1. Semi-open intervals form a semi-ring: *Let $\Omega = \mathbb{R}$. Then,*

$$\mathcal{H} := \{(a, b] : a, b \in \mathbb{Q}, a \le b\}$$

is a semi-ring.
Indeed, if $a_1 \le b_1, a_1' \le b_1'$, then[2] $(a_1, b_1] \cap (a_1', b_1'] = (a_1 \vee a_1', b_1 \wedge b_1']$ *and* $(a_1, b_1] \setminus (a_1', b_1'] = (a_1, a_1' \wedge b_1] \uplus (b_1', b_1]$*, where $(a, b] = \emptyset$, falls $a \ge b$.*

2. Examples for σ-fields: *Trivial examples are $\{\emptyset, \Omega\}$ and 2^Ω. (Recall that both are topologies as well; see Definition A.1.)*

Yet another example will become important in Section 3.1: If $\mathcal{F}'$ is a σ-field on Ω', and $f : \Omega \to \Omega'$. Then,

$$\sigma(f) := \{f^{-1}(A') : A' \in \mathcal{F}'\} \subseteq 2^\Omega \tag{1.1}$$

is a σ-field on Ω.
Indeed: If $A', A_1', A_2', \ldots \in \sigma(f)$, then $(f^{-1}(A'))^c = f^{-1}((A')^c) \in \sigma(f)$ and $\bigcup_{n=1}^\infty f^{-1}(A_n') = f^{-1}\left(\bigcup_{n=1}^\infty A_n'\right) \in \sigma(f)$.

We will frequently use the so-called Borel σ-field (which is the σ-field generated by a topology; see Definition 1.7.

[2] As usual, we write $x \wedge y := \min(x, y)$ and $x \vee y := \max(x, y)$

1.2 Generators and extensions

On the one hand, we want to use σ-fields as much as possible, since they contain the sets we can assign probabilities to. On the other hand, often only semi-rings can be given constructively. However, we can use the (ring or) σ-field generated by a semi-ring., i.e. the smallest σ-field (or smallest ring) which contains the semi-ring.

Remark 1.4 (Generated set-systems). *First, it is easy to see that the intersection of σ-fields (rings) is a σ-field (ring). (For example, since all σ-fields are closed under $\cup$, and if we take A, B, elements of all σ-fields, then $A \cup B$ is an element of all σ-fields and therefore in the intersection.):*
Let $\mathcal{C} \subseteq 2^\Omega$. Then,

$$\mathcal{R}(\mathcal{C}) := \bigcap \Big\{ \mathcal{R} \supseteq \mathcal{C} : \mathcal{R} \text{ ring} \Big\}$$

is the ring generated from $\mathcal{C}$ *and*

$$\sigma(\mathcal{C}) := \bigcap \Big\{ \mathcal{F} \supseteq \mathcal{C} : \mathcal{F}\ \sigma\text{-field} \Big\}$$

is the σ-field generated from $\mathcal{C}$. *Apparently, $\mathcal{R}(\mathcal{R}(\mathcal{H})) = \mathcal{R}(\mathcal{H})$ and $\sigma(\sigma(\mathcal{H})) = \sigma(\mathcal{H})$.*

The next lemma is shown after Example 1.6.

Lemma 1.5 (Ring generated from a semi-ring). *Let $\mathcal{H}$ be a semi-ring. Then,*

$$\mathcal{R}(\mathcal{H}) = \Big\{ \biguplus_{k=1}^{n} A_k : A_1, \ldots, A_n \in \mathcal{H} \text{ disjoint}, n \in \mathbb{N} \Big\}$$

is the ring generated from $\mathcal{H}$.

Example 1.6 (Ring generated from semi-open intervals). *Let $\mathcal{H}$ be the semi-ring of semi-open intervals from Example 1.3. Then,*

$$\mathcal{R}(\mathcal{H}) = \Big\{ \biguplus_{k=1}^{n} (a_k, b_k] : a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{Q}, \\ a_k < b_k, k = 1, \ldots, n \text{ and } a_k < b_{k+1}, k = 1, \ldots, n-1 \Big\}$$

is the ring generated from $\mathcal{H}$.

Proof of Lemma 1.5. It is clear that $\mathcal{R}(\mathcal{H})$ is closed under $\cap$. In order to show that $\mathcal{R}(\mathcal{H})$ is a ring, we start by showing closedness under set-differences. Let $A_1, \ldots, A_n \in \mathcal{H}$ and $B_1, \ldots, B_m \in \mathcal{H}$ be disjoint, respectively. Then,

$$\Big(\biguplus_{i=1}^{n} A_i \Big) \setminus \Big(\biguplus_{j=1}^{m} B_j \Big) = \biguplus_{i=1}^{n} \bigcap_{j=1}^{m} A_i \setminus B_j \in \mathcal{R}(\mathcal{H}).$$

In order to show closedness under $\cup$ of $\mathcal{R}(\mathcal{H})$, let $A, B \in \mathcal{R}(\mathcal{H})$. Then, write $A \cup B = (A \cap B) \uplus (A \setminus B) \uplus (B \setminus A) \in \mathcal{R}(\mathcal{H})$, since we already showed closedness under $\cap$ and under set-differences.

Last, note that there is no smaller ring than $\mathcal{R}(\mathcal{H})$, which contains $\mathcal{H}$. Indeed, such a ring would have to be closed under $\cup$, and clearly $\mathcal{R}(\mathcal{H})$ is the minimal set which contains $\mathcal{H}$ and which is closed under $\cup$. □

Definition 1.7 (Borel σ-algebra)**.** *Let $(\Omega, \mathcal{O})$ be a topological space. Then $\mathcal{B}(\Omega) := \sigma(\mathcal{O})$ denotes the* Borel σ-algebra *on Ω. If $\Omega \subseteq \mathbb{R}^d$, we denote by $\mathcal{B}(\Omega)$ the Borel σ-algebra generated by the Euclidean topology on $\mathbb{R}^d$. If $\Omega \subseteq \overline{\mathbb{R}}$, then $\mathcal{B}(\Omega)$ is the Borel σ-algebra generated by the topology from example A.2. Sets in $\mathcal{B}(\Omega)$ are also called* (Borel-)measurable sets.

Lemma 1.8 (Countable base and Borel σ-algebra)**.** *Let $(\Omega, \mathcal{O})$ be a topological space with countable basis $\mathcal{C} \subseteq \mathcal{O}$. Then, $\sigma(\mathcal{O}) = \sigma(\mathcal{C})$.*

Proof. We only need to show that $\mathcal{O} \subseteq \sigma(\mathcal{C})$. However, this is clear since $A \in \mathcal{O}$ can be represented as a countable union of sets from $\mathcal{C}$. See Lemma A.5. ☐

Lemma 1.9 (Borel σ-algebra is generated by intervals generated)**.** *Let*

$$\begin{aligned}
\mathcal{C}_1 &= \{[-\infty, b] : b \in \mathbb{Q}\} \ \textit{or} \\
\mathcal{C}_2 &= \{(a, b] : a, b \in \mathbb{Q}, a \leq b\} \\
\mathcal{C}_3 &= \{(a, b) : a, b \in \mathbb{Q}, a \leq b\} \\
\mathcal{C}_4 &= \{[a, b] : a, b \in \mathbb{Q}, a \leq b\}.
\end{aligned}$$

Then $\sigma(\mathcal{C}_i) = \mathcal{B}(\overline{\mathbb{R}})$, $i = 1, \ldots, 4$.

Proof. The set system $\mathcal{C}_3$ is a countable basis of Euclidean topology on $\overline{\mathbb{R}}$. So, in this case, the statement follows from Lemma 1.8.

We only show the statement for $\mathcal{C}_1$ and $\mathcal{C}_2$, the statement for $\mathcal{C}_4$ follows analogously. Firstly, $\mathcal{C}_2 := \{A \setminus B : A, B \in \mathcal{C}_1\} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\} \subseteq \sigma(\mathcal{C}_2)$ is the semi-ring generated by $\mathcal{C}_1$ from Example 1.3. Thus $\sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2)$ and it is sufficient to show that $\sigma(\mathcal{C}_2) = \mathcal{B}(\overline{\mathbb{R}})$.

Let $\mathcal{O}$ be as in Definition A.1.8 with $\Omega = \mathbb{R}$. We show (i) that $A \in \mathcal{O}$ implies $A \in \sigma(\mathcal{C}_2)$, and (ii) $A \in \mathcal{C}_2$ implies $A \in \sigma(\mathcal{O})$. It then follows that $\mathcal{O} \subseteq \sigma(\mathcal{C}_2) \subseteq \sigma(\mathcal{O})$, i.e. $\sigma(\mathcal{O}) = \sigma(\mathcal{C}_2)$. For (i) let $A \in \mathcal{O}$. We claim

$$A = \bigcup \{(a, b] : [a, b] \subseteq A, a, b \in \mathbb{Q}\}, \tag{1.2}$$

and note that the right-hand side is an element of $\sigma(\mathcal{C}_2)$. Here, '$\supseteq$' is clear. To see '$\subseteq$', we choose $x \in A$. Then, by definition of $\mathcal{O}$, there is a $\varepsilon > 0$ so that $B_\varepsilon(x) \subseteq A$. However, there are also $a, b \in \mathbb{Q}$ with $a \leq b$ and $x \in (a, b] \subseteq B_\varepsilon(x)$. Thus '$\subseteq$' is shown and (i) follows.
For (ii) we proceed similarly; let $A \in \mathcal{C}_2$. Then obviously

$$A = \bigcap_{n=1}^{\infty} \big(a, b + \tfrac{1}{n}\big).$$

Since $\big(a, b + \frac{1}{n}\big) \in \mathcal{O}$, then $A \in \sigma(\mathcal{O})$. ☐

Example 1.10 (Borel measurable sets)**.** *Of course, all countable intersections and unions of intervals according to Lemma 1.9 in $\mathcal{B}(\overline{\mathbb{R}})$. Let, for example*

$$\begin{aligned}
&A_1 = [0, \tfrac{1}{3}] \cup [\tfrac{2}{3}, 1], \\
&A_2 = [0, \tfrac{1}{9}] \cup [\tfrac{2}{9}, \tfrac{3}{9}] \cup [\tfrac{6}{9}, \tfrac{7}{9}] \cup [\tfrac{8}{9}, 1], \\
&A_3 = [0, \tfrac{1}{27}] \cup [\tfrac{2}{27}, \tfrac{3}{27}] \cup [\tfrac{6}{27}, \tfrac{7}{27}] \cup [\tfrac{8}{27}, \tfrac{9}{27}] \cup [\tfrac{18}{27}, \tfrac{19}{27}] \cup [\tfrac{20}{27}, \tfrac{21}{27}] \cup [\tfrac{24}{27}, \tfrac{25}{27}] \cup [\tfrac{26}{27}, 1], \\
&\cdots,
\end{aligned}$$

then $A = \bigcap_{n=1}^\infty$ denotes Cantor's discontinuum. This set is measurable as a countable intersection of finite unions of intervals. In Example 2.27 we will get to know an example of a non-Borel-measurable set.

1.3 Dynkin systems

In measure theory, it is often necessary to show that a certain set system $\mathcal{F}$ is a σ-algebra and contains a semi-ring $\mathcal{H}$. The Dynkin systems discussed in this section are very helpful here. Because of Theorem 1.13 it is sufficient to show that $\mathcal{F}$ is a $\cap$-stable Dynkin system with $\mathcal{H} \subseteq \mathcal{F}$. This is often easier than showing directly that $\mathcal{F}$ is a σ-algebra.

Definition 1.11 (Dynkin system). *1. A set system $\mathcal{D}$ is called* Dynkin system *(on Ω) if (i) $\Omega \in \mathcal{D}$, (ii) it is set-difference-stable for subsets (i.e. $A, B \in \mathcal{D}$ and $A \subseteq B$ imply $B \setminus A \in \mathcal{D}$ and (iii) $A_1, A_2, \ldots \in \mathcal{D}$ and $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ imply $\bigcup_{n=1}^\infty A_n \in \mathcal{D}$.*

2. For $\mathcal{C} \subseteq 2^\Omega$, we set

$$\lambda(\mathcal{C}) := \bigcap \{\mathcal{D} \supseteq \mathcal{C} \text{ Dynkin-system}\}.$$

Example 1.12 (σ-algebras are Dynkin systems). *1.* Every σ-algebra is a Dynkin system: *Let $\mathcal{F}$ be a σ-algebra. Then $A, B \in \mathcal{F}$ imply $A^c \in \mathcal{F}$ and therefore $\Omega = A \cup A^c \in \mathcal{F}$ and $B \setminus A = B \cap A^c \in \mathcal{F}$.*

2. A Dynkin system $\mathcal{D}$ is complement-stable, since

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

Theorem 1.13 ($\cap$-stable Dynkin systems). *Let $\mathcal{D}$ be a Dynkin system and $\mathcal{C} \subseteq \mathcal{D}$ be $\cap$-stable. Then $\sigma(\mathcal{C}) \subseteq \mathcal{D}$. In particular, every $\cap$-stable Dynkin system is a σ-algebra.*

Proof. Let $\lambda(\mathcal{C})$ be the Dynkin system generated by $\mathcal{C}$ (see Definition 1.11). So, we find $\lambda(\mathcal{C}) \subseteq \mathcal{D}$. We will show that $\lambda(\mathcal{C})$ is a σ-algebra, because then $\sigma(\mathcal{C}) \subseteq \sigma(\lambda(\mathcal{C})) = \lambda(\mathcal{C}) \subseteq \mathcal{D}$. For showing that $\lambda(\mathcal{C})$ is a σ-algebra, it suffices to show that $\lambda(\mathcal{C})$ is $\cap$-stable. Then, since $\lambda(\mathcal{C})$ is complement-stable, writing $A \cup B = (A^c \cap B^c)^c$, we see that $\lambda(\mathcal{C})$ is $\cup$-stable. Hence, for $A_1, A_2, ... \in \lambda(\mathcal{C})$, we find $\bigcup_{n=1}^\infty A_n = \bigcup_{n=1}^\infty \bigcup_{i=1}^n A_i \in \lambda(\mathcal{C})$.

So, it remains to show that $A, B \in \lambda(\mathcal{C})$ imply $A \cap B \in \lambda(\mathcal{C})$: If $A, B \in \mathcal{C}$, this is clear due to the $\cap$-stability of $\mathcal{C}$. For $B \in \mathcal{C}$ we set

$$\mathcal{D}_B := \{A \subseteq \Omega : A \cap B \in \lambda(\mathcal{C})\} \supseteq \mathcal{C}.$$

Then $\mathcal{D}_B$ is a Dynkin system since (i) $\Omega \in \mathcal{D}_B$, (ii) for $A, C \subseteq \mathcal{D}_B$ we have $A \cap B, C \cap B \in \lambda(\mathcal{C})$ and if $A \subseteq C$ we find $A \cap B \subseteq C \cap B$, thus $(C \setminus A) \cap B = (C \cap B) \setminus (A \cap B) \in \lambda(\mathcal{C})$ and (iii) for $A_1, A_2, \ldots \in \mathcal{D}_B$ we have $A_1 \cap B, A_2 \cap B, \ldots \in \lambda(\mathcal{C})$ and with $A_1 \subseteq A_2 \subseteq \cdots$ we have $A_1 \cap B \subseteq A_2 \cap B \subseteq \cdots$, thus $\left(\bigcup_{n=1}^\infty A_n\right) \cap B = \left(\bigcup_{n=1}^\infty A_n \cap B\right) \in \lambda(\mathcal{C})$.

Since $\mathcal{C} \subseteq \mathcal{D}_B$ and $\mathcal{D}_B$ is a Dynkin system, we find that $\lambda(\mathcal{C}) \subseteq \mathcal{D}_B$. This means that $A \in \lambda(\mathcal{C})$ and $B \in \mathcal{C}$ imply $A \cap B \in \lambda(\mathcal{C})$. We now set for an $A \in \lambda(\mathcal{C})$

$$\mathcal{B}_A := \{B \subseteq \Omega : A \cap B \in \lambda(\mathcal{C})\}.$$

As above, we show that $\mathcal{B}_A$ is a Dynkin system with $\mathcal{C} \subseteq \mathcal{B}_A$. Therefore, $\lambda(\mathcal{C}) \subseteq \mathcal{B}_A$. In particular, for $A, B \in \lambda(\mathcal{C})$, we find $A \cap B \in \lambda(\mathcal{C})$, i.e. $\lambda(\mathcal{C})$ is $\cap$-stable. This concludes the proof of the first assertion. The second assertion follows from setting $\mathcal{C} := \mathcal{D}$. □

1.4 Compact systems

In topology, compact subsets of an underlying set play an important role; see Appendix A. Here, we introduce an important connection between compact sets and measure theory. The resulting compact systems play an important role in the proof of Theorem 2.10. Here it is shown that the σ-additivity of the set function follows from the additivity of a set function and an approximation property with respect to a compact system.

Definition 1.14 (Compact system)**.** *A $\cap$-stable set system $\mathcal{K}$ is called* compact system *(on Ω) if $\bigcap_{n=1}^\infty K_n = \emptyset$ with $K_1, K_2, \ldots \in \mathcal{K}$ implies that there is a $N \in \mathbb{N}$ with $\bigcap_{n=1}^N K_n = \emptyset$.*

Example 1.15 (Compact sets)**.** Compact sets form a compact system: *Let (Ω, r) be a metric space and $\mathcal{O}$ the topology generated by r. Then every $\cap$-stable $\mathcal{K} \subseteq \{K \subseteq \Omega : K \text{ compact}\}$ is a compact system.*
Indeed: let $\bigcap_{n=1}^\infty K_n = \emptyset$. Then both K_1 and $L_n := K_1 \cap K_n \subseteq K_1$ are closed for $n = 1, 2, \ldots$ according to Lemma A.8 and because of the compactness of K_1 there is an N with $\bigcap_{n=1}^N K_n = \emptyset$ according to Proposition A.9.

Lemma 1.16 (Extension of compact systems)**.** *Let $\mathcal{K}$ be a compact system. Then*

$$\mathcal{K}_\cup := \Big\{ \bigcup_{i=1}^n K_i : K_1, \ldots, K_n \in \mathcal{K}, n \in \mathbb{N} \Big\}$$

is also a compact system.

Proof. It is clear that $\mathcal{K}_\cup$ is $\cap$-stable. Let $L_1 = \bigcup_{j=1}^{m_1} K_j^1, L_2 = \bigcup_{j=1}^{m_2} K_j^2, \ldots \in \mathcal{K}_\cup$ with $\bigcap_{n=1}^N L_n \neq \emptyset$ for all $N \in \mathbb{N}$. We have to show that $\bigcap_{n=1}^\infty L_n \neq \emptyset$. For this, we use induction over N to show the following:

> For every $N \in \mathbb{N}$ there are sets $K_1, \ldots, K_N \in \mathcal{K}$ with $K_n \subseteq L_n$, $n = 1, \ldots, N$, such that for all $k \in \mathbb{N}_0$ we have $K_1 \cap \cdots \cap K_N \cap L_{N+1} \cap \cdots \cap L_{N+k} \neq \emptyset$.

Let $N = 1$ and $k \in \mathbb{N}_0$ arbitrary. Since $\bigcap_{n=1}^{1+k} L_n = \bigcup_{j=1}^{m_1} \Big(K_j^1 \cap \bigcap_{n=1}^{1+k} L_n\Big) \neq \emptyset$, there is a $j \in \{1, \ldots, m_1\}$ such that $K_j^1 \cap \bigcap_{n=1}^{1+k} L_n \neq \emptyset$. Set $K_1 := K_j^1$, and the assertion is shown for $N = 1$.
For the induction step, assume the assertion holds for $N - 1$ and any $k \in \mathbb{N}_0$. Recall that $L_N = \bigcup_{j=1}^{m_N} K_j^N$ for $K_1^N, \ldots, K_{m_N}^N \in \mathcal{K}$. Thus, according to the induction hypothesis,

$$\begin{aligned} &K_1 \cap \cdots \cap K_{N-1} \cap \Big(\bigcup_{j=1}^{m_N} K_j^N\Big) \cap L_{N+1} \cap \cdots \cap L_{N+k} \\ &= \bigcup_{j=1}^{m_N} K_1 \cap \cdots \cap K_{N-1} \cap K_j^N \cap L_{N+1} \cap \cdots \cap L_{N+k} \neq \emptyset. \end{aligned}$$

Thus there is a j, so that $K_1 \cap \cdots \cap K_{N-1} \cap K_j^N \cap L_{N+1} \cap \cdots \cap L_{N+k} \neq \emptyset$ for all $k \in \mathbb{N}$. Set $K_N := K_j^N$, which completes the induction step.

If we set $k = 0$ in the above assertion, we see that there are $K_1, K_2, \ldots \in \mathcal{K}$ and $K_n \subseteq L_n$, $n \in \mathbb{N}$ with $\bigcap_{n=1}^N K_n \neq \emptyset$ for all $N \in \mathbb{N}$. Since $\mathcal{K}$ is a compact system,

$$\emptyset \neq \bigcap_{n=1}^\infty K_n \subseteq \bigcap_{n=1}^\infty L_n$$

and the assertion is shown. □

2 Set functions

By a set function, we mean a function $m : \mathcal{A} \subseteq 2^\Omega \to \mathbb{R}$. The idea is that $m(A)$ for $A \in \mathcal{A}$ describes the volume of A. Here, *volume* might be an actual volume in space, or something more abstract. In probability theory we think of $m(A)$ as the probability that A occurs. (Mostly, we write $\mathbb{P}$ for the set function.) For any such set function, some requirements seem natural, irrespective of the meaning of *volume*. For example, the empty set (no spatial volume, or an event that never occurs) should be assigned *volume* 0, or m should behave countably additive, see (2.1). In probability theory, Ω consists of all possible outcomes of an experiment, so a natural requirement is $m(\Omega) = 1$. In other words, the probability that there is any outcome of the experiment is 1.

The concept of the probability measure is central to probability theory. As it turns out, measures must be defined on σ-algebras (so usually, $\mathcal{A}$ is a σ-algebra) so that the requirement of countable additivity can be met. In this section we give the most important steps to construct such measures. In *Analysis 3*, the Lebesgue measure was not introduced, which follows along the same lines. However, note that large parts of *Analysis* are dealing with $\Omega \subseteq \mathbb{R}^d$. In probability theory, however, outcomes of experiments might be elements of much larger spaces. When observing a randomly changing quantity (e.g. the position of a particle in space, or a stock price), we might need a probability measure in $\mathcal{C}([0,\infty), \mathbb{R}^d)$ (the set of continuous functions $X : [0,\infty) \to \mathbb{R}$).

2.1 Measures and outer measures

We will now consider functions $\mu : \mathcal{C} \to \overline{\mathbb{R}}_+$ if $\mathcal{C}$ is a semi-ring, ring or σ-algebra. Most important for probability theory is certainly the concept of a *probability measure*, which describes the special case $\mu(\Omega) = 1$.

Definition 2.1 (Measure and outer measure)**.** *For some $\mathcal{F} \subseteq 2^\Omega$, we call any $\mu : \mathcal{F} \to \overline{\mathbb{R}}_+$ a set function.*

1. *The set function μ is called* finitely additive *if for disjoint $A_1, \dots, A_n \in \mathcal{F}$ with $\biguplus_{k=1}^n A_k \in \mathcal{F}$,*

$$\mu\Big(\biguplus_{k=1}^n A_k\Big) = \sum_{k=1}^n \mu(A_k). \tag{2.1}$$

   It is called sub-additive *if for (any, not necessarily disjoint) $A_1, \dots, A_n \in \mathcal{F}$ with $\bigcup_{k=1}^n A_k \in \mathcal{F}$,*

$$\mu\Big(\bigcup_{k=1}^n A_k\Big) \leq \sum_{k=1}^n \mu(A_k). \tag{2.2}$$

2. *A mapping $\mu : \mathcal{F} \to \mathbb{R}_+$ is called σ-*additive *if* (2.1) *also holds for $n = \infty$. It is called σ-*sub-additive *if* (2.2) *also applies for $n = \infty$. It is called* monotonic *if for any $A, B \in \mathcal{F}$ with $A \subseteq B$ we find $\mu(A) \leq \mu(B)$.*

3. *If $\mu(\Omega) < \infty$, then μ is called* finite. *If there is a sequence $\Omega_1, \Omega_2, \dots \in \mathcal{F}$ with $\bigcup_{n=1}^\infty \Omega_n = \Omega$ and $\mu(\Omega_n) < \infty$ for all $n = 1, 2, \dots$, then μ is called σ-*finite.

4. *Let $\mathcal{F}$ be a σ-algebra and $\mu : \mathcal{F} \to \mathbb{R}_+$. If μ is σ-additive, then μ is a* measure *(on $\mathcal{F}$) and $(\Omega, \mathcal{F}, \mu)$ is a* measure space. *If $\mu(\Omega) < \infty$, then μ is called* finite measure *and if $\mu(\Omega) = 1$, then μ is called a* probability measure *or a* probability distribution *or simply a* distribution. *Furthermore, $(\Omega, \mathcal{F}, \mu)$ is then called a* probability space.

5. *Let $(\Omega, \mathcal{O})$ be a topological space and μ a measure on $\mathcal{B}(\mathcal{O})$ (the Borel σ-algebra, see Definition 1.7). Then the smallest closed set F with $\mu(F^c) = 0$ is called the* support of μ[3].

6. *A σ-subadditive, monotone mapping $\mu^* : 2^\Omega \to \mathbb{R}_+$ is called* outer measure *if $\mu^*(\emptyset) = 0$. A set $A \subseteq \Omega$ is called* μ^*-measurable *if*

$$\mu(E) = \mu(E \cap A) + \mu(E \cap A^c) \tag{2.3}$$

for all $E \subseteq \Omega$.

7. *Let $\mathcal{F}$ be $\cap$-stable and $\mathcal{K} \subseteq \mathcal{F}$ a compact system. Then μ is called inner $\mathcal{K}-$regular if for all $A \in \mathcal{K}$*

$$\mu(A) = \sup_{\mathcal{K} \ni K \subseteq A} \mu(K).$$

Example 2.2 (Examples of set functions)**.** 1. *We will often deal with set functions on $\mathcal{H} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$ from Example 1.3. For example, $\mu((a, b]) = b - a$ defines an additive, σ-finite set function on $\mathcal{H}$. We will extend this function uniquely to the Borel σ-algebra $\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{H})$ (see lemma 1.9), which will give the Lebesgue measure, see corollary 2.18.*

2. *Another frequently used example is the* Dirac measures. *If $\omega' \in \Omega$, then*

$$\delta_{\omega'} : \begin{cases} 2^\Omega & \to \{0, 1\} \\ A & \mapsto 1_{\{\omega' \in A\}} \end{cases}$$

is a (probability) measure.

3. *If $\mu_i = \delta_{\omega_i}$, $i \in I$, then $\mu := \sum_{i \in I} \delta_{\omega_i}$ is called a* counting measure.

4. *If $\mu_i, i \in I$ are measures on a σ-algebra $\mathcal{F}$. Then for $a_i \in \mathbb{R}_+, i \in I$, $\sum_{i \in I} a_i \mu_i$ is also a measure. Examples of this are well known from the lecture* Elementary probability theory. *There, for example, with $\mathcal{F} = 2^{\mathbb{N}_0}$ and δ_k as in 2.*

$$\mu_{Poi(\gamma)} := \sum_{k=0}^{\infty} e^{-\gamma} \frac{\gamma^k}{k!} \cdot \delta_k$$

the Poisson distribution on $2^{\mathbb{N}_0}$ with parameter γ,

$$\mu_{geo(p)} := \sum_{k=1}^{\infty} (1 - p)^{k-1} p \cdot \delta_k$$

[3]We will see later that this smallest set indeed exists uniquely.

the geometric distribution with success parameter p and

$$\mu_{B(n,p)} := \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot \delta_k$$

the binomial distribution $B(n,p)$.

Remark 2.3 (Contents and premeasures)**.** *Finite additive set functions are often called* content*, σ-additive set functions that are not defined on σ-algebras are often called premeasures. The measures defined on a Borel σ-algebra that are regular with respect to the compact sets from the inside are called Radon measures. We will not use these terms.*

Lemma 2.4 (Unions written as disjoint unions)**.** *Let $\mathcal{H}$ be a semi-ring, and $A, A_1, ..., A_n \in \mathcal{H}$. Then, there are $m \in \mathbb{N}$ and $B_1, ..., B_m \in \mathcal{H}$ pairwise disjoint and $A \setminus \bigcup_{i=1}^n A_i = \biguplus_{j=1}^m B_j$.*

Proof. We proceed by induction on n. If $n = 1$, the assertion is true by the definition of a semi-ring. Assume the assertion holds for some n, i.e. there is $m \in \mathbb{N}$ and $B_1, ..., B_m$ with $A \setminus \bigcup_{i=1}^n A_i = \biguplus_{j=1}^m B_j$. Then, we can write $B_j \setminus A_{n+1} = \biguplus_{k=1}^{k_j} C_k^j$ for $C_1^j, ..., C_{k_j}^j \in \mathcal{H}$. Then, write

$$A \setminus \bigcup_{i=1}^{n+1} A_i = \Big(A \setminus \bigcup_{i=1}^{n} A_i\Big) \setminus A_{n+1} = \biguplus_{j=1}^{m} B_j \setminus A_{n+1} = \biguplus_{j=1}^{m} \biguplus_{k=1}^{k_j} C_k^j.$$

This concludes the proof, since the latter disjoint union is over a finite set. ☐

Lemma 2.5 (Set-functions on semi-rings)**.** *Let $\mathcal{H}$ be a semi-ring and $\mu : \mathcal{H} \to [0, \infty]$ additive. Then, m is monotone and sub-additive. In addition, μ is σ-additive iff it is σ-sub-additive.*

Proof. We start by monotonicity. Let $A, B \in \mathcal{H}$ with $A \subseteq B$ and $C_1, ..., C_k \in \mathcal{H}$ with $B \setminus A = \biguplus_{i=1}^k C_i$. Therefore, we can write $\mu(A) \leq \mu(A) + \sum_{i=1}^k \mu(C_i) = \mu(B)$.

Next, we claim that for $A \in \mathcal{H}$ and $A_1, ..., A_n \in \mathcal{H}$ disjoint with $\biguplus_{I \in \mathcal{I}} A_i \subseteq A$, we have $\sum_{i=1}^n \mu(A_i) \leq m(A)$. For this, write $A \setminus \biguplus_{i=1}^n A_i = \biguplus_{j=1}^m B_j$ as in Lemma 2.4. Then,

$$\mu(A) = \mu\Big(\biguplus_{i=1}^{n} A_i \uplus \biguplus_{j=1}^{m} B_j\Big) = \sum_{i=1}^{n} \mu(A_i) + \sum_{j=1}^{m} \mu(B_j) \geq \sum_{i=1}^{n} \mu(A_i). \tag{2.4}$$

For sub-additivity, let $A_1, ..., A_n \in \mathcal{H}$ with $\bigcup_{i=1}^n A_i \in \mathcal{H}$. We need to show $\mu\Big(\bigcup_{i=1}^n A_i\Big) \leq \sum_{i=1}^n \mu(A_i)$. For $i = 2, ..., n$, we write

$$\bigcup_{i=1}^{n} A_i = \biguplus_{i=1}^{n} \Big(A_i \setminus \bigcup_{j=1}^{i-1} A_j\Big) = \biguplus_{i=1}^{n} \biguplus_{k=1}^{k_i} C_k^i$$

with C_k^i as in Lemma 2.4. So, since $\biguplus_{k=1}^{k_i} C_k^i \subseteq A_i \in \mathcal{H}$,

$$\mu\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_{k=1}^{n} \sum_{k=1}^{k_i} \mu(C_k^i) \leq \sum_{i=1}^{n} \mu(A_i).$$

Now, we show that μ is σ-additive $\iff$ it is σ-sub-additive.

'$\Rightarrow$': Here, just copy the proof of sub-additivity, but using $n = \infty$. For '$\Leftarrow$', let $A_1, A_2, ... \in \mathcal{H}$ be pairwise disjoint with $A = \biguplus_{i=1}^\infty A_i \in \mathcal{H}$. Since μ is monotone and for any $n \in \mathbb{N}$, we have $\biguplus_{i=1}^n A_i \subseteq A$ (hence $\sum_{i=1}^n \mu(A_i) \leq \mu(A)$ by (2.4)),

$$\sum_{i=1}^\infty \mu(A_i) = \sup_{n\in\mathbb{N}} \sum_{i=1}^n \mu(A_i) \leq \mu(A) \leq \sum_{i=1}^\infty \mu(A_i)$$

by σ-sub-additivity. So, σ-additivity follows. □

Lemma 2.6 (Extension of set functions on semi-rings)**.** *Let $\mathcal{H}$ be a semi-ring, $\mathcal{R}$ the ring generated by $\mathcal{H}$ from Lemma 1.5 and μ an additive function on $\mathcal{H}$. Define $\widetilde{\mu}$ on $\mathcal{R}$ by*

$$\widetilde{\mu}\Big(\biguplus_{i=1}^n A_i\Big) := \sum_{i=1}^n \mu(A_i)$$

for $A_1, \dots, A_n \in \mathcal{H}$ disjoint. Then $\widetilde{\mu}$ is the only additive extension of μ on $\mathcal{R}$ that coincides with μ on $\mathcal{H}$. Moreover, $\widetilde{\mu}$ is σ-additive if and only if μ is σ-additive.

Proof. We only need to show that $\widetilde{\mu}$ is well-defined. All other properties follow by definition of $\widetilde{\mu}$. So, let $A_1, \dots, A_m, B_1, \dots, B_n \in \mathcal{H}$ with $\biguplus_{i=1}^m A_i = \biguplus_{j=1}^n B_j$. Since

$$A_i = \biguplus_{j=1}^n A_i \cap B_j, \qquad B_j = \biguplus_{i=1}^m A_i \cap B_j,$$

we write using additivity of $\widetilde{\mu}$

$$\sum_{i=1}^m \mu(A_i) = \sum_{i=1}^m \sum_{j=1}^n \mu(A_i \cap B_j) = \sum_{j=1}^n \sum_{i=1}^m \mu(A_i \cap B_j) = \sum_{j=1}^n \mu(B_j).$$

□

Proposition 2.7 (Inclusion-exclusion principle)**.** *Let μ be an additive set function on a ring $\mathcal{R}$ and I be finite. Then for $A_i \in \mathcal{R}$, $i \in I$, it holds that*

$$\mu\Big(\bigcup_{i\in I} A_i\Big) = \sum_{J\subseteq I} (-1)^{|J|+1} \mu\Big(\bigcap_{j\in J} A_j\Big)$$

In particular, if $I = \{1, 2\}$,

$$\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2) - \mu(A_1 \cap A_2)$$

and if $I = \{1, 2, 3\}$,

$$\begin{aligned}\mu(A_1 \cup A_2 \cup A_3) = {} & \mu(A_1) + \mu(A_2) + \mu(A_3) \\ & - \mu(A_1 \cap A_2) - \mu(A_1 \cap A_3) - \mu(A_2 \cap A_3) + \mu(A_1 \cap A_2 \cap A_3).\end{aligned}$$

Proof. We use induction over $|I|$. For $|I| = 2$ the assertion is clear because $A_1 \cup A_2 = A_1 \uplus (A_2 \setminus A_1)$ and $(A_2 \setminus A_1) \uplus (A_1 \cap A_2) = A_2$. Assume it applies to all I with $|I| = n$, and consider some I with $|I| = n+1$. Without loss of gnerality,we write $I = \{1, \dots, n+1\}$. By additivity of μ

$$\begin{aligned}\mu\Big(\bigcup_{i=1}^{n+1} A_i\Big) &= \mu\Big(\bigcup_{i=1}^{n}(A_i \cup A_{n+1})\Big) \\ &= \sum_{\emptyset \neq J \subseteq \{1,\dots,n\}} (-1)^{|J|+1} \mu\Big(A_{n+1} \cup \bigcap_{j \in J} A_j)\Big) \\ &= \sum_{\emptyset \neq J \subseteq \{1,\dots,n\}} (-1)^{|J|+1} \Big(\mu(A_{n+1}) + \mu\Big(\bigcap_{j \in J} A_j)\Big) - \mu\Big(\bigcap_{j \in J} A_j \cap A_{n+1})\Big)\Big) \\ &= \mu(A_{n+1}) + \sum_{\emptyset \neq J \subseteq \{1,\dots,n\}} (-1)^{|J|+1} \Big(\mu\Big(\bigcap_{j \in J} A_j)\Big) - \mu\Big(\bigcap_{j \in J} A_j \cap A_{n+1})\Big)\Big) \\ &= \sum_{J \subseteq \{1,\dots,n+1\}} (-1)^{|J|+1} \mu\Big(\bigcap_{j \in J} A_j)\Big).\end{aligned}$$

□

2.2 σ-additivity

The finite additivity of set functions is a requirement that can often be verified. The situation is different with σ-additivity. We will now look at alternative formulations for σ-additivity.

Proposition 2.8 (Continuity of from below and from above)**.** *Let $\mathcal{R}$ be a ring and $\mu : \mathcal{R} \to \overline{\mathbb{R}}_+$ be additive. Consider the following properties:*

1. *μ is σ-additive;*
2. *μ is σ-subadditive;*
3. *μ is continuous from below, i.e. for $A, A_1, A_2, \dots \in \mathcal{R}$ and $A_1 \subseteq A_2 \subseteq \dots$ with $A = \bigcup_{n=1}^\infty A_n$ we have $\mu(A) = \lim_{n\to\infty} \mu(A_n)$;*
4. *μ is continuous from above in $\emptyset$, i.e. for $A_1, A_2, \dots \in \mathcal{R}$, $\mu(A_1) < \infty$ and $A_1 \supseteq A_2 \supseteq \dots$ with $\bigcap_{n=1}^\infty A_n = \emptyset$ we have $\lim_{n\to\infty} \mu(A_n) = 0$.*
5. *μ is continuous from above, i.e. for $A, A_1, A_2, \dots \in \mathcal{R}$, $\mu(A_1) < \infty$ and $A_1 \supseteq A_2 \supseteq \dots$ with $A = \bigcap_{n=1}^\infty A_n$ we have $\mu(A) = \lim_{n\to\infty} \mu(A_n)$.*

Then,

$$1. \iff 2. \iff 3. \implies 4. \iff 5.$$

Furthermore, 4. $\implies$ 3. *holds if $\mu(A) < \infty$ for all $A \in \mathcal{R}$.*

Proof. 1.⇔2. follows from Lemma 2.6, since $\mathcal{R}$ is a semi-ring.

1.⇒3.: Let μ be σ-additive and $A, A_1, A_2, \dots \in \mathcal{R}$ as in 3. Then, with $A_0 = \emptyset$,

$$\mu(A) = \sum_{n=1}^\infty \mu(A_n \setminus A_{n-1}) = \lim_{N\to\infty} \sum_{n=1}^N \mu(A_n \setminus A_{n-1}) = \lim_{N\to\infty} \mu(A_N).$$

3.⇒1.: Let $B_1, B_2, \dots \in \mathcal{R}$ be pairwise disjoint and $B = \biguplus_{n=1}^\infty B_n \in \mathcal{R}$. Then, for $A_N = \biguplus_{n=1}^N B_n$,

$$\mu(B) = \lim_{N\to\infty} \mu(A_N) = \sum_{n=1}^\infty \mu(B_n).$$

4.⇒5.: Let $A, A_1, A_2, \dots \in \mathcal{R}$ be as assumed in 5. Further, let $B_n := A_n \setminus A$. Then $B_1, B_2, \dots$ fulfills the conditions of 4., so $\mu(B_n) \xrightarrow{n\to\infty} 0$, i.e. $\mu(A_n) = \mu(B_n) + \mu(A) \xrightarrow{n\to\infty} \mu(A)$.

5.⇒4.: is clear.

3.⇒4.: Let $A_1, A_2, \dots \in \mathcal{R}$ be as assumed in 4. Set $B_n := A_1 \setminus A_n, n \in \mathbb{N}$. Then $B = A_1, B_1, B_2, \dots \in \mathcal{R}$ fulfills the conditions of 3, and thus $\mu(A_1) = \lim_{n\to\infty} \mu(B_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n)$, from which 4. follows.

4.⇒3. if $\mu(A) < \infty$ for all $A \in \mathcal{R}$. Let $A, A_1, A_2, \dots \in \mathcal{R}$ be as assumed in 3. Set $B_n := A \setminus A_n \in \mathcal{R}, n \in \mathbb{N}$. Then $\bigcap_{n=1}^\infty B_n = \emptyset$, i.e. $0 = \lim_{n\to\infty} \mu(B_n) = \mu(A) - \lim_{n\to\infty} \mu(A_n)$, from which 3. follows. Here, the last equality uses the condition that $\mu(A) < \infty$. □

We now want to take a closer look at set functions which are inner regular with respect to a compact system. For measures, inner regularity with respect to the system of all compact sets (which is a compact system due to Example 1.15) is fulfilled on Polish spaces, as the next result shows. This will play an important role in the theory of weak convergence, an important concept in any course on probability theory.

Lemma 2.9. *If $(\Omega, \mathcal{O})$ is Polish and μ is a finite measure on $\mathcal{B}(\mathcal{O})$, then for every $\varepsilon > 0$ there exists a compact set $K \subseteq \Omega$ with $\mu(\Omega \setminus K) < \varepsilon$.*

Proof. First, note that compact sets are closed according to Lemma A.8, so all compact sets are in $\mathcal{B}(\mathcal{O})$ and thus $\mu(\Omega \setminus K)$ is well-defined.

Let $\varepsilon > 0$. Since Ω is separable (see Definition A.1), there is a countable set $\{\omega_1, \omega_2, \dots\} \subseteq \Omega$ which is dense. In particular, for all n, we find $\Omega = \bigcup_{k=1}^\infty B_{1/n}(\omega_k)$. Since μ is continuous from above (Proposition 2.8),

$$0 = \mu\Big(\Omega \setminus \bigcup_{k=1}^\infty B_{1/n}(\omega_k)\Big) = \lim_{N\to\infty} \mu\Big(\Omega \setminus \bigcup_{k=1}^N B_{1/n}(\omega_k)\Big).$$

Thus there is an $N_n \in \mathbb{N}$ with $\mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k^n)\Big) < \varepsilon/2^n$. Now, consider

$$A := \bigcap_{n=1}^\infty \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k).$$

By definition, this set is totally bounded (i.e. for all radii $\varepsilon > 0$. it can be covered by a finite number of balls of radius $\varepsilon > 0$. Hence, according to Lemma A.9, A is relatively compact. Furthermore, (recall that $\overline{A}$ is the closure of A, which is compact according to Proposition A.9),

$$\mu(\Omega \setminus \overline{A}) \leq \mu(\Omega \setminus A) \leq \mu\Big(\bigcup_{n=1}^\infty \Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k)\Big)\Big) \leq \sum_{n=1}^\infty \mu\Big(\Omega \setminus \bigcup_{k=1}^{N_n} B_{1/n}(\omega_k)\Big) < \varepsilon.$$

This proves the assertion. □

Theorem 2.10 (Inner regular additive set functions are **σ**-additive)**.** *Let $\mathcal{H}$ be a semi-ring and $\mu : \mathcal{H} \to \mathbb{R}_+$ finite, finitely additive and inner regular with respect to a compact system $\mathcal{K} \subseteq \mathcal{H}$. Then μ is σ-additive.*

Proof. As in Lemma 2.6, the set function μ can be uniquely extended to the ring $\mathcal{R}(\mathcal{H})$ generated by $\mathcal{H}$ (see Lemma 1.5). Furthermore, according to Lemma 1.16, the system $\mathcal{K}_\cup \subseteq \mathcal{R}(\mathcal{H})$, which consists of unions of sets in $\mathcal{K}$, is also compact. Choose $\varepsilon > 0$ and $A = \bigcup_{i=1}^n A_i \in \mathcal{R}(\mathcal{H})$ with $A_1, \dots, A_n \in \mathcal{H}$, then there are compact sets $K_1, \dots, K_n \in \mathcal{K} \subseteq \mathcal{H}$ with $\mu(A_i) \leq \mu(K_i) + \frac{\varepsilon}{n}$ for $i = 1, \dots, n$. This means that the extension of μ to the ring $\mathcal{R}(\mathcal{H})$ is inner regular with respect to $\mathcal{K}_\cup$, since

$$\mu\Big(\bigcup_{i=1}^n A_i\Big) = \sum_{i=1}^n \mu(A_i) \leq \Big(\sum_{i=1}^n \mu(K_i)\Big) + \varepsilon = \mu\Big(\bigcup_{i=1}^n K_i\Big) + \varepsilon.$$

This means that μ is $\mathcal{K}_\cup$-regular from the inside. o, without loss of generality, we assume that $\mathcal{H}$ is a ring and $\mathcal{K}$ is $\cup$-stable. We now show that μ is continuous from above in $\emptyset$. This is sufficient according to Proposition 2.8 because of the finiteness of μ on $\mathcal{H}$. Let $A_1, A_2, \dots \in \mathcal{H}$ with $A_1 \supseteq A_2 \supseteq \cdots$ and $\bigcap_{n=1}^\infty A_n = \emptyset$ and $\varepsilon > 0$. Choose $K_1, K_2, \dots \in \mathcal{K}$ with $K_n \subseteq A_n, n \in \mathbb{N}$ and

$$\mu(A_n) \leq \mu(K_n) + \varepsilon 2^{-n}.$$

Then, $\bigcap_{n=1}^\infty K_n \subseteq \bigcap_{n=1}^\infty A_n = \emptyset$, which means that there is a $N \in \mathbb{N}$ with $\bigcap_{n=1}^N K_n = \emptyset$ since $\mathcal{K}$ is a compact system. From this,

$$A_N = A_N \cap \Big(\bigcup_{n=1}^N K_n^c\Big) = \bigcup_{n=1}^N A_N \setminus K_n \subseteq \bigcup_{n=1}^N A_n \setminus K_n.$$

Due to the subadditivity and the monotonicity of μ for all $m \geq N$, it follows that

$$\mu(A_m) \leq \mu(A_N) \leq \sum_{n=1}^N \mu(A_n \setminus K_n) \leq \varepsilon \sum_{n=1}^N 2^{-n} \leq \varepsilon.$$

This shows the assertion, since $\varepsilon > 0$ was arbitrary. □

2.3 Uniqueness and extension of set functions

Suppose an additive set function $\mu : \mathcal{H} \to \overline{\mathbb{R}}_+$ is given, where $\mathcal{H}$ is a semi-ring. We are concerned with the extension of μ to a measure (i.e. an σ-additive set function) to $\sigma(\mathcal{H})$. The aim is to establish conditions when the measure is already uniquely given by μ. The result is summarised in Theorem 2.16. See also Table 2 for an overview of how the results of previous chapters relate to this.

Proposition 2.11 (Uniqueness of measures)**.** *Let $\mathcal{F}$ be a σ-algebra and $\mu, \nu : \mathcal{F} \to \mathbb{R}_+$ measures. Let $\mathcal{H}$ be a $\cap$-stable set system such that $\sigma(\mathcal{H}) = \mathcal{F}$ and $\mu|_\mathcal{H}, \nu|_\mathcal{H}$ are σ-finite. Then $\mu = \nu$ if and only if $\mu(A) = \nu(A)$ holds for all $A \in \mathcal{H}$.*

Corollary 2.12 (Uniqueness of probability measures)**.** *Let $\mathcal{F}$ be a σ-algebra and $\mu, \nu : \mathcal{F} \to [0, 1]$ be probability measures. Let $\mathcal{H}$ be a $\cap$-stable set system with $\sigma(\mathcal{H}) = \mathcal{F}$. Then $\mu = \nu$ if and only if $\mu(A) = \nu(A)$ holds for all $A \in \mathcal{H}$.*

	Lemma 2.5	Theorem 2.10	Theorem 2.16
μ additive	$\circ$	$\circ$	
μ finite		$\circ$	
μ σ-finite			$\circ$
μ defined on semi-ring	$\circ$	$\circ$	$\circ$
μ σ-additive	$\circ/\bullet$	$\bullet$	$\circ$
μ σ-subadditive	$\bullet/\circ$		
μ inner regular wrt a compact system		$\circ$	
μ extends uniquely to $\sigma(\mathcal{H})$			$\bullet$

Table 2: Lemma 2.5 and theorem 2.10 play significant roles in the application of Carathéodory's extension theorem. In the table, the $\circ$'s represent the assumptions of the theorem and $\bullet$ the conclusions. As can easily be seen, Carathéodory's extension theorem applies, for example, if μ is finite and inner regular with respect to a compact system.

Proof. Wlog, $\Omega \in \mathcal{H}$, since $\mu(\Omega) = \nu(\Omega) = 1$. This means that μ and ν are in particular σ-finite and the statement follows from Proposition 2.11. $\square$

Proof of Proposition 2.11. The 'only if' direction is clear. For the 'if' direction, we set for $A \in \mathcal{H}$ with $\mu(A) = \nu(A) < \infty$

$$\mathcal{D}_A := \{B \in \mathcal{F} : \mu(A \cap B) = \nu(A \cap B)\} \supseteq \mathcal{H}.$$

We show that $\mathcal{D}_A$ is a Dynkin system. It is clear that $\Omega \in \mathcal{D}_A$. Furthermore, if $B, C \in \mathcal{D}_A$ and $B \subseteq C$, then $\mu((C \setminus B) \cap A) = \mu(C \cap A) - \mu(B \cap A) = \nu(C \cap A) - \nu(B \cap A) = \nu((C \setminus B) \cap A)$, i.e. $C \setminus B \in \mathcal{D}_A$. If $B_1, B_2, \dots \in \mathcal{D}$ with $B_1 \subseteq B_2 \subseteq B_3 \subseteq \dots \in \mathcal{D}_A$ and $B = \bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$, then because of Proposition 2.8,

$$\mu(B \cap A) = \lim_{n\to\infty} \mu(B_n \cap A) = \lim_{n\to\infty} \nu(B_n \cap A) = \nu(B \cap A),$$

which implies $B \in \mathcal{D}_A$. This means that $\mathcal{D}_A$ is a Dynkin system for all $A \in \mathcal{H}$ with $\mu(A) < \infty$ and thus, due to Theorem 1.13, $\mathcal{F} = \sigma(\mathcal{H}) \subseteq \mathcal{D}_A$. Let $\Omega_1, \Omega_2, \dots \in \mathcal{H}$ with $\Omega_n \uparrow \Omega$ and $\mu(\Omega_n), \nu(\Omega_n) < \infty, n = 1, 2, \dots$ Then for all $n = 1, 2, \dots$ it holds that $\mu(B \cap \Omega_n) = \nu(B \cap \Omega_n)$ for all $B \in \mathcal{F}$. This implies $B \in \mathcal{F}$, since μ and ν are continuous from below, thus

$$\mu(B) = \lim_{n\to\infty} \mu(B \cap \Omega_n) = \lim_{n\to\infty} \nu(B \cap \Omega_n) = \nu(B),$$

i.e. $\mu = \nu$. $\square$

The following theorem explains why the notion of a σ-algebra is so important.

Theorem 2.13 (**$\boldsymbol{\mu^*}$-measurable sets are a $\boldsymbol{\sigma}$-algebra**)**.** *Let μ^* be an outer measure on Ω and $\mathcal{F}^*$ the set of μ^*-measurable sets; recall from* (2.3)*. Then $\mathcal{F}^*$ is a σ-algebra and $\mu := \mu^*|_{\mathcal{F}^*}$ is a measure. Furthermore, $\mathcal{N} := \{N \subseteq \Omega : \mu^*(N) = 0\} \subseteq \mathcal{F}^*$.*

Remark 2.14 (Null-sets and properties almost everywhere)**.** *Sets $N \subseteq \Omega$ with $\mu(N) = 0$ are called* (μ-)null sets*. We further say that $A \subseteq \Omega$ (μ)-*almost everywhere *holds if $A^c \in \mathcal{N}$. If μ is a probability measure, we say* almost surely *instead of 'almost everywhere'.*

Proof of Theorem 2.13. We first show that $\mathcal{F}^*$ is a σ-algebra. It is clear that

$$\mu^*(E) = \mu^*(\emptyset) + \mu^*(E) = \mu^*(E \cap \emptyset) + \mu^*(E \cap \Omega),$$

i.e. $\emptyset \in \mathcal{F}^*$. It is also clear that $A^c \in \mathcal{F}^*$ follows from $A \in \mathcal{F}^*$. Next, let us show that $\mathcal{F}^*$ is $\cap$-stable. For $A, B, E \subseteq \Omega$, note that $(E \cap A \cap B^c) \uplus (E \cap A^c) = E \cap (A \cap B)^c$. So, using the sub-additivity of μ^*, for $A, B \in \mathcal{F}^*$,

$$\begin{aligned}\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) &= \mu^*((E \cap A) \cap B) + \mu^*((E \cap A) \cap B^c) + \mu^*(E \cap A^c)\\ &\geq \mu^*(E \cap (A \cap B)) + \mu^*(E \cap (A \cap B)^c) \geq \mu^*(E),\end{aligned}$$

and we have shown $A \cap B \in \mathcal{F}^*$. Now let $A_1, A_2, \dots \in \mathcal{F}^*$ be disjoint and $B_n = \biguplus_{k=1}^n A_k$ and $B = \bigcup_{n=1}^\infty B_n = \biguplus_{k=1}^\infty A_k$. Since $\mathcal{F}^*$ is $\cap$- and complement stable, it is also $\cup$-stable, so $B_1, B_2, ... \in \mathcal{F}^*$. We further show that $\mu^*(E \cap B_n) = \sum_{k=1}^n \mu^*(E \cap A_k)$ applies to all $E \subseteq \Omega$. For $n = 1$ this is clear, and if it applies to n, it follows that

$$\begin{aligned}\mu^*(E \cap B_{n+1}) &= \mu^*(E \cap B_{n+1} \cap B_n) + \mu^*(E \cap B_{n+1} \cap B_n^c)\\ &= \mu^*(E \cap B_n) + \mu^*(E \cap A_{n+1}) = \sum_{k=1}^{n+1} \mu^*(E \cap A_k).\end{aligned}$$

Therefore, since μ^* is sub-additive and monotone,

$$\mu^*(E \cap B) \leq \sum_{k=1}^\infty \mu^*(E \cap A_k) = \lim_{n\to\infty} \sum_{k=1}^n \mu^*(E \cap A_k) = \lim_{n\to\infty} \mu^*(E \cap B_n) \leq \mu^*(E \cap B),$$

thus

$$\mu^*(E \cap B) = \lim_{n\to\infty} \mu^*(E \cap B_n) = \sum_{k=1}^\infty \mu^*(E \cap A_k). \tag{2.5}$$

Next, we show that $B \in \mathcal{F}^*$, which implies that $\mathcal{F}^*$ is a σ-algebra. For any $E \subseteq \Omega$, since $B_1, B_2, ... \in \mathcal{F}^*$, (2.5) holds,

$$\mu^*(E) = \lim_{n\to\infty} \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \mu^*(E \cap B) + \mu^*(E \cap B^c) \geq \mu^*(E).$$

So, $B \in \mathcal{F}^*$ follows. Furthermore, it follows from (2.5) that μ^* is σ-additive, i.e. $\mu = \mu^*|_{\mathcal{F}^*}$ is a measure.

Now let $N \subseteq \Omega$ be such that $\mu^*(N) = 0$ and $E \subseteq \Omega$. Then, due to the monotonicity of μ^*, $\mu^*(E \cap N) = 0$, i.e.

$$\mu^*(E \cap N^c) + \mu^*(E \cap N) \geq \mu^*(E) \geq \mu^*(E \cap N^c) = \mu^*(E \cap N^c) + \mu^*(E \cap N)$$

and therefore $N \in \mathcal{F}^*$. ∎

Proposition 2.15 (Outer measure generated by finite additive set function)**.** *Let $\mathcal{H}$ be a semi-ring and $\mu : \mathcal{H} \to \mathbb{R}_+$ additive. For $A \subseteq \Omega$ let*

$$\mu^*(A) := \inf_{\mathcal{G} \in \mathcal{U}(A)} \sum_{G \in \mathcal{G}} \mu(G),$$

where

$$\mathcal{U}(A) := \big\{\mathcal{G} \subseteq \mathcal{H} \text{ at most countable, } A \subseteq \bigcup_{G \in \mathcal{G}} G\big\}$$

is the set of at most countable covers of A and $\mu^*(A) = \infty$ if $\mathcal{U}(A) = \emptyset$. Then μ^* is an outer measure.

Proof. The mapping μ^* is monotone (by definition) with $\mu^*(\emptyset) = 0$ (note that $\emptyset \in \mathcal{H}$ and, using finite additivity of $\mu(\emptyset) = \mu(\emptyset) + \mu(\emptyset)$, from which $\mu(\emptyset) = 0$ follows). To check the σ-sub-additivity of μ^*, we choose $A_1, A_2, \dots \subseteq \Omega$. For $n = 1, 2, \dots$ and $\epsilon > 0$ there are sets $G_{nk} \in \mathcal{H}$, $k \in \mathcal{K}_n$ at most countable with

$$A_n \subseteq \bigcup_{k \in \mathcal{K}_n} G_{nk},$$
$$\mu^*(A_n) \geq \sum_{k \in \mathcal{K}_n} \mu(G_{nk}) - \varepsilon 2^{-n}.$$

Since $\bigcup_{n=1}^\infty A_n \subseteq \bigcup_{n=1}^\infty \bigcup_{k \in \mathcal{K}_n} G_{nk}$, and by the monotonicity of μ^* and the definition of μ^*,

$$\mu^*\Big(\bigcup_{n=1}^\infty A_n\Big) \leq \sum_{n=1}^\infty \sum_{k \in \mathcal{K}_n} \mu(G_{nk}) \leq \varepsilon + \sum_{n=1}^\infty \mu^*(A_n).$$

With $\varepsilon \to 0$ the σ-sub-additivity of μ^* follows, i.e. μ^* is an outer measure. □

Theorem 2.16 (Extension of a $\boldsymbol{\sigma}$-additive set function)**.** *Let $\mathcal{H}$ be a semi-ring and $\mu : \mathcal{H} \to \mathbb{R}_+$ σ-finite and σ-additive. Furthermore, let $\widetilde{\mu} = \mu^*|_{\mathcal{F}^*}$ with μ^* from Proposition 2.15 and $\mathcal{F}^*$ from Theorem 2.13. Then $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$ and $\widetilde{\mu}|_{\sigma(\mathcal{H})}$ is the only measure that agrees with μ on $\mathcal{H}$.*

Proof. First we note that μ is both finitely additive and σ-subadditive according to Lemma 2.5. According to Proposition 2.15, μ^* is an outer measure and according to Theorem 2.13, $\mathcal{F}^*$ is a σ-algebra.

Step 1: μ^* coincides with μ on $\mathcal{H}$: Let $H \in \mathcal{H}$. Choose $\mathcal{K}$ at most countable and $H_k \in \mathcal{H}$, $k \in \mathcal{K}$ with $H \subseteq \bigcup_{k \in \mathcal{K}} H_k$ and

$$\mu^*(H) \geq \sum_{k \in \mathcal{K}} \mu(H_k) - \varepsilon.$$

Then, because of $H = \bigcup_{k \in \mathcal{K}} H_k \cap H$ and the σ-sub-additivity of μ

$$\mu^*(H) \leq \mu(H) \leq \sum_{k \in \mathcal{K}} \mu(H_k \cap H) \leq \sum_{k \in \mathcal{K}} \mu(H_k) \leq \mu^*(H) + \varepsilon,$$

where we have used the σ-additivity of μ in the second '$\leq$'. With $\varepsilon \to 0$, we find $\mu^*(H) = \mu(H)$.

Step 2: $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$: Let $E \subseteq \Omega, H \in \mathcal{H}$ and $\varepsilon > 0$. Choose $\mathcal{K}$ at most countable and $H_k \in \mathcal{H}, k \in \mathcal{K}$ with $E \subseteq \bigcup_{k\in\mathcal{K}} H_k$ and $\mu^*(E) \geq \sum_{k\in\mathcal{K}} \mu(H_k) - \varepsilon$. Then, due to σ-additivity of μ

$$\mu^*(E) + \varepsilon \geq \sum_{k\in\mathcal{K}} \mu(H_k) = \sum_{k\in\mathcal{K}} \mu(H_k \cap H) + \sum_{k\in\mathcal{K}} \mu(H_k \cap H^c) \geq \mu^*(E \cap H) + \mu^*(E \cap H^c).$$

With $\varepsilon \to 0$ and the σ-sub-additivity of μ^*, $\mu^*(E) = \mu^*(E \cap H) + \mu^*(E \cap H^c)$, i.e. H is μ^*-measurable and therefore $\mathcal{H} \subseteq \mathcal{F}^*$. Since $\mathcal{F}^*$ is a σ-algebra according to Theorem 2.13, $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$.

Step 3: Uniqueness: According to Theorem 2.13, $\widetilde{\mu}$ is a measure. Since $\widetilde{\mu}$ coincides with μ on $\mathcal{H}$, which in turn coincides with μ^* on $\mathcal{H}$, we find $\widetilde{\mu}|_{\sigma(\mathcal{H})} = \mu^*|_{\sigma(\mathcal{H})}$. Let $\nu : \sigma(\mathcal{H}) \to \mathbb{R}_+$ another measure that is equal to μ on $\mathcal{H}$. Since $\mu = \widetilde{\mu}|_{\mathcal{H}}$ was assumed to be σ-finite, $\nu|_{\mathcal{H}}$ is also σ-finite. With Proposition 2.11 it follows that $\widetilde{\mu} = \nu$ applies to $\sigma(\mathcal{H})$ due to the $\cap$-stability of $\mathcal{H}$.

Now all assertions are proven. □

The above theorem only makes it clear that $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$. The next result shows how sets in $\mathcal{F}^*$ differ from sets in $\sigma(\mathcal{H})$.

Proposition 2.17 (Characterisation of $\boldsymbol{\mathcal{F}^*}$ from Proposition 2.15)**.** *Let $\mathcal{H}$ be a semi-ring, $\mu : \mathcal{H} \to \mathbb{R}_+$ σ-finite and σ-additive, μ^* as in Proposition 2.15 and $\mathcal{F}^*, \mathcal{N}$ as in Theorem 2.13. Then*

$$\mathcal{F}^* = \{A \setminus N : A \in \sigma(\mathcal{H}), N \in \mathcal{N}\}.$$

In particular, the right-hand side is a σ-algebra.

Proof. '$\supseteq$': On the one hand we have $\sigma(\mathcal{H}) \subseteq \mathcal{F}^*$ according to theorem 2.16, on the other hand, there is $\mathcal{N} \subseteq \mathcal{F}^*$ from Theorem 2.13. This implies '$\supseteq$', since $\mathcal{F}^*$ is complement stable.

'$\subseteq$': Let $B \in \mathcal{F}^*$. Further, let $\Omega_1, \Omega_2, \dots \in \mathcal{H}$ with $\mu(\Omega_n) < \infty, n = 1, 2, \dots$ and $\Omega = \bigcup_{n=1}^\infty \Omega_n$. Let $\varepsilon_1, \varepsilon_2, \dots > 0$ with $\varepsilon_i \downarrow 0$. For $B_n := B \cap \Omega_n$ and $i = 1, 2, ...$, we choose $\mathcal{K}_{ni}$ at most countable, $A_{nik} \in \mathcal{H}, n \in \mathbb{N}, k \in \mathcal{K}_{ni}$, $B_n \subseteq \bigcup_{k\in\mathcal{K}_{ni}} A_{nik}$ and

$$\mu^*(B_n) \geq \sum_{k\in\mathcal{K}_{ni}} \mu(A_{nik}) - 2^{-n}\varepsilon_i.$$

Clearly, $A_i := \bigcup_{n=1}^\infty \bigcup_{k\in\mathcal{K}_{ni}} A_{nik} \in \sigma(\mathcal{H})$, $B \subseteq A_i$ for all $i = 1, 2, ...$ and $A_i \setminus B = \bigcup_{n=1}^\infty \bigcup_{k\in\mathcal{K}_{ni}} A_{nik} \setminus B_n$. This means that

$$\mu^*(A_i \setminus B) \leq \sum_{n=1}^\infty 2^{-n}\varepsilon_i = \varepsilon_i.$$

Set $A = \bigcap_{i=1}^\infty A_i \in \sigma(\mathcal{H})$. Then $B \subseteq A$, $N := A \setminus B \subseteq A_n \setminus B$ for all $n = 1, 2, ...$ and

$$\mu^*(N) = \mu^*(A \setminus B) \leq \limsup_{i\to\infty} \mu^*(A_i \setminus B) \leq \limsup_{i\to\infty} \varepsilon_i = 0.$$

Thus the assertion follows, since $B = A \setminus N$. □

2.4 Measures on $\mathcal{B}(\mathbb{R})$

From the lecture *Stochastik 1* you already know probability distributions with density. These are measures on $\mathcal{B}(\mathbb{R})$, the Borel σ-algebra on $\mathbb{R}$ (recall from Definition 1.7). We will apply the general theory developed in the last chapters to characterise such measures.

Proposition 2.18 (Lebesgue measure on $\mathbb{R}$)**.** *There is exactly one measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with*

$$\lambda((a, b]) = b - a \tag{2.6}$$

for $a, b \in \mathbb{Q}$ with $a \leq b$.

Proof. Consider $\widetilde{\mathcal{H}} := \{(a, b], [a, b), (a, b), [a, b] : a, b \in \mathbb{Q}, a \leq b\}$, which is a semi-ring with $\sigma(\widetilde{\mathcal{H}}) = \mathcal{B}(\mathbb{R})$. We define $\widetilde{\lambda}$ on $\widetilde{\mathcal{H}}$ by

$$\widetilde{\lambda}((a, b]) = \widetilde{\lambda}([a, b)) = \widetilde{\lambda}((a, b)) = \widetilde{\lambda}([a, b]) = b - a.$$

(Note that $\widetilde{\lambda}$ is the only monotone extension of λ to $\mathcal{H}$.) Then, $\widetilde{\lambda}$ is clearly σ-finite. It is $\mathcal{K} = \{[a, b] : a, b \in \mathbb{Q}, a \leq b\} \subseteq \widetilde{\mathcal{H}}$ a compact system according to Example 1.3. Furthermore, $\widetilde{\lambda}$ is inner $\mathcal{K}$-regular and thus σ-additive according to Theorem 2.10. Hence, Theorem 2.16 gives the only extension of $\widetilde{\lambda}$ to $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R})$. □

Proposition 2.19 (Characterisation of σ-finite measures on $\mathbb{R}$)**.** *A function $\mu : \mathcal{B}(\mathbb{R}) \to \overline{\mathbb{R}}_+$ is a σ-finite measure if and only if there is a non-decreasing and right-continuous function $G : \mathbb{R} \to \mathbb{R}$ with*

$$\mu((a, b]) = G(b) - G(a) \tag{2.7}$$

for $a, b \in \mathbb{Q}$ with $a \leq b$. If $\widetilde{G} : \mathbb{R} \to \mathbb{R}$ is another right-continuous function satisfying (2.7), then $\widetilde{G} = G + c$ for some $c \in \mathbb{R}$.

Corollary 2.20 (Characterisation of probability measures on $\mathbb{R}$)**.** *A function $\mu : \mathcal{B}(\mathbb{R}) \to [0, 1]$ is a probability measure if and only if there is a non-decreasing and right-continuous function $F : \mathbb{R} \to [0, 1]$ with $\lim_{a \to -\infty} F(a) = 0$, $\lim_{b \to \infty} F(b) = 1$ and*

$$\mu((a, b]) = F(b) - F(a) \tag{2.8}$$

for $a, b \in \mathbb{Q}$ with $a \leq b$. In this case, F is uniquely defined by μ.

Proof. The assertion follows directly from Proposition 2.19, since the limit condition uniquely defines c. □

Proof of Proposition 2.19. '$\Rightarrow$': Let μ be a σ-finite measure on $\mathcal{B}(\mathbb{R})$. Define $G(0) := 0$, and $G(x) := \mu((0, x])$ for $x > 0$ and $G(x) := -\mu((x, 0])$ for $x < 0$. Then G is right-continuous, non-decreasing, and (for example for $0 < a < b$) $\mu((a, b]) = \mu((0, b]) - \mu((0, a]) = G(b) - G(a)$.

'$\Leftarrow$': The proof is similar to the proof of Proposition 2.18. Let $\mathcal{H} = \{(a, b] : a, b \in \mathbb{Q}, a \leq b\}$ be the semi-ring of half-open intervals with ends in rational numbers. We show that (2.7) defines a σ-additive set function μ on $\mathcal{H}$. Then, using Theorem 2.16, we see that μ can be uniquely extended to a σ-finite measure on $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R})$. Let $a_1, a_2, \ldots$ be such that

$\bigcup_{n=1}^{\infty}(a_{n+1}, a_n] = (a, b] \in \mathcal{H}$. Without loss of generality, $a_1 \geq a_2 \geq \dots$ Then $b = a_1$ and $a_n \xrightarrow{n\to\infty} a$. Due to the right continuity of G,

$$\mu(a,b] = G(b) - G(a) = G(a_1) - \lim_{N\to\infty} G(a_N) = \sum_{n=1}^{\infty} G(a_n) - G(a_{n+1}) = \sum_{n=1}^{\infty} \mu((a_{n+1}, a_n]),$$

and we have shown the σ-additivity of μ.

Now, let $\widetilde{G}$ be another function for which (2.7) applies. Then for all $a \in \mathbb{R}$,

$$\widetilde{G}(b) = \widetilde{G}(a) + \mu((a,b]) = G(b) + \widetilde{G}(a) - G(a),$$

and the assertion follows with $c = \widetilde{G}(a) - G(a)$. □

Definition 2.21 (Lebesgue measure and distribution functions)**.** *1. The uniquely defined measure λ from Corollary 2.18 is called* one-dimensional Lebesgue measure*.*

2. For a probability measure μ on $\mathcal{B}(\mathbb{R})$, the function F from Corollary 2.20 is called distribution function*.*

Example 2.22 (Some distribution functions)**.** *Let $f : \mathbb{R} \to \mathbb{R}_+$ be piecewise continuous, and*[4] *$\int_{-\infty}^{\infty} f(x)dx = 1$. As known from the lecture* Stochastik 1*, such a function is called a* density*. On the one hand, such density functions define a distribution function by means of*

$$F(x) := \int_{-\infty}^{x} f(a)da.$$

On the other hand, each of these distribution functions defines a probability measure in a unique way due to Corollary 2.20. We will look at distributions with densities in more detail in the Radon-Nikodým theorem (see section 4.4).

As already known,

$$F_{U(0,1)}(x) = \int_{-\infty}^{x} 1_{[0,1]}(a)da = \begin{cases} 0, & x \leq 0, \\ x, & 0 < x \leq 1, \\ 1, & x > 1 \end{cases} \tag{2.9}$$

is the distribution function of the uniform distribution on $[0,1]$. Further, for $x \geq 0$

$$F_{exp(\lambda)}(x) = \int_{-\infty}^{x} 1_{[0,\infty)}(a)\lambda e^{-\lambda a}da = 1 - e^{-\lambda x} \tag{2.10}$$

is the distribution function of the exponential distribution with parameter λ. Furthermore,

$$F_{N(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\Big(-\frac{(a-\mu)^2}{2\sigma^2}\Big)da =: \Phi(x) \tag{2.11}$$

is the distribution function of the normal distribution $N(\mu, \sigma^2)$ with the expected value μ and the variance σ^2.

[4] We assume here that the Riemann integral $\int_a^b f(x)dx$ is known. (See also definition 3.22.) We will get to know another integral term, the Lebesgue integral, in Chapter 3.

2.5 Image measures

Let μ be a measure on some σ-algebra $\mathcal{F}$. If we transform the base space by means of a function $f : \Omega \to \Omega'$, you can define a measure corresponding to the transformation on Ω', the so-called image measure. Let $\Omega := [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$ and $f : u \mapsto -\log u$. We will then see that the image measure of $\mu_{U(0,1)}$ under f is $\mu_{\exp(1)}$. We first recall the situation from example 1.3.2 and define the image measure.

Definition 2.23 (Image measure)**.** *If $(\Omega, \mathcal{F}, \mu)$ is a measure space, $(\Omega', \mathcal{F}')$ is a measurable space and $f : \Omega \to \Omega'$ such that $\sigma(f) \subseteq \mathcal{F}$ for $\sigma(f)$ from* (1.1)*. Then we define a set function $f_*\mu : \mathcal{F}' \to \mathbb{R}_+$ by*

$$f_*\mu(A') := \mu(f^{-1}(A')) = \mu(f \in A'), \qquad A' \in \mathcal{F}'.$$

Here $f_*\mu$ is also called image measure *of μ under f.*

If μ is a probability measure, then $f_*\mu$ is also called distribution *of f (under μ).*

Remark 2.24 (Measurable functions)**.** *If $\sigma(f) \subseteq \mathcal{F}$ as in the definition above, we say that f is measurable (with respect to $\mathcal{F}/\mathcal{F}'$) . This concept will be discussed further in the next section.*

Proposition 2.25 (Image measure is a measure)**.** *Let $(\Omega, \mathcal{F}, \mu)$, $(\Omega', \mathcal{F}')$, $f : \Omega \to \Omega'$ and $f_*\mu$ as in Definition 2.23. Then, $f_*\mu$ is a measure on $\mathcal{F}'$.*

Proof. If $A'_1, A'_2, \dots \in \mathcal{F}'$ are disjoint, then

$$f_*\mu\Big(\biguplus_{n=1}^\infty A'_n\Big) = \mu\Big(f^{-1}\Big(\biguplus_{n=1}^\infty A'_n\Big)\Big) = \mu\Big(\biguplus_{n=1}^\infty (f^{-1}(A'_n)\Big) = \sum_{n=1}^\infty \mu(f^{-1}(A'_n)) = \sum_{n=1}^\infty f_*\mu(A'_n).$$

This means that $f_*\mu$ is σ-additive and the assertion is shown. ☐

Example 2.26 (Some transformations)**.**

1. *For $\Omega = [0,1]$, $\{[0,b) : 0 \le b \le 1\}$ is a $\cap$-stable generating system of $\mathcal{B}([0,1])$. Let $\mu = \mu_{U(0,1)}$ be the uniform distribution on $[0,1]$ with distribution function $F_{U(0,1)}$ from* (2.9) *and $f : u \mapsto 1-u$. Then $f_*\mu = \mu$, because*

$$f_*\mu([0,b)) = \mu(f^{-1}([0,b))) = \mu([1-b,1]) = F_{U(0,1)}(1) - F_{U(0,1)}(1-b) = b.$$

Thus, μ and $f_*\mu$ agree on a $\cap$-stable generator and the statement follows with Proposition 2.11.

2. *Let $\Omega = \mathbb{R}$, $f_y : x \mapsto x+y$ for a $y \in \mathbb{R}$ and λ the Lebesgue measure from Corollary 2.18. Then $(f_y)_*\lambda = \lambda$, because*

$$(f_y)_*\lambda([a,b]) = \lambda(f_y^{-1}([a,b])) = \lambda([a-y,b-y]) = b-a.$$

We say that the Lebesgue measure is translation invariant.

3. *Let $\Omega = [0,1], \Omega' = \mathbb{R}_+$, each equipped with Borel's σ-algebra. Further, let $\mu = \mu_{U(0,1)}$ with distribution function $F_{U(0,1)}$ from* (2.9) *and $f : x \mapsto -\frac{1}{\lambda}\log(x)$ for a $\lambda > 0$. Then $f_*\mu = \mu_{\exp(\lambda)}$, where $\mu_{\exp(\lambda)}$ has the distribution function $F_{\exp(\lambda)}$ from* (2.10)*. This is because for $x \ge 0$*

$$f_*\mu([0,x]) = \mu(f^{-1}([0,x])) = \mu([e^{-\lambda x},1]) = 1 - e^{-\lambda x}.$$

Example 2.27 (Example of a non Borel-measurable set (Vitali's set))**.**
So far, there has not yet been an example of a set that is not in $\mathcal{B}(\mathbb{R})$*. We will now construct such a set. It is known as* Vitali's set*. For this purpose, we define an equivalence relation on* $\mathbb{R}$ *by* $x \sim y \iff y - x \in \mathbb{Q}$*. With respect to this equivalence relation,* $\mathbb{R}$ *decomposes into equivalence classes of the form* $\{x + q : q \in \mathbb{Q}\}$*. We select a number from* $[0, 1]$ *from each equivalence class, and put all such numbers into the set* V*. (It should be noted here that this selection is made using the axiom of choice and is therefore not a trivial step). Further, now for* $q \in \mathbb{Q} \cap [-1, 1]$

$$V_q := \{x + q : x \in V\}.$$

Then $[0, 1] \subseteq \biguplus_{q\in\mathbb{Q}\cap[-1,1]} V_q \subseteq [-1, 2]$.

Assume that the quantity V *is measurable. Then the quantities* V_q *would also be measurable and, due to the translation invariance of the Lebesgue measure from Example 2.26.2,* $\lambda(V_q)$ *would not depend on* q*. So let* $\lambda(V_q) =: a \geq 0$*. Furthermore, due to the monotonicity of the Lebesgue measure*

$$1 \leq \sum_{q\in\mathbb{Q}\cap[-1,1]} \lambda(V_q) = \sum_{q\in\mathbb{Q}\cap[-1,1]} a \leq 3.$$

However, this is not possible, neither for $a = 0$ *nor for* $a > 0$*. Because of this contradiction,* $V \notin \mathcal{B}(\mathbb{R})$ *must therefore apply.*

3 Measurable functions and the integral

In this chapter, let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We can now use the measure μ to measure the content of sets of $\mathcal{F}$. The aim of introducing the integral is to weight the elements of Ω differently in such a measurement. This weighting is carried out with a function $f : \Omega \to \mathbb{R}$. Such functions must fulfil the minimal requirement of measurability. The result of this weighting leads to the concept of the integral.

3.1 Measurable functions

We already know what a measurable set (with respect to the σ-algebra $\mathcal{F}$) is, i.e. $A \subseteq \Omega$ is $(\mathcal{F}-)$measurable iff $A \in \mathcal{F}$. We will extend this notion to functions $f : \Omega \to \Omega'$ (for some measurable space $(\Omega', \mathcal{F}')$. Note that for $A \in \mathcal{F}$, there is the indicator function $\omega \mapsto 1_A(\omega)$, which is the simplest form of a measurable function in Definition 3.3. We will call the linear combination of such indicator functions a simple function, which will be measurable as well. These are of particular importance due to Theorem 3.9, which shows that every non-negative measurable function – see below – can be approximated from below (in the sense of pointwise convergence) by simple functions.

Remark 3.1 (Notation)**.** *Let Ω, Ω' be sets, $f : \Omega \to \Omega'$ and I be arbitrary.*

1. *We write $f(A) := \{f(\omega) : \omega \in A\}$ for $A \subseteq \Omega$ and $f^{-1}(A') := \{f^{-1}(\omega') : \omega' \in A'\}$ for $A' \subseteq \Omega'$. We note that the following rules apply to $A', A'_i \subseteq \Omega', i \in I$:*

$$f^{-1}((A')^c) = (f^{-1}(A'))^c, \quad f^{-1}\Big(\bigcap_{i\in I} A'_i\Big) = \bigcap_{i\in I} f^{-1}(A'_i), \quad f^{-1}\Big(\bigcup_{i\in I} A'_i\Big) = \bigcup_{i\in I} f^{-1}(A'_i).$$

   However, some caution is required, since for $A, A_i \subseteq \Omega, i \in I$ only $f(\bigcup_{i\in I} A_n) = \bigcup_{i\in I} f(A_i)$, in general, however, $f(A^c) \neq (f(A))^c$ and $f(\bigcap_{i\in I} A_i) \neq \bigcap_{i\in I} f(A_i)$.

2. *For $\mathcal{C} \subseteq 2^{\Omega'}$ we write analogously*

$$f^{-1}(\mathcal{C}) := \{f^{-1}(A') : A' \in \mathcal{C}\}.$$

Lemma 3.2 (Pre-image of σ-algebras)**.** *Let Ω be a set and $(\Omega', \mathcal{F}')$ a measurable space, $f : \Omega \to \Omega'$ and $\mathcal{C}' \subseteq \mathcal{F}'$ with $\sigma(\mathcal{C}') = \mathcal{F}'$. Then $\sigma(f^{-1}(\mathcal{C}')) = f^{-1}(\sigma(\mathcal{C}'))$. In particular, $f^{-1}(\mathcal{F}')$ is a σ-algebra on Ω.*

Proof. '$\subseteq$': From Remark 3.1, it is clear that $f^{-1}(\sigma(\mathcal{C}'))$ is a σ-algebra. This means that $\sigma(f^{-1}(\mathcal{C}')) \subseteq \sigma(f^{-1}(\sigma(\mathcal{C}'))) = f^{-1}(\sigma(\mathcal{C}'))$.
'$\supseteq$': We define

$$\widetilde{\mathcal{F}}' = \{A' \in \sigma(\mathcal{C}') : f^{-1}(A') \in \sigma(f^{-1}(\mathcal{C}'))\} \subseteq \sigma(\mathcal{C}').$$

Then, again due to Remark 3.1, $\widetilde{\mathcal{F}}'$ is a σ-algebra and $\mathcal{C}' \subseteq \widetilde{\mathcal{F}}' \subseteq \sigma(\mathcal{C}')$. Thus, $\widetilde{\mathcal{F}}' = \sigma(\mathcal{C}')$. For $A' \in \sigma(\mathcal{C}')$, we find $f^{-1}(A') \in \sigma(f^{-1}(\mathcal{C}'))$, which is equivalent to $f^{-1}(\sigma(\mathcal{C}')) \subseteq \sigma(f^{-1}(\mathcal{C}'))$. □

Definition 3.3 (Measurable functions)**.** *Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be measurable spaces and $f : \Omega \to \Omega'$.*

1. *The function f is called* $\mathcal{F}/\mathcal{F}'$-measurable *if $f^{-1}(\mathcal{F}') \subseteq \mathcal{F}$. The σ-algebra $f^{-1}(\mathcal{F}')$ (recall from Lemma 3.2 that this is in fact a σ-algebra) is called the* σ-algebra (on Ω) generated by f *and is denoted $\sigma(f)$.*

2. *If* $(\Omega, \mathcal{F}, \mathbf{P})$ *is a probability space and* $X : \Omega \to \Omega'$ *measurable, then* X *is called an* Ω'-valued random variable. *The image measure* $X_*\mathbf{P}$ *from Definition 2.23 is called the* distribution of X.

3. *If* $(\Omega', \mathcal{F}') = (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$*, then* f *is called a* real-valued function. *If* f *is measurable according to* $\mathcal{F}/\mathcal{B}(\overline{\mathbb{R}})$*, we say that the function* f *is (Borel-)measurable.*

4. *If* $\Omega' = \overline{\mathbb{R}}$ *and* $f = 1_A$ *for* $A \subseteq \Omega$*, then* f *is called* indicator function. *If* $f = \sum_{k=1}^n c_k 1_{A_k}$ *for* $c_1, \dots, c_n \in \overline{\mathbb{R}}$ *pairwise different and* $A_1, \dots, A_n \subseteq \Omega$*, then* f *is called* simple.

Example 3.4. *Let* $(\Omega, \mathcal{F})$ *be a measurable space.*

1. *The vast majority of functions* $f : \Omega \to \overline{\mathbb{R}}$ *that one can imagine are (Borel-)measurable. For example, the identity* $f : \omega \mapsto \omega$ *is measurable, since* $f^{-1}(\mathcal{F}) = \mathcal{F}$.

2. *Let* $(\Omega, \mathcal{O})$ *and* $(\Omega'.\mathcal{O}')$ *be topological spaces and* $f : \Omega \to \Omega'$ *continuous. Then* f *is* $\mathcal{B}(\Omega)/\mathcal{B}(\Omega')$*-measurable.*
   Indeed, by continuity we have that $f^{-1}(\mathcal{O}') \subseteq \mathcal{O}$*. Therefore, using Lemma 3.2,*
   $$f^{-1}(\mathcal{B}(\Omega')) = f^{-1}(\sigma(\mathcal{O}')) = \sigma(f^{-1}(\mathcal{O}') \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\Omega).$$

3. *It is important to see that for many measurable functions* f *it is true that* $\sigma(f) \subsetneq \mathcal{F}$*, see for example the next example.*

4. *A function* $f : \Omega \to \{0, 1\}$ *is measurable if and only if* $f^{-1}(\{1\}) \in \mathcal{F}$*. In this case,* $\sigma(f) = \{\emptyset, f^{-1}(\{1\}), (f^{-1}(\{1\}))^c, \Omega\}$.

5. *Let* $\mathcal{F} = \mathcal{B}(\mathbb{R})$*. To specify a non* $\mathcal{F}$*-measurable function, you have to make the same effort as to construct a non Borel-measurable set. For example, the function* 1_V *is not measurable for the Vitali set* V *from Example 2.27.*

Example 3.5 (Random variables)**.** *Let* (E, r) *be some metric space (equipped with the Borel* σ*-algebra).*

1. *Let* X *be an* E*-valued random variable on a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*, and* Y *an* E*-valued random variable on* $(\Omega', \mathcal{A}, \mathbf{Q})$*. If* $X_*\mathbf{P} = Y_*\mathbf{Q}$*, we say that* X *and* Y *are* identically distributed *and write* $X \sim Y$*. Note, however, since* X *and* Y *need not be defined on the same probability space, it does not make sense to write something like* $X - Y$*. If* μ *is a measure on* $\mathcal{B}(E)$ *and* $X_*\mathbf{P} = \mu$*, we write* $X \sim \mu$.

2. *Let* $(X_i)_{i \in I}$ *be a family of random variables on a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$*. Then, the distribution of* $((X_i)_{i \in I})_*\mathbf{P}$ *is called the* joint distribution of $(X_i)_{i \in I}$.

Lemma 3.6 (Properties of measurability)**.** *Let* $(\Omega, \mathcal{F})$*,* $(\Omega', \mathcal{F}')$ *and* $(\Omega'', \mathcal{F}'')$ *be measurable spaces.*

1. *If* $\mathcal{C}' \subseteq \mathcal{F}'$ *with* $\mathcal{F}' = \sigma(\mathcal{C}')$*, then* $f : \Omega \to \Omega'$ *is* $\mathcal{F}/\mathcal{F}'$*-measurable if and only if* $f^{-1}(\mathcal{C}') \subseteq \mathcal{F}$.

2. *If* $f : \Omega \to \Omega'$ *is* $\mathcal{F}/\mathcal{F}'$*-measurable and* $g : \Omega' \to \Omega''$ *is* $\mathcal{F}'/\mathcal{F}''$*-measurable, then* $g \circ f : \Omega \to \Omega''$ *is* $\mathcal{F}/\mathcal{F}''$*-measuarble.*

3. *Let $(\Omega, \mathcal{O})$ and $(\Omega', \mathcal{O}')$ be topological spaces, $f : \Omega \to \Omega'$ continuous and $\mathcal{F} = \sigma(\mathcal{O})$ and $\mathcal{F}' = \sigma(\mathcal{O}')$ the Borel σ-algebras. Then f is $\mathcal{F}/\mathcal{F}'$-measurable.*

4. *A real-valued function f (i.e. $f : \Omega \to \mathbb{R}$) is measurable (with respect to $\mathcal{F}/\mathcal{B}(\mathbb{R})$) if and only if $\{\omega : f(\omega) \le x\} \in \mathcal{F}$ for all $x \in \mathbb{Q}$.*

5. *A simple function $f = \sum_{k=1}^n c_k 1_{A_k}$ with pairwise different $c_1, \dots, c_n \in \overline{\mathbb{R}}$ and $A_1, \dots, A_n \subseteq \Omega$ is measurable if and only if $A_1, \dots, A_n \in \mathcal{F}$.*

Proof. 1. the 'only if' direction is clear. For the 'if' direction, we use Lemma 3.2 and obtain $f^{-1}(\mathcal{F}') = f^{-1}(\sigma(\mathcal{C}')) = \sigma(f^{-1}(\mathcal{C}')) \subseteq \sigma(\mathcal{F}) = \mathcal{F}$. This means that f is $\mathcal{F}/\mathcal{F}'$-measurable.
2. We write directly $(g \circ f)^{-1}(\mathcal{F}'') = f^{-1}(g^{-1}(\mathcal{F}'')) \subseteq f^{-1}(\mathcal{F}') \subseteq \mathcal{F}$, which already shows the assertion.
3. By definition of Borel's σ-algebra, $\mathcal{O}'$ is a generator for $\mathcal{B}(\Omega')$. Since f is continuous (i.e. $f^{-1}(\mathcal{O}') \subseteq \mathcal{O}$), $f^{-1}(\mathcal{O}') \subseteq \mathcal{O} \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\Omega)$ follows. According to 1. f is therefore $\mathcal{B}(\Omega)/\mathcal{B}(\Omega')$-measurable.
4. We use 1 with $\mathcal{C} = \{(-\infty, x] : x \in \mathbb{Q}\}$: If $\Omega' = \mathbb{R}$, then according to Lemma 3.2, $\mathcal{B}(\Omega')$ is generated by $\mathcal{C}$. Therefore, a f is measurable if $f^{-1}(\mathcal{C}') = \{\{\omega : f(\omega) \le x\} : x \in \mathbb{R}\} \subseteq \mathcal{F}$.
5. Let $A := \Big(\bigcup_{k=1} A_k\Big)^c \in \mathcal{F}$. Then $f^{-1}(\mathcal{B}(\overline{\mathbb{R}})) = \Big\{A \cup \bigcup_{j \in J} A_j, \bigcup_{j \in J} A_j : J \subseteq \{1, \dots, n\}\Big\}$, from which the assertion follows. □

Lemma 3.7 (Algebraic structure of measurability)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space.*

1. *Let $f, g : \Omega \to \overline{\mathbb{R}}$ be measurable. Then fg, as well as $af + bg$ for $a, b \in \mathbb{R}$ are measurable. In addition, f/g is measurable if $g(\omega) \neq 0$ for all $\omega \in \Omega$.*

2. *Let $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ be measurable. Then,*

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

are measurable as well. If it exists, $\lim_{n \to \infty} f_n$ is also measurable.

Proof. 1. Consider $\psi : \Omega \to \mathbb{R}^2$, defined by $\psi(\omega) = (f(\omega), g(\omega))$. It is easy to see that ψ is $\mathcal{F}/\mathcal{B}(\mathbb{R}^2)$-measurable. Furthermore, $(x, y) \mapsto ax + by$ and $(x, y) \mapsto xy$ are continuous on $\mathbb{R}$ and $(x, y) \mapsto x/y$ on $\mathbb{R} \times (\mathbb{R} \setminus \{0\})$, i.e. measurable according to Lemma 3.6.3. Thus the assertions according to Lemma 3.6.2 follow.
2. We only show the measurability of $\sup_{n \in \mathbb{N}} f_n$. The other statements then follow using

$$\inf_{n \in \mathbb{N}} f_n = -\sup_{n \in \mathbb{N}}(-f_n), \quad \limsup_{n \to \infty} f_n = \inf_{n \in \mathbb{N}} \sup_{k \ge n} f_k, \quad \liminf_{n \to \infty} f_n = \sup_{n \in \mathbb{N}} \inf_{k \ge n} f_k.$$

We write, for $x \in \mathbb{R}$, according to Lemma 3.6.4

$$\Big\{\omega : \sup_{n \in \mathbb{N}} f_n(\omega) \le x\Big\} = \bigcap_{n=1}^{\infty} \underbrace{\Big\{\omega : f_n(\omega) \le x\Big\}}_{\in \mathcal{F}} \in \mathcal{F}$$

and the assertion is shown. □

Corollary 3.8 (Measurability of positive and negative part)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \overline{\mathbb{R}}$. Then f is measurable if and only if $f^+ := f \vee 0$ and $f^- := (-f) \vee 0$ are measurable. Then $|f|$ is also measurable.*

Proof. Note that $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Thus the assertion follows from Lemma 3.7.2. □

Theorem 3.9 (Approximation with measurable functions)**.** *Let $(\Omega, \mathcal{F})$ be a measurable space and $f : \Omega \to \overline{\mathbb{R}}_+$ measurable. Then there is a sequence $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ of simple functions with*[5] *$f_n \uparrow f$.*

Proof. We write for[6] $\omega \in \Omega, n \in \mathbb{N}$

$$f_n(\omega) = n \wedge 2^{-n}[2^n f(\omega)],$$

and note that $f_n \uparrow f$ holds by construction. Furthermore, $\omega \mapsto [2^n f(\omega)]$ is measurable according to Lemma 3.6.4 if f. □

3.2 Definition

The construction of the integral of a function f according to a measure will take place in several steps. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space. For the integral of $f : \Omega \to \mathbb{R}$ with respect to μ we use different synonymous notations, namely

$$\mu[f] = \int f d\mu = \int f(\omega)\mu(d\omega). \tag{3.1}$$

The integral is first defined for simple functions $f$and then (see Theorem 3.9) by approximation for general non-negative measurable functions. The integral for (not necessarily non-negative) measurable functions is then defined by the integral of the positive and negative parts; see Definition 3.17.

The application in probability theory is as follows: Recall the notion of a *probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ from Definition 2.1. Here, any measurable $X : \Omega \to \mathbb{R}$ is called a *random variable* (recall from Definition def:measurable). Then, we use the notation

$$\mathbf{E}[X] := \mathbf{P}[X],$$

where $\mathbf{P}[X]$ is defined as in (3.1) and denote this by the *expectation of X (with respect to $\mathbf{P}$).*

Definition 3.10 (Integral of simple functions)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f = \sum_{k=1}^m c_k 1_{A_k}$ a simple function with $c_1, \dots, c_m \geq 0, A_1, \dots, A_m \in \mathcal{F}$. Then,*

$$\mu[f] := \int f d\mu := \sum_{k=1}^n c_k \mu(A_m)$$

is the integral of f with respect to μ.

Remark 3.11 (Integral is well-defined)**.** *We must make sure that the above integral is well-defined. Let $f = \sum_{l=1}^n d_l 1_{B_l}$ be another representation of f with $d_1, \dots, d_n \geq 0$ and $B_1, \dots, B_n \in \mathcal{F}$. Then,*

$$\sum_{k=1}^m c_k\mu(A_k) = \sum_{k=1}^m \sum_{l=1}^n c_k\mu(A_k \cap B_l) = \sum_{k=1}^m \sum_{l=1}^n d_l\mu(A_k \cap B_l) = \sum_{l=1}^n d_l\mu(B_l),$$

so the integral of simple functions is well-defined.

[5] Analogously to '↓', we write for $x, x_1, x_2, \dots \in \mathbb{R}$ that $x_n \uparrow x$ if $x_1 \leq x_2 \leq \dots$ and $x_n \xrightarrow{n\to\infty} x$. For functions $f, f_1, f_2, \dots : \Omega \to \mathbb{R}$, $f_n \uparrow f$ means that $f_n(\omega) \uparrow f(\omega)$ for all $\omega \in \Omega$.

[6] here $[x]$ for $x \in \mathbb{R}$ is the largest integer smaller than x, so $[x] := \sup\{n \in \mathbb{Z} : n \leq x\}$.

Lemma 3.12 (Simple properties)**.** *Let f, g be non-negative, simple functions and $\alpha \geq 0$. Then*[7]

$$\mu[af + bg] = a\mu[f] + b\mu[g], \qquad f \leq g \Rightarrow \mu[f] \leq \mu[g].$$

Proof. Clear. □

Example 3.13 (The integral of indicator functions and the Riemann integral)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $A \in \mathcal{F}$. Then $f = 1_A$ is a simple function and the following applies*

$$\mu[f] = \mu(A)$$

according to definition 3.10. It should be noted that the function $f = 1_A$ no longer has to be piecewise continuous. (Let $A = \mathbb{Q}$ or A be the Cantor continuum considered in Example 1.10). Therefore, it is not clear that the function 1_A is integrable in the sense of Riemann.

Definition 3.14 (The integral of measurable, non-negative functions)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measue space and $f : \Omega \to \overline{\mathbb{R}}_+$ measurable. The integral of f with respect to μ is given by*

$$\mu[f] := \int f(\omega)\mu(d\omega) := \int f d\mu := \sup\{\mu[g] : g \text{ simple, non-negative, } g \leq f\}. \tag{3.2}$$

Remark 3.15 (The integral as an extension)**.** *From Lemma 3.12 it is clear that the definition of $\mu[f]$ for simple, non-negative functions f from Definition 3.10 and Definition 3.14 is the same. The above definition is therefore an extension of $\mu[f]$ to the space of non-negative, measurable functions.*

It is also important to note that, according to Theorem 3.9, each of the functions occurring in Definition 3.14 can be approximated (pointwise) by simple functions. In particular, the supremum is in (3.2) *is over simple functions g which are arbitrarily close to f.*

Proposition 3.16 (Properties of the integral)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g, f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}_+$ measurable. Then the following applies:*

1. *If $f \leq g$, then $\mu[f] \leq \mu[g]$.*
2. *If*
$$f_n \uparrow f, \quad \text{then} \quad \mu[f_n] \uparrow \mu[f].$$
We say that the integral obeys monotone convergence.
3. *If $a, b \geq 0$, then $\mu[af + bg] = a\mu[f] + b\mu[g]$.*

Proof. 1. is clear from the definition of the integral. 2. From 1., it is clear that $\mu[f_1], \mu[f_2], ...$ is increasing. In particular, $\lim_{n\to\infty} \mu[f_n]$ exists. We need to show $\lim_{n\to\infty} \mu[f_n] \leq \mu[f]$ as well as $\mu[f] \leq \lim_{n\to\infty} \mu[f_n]$. First, since $f_1, f_2, ... \leq f$,

$$\lim_{n\to\infty} \mu[f_n] = \sup_{n\in\mathbb{N}} \mu[f_n] \leq \mu[f].$$

Second, it is sufficient to show that

$$\mu[g] \leq \sup_{n\in\mathbb{N}} \mu[f_n] \tag{3.3}$$

[7] For $f, g : \Omega \to \overline{\mathbb{R}}$, we write $f \leq g$ if $f(\omega) \leq g(\omega)$ holds for all $\omega \in \Omega$.

for all simple functions $g \leq f$. Let $g = \sum_{k=1}^m c_k 1_{A_k} \leq f$ for disjoint sets $A_1, \dots, A_m$ and $c_1, \dots, c_m > 0$. For $\varepsilon > 0$ and $n = 1, 2, \dots$ let $B_n^\varepsilon := \{f_n \geq (1-\varepsilon)g\}$. Since $f_n \uparrow f$ and $g \leq f$, $\bigcup_{n=1}^\infty B_n^\varepsilon = \Omega$ for all $\varepsilon > 0$. Therefore

$$\begin{aligned}\mu[f_n] \geq \mu[(1-\varepsilon)g1_{B_n^\varepsilon}] &= \sum_{k=1}^m (1-\varepsilon)c_k\mu(A_k \cap B_n^\varepsilon)\\ &\xrightarrow{n\to\infty} \sum_{k=1}^m (1-\varepsilon)c_k\mu(A_k) = (1-\varepsilon)\mu[g].\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, (3.3) follows.
For 3., let $f_1, g_1, f_2, g_2, \dots$ be simple functions with $f_n \uparrow f$ and $g_n \uparrow g$. Then, $af_n + bg_n \uparrow af + bg$ and it follows

$$\mu[af + bg] = \lim_{n\to\infty} \mu[af_n + bg_n] = \lim_{n\to\infty} a\mu[f_n] + b\mu[g_n] = a\mu[f] + b\mu[g]$$

from 2. because of Lemma 3.12. □

We can now define the integral for measurable functions. First, we note that $f^+, f^- \leq |f|$ for any $f : \Omega \to \overline{\mathbb{R}}$. In particular, if f is measurable with $\mu[|f|] < \infty$, then $\mu[f^+], \mu[f^-] < \infty$.

Definition 3.17 (Integral of measurable functions)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f : \Omega \to \overline{\mathbb{R}}$ measurable. Then f is said to be μ-integrable if $\mu[|f|] < \infty$ and we define*

$$\mu[f] := \int f(\omega)\mu(d\omega) := \int f d\mu := \mu[f^+] - \mu[f^-]. \tag{3.4}$$

We also set

$$\mathcal{L}^1(\mu) := \Big\{f : \Omega \to \overline{\mathbb{R}} : \mu[|f|] < \infty\Big\}$$

For $A \in \mathcal{F}$ we also write

$$\mu[f, A] := \int_A f d\mu := \mu[f1_A].$$

Remark 3.18 (Extension of the integral and $\mathcal{L}^p$-spaces)**.**

1. *If at most one of the two terms $\mu[f^+]$ or $\mu[f^-]$ is infinite, we continue to define the integral $\mu[f]$ using* (3.4)*. In other cases, the integral remains undefined.*
2. *The function spaces $\mathcal{L}^p(\mu) := \Big\{f : \Omega \to \overline{\mathbb{R}} : \mu[|f|^p] < \infty\Big\}$, $p > 0$, will play a special role in Section 4.*

3.3 Properties of the integral

We first establish some properties of the integral. These are, for example, monotonicity and linearity. We will also see that the integral of a function does not change if it is modified on a null-set; see Proposition 3.21.

Proposition 3.19 (Simple properties of the integral)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f, g \in \mathcal{L}^1(\mu)$. Then, the following holds:*

1. *The integral is monotone, i.e.*

$$f \leq g \text{ almost everywhere} \qquad \Longrightarrow \qquad \mu[f] \leq \mu[g].$$

2. *As a special case of 1., since* $-f, f \leq |f|$,

$$|\mu[f]| \leq \mu[|f|].$$

3. *The integral is linear, so if* $a, b \in \mathbb{R}$, *then* $af + bg \in \mathcal{L}^1(\mu)$ *and*

$$\mu[af + bg] = a\mu[f] + b\mu[g].$$

Proof. All properties follow from Proposition 3.16.1 and 3, and the definition of the integral for measurable functions. □

Proposition 3.20 (Substitution theorem)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space,* $(\Omega', \mathcal{F}')$ *a measurable space,* $f : \Omega \to \Omega'$ *measurable and* $f_*\mu$ *the image measure of* f *from Definition 2.23. Then for* $g \in \mathcal{L}^1(f_*\mu)$ *it is true that* $g \circ f \in \mathcal{L}^1(\mu)$ *and*

$$\mu[g \circ f] = f_*\mu[g].$$

Proof. It is sufficient to show the assertion for simple, non-negative functions g. The general case then follows by means of approximation by simple functions. Let $g = \sum_{k=1}^m c_k 1_{A'_k}$ with $A'_k \in \mathcal{F}'$. Then $g \circ f = \sum_{k=1}^m c_k 1_{f \in A'_k}$ and

$$\mu[g \circ f] = \sum_{k=1}^m c_k \mu(f \in A'_k) = \sum_{k=1}^m c_k f_*\mu(A'_k) = f_*\mu[g].$$

□

Proposition 3.21 (Integrals and properties almost everywhere)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* $f : \Omega \to \overline{\mathbb{R}}_+$ *measurable.*

1. *It is* $f = 0$ *almost everywhere iff* $\mu[f] = 0$.

2. *If* $\mu[f] < \infty$, *then* $f < \infty$ *almost everywhere.*

Proof. 1. Let $N := \{f > 0\} \in \mathcal{F}$. '⇒': Since $\mu(N) = 0$, we find $f \leq \infty \cdot 1_N$, so because of Proposition 3.16.2,

$$0 \leq \mu[f] \leq \mu[\infty, N] = \lim_{n\to\infty} \mu[n, N] = 0.$$

For '⇐', let $N_n := \{f \geq 1/n\}$ and thus $N_n \uparrow N$ and $nf \geq 1_{N_n}$, i.e.

$$0 = \mu[f] \geq \tfrac{1}{n}\mu(N_n).$$

This means that $\mu(N_n) = 0$ and therefore $\mu(N) = \mu(\bigcup_{n=1}^\infty N_n) = 0$ by σ-sub-additivity of μ. For 2., let $A := \{f = \infty\}$. Since $f1_{f\geq n} \geq n1_{f\geq n}$,

$$\mu(A) = \mu[1_A] \leq \mu[1_{f\geq n}] \leq \tfrac{1}{n}\mu[f, 1_{f\geq n}] \leq \tfrac{1}{n}\mu[f] \xrightarrow{n\to\infty} 0.$$

This means that $\mu(f = \infty) = 0$, i.e. $f < \infty$ almost everywhere; see Remark 2.14. □

To conclude this section, we show the relationship between the (Lebesgue) integral and the Riemann integral.

Definition 3.22 (Piece-wise constant function and Riemann integral)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a piece-wise constant function, i.e.*

$$f(x) = \sum_{j=-\infty}^{\infty} a_j 1_{[x_{j-1}, x_j)}(x) \tag{3.5}$$

with $x_{j-1} \leq x_j, j \in \mathbb{Z}$, where $a_j \in \mathbb{R}, j \in \mathbb{Z}$. Some $f : [a, b] \to \mathbb{R}$ (with $a < b$) is called Riemann-integrable *if $\lambda[|f|] < \infty$ and there are piece-wise constant functions $f_1^+, f_1^-, f_2^+, f_2^-, ...$ with $f_n^- \leq f \leq f_n^+$ and $\lambda[f_n^+ - f_n^-] \xrightarrow{n\to\infty} 0$. The Riemann integral of f is then defined by $\lambda[f]$. (In particular, the Riemann integral and Lebesgue integral then coincide.*

A function $f : \mathbb{R} \to \mathbb{R}$ is called Riemann-integrable *if $f 1_K$ is Riemann-integrable for all compact intervals $K \subseteq \mathbb{R}$ and $\lambda[f 1_{[-n,n]}]$ converges. This limit is then the Riemann integral of f with respect to λ.*

Proposition 3.23 (Riemann integrability)**.** *Let $f : [0, \infty) \to \mathbb{R}$ have a discrete set of jump points. Then f is integrable, Riemann-integrable, and*

$$\lambda[f] = \lim_{n\to\infty} \sum_{k=1}^{\infty} f(y_{n,k})(x_{n,k} - x_{n,k-1}) \tag{3.6}$$

for $0 = x_{n,0} \leq ... \leq x_{n,k_n} = t$ with $\max_k |x_{n,k} - x_{n,k-1}| \xrightarrow{n\to\infty} 0$ and any $x_{n,k-1} \leq y_{n,k} \leq x_{n,k}$.

Proof. It is sufficient to show the assertion for continuous f. Otherwise, f can be broken down into the continuous pieces. It is also sufficient to show the assertion for f with compact support K. Since f is uniformly continuous on K, first choose $\varepsilon_n \downarrow 0$ and $x_{n,0} \leq ... \leq x_{n,k_n}$ such that $K \subseteq [x_{n,0}, x_{n,k_n}]$ and $\max_{x_{n,k-1} \leq y < x_{n,k}} |f(x_{n,k-1}) - f(y)| < \varepsilon_n$. Now it is easy to find piecewise constant functions f_n^+ and f_n^- such that $f_n^- \leq f \leq f_n^+$ and $||f_n^+ - f_n^-|| \leq \varepsilon_n$. Integrability and Riemann-integrability follows. The formula (3.6) is valid due to the uniform approximation of the function f. □

Example 3.24 (Differences between Riemann and Lebesgue integral)**.**

1. *We start with a function that is Lebesgue-integrable but not Riemann-integrable. Let $f = 1_{[0,1]\cap\mathbb{Q}}$. Then $1_{[0,1]} \leq f^+$ for every piece-wise constant function $f^+ \geq f$ and $f^- \leq 0$ for every piece-wise constant function $f^- \leq f$. In particular, f is not Riemann-integrable.*

2. *As can be seen from the definition of the Riemann integral, every piece-wise constant function is simple, so every Riemann-integrable function is also Lebesgue-integrable. The situation is different for functions on unbounded domains. Let f be given by $f(t) = \frac{(-1)^{\lceil t\rceil+1}}{\lceil t\rceil}$. Then*

$$\lambda[f 1_{[0,2n]}] = \sum_{k=1}^{2n} \frac{(-1)^{k+1}}{k} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots = \sum_{k=1}^{n} \frac{1}{2k-1} - \frac{1}{2k} = \sum_{k=1}^{n} \frac{1}{(2k-1)2k}$$

and we see that the limit is finite, thus f is Riemann-integrable. However, the following applies

$$\lambda[|f|] = \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

So, according to Definition 3.14, f is not Lebesgue-integrable.

3.4 Convergence results

You may ask whether it is really so important that you can integrate more functions with respect to the Lebesgue integral than with respect to the Riemann integral. After all, most applications involve Riemann-integrable functions. However, there is another advantage of the Lebesgue integral, which we will now discuss. In calculus, the following convergence result for the Riemann integral is frequently given:

Theorem 3.25 (Riemann integral convergence result)**.** *Let $a, b \in \mathbb{R}$ with $a < b$, and $f, f_1, f_2, ... : [a, b] \to \mathbb{R}$ be piecewise continuous. If $f_n \xrightarrow{n\to\infty} f$ uniformly, then (using $\int$ for the Riemann integral)*

$$\int_a^b f_n(x)dx \xrightarrow{n\to\infty} \int_a^b f(x)dx.$$

As you see, this result is concerning the interchange of limits limits and integrals, which requires a uniform convergence in this case. For convergence results with respect to the Lebesgue integral, however, we need much weaker conditions for the exchange of integral and limit. The most important of these are the theorem of monotone convergence and the theorem of dominated convergence.

Theorem 3.26 (Monotone convergence)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, $f_1, f_2, \dots \in \mathcal{L}^1(\mu)$ and $f : \Omega \to \overline{\mathbb{R}}$ measurable with $f_n \uparrow f$ almost everywhere. Then,*

$$\lim_{n\to\infty} \mu[f_n] = \mu[f],$$

where both sides can take the value ∞.

Proof. Let $N \in \mathcal{F}$ be such that $\mu(N) = 0$ and $f_n(\omega) \uparrow f(\omega)$ for $\omega \notin N$. Set $g_n := (f_n - f_1)1_{N^c} \geq 0$. This means that $g_n \uparrow (f - f_1)1_{N^c} =: g$ and with Proposition 3.19, Proposition 3.21 and Proposition 3.16.2,

$$\mu[f_n] = \mu[f_1] + \mu[g_n] \xrightarrow{n\to\infty} \mu[f_1] + \mu[g] = \mu[f].$$

□

Theorem 3.27 (Lemma of Fatou)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}_+$ measurable. Then,*

$$\liminf_{n\to\infty} \mu[f_n] \geq \mu[\liminf_{n\to\infty} f_n].$$

Proof. For all $k \geq n$, $f_k \geq \inf_{\ell\geq n} f_\ell$ and thus, for all n,

$$\inf_{k\geq n} \mu[f_k] \geq \mu[\inf_{\ell\geq n} f_\ell]$$

by Proposition 3.16.1 Therefore, with $n \to \infty$

$$\liminf_{n\to\infty} \mu[f_n] = \sup_{n\in\mathbb{N}} \inf_{k\geq n} \mu[f_k] \geq \sup_{n\in\mathbb{N}} \mu[\inf_{k\geq n} f_k] = \mu[\liminf_{n\to\infty} f_n]$$

by monotone convergence, Theorem 3.26, since $\inf_{k\geq n} f_k \uparrow \sup_{n\in\mathbb{N}} \inf_{k\geq n} f_k = \liminf_{n\to\infty} f_n$.
□

Theorem 3.28 (Dominated convergence)**.** *Let* $(\Omega, \mathcal{F}, \mu)$ *be a measure space and* $f, g, f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ *measurable with* $|f_n| \le g$ *almost everywhere,* $\lim_{n\to\infty} f_n = f$ *almost everywhere, and* $g \in \mathcal{L}^1(\mu)$*. Then,*

$$\lim_{n\to\infty} \mu[f_n] = \mu[f].$$

Proof. Without loss of generality, $|f_n| \le g$ and $\lim_{n\to\infty} f_n = f$ holds everywhere. (Otherwise, restrict to a set of full measure.) We use Fatou's lemma and $g - f_n, g + f \ge 0$, i.e.

$$\begin{aligned}
\mu[g+f] &\le \liminf_{n\to\infty} \mu[g + f_n] = \mu[g] + \liminf_{n\to\infty} \mu[f_n], \\
\mu[g-f] &\le \liminf_{n\to\infty} \mu[g - f_n] = \mu[g] - \limsup_{n\to\infty} \mu[f_n].
\end{aligned}$$

After subtracting $\mu[g]$,

$$\mu[f] \le \liminf_{n\to\infty} \mu[f_n] \le \limsup_{n\to\infty} \mu[f_n] \le \mu[f].$$

□

Example 3.29.

1. *Fatou's lemma does not require that any of the* f_n *is integrable. We now give an example to show that in Fatou's lemma '<' rather than '=' holds in general. Let* λ *be the Lebesgue measure and* $f_n = 1/n$ *(i.e. in particular,* f_n *constant),* $n = 1, 2, \dots$*. Then* $f_n \downarrow 0$*, but*

$$\liminf_{n\to\infty} \mu[f_n] = \infty > 0 = \mu[0] = \mu[\liminf_{n\to\infty} f_n].$$

2. *In the theorem of dominated convergence, the condition that* $|f_n| \le g$ *and* $g \in \mathcal{L}^1(\mu)$ *is necessary. For example, let* λ *be the Lebesgue measure on* $[0,1]$ *and* $f_n = n \cdot 1_{[0,1/n]}$*. Then* $\sup_{n\in\mathbb{N}} f_n(x) = \sup\{n : x \le 1/n\} = \left[\frac{1}{x}\right]$[8]*. So there is no* $g \in \mathcal{L}^1(\lambda)$ *with* $f_n \le g$*. Moreover,* $\lim_{n\to\infty} f_n = 0$ *almost everywhere (since* $\{0\}$ *is a null-set) and*

$$\lim_{n\to\infty} \mu[f_n] = 1 \ne 0 = \mu[\lim_{n\to\infty} f_n].$$

The situation is different for $f_n = n \cdot 1_{[0,1/n^2]}$*. Here,*

$$\sup_{n\in\mathbb{N}} f_n(x) = \sup\{n : x \le 1/n^2\} = \left[\frac{1}{\sqrt{x}}\right] \le \frac{1}{\sqrt{x}} =: g(x).$$

On the one hand, $g \in \mathcal{L}^1(\lambda)$*, so dominated convergence applies. On the other hand,* $\lim_{n\to\infty} f_n = 0$ *almost everywhere and*

$$\lim_{n\to\infty} \mu[f_n] = \lim_{n\to\infty} \tfrac{1}{n} = 0 = \mu[0] = \mu[\lim_{n\to\infty} f_n].$$

[8] With $[x] := \sup\{n \in \mathbb{Z} : n \le x\}$ we denote the rounding function.

4 $\mathcal{L}^p$-spaces

Throughout the following section, let $(\Omega, \mathcal{F}, \mu)$ be a measure space. We will now deal with the set of measurable functions $f : \Omega \to \overline{\mathbb{R}}$ satisfying $\mu[|f|^p] < \infty$. We will recognise the resulting function spaces $\mathcal{L}^p(\mu)$ as normed, complete spaces (Proposition 4.8 and Remark 4.4), which also leads to a new concept of convergence. Furthermore, the space $\mathcal{L}^2(\mu)$ will play a special role. It is equipped with a scalar product (namely $\langle f, g\rangle := \mu[fg]$), so general statements are available here, such as the Riesz-Fréchet's theorem (Proposition 4.11). We will use this to characterise σ-finite measures with density by the Radon-Nikodým Theorem (Corollary 4.17).

4.1 Basics

We have already mentioned the spaces $\mathcal{L}^p(\mu)$ in Remark 3.18. By defining the integral in the last section, we can now take a closer look at them. In particular, we show the important Hölder and Minkowski inequalities; see Proposition 4.2. Note that the notation $||.||$ in (4.1) is reminiscent of a norm. As we will discuss in Remark 4.4, it is almost true that $\mathcal{L}^p$, equipped with $||.||_p$ for $p \geq 1$, is a normed space.

Definition 4.1 (**$\mathcal{L}^p(\mu)$**-spaces)**.** *Let $0 < p \leq \infty$. We set*

$$\mathcal{L}^p := \mathcal{L}^p(\mu) := \{f : \Omega \to \overline{\mathbb{R}} \text{ measurable with } ||f||_p < \infty\}$$

for

$$||f||_p := (\mu[|f|^p])^{1/p}, \qquad 0 < p < \infty \tag{4.1}$$

and

$$||f||_\infty := \inf\{K : \mu(|f| > K) = 0\}.$$

On the spaces $\mathcal{L}^p$, $p \geq 1$ we now show a triangle inequality, the Minkowski inequality. It should also be noted that the Hölder inequality in the special case $p = q = 2$ is also called the Cauchy-Schwartz inequality.

Proposition 4.2 (Hölder's and Minkowski's inequality)**.** *Let f, g be measurable.*

1. *Let $0 < p, q, r \leq \infty$ such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$. Then,*

$$||fg||_r \leq ||f||_p ||g||_q \qquad \textit{(Hölder inequality)} \tag{4.2}$$

2. *For $1 \leq p \leq \infty$,*

$$||f + g||_p \leq ||f||_p + ||g||_p. \qquad \textit{(Minkowski inequality)} \tag{4.3}$$

Proof. We start with the proof of Hölder's inequality. In the case $p = \infty$ or $q = \infty$, the statement is clear, so let $p, q < \infty$. If either $||f||_p = 0$, $||f||_p = \infty$, $||g||_q = 0$ or $||g||_q = \infty$, the statement is clear as well. Let $f, g \geq 0$ and $0 < ||f|||_p, ||g|||_q < \infty$ and

$$\widetilde{f} := \frac{f}{||f||_p}, \qquad \widetilde{g} = \frac{g}{||g||_q}.$$

Then, we have to show that $||\widetilde{f}\widetilde{g}||_r \le 1$. Due to the convexity of the exponential function

$$(xy)^r = \exp\left(\tfrac{r}{p} p \log x + \tfrac{r}{q} q \log y\right) \le \tfrac{r}{p} x^p + \tfrac{r}{q} y^q,$$

and thus

$$||\widetilde{f}\widetilde{g}||_r^r = \mu[(\widetilde{f}\widetilde{g})^r] \le \tfrac{r}{p}\mu[\widetilde{f}^p] + \tfrac{r}{q}\mu[\widetilde{g}^q] = 1$$

and the assertion follows.

To prove Minkowski's inequality, we first note that in the cases $p = 1$ and $p = \infty$ the assertion is clear. In the case $1 < p < \infty$, $q = p/(p-1)$ and $r = 1/p + 1/q = 1$ with Hölder's inequality

$$\begin{aligned} ||f+g||_p^p &\le \mu[|f| \cdot |f+g|^{p-1}] + \mu[|g| \cdot |f+g|^{p-1}] \\ &\le ||f||_p \cdot ||(f+g)^{p-1}||_q + ||g||_p \cdot ||(f+g)^{p-1}||_q \\ &= (||f||_p + ||g||_p) \cdot ||f+g||_p^{p-1}, \end{aligned}$$

since $||(f+g)^{p-1}||_q = ||(f+g)^{q(p-1)}||_1^{1/q} = ||(f+g)^p||_1^{(p-1)/p} = ||f+g||_p^{p-1}$. Dividing by $||f+g||_p^{p-1}$ gives the result. □

Proposition 4.3 (Relationship between $\boldsymbol{\mathcal{L}^r}$ and $\boldsymbol{\mathcal{L}^q}$)**.** *Let μ be finite and $1 \le r < q \le \infty$. Then $\mathcal{L}^q(\mu) \subseteq \mathcal{L}^r(\mu)$.*

Proof. The assertion is clear for $q = \infty$. So let $q < \infty$. We use Hölder's inequality. It applies to $f \in \mathcal{L}^q$, since $||1||_p < \infty$ due to the finiteness of μ,

$$||f||_r = ||1 \cdot f||_r \le ||1||_p \cdot ||f||_q < \infty \tag{4.4}$$

for $\frac{1}{p} = \frac{1}{r} - \frac{1}{q} > 0$, from which the assertion immediately follows. □

Remark 4.4 ($\mathcal{L}^p(\mu)$ as a normed space)**.** *For every $p > 0$, we have $||af||_p = |a| \cdot ||f||_p$ for $a \in \mathbb{R}$. Together with Minkowski's inequality (which we have only shown for $1 \le p \le \infty$), this means that $\mathcal{L}^p(\mu)$ is a real vector space. It is crucial to note that the mapping $f \mapsto ||f||_p$ is a pseudo-norm, but not a full norm.*[9] *Indeed, because $||f||_p = 0$ according to Proposition 3.21 only implies that $\mu(f \neq 0) = 0$, but not that $f = 0$, we have $f \neq 0$ with $||f||_p = 0$. In the following, we will therefore identify functions f and g if $f = g$ applies μ almost everywhere. (More precisely, we introduce equivalence classes, where for $f \in \mathcal{L}^p$, the set $\{g \in \mathcal{L}^p : f = g$ almost everywhere$\}$ is the equivalence class of f.) According to the above, $(\{$ equivalence class of $f : f \in \mathcal{L}^p\}, ||\cdot||_p)$ is a normalised space. We will show below that $||\cdot||_p$ is complete (Proposition 4.8), so $(\mathcal{L}^p, ||\cdot||_p)$ is even a Banach space for every $1 \le p \le \infty$. However, we will not make the distinction between $f \in \mathcal{L}^p(\mu)$ and its equivalence class in the sequel.*

Remark 4.5 (Counterexample for $\boldsymbol{\sigma}$-finite $\boldsymbol{\mu}$)**.** *We stress that Proposition 4.3 does not hold if μ is not finite. For example, let λ be the one-dimensional Lebesgue measure and $f : x \mapsto \frac{1}{x} \cdot 1_{x>1}$. Then $f \in \mathcal{L}^2(\lambda)$, but $f \notin \mathcal{L}^1(\lambda)$.*

[9] If V is a real vector space, a mapping $||\cdot|| : V \to \mathbb{R}$ is called *norm* if (i) $||x|| = 0$ iff $x = 0$, (ii) $||a \cdot x|| = |a| \cdot ||x||$ for all $a \in \mathbb{R}$ and $x \in V$, and (iii) $||x+y|| \le ||x|| + ||y||$ for all $x, y \in V$. Then the pair $(V, ||\cdot||)$ is called a normed space. If (i) fails, $||\cdot||$ is called *pseudo-norm*.

4.2 $\mathcal{L}^p$-convergence

We have seen in the theorem of dominated convergence (Theorem 3.28) that for a sequence of functions that converges almost everywhere, their integrals often converge as well. The $\mathcal{L}^p$-convergence considered here now assumes convergence of integrals. We will see that the resulting notion of convergence means that every Cauchy sequence (with respect to $||.||_p$), see Definition 4.1) converges (Proposition 4.8).

Definition 4.6 (Convergence in the ***p***-th mean)**.** *A sequence $f_1, f_2, \dots$ in $\mathcal{L}^p(\mu)$* converges to *$f \in \mathcal{L}^p(\mu)$ iff*

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

We then also write $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} f$.

Proposition 4.7 (Convergence in $\mathcal{L}^p$ and in $\mathcal{L}^q$)**.** *Let $\mu(\Omega) < \infty$, $1 \leq r < q \leq \infty$ and $f, f_1, f_2, \dots \in \mathcal{L}^q$. If $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^q} f$, then also $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^r} f$.*

Proof. The assertion is clear for $q = \infty$, so let $q < \infty$. From (4.4) we have $||f-g||_r \leq ||f-g||_q$, from which the assertion already follows. ☐

Proposition 4.8 (Completeness of $\mathcal{L}^p$)**.** *Let $p \geq 1$ and $f_1, f_2, \dots$ be a Cauchy sequence in $\mathcal{L}^p$. (That is, for every $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that $||f_n - f_m||_p < \varepsilon$ for all $m, n \geq N$.) Then there is an $f \in \mathcal{L}^p$ with $||f_n - f||_p \xrightarrow{n\to\infty} 0$.*

Proof. Let $\varepsilon_1, \varepsilon_2, \dots$ be summable, e.g. $\varepsilon_n := 2^{-n}$. Since $f_1, f_2, \dots$ is a Cauchy sequence, there is an index n_k for each k with $||f_m - f_n||_p \leq \varepsilon_k$ for all $m, n \geq n_k$. In particular, the following applies

$$\sum_{k=1}^{\infty} ||f_{n_{k+1}} - f_{n_k}||_p \leq \sum_{k=1}^{\infty} \varepsilon_k < \infty.$$

With monotone convergence and Minkowski's inequality,

$$\Big|\Big| \sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| \Big|\Big|_p \leq \sum_{k=1}^{\infty} ||f_{n_{k+1}} - f_{n_k}||_p < \infty.$$

In particular $\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}| < \infty$ almost everywhere, i.e. for almost all $\omega \in \Omega$, the sequence $f_{n_1}(\omega), f_{n_2}(\omega), \dots$ is Cauchy in $\mathbb{R}$. Thus, there is a measurable mapping f with $f_{n_k} \xrightarrow{k\to\infty} f$ almost everywhere. According to Fatou's lemma

$$||f_n - f||_p \leq \liminf_{k\to\infty} ||f_{n_k} - f_n||_p \leq \sup_{m\geq n} ||f_m - f_n||_p \xrightarrow{n\to\infty} 0,$$

i.e. $f_n \xrightarrow{n\to\infty}_{\mathcal{L}^p} f$. ☐

4.3 The space $\mathcal{L}^2$

Recall from Remark 4.4, that $\mathcal{L}^p(\mu)$ is in fact a Banach space for all $p \geq 1$. Let us consider the special case $p = 2$. We define a mapping $\mathcal{L}^2 \times \mathcal{L}^2 \to \mathbb{R}$ by

$$\langle f, g \rangle := \mu[fg].$$

Then $\langle\cdot,\cdot\rangle$ is obviously linear, symmetric and positive semi-definite, i.e. a scalar product[10] Consequently, we write $f \perp g$ if and only if $\mu[fg] = 0$. Using

$$||f|| := ||f||_2 = \langle f, f\rangle^{1/2}$$

in this section, $(\mathcal{L}^2, \langle\cdot,\cdot\rangle)$ is therefore a Hilbert space.

Lemma 4.9 (Parallelogram identity)**.** *Let be $f, g \in \mathcal{L}^2$. Then*

$$||f+g||^2 + ||f-g||^2 = 2||f||^2 + 2||g||^2.$$

Proof. From the definition of $||\cdot||$ and the symmetry and bilinearity of $\langle\cdot,\cdot\rangle$,

$$||f+g||^2 + ||f-g||^2 = \langle f+g, f+g\rangle + \langle f-g, f-g\rangle = 2\langle f, f\rangle + 2\langle g, g\rangle = 2||f||^2 + 2||g||^2.$$

□

Proposition 4.10 (Decomposition of $\boldsymbol{f \in \mathcal{L}^2}$)**.** *Let M be a closed, linear subspace of $\mathcal{L}^2$. Then every function $f \in \mathcal{L}^2$ has an almost everywhere unique decomposition $f = g + h$ with $g \in M, h \perp M$.*

Proof. For $f \in \mathcal{L}^2$, we define

$$d_f := \inf_{g \in M}\{||f-g||\}.$$

Choose $g_1, g_2, \dots$ with $||f - g_n|| \xrightarrow{n\to\infty} d_f$. According to the parallelogram identity

$$4d_f^2 + ||g_m - g_n||^2 \le ||2f - g_m - g_n||^2 + ||g_m - g_n||^2 = 2||f - g_m||^2 + 2||f - g_n||^2 \xrightarrow{m,n\to\infty} 4d_f^2.$$

Thus $||g_m - g_n||^2 \xrightarrow{m,n\to\infty} 0$, i.e. $g_1, g_2, \dots$ is a Cauchy sequence. According to Proposition 4.8, there is some $g \in \mathcal{L}^2$ with $||g_n - g|| \xrightarrow{n\to\infty} 0$. Since M is closed, we find $g \in M$ as well as $||h|| = d_f$ for $h := f - g$. So, for all $t > 0, l \in M$, due to the definition of d_f,

$$d_f^2 \le ||h + tl||^2 = d_f^2 + 2t\langle h, l\rangle + t^2||l||^2.$$

Since this applies to all t, $\langle h, l\rangle = 0$, i.e. $h \perp M$.

To prove uniqueness, let $g' + h'$ be a further decomposition of f. Then, due to the linearity of M, on the one hand $g - g' \in M$, on the other hand, almost everywhere, $g - g' = h - h' \perp M$, i.e. $g - g' \perp g - g'$. This means $||g - g'|| = \langle g - g', g - g'\rangle = 0$, i.e. $g = g'$ almost everywhere. □

Proposition 4.11 (Riesz-Fréchet)**.** *A mapping $F : \mathcal{L}^2 \to \mathbb{R}$ is continuous and linear if and only if there exists some $h \in \mathcal{L}^2$ with*

$$F(f) = \langle f, h\rangle, \qquad f \in \mathcal{L}^2.$$

Then, $h \in \mathcal{L}^2$ is almost everywhere uniquely determined.

[10]If V is a real vector space. Then a mapping is called $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{R}$ is a *scalar product* if (i) $\langle x, \alpha y + z\rangle = \alpha\langle x, y\rangle + \langle x, z\rangle$ for all $x, y, z \in V$ and $\alpha \in \mathbb{R}$ (linearity), (ii) $\langle x, y\rangle = \langle y, x\rangle$ (symmetry) and (iii) $\langle x, x\rangle > 0$ for every $x \in V \setminus \{0\}$ (positive definiteness). The norm $||x|| := \langle x, x\rangle^{1/2}$ on V is defined by a scalar product. If $(V, ||\cdot||)$ is complete, then $(V, \langle\cdot,\cdot\rangle)$ is called an *Hilbert space*.

Proof. '⇐': The linearity of $f \mapsto \langle f, h\rangle$ follows from the bilinearity of $\langle\cdot,\cdot\rangle$. The continuity follows from the Cauchy-Schwartz inequality using

$$|\langle |f - f'|, h\rangle| \leq ||f - f'|| \cdot ||h||.$$

'⇒': If $F \equiv 0$, choose $h = 0$. If $F \not\equiv 0$, $M = F^{-1}\{0\}$ is (due to the continuity of F) a closed and (due to the linearity of F) linear subspace of $\mathcal{L}^2$. Choose $f' \in \mathcal{L}^2 \setminus M$ with the (according to Proposition 4.10 almost everywhere unique) orthogonal decomposition $f' = g' + h'$ with $g' \in M$ and $h' \perp M$. Since $f' \notin M$, we have $h' \not\equiv 0$, and $F(h') = F(f') - F(g') = F(f') \neq 0$. We set $h'' = \frac{h'}{F(h')}$, so that $h'' \perp M$ and $F(h'') = 1$ as well as, for all $f \in \mathcal{L}^2$

$$F(f - F(f)h'') = F(f) - F(f)F(h'') = 0.$$

i.e. $f - F(f)h'' \in M$, in particular $\langle F(f)h'', h''\rangle = \langle f, h''\rangle$ and

$$F(f) = \tfrac{1}{||h''||^2} \cdot \langle F(f)h'', h''\rangle = \tfrac{1}{||h''||^2} \cdot \langle f, h''\rangle = \langle f, \tfrac{h''}{||h''||^2}\rangle.$$

Now, the assertion follows with $h := \frac{h''}{||h''||^2}$.

For uniqueness, let $\langle f, h_1 - h_2\rangle = 0$ for all $f \in \mathcal{L}^2$; in particular, with $f = h_1 - h_2$

$$||h_1 - h_2||^2 = \langle h_1 - h_2, h_1 - h_2\rangle = 0,$$

thus $h_1 = h_2$ μ-almost everywhere. □

Remark 4.12 (Generality of the last statements)**.** *Lemma 4.9, as well as the propositions 4.10 and 4.11 also apply if $\mathcal{L}^2$ is replaced by any other Hilbert space.*

4.4 Theorem of Radon-Nikodým

Probability measures with density are already known from the lecture *Elementare probability 1*. This concept is now taken up and embedded in the context of integrals. Let ν be another measure on $\mathcal{F}$. The aim is to specify conditions when the measure ν can be represented by a density. The answer can be found in the Radon-Nikodým theorem (Corollary 4.17). It is a special case of Lebesgue's decomposition theorem, Theorem 4.16. This shows that for every two σ-finite measures μ, ν, the measure ν can be (additively) decomposed into two parts: one absolute continuous with respect to μ and one singular with respect to μ. The absolutely continuous part has a density with respect to μ. First we have to explain all terms.

Definition 4.13 (Absolutely continuous measures)**.**

1. *We say that ν has a* density *f with respect to μ if for all $A \in \mathcal{F}$*

   $$\nu(A) = \mu[f; A].$$

   We then write $f = \frac{d\nu}{d\mu}$ and $\nu = f \cdot \mu$.

2. *The measure ν is called* absolutely continuous with respect to μ *if all μ-zero sets are also ν-zero sets. We then write $\nu \ll \mu$. If both $\nu \ll \mu$ and $\mu \ll \nu$, then μ and ν are called* equivalent.

3. *The measures μ and ν are called* singular *if there is an $A \in \mathcal{F}$ with $\mu(A) = 0$ and $\nu(A^c) = 0$. We then write $\mu \perp \nu$.*

Lemma 4.14 (Chain rule and uniqueness). *Let μ be a measure on $\mathcal{F}$.*

1. *Let ν be a σ-finite measure. If g_1 and g_2 are densities of ν with respect to μ, then $g_1 = g_2$, μ-almost everywhere.*

2. *Let $f : \Omega \to \mathbb{R}_+$ and $g : \Omega \to \mathbb{R}$ be measurable. Then,*

$$(f \cdot \mu)[g] = \mu[fg],$$

if one of the two sides exists.

Proof. 1. Let $\Omega_1, \Omega_2, \dots \in \mathcal{F}$ be such that $\Omega_n \uparrow \Omega$ and $\nu(\Omega_n) < \infty$. Set $A_n := \Omega_n \cap \{g_1 > g_2\}$. Since both g_1 and g_2 are densities of ν with respect to μ,

$$\mu[g_1 - g_2; A_n] = 0.$$

Since only $g_1 > g_2$ is possible on A_n, $g_1 = g_2$ is $1_{A_n}\mu$-almost everywhere. Furthermore,

$$\mu\{g_1 > g_2\} = \mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = 0.$$

Analogously, $\mu\{g_1 < g_2\} = 0$ and thus $g_1 = g_2$ μ-almost everywhere.

2. The statement is clear for $g = 1_A$ with $A \in \mathcal{F}$. This extends step by step to simple functions, positive measurable functions and finally to the general case. □

Example 4.15 (Known densities). 1. *Some density functions are already known from the lecture* Elementary probability 1. *For example, let $\mu \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+$ be*

$$f_{N(\mu,\sigma^2)}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2}\Big)$$

and λ is the one-dimensional Lebesgue measure. Then the probability measure $f_{N(\mu,\sigma^2)} \cdot \lambda$ is called normal distribution with expected value μ and variance σ^2. *We can compute for some $X \sim N_{\mu,\sigma^2}$ and $h : x \mapsto (x-\mu)/\sigma$*

$$\mathbf{P}(h(X) \le x) = \mathbf{P}(X \le \mu + x\sigma) = \int_{-\infty}^{\mu+x\sigma} f_{N(\mu,\sigma^2)}(y)dy$$
$$\overset{z=(y-\mu)/\sigma}{=} \int_{-\infty}^{x} f_{N(0,1)}(z)dz,$$

which shows that $(X-\mu)/\sigma \sim N_{(0,1)}$.

For $\gamma \ge 0$, let

$$f_{\exp(\gamma)}(x) := 1_{x\ge 0} \cdot \gamma e^{-\gamma x},$$

the probability measure $f_{\exp(\gamma)} \cdot \lambda$ is called exponential distribution with parameter γ. *For example, you can now use Lemma 4.14 to calculate for some $X \sim \exp(\gamma)$*

$$\mathbf{E}[X] = f_{\exp(\gamma)} \cdot \lambda[id] = \int_0^\infty \gamma e^{-\gamma x} x dx = -e^{-\gamma x}x\Big|_0^\infty + \int_0^\infty e^{-\gamma x}dx = \frac{1}{\gamma}.$$

So, we have computed the expected value of the exponential distribution for the parameter γ.

2. *Of course, there are not only densities with respect to the Lebesgue measure. Let, for example*

$$\mu = \sum_{n=0}^{\infty} \delta_n$$

be the counting measure on $\mathbb{N}_0$ *(see Example 2.2) and* $f : \mathbb{N}_0 \to \mathbb{R}_+$*, given for a* $\gamma \geq 0$ *by*

$$f(k) = e^{-\gamma} \frac{\gamma^k}{k!}.$$

Then $f \cdot \mu$ *is the Poisson distribution for the parameter* γ *on* $2^{\mathbb{N}_0}$ *according to Example 2.2.*

Theorem 4.16 (Lebesgue's decomposition theorem)**.** *Let* μ, ν *be* σ*-finite measures on* $(\Omega, \mathcal{F})$*. Then* ν *can be written uniquely as*

$$\nu = \nu_a + \nu_s \qquad \text{with} \qquad \nu_a \ll \mu, \nu_s \perp \mu.$$

The measure ν_a *has a density with respect to* μ *that is* μ*-almost everywhere finite.*

Proof. Since μ, ν are σ-finite, we find $\Omega_1, \Omega_2, \dots \in \mathcal{F}$ with $\Omega_n \uparrow \Omega$ and $\nu(\Omega_n), \mu(\Omega_n) < \infty$. In particular, without loss of generality, we can assume that μ, ν are finite measures. With Proposition 4.7. the linear mapping

$$\begin{cases} \mathcal{L}^2(\mu + \nu) & \to \mathbb{R}) \\ f & \mapsto \nu[f] \end{cases}$$

is continuous. According to Proposition 4.11, there is some $h \in \mathcal{L}^2(\mu + \nu)$ with

$$\nu[f] = (\mu + \nu)[fh], \tag{4.5}$$

thus

$$\nu[f(1 - h)] = \mu[fh] \tag{4.6}$$

for each $f \in \mathcal{L}^2(\mu + \nu)$. If one chooses $f = 1_{\{h<0\}}$ in (4.5), we find

$$0 \leq \nu\{h < 0\} = (\mu + \nu)[h; h < 0] \leq 0,$$

i.e. $h \geq 0$ $(\mu + \nu)$-almost everywhere. Similarly, $f = 1_{\{h>1\}}$ can be used to deduce from (4.6) that

$$0 \leq \mu[h; \{h > 1\}] = \nu[1 - h; \{h > 1\} \leq 0,$$

so $h \leq 1$ $(\mu + \nu)$-almost everywhere. Now, let $f \geq 0$ be measurable and $f_1, f_2, \dots \in \mathcal{L}^2(\mu + \nu)$ with $f_n \uparrow f$. With monotone convergence,

$$\nu[f(1 - h)] = \lim_{n\to\infty} \nu[f_n(1 - h)] = \lim_{n\to\infty} \mu[f_n h] = \mu[fh],$$

i.e. (4.6) applies to all measurable $f \geq 0$.

Now let $E := h^{-1}\{1\}$. From (4.6) it follows with $f = 1_E$ that

$$\mu(E) = \mu[h; E] = \nu[1 - h; E] = 0.$$

We define two measures ν_a and ν_s for $A \in \mathcal{F}$ by

$$\nu_a(A) = \nu(A \setminus E), \qquad \nu_s(A) = \nu(A \cap E),$$

so that $\nu = \nu_a + \nu_s$ and $\nu_s \perp \mu$. To show that $\nu_a \ll \mu$ choose $A \in \mathcal{F}$ with $\mu(A) = 0$. This means that after (4.6)

$$\nu[1 - h; A \setminus E] = \mu[h; A \setminus E] = 0.$$

Since $h < 1$ on $A \setminus E$, $\nu_a(A) = \nu(A \setminus E) = 0$, i.e. $\nu_a \ll \mu$.

We claim that $g := \frac{h}{1-h} 1_{\Omega \setminus E}$ is the density of ν_a with respect to μ. Indeed, using (4.6),

$$\mu[g; A] = \mu\Big[\frac{h}{1-h}; A \setminus E\Big] = \nu(A \setminus E) = \nu_a(A).$$

To show the uniqueness of the decomposition, let $\nu = \nu_a + \nu_s = \widetilde{\nu}_a + \widetilde{\nu}_s$ for $\nu_a, \widetilde{\nu}_a \ll \mu$, $\nu_s, \widetilde{\nu}_s \perp \mu$. Choose $A, \widetilde{A} \in \mathcal{A}$ with $\nu_s(A) = \mu(A^c) = \widetilde{\nu}_s(\widetilde{A}) = \mu(\widetilde{A}^c) = 0$. Then,

$$\nu_s(A \cap \widetilde{A}) = \widetilde{\nu}_s(A \cap \widetilde{A}) = \nu_a(A^c \cup \widetilde{A}^c) = \widetilde{\nu}_a(A^c \cup \widetilde{A}^c) = 0$$

and therefore

$$\begin{aligned} \nu_a &= 1_{A\cap\widetilde{A}} \cdot \nu_a = 1_{A\cap\widetilde{A}} \cdot \nu = 1_{A\cap\widetilde{A}} \cdot \widetilde{\nu}_a = \widetilde{\nu}_a, \\ \nu_s &= \nu - \nu_a = \nu - \widetilde{\nu}_a = \widetilde{\nu}_s. \end{aligned}$$

□

Corollary 4.17 (Theorem of Radon-Nikodým)**.** *Let μ and ν be σ-finite measures. Then, ν has a density with respect to μ if and only if $\nu \ll \mu$.*

Proof. '$\Rightarrow$': clear.
'$\Leftarrow$': According to Theorem 4.16, there is a unique decomposition $\nu = \nu_a + \nu_s$ with $\nu_a \ll \mu, \nu_s \perp \mu$. Since $\nu \ll \mu$, $\nu_s = 0$ must apply and therefore $\nu = \nu_a$. In particular, the density of ν exists with respect to μ. □

Example 4.18. *In Lebesgue's decomposition Theorem 4.16 and in the Theorem of Radon-Nikodým 4.17, the condition that μ and ν are σ-finite cannot be omitted, as the following example shows:*

Let $(\Omega, \mathcal{F})$ be a measure space with uncountable Ω and

$$\mathcal{F} := \{A : A \text{ or } A^c \text{ countable}\}.$$

Let μ and ν be infinite measures on $(\Omega, \mathcal{F})$, given by

$$\nu(A) := \begin{cases} 0, & A \text{ countable}, \\ \infty, & \text{otherwise}, \end{cases} \qquad \mu(A) := \begin{cases} |A|, & A \text{ finite}, \\ \infty, & \text{otherwise}. \end{cases}$$

Then obviously $\nu \ll \mu$. Assume there is a $\mathcal{F}$-measurable density of ν with respect to μ. Then, for all $\omega \in \Omega$

$$0 = \nu\{\omega\} = \mu[f; \{\omega\}] = f(\omega)\mu(\{\omega\}) = f(\omega).$$

Thus $f = 0$ and $\nu = 0$ would contradict the definition of ν.

5 Product spaces

Let $(\Omega_i)_{i\in I}$ be a family of sets. Then,

$$\Omega := \bigtimes_{i\in I} \Omega_i := \{(\omega_i)_{i\in I} : \omega_i \in \Omega_i\}$$

is the product space of $(\Omega_i)_{i\in I}$. We further define the projections for $H \subseteq J \subseteq I$

$$\pi_H^J : \bigtimes_{i\in J} \Omega_i \to \bigtimes_{i\in H} \Omega_i,$$

as well as $\pi_H := \pi_H^I$ and $\pi_i := \pi_{\{i\}}$, $i \in I$. In this chapter, we will apply all the concepts in the context of measurability. Of particular importance is the theorem on projective limits of probability measures, Theorem 5.24, which will play a fundamental role in the theory of stochastic processes.

5.1 Topology

We start with the definition of a topology on product spaces. In short, this topology is made such that projections are continuous.

Definition 5.1 (Product space and product topology). *If $(\Omega_i, \mathcal{O}_i)_{i\in I}$ is a family of topological spaces, then the topology $\mathcal{O}$, generated by (recall from Definition A.1.7)*[11]

$$\mathcal{C} := \{A_i \times \bigtimes_{j\in I, j\neq i} \Omega_j; i \in I, A_i \in \mathcal{O}_i\}$$

is called the product topology *on Ω.*

Remark 5.2 (Continuity of projections). *All projections $\pi_i, i \in I$ are continuous with respect to the product topology.*
Indeed, it is

$$\pi_i^{-1}(A_i) = A_i \times \bigtimes_{I\ni j\neq i} \Omega_j \in \mathcal{C} \subseteq \mathcal{O}$$

for $A_i \in \mathcal{O}_i$. The projection is therefore continuous (see Definition A.1.10).

5.2 Semi-rings, rings and σ-algebras

Analogous to topology, the product σ-algebra is just such that projections are measurable functions.

Definition 5.3 (Product-$\boldsymbol{\sigma}$-algebra). *If $(\Omega_i, \mathcal{F}_i)_{i\in I}$ is a family of measurable spaces, the σ-algebra*

$$\bigotimes_{i\in I} \mathcal{F}_i := \sigma(\mathcal{E}), \qquad \mathcal{E} := \{A_i \times \bigtimes_{j\in I, j\neq i} \Omega_j : i \in I, A_i \in \mathcal{F}_i\} \tag{5.1}$$

is called the product-σ-algebra *on $\Omega := \bigtimes_{i\in I} \Omega_i$. If $(\Omega_i, \mathcal{F}_i) = (\Omega, \mathcal{F}), i \in I$, we set $\mathcal{F}^I := \bigotimes_{i\in I} \mathcal{F}$.*

[11]We write $A \subseteq_f B$ if $\subseteq B$ and A is finite.

Remark 5.4 (Measurability of projections)**.** *Analogous to the product topology, the projections π_i are measurable with respect to $\bigotimes_{i\in I}\mathcal{F}_i$. This is because for $A_i \in \mathcal{F}_i$,*

$$\pi_i^{-1}(A_i) = A_i \times \bigtimes_{I\ni j\neq i} \Omega_j \in \bigotimes_{i\in I}\mathcal{F}_i.$$

Lemma 5.5 (Product-σ-algebra for countable products)**.** *Let I be arbitrary and $(\Omega_i, \mathcal{O}_i)_{i\in I}$ a family of topological spaces, and $(\Omega, \mathcal{O})$ the product space, equipped with the product topology from Definition 5.1. Then $\bigotimes_{i\in I}\mathcal{B}(\Omega_i) \subseteq \mathcal{B}(\Omega)$. Moreover, if I is countable and $(\Omega_i, \mathcal{O}_i)_{i\in I}$ a family of separable metric spaces, then $\mathcal{B}(\Omega) = \bigotimes_{i\in I}\mathcal{B}(\Omega_i)$. In particular, $\mathcal{B}(\mathbb{R}^d) = \bigotimes_{i=1}^d \mathcal{B}(\mathbb{R})$.*

Proof. Let $\mathcal{C}$ be as in Definition 5.1, $\mathcal{O}$ the product topology (i.e. $\sigma(\mathcal{O}) = \mathcal{B}(\Omega)$), and $\mathcal{E}$ as in Definition 5.3 with $\mathcal{F}_i$ replaced by $\mathcal{B}(\Omega_i)$. Clearly, $\mathcal{C} \subseteq \mathcal{O}(\mathcal{C})$ as well as $\mathcal{C} \subseteq \mathcal{E}$ by definition. In addition, $\mathcal{E} \subseteq \sigma(\mathcal{C})$ by definition of $\mathcal{B}(\Omega_i)$. This leads to

$$\bigotimes_{i\in I}\mathcal{B}(\Omega_i) = \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{C}) \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\Omega).$$

In case of a countable union of separable spaces, every set in $\mathcal{O}(\mathcal{C})$ is a countable union of sets in $\mathcal{C}$ (see Lemma 1.8), leading to

$$\mathcal{O}(\mathcal{C}) \subseteq \sigma(\mathcal{C}), \qquad \text{so} \qquad \sigma(\mathcal{O}(\mathcal{C})) \subseteq \sigma(\sigma(\mathcal{C})) = \sigma(\mathcal{C}).$$

Hence, all assertions are shown. □

Remark 5.6. *If I is uncountable, by using countable intersections and unions, $\bigotimes_{i\in I}\mathcal{B}(\Omega_i)$ only contains sets which depends on a countable number of coordinates. In contrast, $\sigma(\mathcal{O}(\mathcal{B}))$ contains sets which arise as uncountable intersections of closed sets, which in general depend on an uncountable number of coordinates. This shows that for uncountable product spaces, in general $\bigotimes_{i\in I}\mathcal{B}(\Omega_i) \subsetneq \mathcal{B}(\Omega))$.*

Lemma 5.7 (Products of generators/semi-rings are generators/semi-rings)**.** *Let $(\Omega_i, \mathcal{F}_i)$ be measurable spaces and $\Omega = \bigtimes_{i\in I}\Omega_i$.*

1. *Let I be finite and $\mathcal{H}_i$ a semi-ring with $\sigma(\mathcal{H}_i) = \mathcal{F}_i$. Then*

$$\mathcal{H} := \{\bigtimes_{i\in I} A_i : A_i \in \mathcal{H}_i, i \in I\} \tag{5.2}$$

   is a semi-ring with $\sigma(\mathcal{H}) = \bigotimes_{i\in I}\mathcal{F}_i$.

2. *Let I be arbitrary and $\mathcal{H}_i$ a $\cap$-stable generator of $\mathcal{F}_i$, $i \in I$. Then*

$$\mathcal{H} := \{\bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega_i : J \subseteq_f I, A_i \in \mathcal{H}_i, i \in J\}$$

   is a $\cap$-stable generator of $\bigotimes_{i\in I}\mathcal{F}_i$.

Proof. For 1., let $I = \{1, \dots, d\}$ without loss of generality. It is clear that $\mathcal{H}$ is $\cap$-stable. Property (ii) for semi-rings is shown by induction over d. The assertion is clear for $d = 1$, since $\mathcal{H}_1$ is a half-ring. If it holds to $d - 1$, then

$$\begin{aligned}&(A_1 \times \cdots \times A_d) \setminus (B_1 \times \cdots \times B_d)\\ &\quad = \big(A_1 \times \cdots \times A_{d-1} \times (A_d \setminus B_d)\big) \uplus \big((A_1 \times \cdots \times A_{d-1}) \setminus (B_1 \times \cdots \times B_{d-1})\big) \times (A_d \cap B_d)\end{aligned}$$

The first term of the last line can be represented as a disjoint union of sets from $\mathcal{H}$, since $\mathcal{H}_d$ is a half-ring. The second term can be represented as a disjoint union, since by the induction hypothesis, $(A_1 \times \cdots \times A_{d-1}) \setminus (B_1 \times \cdots \times B_{d-1})$ can be represented as a disjoint union of sets of the form $H_1 \times \cdots \times H_{d-1}$ with $H_i \in \mathcal{H}_i, i = 1, \dots, d - 1$.

For 2. it is again clear that $\mathcal{H}$ is $\cap$-stable. From (5.1) it immediately follows that $\mathcal{H} \subseteq \bigotimes_{i \in I} \mathcal{F}_i$, therefore $\sigma(\mathcal{H}) \subseteq \bigotimes_{i \in I} \mathcal{F}_i$. Conversely, it is clear that for $A_i \in \mathcal{F}_i$

$$A_i \times \bigtimes_{j \neq i} \Omega_j \in \sigma\Big(\big\{A_i \times \bigtimes_{j \neq i} \Omega_j : A_i \in \mathcal{H}_i\big\}\Big) \subseteq \sigma(\mathcal{H}),$$

from which $\bigotimes_{i \in I} \mathcal{F}_i \subseteq \sigma(\mathcal{H})$ and thus the assertion follows. □

Corollary 5.8 (Borel's $\boldsymbol{\sigma}$-algebra on $\mathbb{R}^{\boldsymbol{d}}$ is generated by cylinders)**.** *Let $\Omega = \mathbb{R}^d$. For $\underline{a} = (a_1, \dots, a_d), \underline{b} = (b_1, \dots, b_d) \in \mathbb{R}^d$ we set $\underline{a} \leq \underline{b}$ if and only if $a_i \leq b_i, i = 1, \dots, d$, and with*

$$(\underline{a}, \underline{b}] = (a_1, b_1] \times \cdots \times (a_d, b_d]$$

the half-open cylinder. Then,

$$\mathcal{H} := \{(\underline{a}, \underline{b}] : \underline{a}, \underline{b} \in \mathbb{Q}, \underline{a} \leq \underline{b}\}$$

is a semi-ring with $\sigma(\mathcal{H}) = \mathcal{B}(\mathbb{R}^d)$.

Proof. According to Example 1.3.1 and Lemma 5.7.1, $\mathcal{H}$ is a semi-ring that generates $\bigotimes_{i=1}^d \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d)$; see Lemma 5.5. □

5.3 Measures and integrals

Integrals in multi-dimensional spaces are already known from calculus. We now first define measures on product spaces and the corresponding (multiple) integrals. Fubini's theorem (Theorem 5.13) can then be used to interpret and analyse integrals according to measures on product spaces as multiple integrals. For this purpose, it is necessary that the integrands appearing in the multiple integrals are measurable. This is ensured in Lemma 5.11. In order to be able to define measures on product spaces in sufficient generality, we first need the concept of the transition kernel.

Definition 5.9 (Transition kernel)**.** *Let $(\Omega_i, \mathcal{F}_i), i = 1, 2$ be measurable spaces. A mapping $\kappa : \Omega_1 \times \mathcal{F}_2 \to \mathbb{R}_+$ is called a* transition kernel from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$ *if (i) for all $\omega_1 \in \Omega_1$, the map $\kappa(\omega_1, .)$ is a measure on $\mathcal{F}_2$ and (ii) for all $A_2 \in \mathcal{F}_2$ $\kappa(., A_2)$ is $\mathcal{F}_1$-measurable.*

A transition kernel is called σ-finite if there is a sequence $\Omega_{21}, \Omega_{22}, \dots \in \mathcal{F}_2$ with $\Omega_{2n} \uparrow \Omega_2$ and $\sup_{\omega_1} \kappa(\omega_1, \Omega_{2n}) < \infty$ for all $n = 1, 2, \dots$ It is called stochastic kernel *or* Markov kernel *if for all $\omega_1 \in \Omega_1$ the map $\kappa(\omega_1, .)$ is a probability measure.*

Example 5.10 (Markov chain)**.** *Let $\Omega = \{\omega_1, \dots, \omega_n\}$ be finite and $P = (p_{ij})_{1 \le i,j \le n}$ with $p_{ij} \in [0,1]$ and $\sum_{j=1}^n p_{ij} = 1$. Then,*

$$\kappa(\omega_i, .) := \sum_{j=1}^n p_{ij} \cdot \delta_{\omega_j}$$

is a Markov kernel from $(\Omega, 2^\Omega)$ to $(\Omega, 2^\Omega)$. Here, P as a stochastic matrix is the transition matrix of a homogeneous, Ω-valued Markov chain.

Lemma 5.11 (Measurability of integrable sections)**.** *Let $(\Omega_i, \mathcal{F}_i), i = 1, 2$ be measurable spaces, κ be a σ-finite transition kernel from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$ and $f : \Omega_1 \times \Omega_2 \to \mathbb{R}_+$ to $\mathcal{F}_1 \otimes \mathcal{F}_2$ measurable. Then*

$$\omega_1 \mapsto \kappa(\omega_1, .)[f] := \int \kappa(\omega_1, d\omega_2) f(\omega_1, \omega_2)$$

to $\mathcal{F}_1$-measurable.

Proof. We assume that $\kappa(\omega_1, \Omega_2) < \infty$ for all $\omega_1 \in \Omega_1$. (The general case is then performed using a sequence $\Omega_{11}, \Omega_{12}, \dots \in \mathcal{F}_1$ with $\Omega_{1n} \uparrow \Omega_1$.) Let

$$\mathcal{D} := \{A \in \mathcal{F}_1 \otimes \mathcal{F}_2 : \omega_1 \mapsto \kappa(\omega_1, .)[1_A] \text{ is } \mathcal{F}_1\text{-measurable}\}.$$

Then it is easy to check that $\mathcal{D}$ is a $\cap$-stable Dynkin system. Furthermore, $\mathcal{H} \subseteq \mathcal{D}$, where $\mathcal{H}$ is defined as in (5.2). Thus, according to Theorem 1.13, $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{H}) \subseteq \mathcal{D} \subseteq \mathcal{F}_1 \otimes \mathcal{F}_2$. Therefore, $\omega_1 \mapsto \kappa(\omega_1, .)[1_A]$ is measurable for all $A \in \mathcal{F}_1 \otimes \mathcal{F}_2$ with respect to $\mathcal{F}_1$. This statement can be extended immediately by using a simple function instead of 1_A. By monotonic convergence, it then also follows that $\omega_1 \mapsto \kappa(\omega_1, .)[f]$ is measurable for all measurable, non-negative functions according to $\mathcal{F}_1$. ☐

Theorem 5.12 (Theorem of Ionescu-Tulcea)**.** *Let $(\Omega_i, \mathcal{F}_i), i = 0, \dots, n$ measurable spaces, μ a σ-finite measure on $\mathcal{F}_0$ and κ_i a σ-finite transition kernel of $\left(\bigtimes_{j=0}^{i-1} \Omega_j, \bigotimes_{j=0}^{i-1} \mathcal{F}_j\right)$ to $(\Omega_i, \mathcal{F}_i)$, $i = 1, \dots, n$. Then there is exactly one σ-finite measure $\mu \bigotimes_{i=1}^n \kappa_i$ on $\left(\bigtimes_{i=0}^n \Omega_i, \bigotimes_{i=0}^n \mathcal{F}_i\right)$ with*

$$\Big(\mu \bigotimes_{i=1}^n \kappa_i\Big)(A_0 \times \dots \times A_n) = \int_{A_0} \mu(d\omega_0) \Big(\int_{A_1} \kappa_1(\omega_0, d\omega_1) \cdots \Big(\int_{A_n} \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) \Big) \cdots \Big). \tag{5.3}$$

Proof. We show the theorem only for $n = 1$, the general case is then done by induction.

The proof is an application of Theorem 2.16. First we establish that according to Lemma 5.7, the set system $\mathcal{H}$ defined in (5.2) is a semi-ring on $\bigtimes_{i=1}^n \Omega_i$. We first show that the given set function is σ-finite on $\mathcal{H}$. Namely, there is $\Omega_{i1}, \Omega_{i2}, \in \mathcal{F}_i$ with $\Omega_{in} \uparrow \Omega_i, i = 0, 1$ with $\mu(\Omega_{0n}) < \infty, \kappa_1(\omega_0, \Omega_{1n}) < \infty, n = 1, 2, \dots, \omega_0 \in \Omega_0$ and $\sup_{\omega_0 \in \Omega_0} \kappa_1(\omega_0, \Omega_{1n}) =: C_n < \infty$. This means that $\mu \otimes \kappa_1(\Omega_{0n} \times \Omega_{1n}) \le C_n \cdot \mu(\Omega_{0n}) < \infty$ and $\Omega_{0n} \times \Omega_{1n} \uparrow \Omega_0 \times \Omega_1$. This means that $\mu \otimes \kappa_1$ is also σ-finite. If we define $\widetilde{\mu}$ on $\mathcal{H}$ using (5.3), this is therefore a σ-finite set function.

We now show that $\widetilde{\mu}$ is σ-subadditive and finitely additive on $\mathcal{H}$. For $A_1, \dots, A_n \in \mathcal{H}$ and $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{H}$, by σ-subadditivity of $\kappa_1(\omega_0, .)$ for all $\omega_0 \in \Omega_0$

$$\begin{aligned}\widetilde{\mu}(A) &= \int \mu(d\omega_0) \int \kappa_1(\omega_0, d\omega_1) 1_A(\omega_0, \omega_1) \\ &\leq \sum_{n=1}^{\infty} \int \mu(d\omega_0) \int \kappa_1(\omega_0, d\omega_1) 1_{A_n}(\omega_0, \omega_1) = \sum_{n=1}^{\infty} \widetilde{\mu}(A_n).\end{aligned}$$

Similarly, finite additivity is shown. According to Lemma 2.5, $\widetilde{\mu}$ is therefore σ-additive. From Theorem 2.16 it now follows that there is exactly one extension of $\widetilde{\mu}$ to $\sigma(\mathcal{H}) = \bigotimes_{i=1}^n \sigma(\mathcal{H}_i)$, which is the one given in the theorem. □

We now deal with the measure defined in Theorem 5.12.

Theorem 5.13 (Fubini's theorem). *Let* $(\Omega_i, \mathcal{F}_i)$, μ, κ_i *and* $\mu \bigotimes_{i=1}^n \kappa_i$ *be as in Theorem 5.12. Further, let* $f : \times_{i=0}^n \Omega_i \to \mathbb{R}_+$ *be measurable with respect to* $\bigotimes_{i=0}^n \mathcal{F}_i$. *Then,*

$$\int f d\big(\mu \bigotimes_{i=0}^n \kappa_i\big) = \int \mu(d\omega_0)\Big(\int \kappa_1(\omega_1, d\omega_2) \cdots \Big(\int \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) f(\omega_0, \dots, \omega_n)\Big) \cdots \Big). \tag{5.4}$$

This equality also applies if $f : \times_{i=0}^n \Omega_i \to \mathbb{R}$ *is measurable with* $\int |f| d\big(\mu \bigotimes_{i=0}^n \kappa_i\big) < \infty$.

Proof. Consider the set function $\widetilde{\mu}$ on $\bigotimes_{i=0}^n \mathcal{F}_i$, given by

$$\widetilde{\mu} : A \mapsto \int \mu(d\omega_0)\Big(\int \kappa_1(\omega_1, d\omega_2) \cdots \Big(\int \kappa_n(\omega_0, \dots, \omega_{n-1}, d\omega_n) 1_A(\omega_0, \dots, \omega_n)\Big) \cdots \Big).$$

You can see that $\widetilde{\mu}$ corresponds on $\mathcal{H}$ from (5.2) with $\mu \bigotimes_{i=1}^n \kappa_i$. Since $\mathcal{H}$ is $\cap$-stable, the equality (5.4) for indicator functions follows due to Proposition 2.11. By means of linearity of the integral, (5.4) is first extended to simple functions and then using monotonicity to any non-negative, measurable function. Note that all occurring integrands are measurable according to Lemma 5.11. □

Corollary 5.14 (Product measures). *Let* $\Omega = \times_{i=1}^n \Omega_i$ *and* $\mathcal{H}_i \subseteq 2^{\Omega_i}$ *be a semi-ring,* $i = 1, \dots, n$, *and* $\mu_i : \mathcal{H}_i \to \mathbb{R}_+$ σ*-finite and,* σ*-additive,* $i = 1, \dots, n$. *Then there is exactly one measure* $\mu_1 \otimes \cdots \otimes \mu_n$ *on* $\bigotimes_{i=1}^n \sigma(\mathcal{H}_i)$ *with*

$$\mu_1 \otimes \cdots \otimes \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \cdots \mu_n(A_n). \tag{5.5}$$

For a measurable function $f : \Omega \to \mathbb{R}_+$, *the value of the integral does not depend on the order of integration of the coordinates* $\omega, ..., \omega_n$, *i.e. for every permutation* π *on* $\{1, \dots, n\}$,

$$\int f d\mu_1 \otimes \cdots \otimes \mu_n = \int \Big(\cdots \Big(\int f(\omega_1, \dots, \omega_n) \mu_{\pi(1)}(d\omega_{\pi_1})\Big) \cdots \Big) \mu_{\pi(n)}(d\omega_{\pi(n)}).$$

This formula also applies to $f : \Omega \to \mathbb{R}$, *if* $\int |f| d\mu_1 \otimes \cdots \otimes \mu_n < \infty$.

Proof. The corollary follows directly from Theorem 5.12 and Theorem 5.13 if you set $\kappa_i(\omega_0, \dots, \omega_{i-1}, .) = \mu_i(.)$ for all $\omega_0, \dots, \omega_{i-1}$. □

Definition 5.15 (Finite product measure). *Consider the same situation as in Corollary 5.14. Then, the unique measure $\mu_1 \otimes \cdots \otimes \mu_n$ from Corollary 5.14 is called the* product measure of $\mu_1, \dots, \mu_n$. *We also write*

$$\bigotimes_{i=1}^{n} \mu_i := \mu_1 \otimes \cdots \otimes \mu_n.$$

If $(\Omega_i, \mathcal{H}_i, \mu_i) = (\Omega_0, \mathcal{H}_i, \mu_0)$, $i = 1, \dots, n$, i.e. all spaces are equal, we also denote it by

$$\mu_0^{\otimes n} := \mu_1 \otimes \cdots \otimes \mu_n.$$

Example 5.16 (multidimensional Lebesgue measure).

1. *Let λ be the one-dimensional Lebesgue measure on $\mathcal{B}(\mathbb{R})$ from Proposition 2.18. Then $\lambda^{\otimes d}$ is the d-dimensional Lebesgue measure.*

2. *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by*

$$f(x, y) = \frac{xy}{(x^2 + y^2)^2}.$$

Then, for every $x \in \mathbb{R}$

$$\int \lambda(dy) f(x, y) = 0,$$

since $f(x, .) \in \mathcal{L}^1(\lambda)$ and $f(x, y) = -f(x, -y)$. Therefore, in particular

$$\int \lambda(dx)\Big(\int \lambda(dy) f(x, y)\Big) = \int \lambda(dy)\Big(\int \lambda(dx) f(x, y)\Big) = 0.$$

However, $|f|$ is not integrable with respect to $\lambda^{\otimes 2}$ because f has a non-integrable pole in $(0, 0)$. As this example shows, we have to be careful with multiple integrals. In particular, it does not follow from the equality and finiteness of multiple integrals that the integrand is integrable.

5.4 Convolution of measures

We now consider a simple combination of product dimensions and image measure. To convolve measures μ, ν on $\mathcal{B}(\mathbb{R})$, we first consider the product measure $\mu \otimes \nu$. The image measure under summation is then the convolution of μ, ν. We will later identify this convolution as the distribution of $X + Y$ if X, Y are *independent* random variables with distribution μ and ν, respectively. Sometimes, for example with Poisson distributions and normal distributions, the convolution is again a Poisson or normal distribution.

Definition 5.17 (Convolution of measures). *Let $\mu_1, \dots, \mu_n$ be σ-finite measures on $\mathcal{B}(\mathbb{R})$ and $\mu_1 \otimes \cdots \otimes \mu_n$ their product measure. Further, let $S(x_1, \dots, x_n) := x_1 + \cdots + x_n$. Then the image measure $S_*(\mu_1 \otimes \cdots \otimes \mu_n)$ is called the* convolution *of the measures $\mu_1, \dots, \mu_n$ and is denoted by $\mu_1 * \cdots * \mu_n$ or $*_{i=1}^n \mu_i$.*

Example 5.18 (Convolution of Poisson and geometric distributions).

1. *For $\gamma_1, \gamma_2 \geq 0$ let $\mu_{Poi(\gamma_1)}$ and $\mu_{Poi(\gamma_2)}$ be two Poisson distributions from Example 2.2. We calculate*

the convolution of the two distributions by

$$\begin{aligned}
\mu_{Poi(\gamma_1)} * \mu_{Poi(\gamma_2)} &= \sum_{m,n} 1_{m+n=k} e^{-(\gamma_1+\gamma_2)} \frac{\gamma_1^m \gamma_2^n}{m!n!} \cdot \delta_k \\
&= \sum_{m=0}^{k} e^{-(\gamma_1+\gamma_2)} \frac{\gamma_1^m \gamma_2^{k-m}}{m!(k-m)} \cdot \delta_k \\
&= e^{-(\gamma_1+\gamma_2)} \frac{(\gamma_1+\gamma_2)^k}{k!} \cdot \delta_k \sum_{m=0}^{k} \binom{k}{m} \frac{\gamma_1^m \gamma_2^{k-m}}{(\gamma_1+\gamma_2)^k} \\
&= \mu_{Poi(\gamma_1+\gamma_2)}.
\end{aligned}$$

2. *The geometric distribution for the parameter $p \in [0,1]$ is as well known from Example 2.2. The convolution of two measures $\mu_{geom(p)}$ is given by*

$$\begin{aligned}
\mu_{geom(p)} * \mu_{geom(p)} &= \sum_{m=2}^{k} (1-p)^{m-1} p (1-p)^{k-m-1} p \cdot \delta_k \\
&= (k-1)(1-p)^{k-2} p^2 \cdot \delta_k.
\end{aligned}$$

This is a negative binomial distribution for the parameters p and 2.

Lemma 5.19 (Convolution of distributions with densities)**.** *Let λ be a measure on $\mathcal{B}(\mathbb{R})$, $\mu = f_\mu \cdot \lambda$ and $\nu = f_\nu \cdot \lambda$ for measurable densities $f_\mu, f_\nu : \mathbb{R} \to \mathbb{R}_+$. Then $\mu * \nu = f_{\mu*\nu} \cdot \lambda$ with*

$$f_{\mu*\nu}(t) = \int f_\mu(s) f_\nu(t-s) \lambda(ds).$$

Proof. The proof is a simple application of Fubini's theorem, Theorem 5.13. □

Example 5.20 (Convolution of normal distributions)**.** *Let $f_{N(\mu_1,\sigma_1^2)}$ and $f_{N(\mu_2,\sigma_2^2)}$ be the density functions of two normal distributions with expected value μ_1, μ_2 and variance σ_1^2 and σ_2^2, respectively. Let further $\mu := \mu_1+\mu_2$ and $\sigma^2 = \sigma_1^2+\sigma_2^2$. Then the density of the convolution is given by*

$$\begin{aligned}
x \mapsto & \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} \int \exp\Big(- \frac{(y-\mu_1)^2}{2\sigma_1^2} - \frac{(x-y-\mu_2)^2}{2\sigma_2^2} \Big) dy \\
& \overset{y \to (y-\mu_1)\sigma/(\sigma_1\sigma_2)}{=} \frac{1}{2\pi\sigma} \int \exp\Big(- \frac{\sigma_2^2 y^2}{2\sigma^2} - \frac{\Big((x-\mu) - y\frac{\sigma_1\sigma_2}{\sigma}\Big)^2}{2\sigma_2^2} \Big) dy \\
&= \frac{1}{2\pi\sigma} \int \exp\Big(- \frac{\sigma_2^2 y^2 + \Big((x-\mu)\frac{\sigma}{\sigma_2} - \sigma_1 y\Big)^2}{2\sigma^2} \Big) dy \\
&= \frac{1}{2\pi\sigma} \int \exp\Big(- \frac{(\sigma y - \frac{\sigma_1}{\sigma_2}(x-\mu))^2}{2\sigma^2} - \frac{(x-\mu)^2\Big(\frac{\sigma^2}{\sigma_2^2} - \frac{\sigma_1^2}{\sigma_2^2}\Big)}{2\sigma^2} \Big) dy \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-\mu)^2}{2\sigma^2} \Big).
\end{aligned}$$

So, the convolution is again a normal distribution. This now has expected value μ and variance σ^2.

5.5 Projective families of probability measures

So far we have defined σ-finite measures on finite product spaces. This is not sufficient for the probability theory to be discussed later. To understand this, let us recall the infinite coin toss, which was already considered in the lecture *Elementary probability 1*. Here, we would say that $\Omega = \{\text{head}, \text{tail}\}^{\mathbb{N}}$ and the corresponding probability measure is the product measure $\mathbf{P}^{\otimes\infty}$ of $\mathbf{P} = \frac{1}{2}\delta_{\text{head}} + \frac{1}{2}\delta_{\text{tail}}$. However, this is an infinite (but still countable) product measure whose existence we have not yet shown. More generally, a large part of the lecture *Stochastic Processes* will contain such measures (even on uncountable product spaces). We now give the general construction of probability measures on product measures, which goes back to Kolmogorov (and Daniell). It should be mentioned here that in the resulting theorem of Kolmogorov (theorem 5.24) the assumption is made that Ω is Polish.

Definition 5.21 (Projective limit)**.**

1. *Let $(\Omega, \mathcal{F})$ be a measurable space, I an arbitrary index set and $(\Omega^J, \mathcal{F}^J)_{J\subseteq_f I}$ be a family of measurable product spaces, equipped with the product σ-algebra, as in Definition 5.3. A family of probability measures $(\mathbf{P}_J)_{J\subseteq_f I}$, where $\mathbf{P}_J$ is a probability measure on $\mathcal{F}^J$, is called a projective family if*

$$\mathbf{P}_H = (\pi^J_H)_*\mathbf{P}_J$$

for all $H \subseteq J \subseteq_f I$. (In other words, projection of coordinates in J to coordinates in H under $\mathbf{P}_J$ leads to $\mathbf{P}_H$.)

2. *If for a projective family $(\mathbf{P}_J)_{J\subseteq_f I}$ of probability measures there exists a probability measure $\mathbf{P}_I$ on $\mathcal{F}^I$ with $\mathbf{P}_J = (\pi_J)_*\mathbf{P}_I$ for all $J \subseteq_f I$, then $\mathbf{P}_I$ is called the projective limit of the projective family. We then write*

$$\mathbf{P}_I = \varprojlim_{J\subseteq_f I} \mathbf{P}_J.$$

Example 5.22 (Projective limits and stochastic processes)**.** *Projective families play a major role in at least two situations.*

1. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and I an infinite index set. In Definition 5.15 we have defined the product measure $\mathbf{P}^{\otimes J}$ on $\mathcal{F}^J$ for each $J \subseteq_f I$. The family $(\mathbf{P}^{\otimes J})_{J\subseteq_f I}$ is projective. If $H \subseteq J \subseteq_f I$, then for $A_i \in \mathcal{F}, i \in H$,*

$$\begin{aligned}(\pi^J_H)_*\mathbf{P}^{\otimes J}\big(\bigtimes_{i\in H} A_i\big) &= \mathbf{P}^{\otimes J}\big((\pi^J_H)^{-1}\big(\bigtimes_{i\in H} A_i\big)\big)\\ &= \mathbf{P}^{\otimes J}\big(\bigtimes_{i\in H} A_i \times \bigtimes_{i\in J\setminus H} \Omega\big)\\ &= \prod_{i\in H}\mathbf{P}(A_i)\cdot\prod_{i\in J\setminus H}\mathbf{P}(\Omega)\\ &= \prod_{i\in H}\mathbf{P}(A_i)\\ &= \mathbf{P}^{\otimes H}\big(\bigtimes_{i\in H} A_i\big).\end{aligned}$$

However, we have not yet shown that the projective limit of $(\mathbf{P}^{\otimes J})_{J\subseteq_f I}$ exists. We would then call this the infinite product measure $\mathbf{P}^{\otimes I}$. (In particular, this would give the probability space for the infinite coin toss from the beginning of this section.)

2. *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, I an arbitrary index set, $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ a measurable space and $X_i : \Omega \to \widetilde{\Omega}, i \in I$ a random variable (i.e. a function measurable with respect to $\mathcal{F}/\mathcal{F}_i$). We will call the family $\mathcal{X} := (X_i)_{i\in I}$ a stochastic process. So $\mathcal{X} : \Omega \to \widetilde{\Omega}^I$ with $\mathcal{X}(\omega) = (X_i(\omega))_{i\in I}$. One can now ask whether the distribution of $\mathcal{X}$ (i.e. the image measure $\mathcal{X}_*\mathbf{P}$) exists as a distribution on $\widetilde{\mathcal{F}}^I$.*

   It should be noted that $\widetilde{\mathbf{P}}_J := ((X_j)_{j\in J})_*\mathbf{P}, J \subseteq_f I$ is a projective family. If $H \subseteq J \subseteq_f I$ and $\widetilde{A}_i \in \widetilde{\mathcal{F}}, i \in H$, then

$$
\begin{aligned}
(\pi_H^J)_*\widetilde{\mathbf{P}}_J\big(\bigtimes_{j\in H} \widetilde{A}_j\big) &= \widetilde{\mathbf{P}}_J\big((\pi_H^J)^{-1} \bigtimes_{j\in H} \widetilde{A}_j\big) \\
&= \widetilde{\mathbf{P}}_J\big(\bigtimes_{j\in H} \widetilde{A}_j \times \bigtimes_{j\in J\setminus H} \widetilde{\Omega}\big) \\
&= \mathbf{P}\big(X_j \in \widetilde{A}_j, j \in H \text{ and } X_j \in \widetilde{\Omega}, j \in J \setminus H\big) \\
&= \mathbf{P}\big(X_j \in \widetilde{A}_j, j \in H\big) \\
&= \widetilde{\mathbf{P}}_H\big(\bigtimes_{j\in H} \widetilde{A}_j\big).
\end{aligned}
$$

   As Theorem 5.24 below shows, the distribution $\mathcal{X}_*\mathbf{P}$ (which is then the projective limit of $(\widetilde{\mathbf{P}}_J)_{J\subseteq_f I}$) exists at least if $\widetilde{\mathcal{F}}$ is the Borel's σ-algebra of a Polish space.

Remark 5.23 (Uniqueness of the projective limit)**.** *For each projective family $(\mathbf{P}_J)_{J\subseteq_f I}$ there is at most one projective limit: If $\mathbf{P}_I$ and $\widetilde{\mathbf{P}}_I$ are two projective limits, then for*

$$\mathcal{H}' := \Big\{ \bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega_i, A_i \in \mathcal{F}_i, i \in J \subseteq_f I \Big\},$$

we see that $\mathcal{H}'$ generates $\mathcal{F}^I$ (compare with $\mathcal{H}$ from Lemma 5.7), and is $\cap$-stable. Hence, for $A = \bigtimes_{i\in J} A_i \times \bigtimes_{i\in I\setminus J} \Omega_i \in \mathcal{H}'$,

$$\mathbf{P}_I(A) = \mathbf{P}_J\Big(\bigtimes_{i\in J} A_i\Big) = \widetilde{\mathbf{P}}_J\Big(\bigtimes_{i\in J} A_i\Big) = \widetilde{\mathbf{P}}_I(A).$$

This means that $\mathbf{P}_I$ and $\widetilde{\mathbf{P}}_I$ coincide on the $\cap$-stable generator and according to Proposition 2.11, $\mathbf{P}_I = \widetilde{\mathbf{P}}_I$. The content of the next theorem is that there is exactly one projective limit for Polish spaces.

Theorem 5.24 (Existence of processes, Kolmogorov)**.** *Let $(\Omega, \mathcal{O})$ be Polish, $\mathcal{F} = \mathcal{B}(\mathcal{O})$ and $(\mathbf{P}_J)_{J\subseteq_f I}$ a projective family of probability measures on $\mathcal{F}$. Then there is the projective limit $\varprojlim_{J\subseteq_f I} \mathbf{P}_J$.*

Proof. Let $\mathcal{H}'$ be as in Remark 5.23 and μ be a finite additive set function on $\mathcal{H}'$, defined by the projective family using

$$\mu\Big(\bigtimes_{j\in J} A_j \times \bigtimes_{i\in I\setminus J} \Omega\Big) := \mathbf{P}_J\Big(\bigtimes_{j\in J} A_j\Big).$$

According to Lemma 5.7, $\mathcal{H}$ is a semi-ring and μ is a well-defined content on $\mathcal{H}$. Further,

$$\mathcal{K} := \{\bigtimes_{j\in J} K_j \times \bigtimes_{i\in I\setminus J} \Omega : J \subseteq_f I, K_j \text{ compact}\} \subseteq \mathcal{H}$$

is a compact system.

We now show that μ is inner regular with respect to $\mathcal{K}$. Let $\varepsilon > 0$, $\bigtimes_{i \in J} A_i \times \bigtimes_{i \in I \setminus J} \Omega \in \mathcal{H}$ for $J \subseteq_f I$ and $A_i \in \mathcal{F}, i \in J$. Since $\mathbf{P}_j$ is a measure for $j \in I$, according to Lemma 2.9 there are compact sets $K_j \in \mathcal{F}$ with $K_j \subseteq A_j$ and $\mathbf{P}_j(A_j \setminus K_j) \leq \varepsilon$. This means that

$$\begin{aligned}
\mu\Big(\Big(\bigtimes_{i \in J} A_i \times \bigtimes_{i \in I \setminus J} \Omega\Big) \setminus \Big(\bigtimes_{i \in J} K_i \times \bigtimes_{i \in I \setminus J} \Omega\Big)\Big) &= \mu\Big(\Big((\bigtimes_{i \in J} A_i) \setminus (\bigtimes_{i \in J} K_i)\Big) \times \bigtimes_{i \in I \setminus J} \Omega\Big) \\
&= \mathbf{P}_J\big((\bigtimes_{j \in J} A_j) \setminus (\bigtimes_{j \in J} K_j)\big) \\
&\leq \mathbf{P}_J\Big(\bigcup_{j \in J} (A_j \setminus K_j) \times \bigtimes_{i \neq j} \Omega\Big) \\
&\leq \sum_{j \in J} \mathbf{P}_J\Big((A_j \setminus K_j) \times \bigtimes_{i \neq j} \Omega\Big) \\
&= \sum_{j \in J} \mathbf{P}_j(A_j \setminus K_j) \\
&\leq |J|\varepsilon.
\end{aligned}$$

Since J was finite and $\varepsilon > 0$ was arbitrary, we have shown inner regularity of μ with respect to $\mathcal{K}$. According to Theorem 2.10, μ is σ-additive. Furthermore, $\mu(\Omega^I) = 1$, so μ can be uniquely extended to a measure $\mathbf{P}$ on $\sigma(\mathcal{H}') = \mathcal{F}^I$ according to Theorem 2.16. This must be the projective limit of $(\mathbf{P}_J)_{J \subseteq_f I}$. □

A Some Topology

A topology is used in mathematics whenever a notion of convergence is introduced. Even if topologies have only been treated as a sideline in most of your lectures so far, some concepts of convergence are well known. There are also many connections between measure theory and topology; see for example the notion of a Borel σ-algebra in Definition 1.7. Therefore, we repeat basic notions of topology here.

A.1 Basics

By a *topology* we understand a family of open subsets of a space Ω.[12] In metric spaces one calls a set A *open* if for every $\omega \in A$ there is an open ball[13] $B_\varepsilon(\omega) \subseteq A$ for some $\varepsilon > 0$. This case of metric spaces is in practice the most important.

In measure theory, the case of separable topologies, which are generated by complete metrics, is of particular importance. Such spaces are called *Polish*.

Definition A.1 (Metric space, topological space)**.** *Let Ω be some set.*

1. *A function $r : \Omega \times \Omega \to \mathbb{R}_+$ is called a* metric *if (i) $r(\omega, \omega') \neq 0$ for $\omega \neq \omega'$, (ii) $r(\omega, \omega') = r(\omega', \omega)$ for all $\omega, \omega' \in \Omega$, and (iii) $r(\omega, \omega'') \leq r(\omega, \omega') + r(\omega', \omega'')$ for all $\omega, \omega', \omega'' \in \Omega$. The pair (Ω, r) is a* metric space*.*

   For $\omega \in \Omega$ and $\varepsilon > 0$, we denote by $B_\varepsilon(\omega) := \{\omega' \in \Omega : r(\omega, \omega') < \varepsilon\}$ the open ball *around ω with radius ε.*

2. *A metric r on Ω is called* complete *if every Cauchy sequence converges. That is, if $\omega_1, \omega_2, \ldots \in \Omega$ with*

   $$\forall \varepsilon > 0 \ \exists N \in \mathbb{N} \ \forall n, m \geq N : \ \ r(\omega_n, \omega_m) < \varepsilon,$$

   then there is $\omega \in \Omega$ with $r(\omega_n, \omega) \xrightarrow{n\to\infty} 0$.

3. *A set system $\mathcal{O} \subseteq 2^\Omega$ is called* topology *if (i) $\emptyset, \Omega \in \mathcal{O}$; (ii) if $A, B \in \mathcal{O}$, then $A \cap B \in \mathcal{O}$; (iii) if I is arbitrary and if $A_i \in \mathcal{O}, i \in I$, then $\bigcup_{i\in I} A_i \in \mathcal{O}$. The pair $(\Omega, \mathcal{O})$ is called* topological space*. Its members, i.e. every $A \in \mathcal{O}$, is called* open*; any set $A \subseteq \Omega$ with $A^c \in \mathcal{O}$ is called* closed*.*

4. *Let $(\Omega, \mathcal{O})$ be a topological space and $A \subseteq \Omega$. Then*

   $$A^\circ := \bigcup\{O \subseteq A : O \in \mathcal{O}\}$$

   is called the interior of A *and*

   $$\overline{A} := \bigcap\{F \supseteq A : F^c \in \mathcal{O}\}$$

   is called closure of A*.*

5. *A topological space $(\Omega, \mathcal{O})$ is called* separable *if there is a countable set $\Omega' \subseteq \Omega$ with $\overline{\Omega}' = \Omega$.*

[12] We will write 2^Ω for the set of all subsets of Ω.
[13] We define $B_\varepsilon(\omega) := \omega' : r(\omega, \omega') < \varepsilon\}$.

6. *Let $(\Omega, \mathcal{O})$ be a topological space and $\mathcal{B} \subseteq \mathcal{O}$. Then $\mathcal{B}$ is called a* base *of $\mathcal{O}$ if*
$$\forall A \in \mathcal{O} \;\; \forall \omega \in A \;\; \exists B \in \mathcal{B} : \omega \in B \subseteq A.$$
This is exactly the case if
$$\mathcal{O} = \{A \subseteq \Omega : \;\; \forall \omega \in A \, \exists B \in \mathcal{B} : \omega \in B \subseteq A\}. \tag{A.1}$$
or (equivalently)
$$\mathcal{O} = \Big\{ \bigcup_{B \in \mathcal{C}} B : \mathcal{C} \subseteq \mathcal{B} \Big\}. \tag{A.2}$$

7. *Let $\mathcal{B} \subseteq 2^\Omega$. Then, the right hand sides of* (A.1) *and* (A.2) *define the topology* generated by $\mathcal{B}$*, which we denote by $\mathcal{O}(\mathcal{B})$.*

8. *Let (Ω, r) be a metric space and*
$$\mathcal{B} := \{B_\varepsilon(\omega) : \varepsilon > 0, \omega \in \Omega\}. \tag{A.3}$$
Then $\mathcal{O}(\mathcal{B})$ is the topology generated by r*. If specifically $\Omega \subseteq \mathbb{R}^d$ and r is the Euclidean distance, then the topology generated in* (A.1) *or* (A.2) *is called the* euclidean topology*.*

9. *The space $(\Omega, \mathcal{O})$ is called (completely) metrizable if there exists a (complete) metric r on Ω such that* (A.1) *holds with $\mathcal{B}$ from* (A.3)*. The space $(\Omega, \mathcal{O})$ is called* Polish *if it is separable and completely metrizable.*

10. *Let $(\Omega, \mathcal{O})$ and $(\Omega', \mathcal{O}')$ be topological spaces. Then a mapping $f : \Omega \to \Omega'$ is called* continuous *if $f^{-1}(A') \in \mathcal{O}$ for all $A' \in \mathcal{O}'$.*

Example A.2 (The space $\overline{\mathbb{R}}$)**.** *We will often use functions with values in*[14]
$$\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\} \quad \textit{or} \quad \overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$$
In order to be able to consider these spaces as topological spaces, we set
$$\varphi : \begin{cases} \overline{\mathbb{R}} & \to [-1, 1], \\ x & \mapsto \begin{cases} \frac{2}{\pi}\arctan(x), & x \in \mathbb{R}, \\ 1, & x = \infty, \\ -1, & x = -\infty \end{cases} \end{cases}$$
and define the metric
$$r_{\overline{\mathbb{R}}}(x, y) := |\varphi(x) - \varphi(y)|, \qquad x, y \in \overline{\mathbb{R}}.$$
The topological space defined by $r_{\overline{\mathbb{R}}}(\overline{\mathbb{R}}, \overline{\mathcal{O}})$ extends the Euclidean topology $(\mathbb{R}, \mathcal{O})$ to $\mathbb{R}$ in the sense that $\{A \cap \mathbb{R} : A \in \overline{\mathcal{O}}\} = \mathcal{O}$. This is true because φ is continuous on $\mathbb{R}$ with a continuous inverse function. It further holds that $(\overline{\mathbb{R}}, \overline{\mathcal{O}})$ is separable and $r_{\overline{\mathbb{R}}}$ is a complete metric.

On $\overline{\mathbb{R}}$ one can calculate as usual in calculus. For example, $a \cdot \infty = \infty$ for $a > 0$. However, expressions like $\infty - \infty$ and ∞/∞ are not defined.

[14] The notation $\overline{\mathbb{R}}$ suggests that the termination of $\mathbb{R}$ is meant here. This is not true, since the added elements $-\infty, \infty$ do not lie in $\mathbb{R}$, but closures of sets always contain at most the elements of the basic space can contain. Topologically, $\overline{\mathbb{R}}$ is the two-point compactification of $\mathbb{R}$

Remark A.3 (Metric and topological spaces). *Let $(\Omega, \mathcal{O})$ be a topological space and $\omega, \omega_1, \omega_2, \dots \in \Omega$. We define*

$$\omega_n \xrightarrow{n\to\infty} \omega :\iff (\forall O \in \mathcal{O} : \omega \in O \Rightarrow \omega_n \in O \text{ for almost all } n \in \mathbb{N}). \tag{A.4}$$

In particular, this gives any topology on Ω a notion of convergence for sequences in Ω.

This notion of convergence agrees with the well-known notion on metric spaces: namely, if r is a metric on Ω, which generates $\mathcal{O}$, then the right-hand side of (A.4) holds if and only if for all $\varepsilon > 0$, we have $r(\omega_n, \omega) < \varepsilon$ for almost all $n \in \mathbb{N}$.

Using the notion of convergence from (A.4), we state the following well-known property:

Lemma A.4 (Closure of a metric space). *Let (Ω, r) be a metric space and $\mathcal{O}$ be the topology generated by r. For $F \subseteq \Omega$ the following are equivalent:*

1. *F is closed.*
2. *If $\omega_1, \omega_2, \dots \in F$ and $\omega \in \Omega$ are such that $\omega_n \xrightarrow{n\to\infty} \omega$, then $\omega \in F$.*

In particular, for every $A \subseteq \Omega$ there exists the closure $\overline{A}$ consists exactly of the cluster points[15] of A.

Proof. '1.⇒2.' Assume there is a sequence $\omega_1, \omega_2, \dots \in F$ with $\omega_n \xrightarrow{n\to\infty} \omega \in F^c$. Then, since $F^c \in \mathcal{O}$, we find $\omega_n \in F^c$ for almost all n. This is in contradiction with the assumption.
'2.⇒1.': Suppose F was not closed, i.e. F^c, is not open. Then there is $\omega \in F^c$ such that for all $\varepsilon > 0$ it holds that $B_\varepsilon(\omega) \not\subseteq F^c$. Choose $\varepsilon_1, \varepsilon_2, \dots > 0$ with[16] $\varepsilon_n \downarrow 0$ and $\omega_n \in B_{\varepsilon_n}(\omega) \cap F$. Then $\omega_1, \omega_2, \dots \in F$ with $\omega_n \xrightarrow{n\to\infty} \omega$, but $\omega \in F^c$. □

Lemma A.5 (Countable base and separable spaces). *Let (Ω, r) be a separable metric space, $\mathcal{O}$ be the topology generated by r, Ω' countable with $\overline{\Omega'} = \Omega$ and*

$$\widetilde{\mathcal{B}} := \{B_\varepsilon(\omega) : \varepsilon \in \mathbb{Q}_+, \omega \in \Omega'\}.$$

Then $\widetilde{\mathcal{B}}$ is countable and $\mathcal{O}(\widetilde{\mathcal{B}}) = \mathcal{O}$.

Proof. Clearly, $\widetilde{\mathcal{B}}$ is countable and $\mathcal{O}(\widetilde{\mathcal{B}}) \subseteq \mathcal{O}$. Let $\mathcal{B}$ as in (A.3). Then for $B_\varepsilon(\omega) \in \mathcal{B}$

$$B_\varepsilon(\omega) = \bigcup_{\widetilde{\mathcal{B}} \ni B \subseteq B_\varepsilon(\omega)} B,$$

thus $\mathcal{B} \subseteq \mathcal{O}(\widetilde{\mathcal{B}})$ and thus $\mathcal{O} = \mathcal{O}(\mathcal{B}) \subseteq \mathcal{O}(\widetilde{\mathcal{B}})$. □

Example A.6 (Two Polish spaces).

1. *Let $\mathcal{O}$ be the Euclidean topology on $\mathbb{R}^d$, as given in Definition A.1.9 by the Euclidean metric. From your lecture* Analysis I, *it is known that this metric is complete. Further, $\mathbb{Q}^d$ is countable and every $\omega \in \mathbb{R}^d$ is a cluster point of a sequence in $\mathbb{Q}^d$. Thus, in particular, $\overline{\mathbb{Q}^d} = \mathbb{R}^d$ by Lemma A.4, so $\mathbb{R}^d$ is separable. So overall, $(\mathbb{R}^d, \mathcal{O})$ is Polish.*

---

[15] A cluster point of A is any limit of a convergent sequence $\omega_1, \omega_2, \dots \in A$.
[16] We write $\varepsilon_n \downarrow 0$ if $\varepsilon_1 \geq \varepsilon_2 \geq \dots$ and $\varepsilon_n \xrightarrow{n\to\infty} 0$

2. *Let $K \subseteq \mathbb{R}$ be compact (i.e. closed and bounded) and $\Omega = \mathcal{C}_{\mathbb{R}}(K)$ be the set of continuous functions $\omega : K \to \mathbb{R}$. On Ω let*

$$r(\omega_1, \omega_2) := \sup_{x\in K} |\omega_1(x) - \omega_2(x)|$$

be the supremum distance. It is known from Analysis II *that r is complete is complete. Furthermore, every $\omega \in \Omega$ can be calculated according to the Weierstrass' approximation theorem can be uniformly approximated by polynomials by polynomials. Let Ω' be the countable set of polynomials with rational coefficients. Then it also holds that $\overline{\Omega'} = \Omega$. Thus $(\Omega, \mathcal{O})$ is separable, i.e. Polish.*

A.2 Compact sets

Topological spaces can be very large. Just think of the space $\mathbb{R}$, in which there are sequences that diverge. Now *compact set* are considered as smaller subsets of a topological space. In such compact sets there are always convergent subsequences.

Definition A.7 (Relatively compact, compact, relatively sequenctially compact, totally restricted)**.** *Let $(\Omega, \mathcal{O})$ be a topological space and $K \subseteq \Omega$.*

1. *The set K is called* compact *if every open cover has a finite partial cover. That is: If $O_i \in \mathcal{O}, i \in I$ and $K \subseteq \bigcup_{i\in I} O_i$, then there is*[17]*. $J \subset I$ with $K \subseteq \bigcup_{i\in J} O_i$.*
2. *The set K is called* relatively compact *if $\overline{K}$ is compact.*
3. *The set K is called* relatively sequentially compact *if for every sequence $\omega_1, \omega_2, \ldots \in K$ there is a convergent subsequence, i.e. there is an increasing sequence $k_1, k_2, ... \uparrow \infty$ and $\omega \in \Omega$ with $\omega_{k_n} \xrightarrow{n\to\infty} \omega$ as in* (A.4)*.*
4. *Let r be a metric that generates $\mathcal{O}$. Then we call $K \subseteq \Omega$* totally bounded *if for every $\varepsilon > 0$ there is an $N \in \mathbb{N}$ and $\omega_1, \ldots, \omega_N \in K$ such that $K \subseteq \bigcup_{n=1}^N B_\varepsilon(\omega_n)$. In other words, for every radius $\varepsilon > 0$, there is a finite number of balls with this radius covering K.*

Lemma A.8 (Compact sets are closed)**.** *. Let (Ω, r) be a metric space and $\mathcal{O}$ the topology generated by r. If $K \subseteq \Omega$ is compact, then K is also closed.*

Proof. We show that K^c is open. For this, let $\omega \in K^c$. For all $\omega' \in K$ we choose $\delta_{\omega'}$ and $\varepsilon_{\omega'}$ such that $B_{\delta_{\omega'}}(\omega) \cap B_{\varepsilon_{\omega'}}(\omega') = \emptyset$. Then obviously $\bigcup_{\omega'\in K} B_{\varepsilon_{\omega'}}(\omega') \supseteq K$, so there is $J \subseteq_f K$ with $K \subseteq \bigcup_{\omega'\in J} B_{\varepsilon_{\omega'}}(\omega')$. Set $\delta := \min_{\omega'\in J} \delta_{\omega'} > 0$. Then $B_\delta(\omega) is \cap K \subseteq B_\delta(\omega) \cap \bigcup_{\omega'\in J} B_{\varepsilon_{\omega'}}(\omega') = \emptyset$, i.e. $B_\delta(\omega) \subseteq K^c$. Since $\omega \in K^c$ was arbitrary, K^c is open, so K is closed. □

The following theorem about compact sets gives a complete characterization of compact sets in Polish spaces.

Proposition A.9 (Characterising relatively compact sets)**.** *. Let (Ω, r) be a metric space, $\mathcal{O}$ be the topology generated by r and $K \subseteq \Omega$. Consider the following statements:*

1. *K is relatively compact.*

[17] We write $J \subseteq_f I$ if $J \subseteq I$ and J is finite

2. If $F_i \subseteq \overline{K}$ is closed, $i \in I$, and $\bigcap_{i\in I} F_i = \emptyset$, then there is $J \subseteq_f I$ with $\bigcap_{i\in J} F_i = \emptyset$.

3. K is relatively sequentially compact.

4. K is totally bounded.

Then

$$4. \Longleftarrow 1. \iff 2. \Longrightarrow 3.$$

Furthermore, 3. $\Longrightarrow$ 2. *also holds if $(\Omega, \mathcal{O})$ is separable and* 4. $\Longrightarrow$ 3. *if (Ω, r) is complete. In particular, all four statements are equivalent, if $(\Omega, \mathcal{O})$ is Polish.*

Corollary A.10. *Let (Ω, r) be a metric space, $\mathcal{O}$ the topology generated by r. Then closed subsets of compact sets are compact.*

Proof. Let $K \subseteq \Omega$ be compact and $A \subseteq K$ closed. A closed set is compact if and only if it is relatively compact. From Proposition A.9.2 one reads, because of the relative compactness of K, that for F_i closed, $i \in I$, with $F_i \subseteq A \subseteq K$ and $\bigcap_{i\in I} F_i = \emptyset$ a $J \subset I$ exists with $\bigcap_{i\in J} F_i = \emptyset$. Again with Proposition A.9.2 it follows that A is relatively compact, i.e. compact. □

Proof of Proposition A.9. '1.⇒4.' Let $\overline{K}$ be compact and $\varepsilon > 0$. Obviously, $\bigcup_{\omega\in K} B_\varepsilon(\omega) \supseteq \overline{K}$ is an open covering. Thus, since $\overline{K}$ is compact, there is a finite subcover, i.e. there is $\omega_1, \ldots, \omega_N$ with $\overline{K} \subseteq \bigcup_{n=1}^N B_\varepsilon(\omega_n)$. Since $\varepsilon > 0$ was arbitrary, the assertion follows.

'1.⇒2.' Now let $F_i, i \in I$ be as stated. Then $\bigcup_{i\in I} F_i^c = \left(\bigcap_{i\in I} F_i\right)^c = \Omega \supseteq \overline{K}$. Since $\overline{K}$ is compact, there is $J \subseteq_f I$ with $\overline{K} \subseteq \bigcup_{i\in J} F_i^c$. Thus $\bigcap_{i\in J} F_i = \left(\bigcup_{i\in J} F_i^c\right) \subseteq \overline{K}^c$. But since $F_i \subseteq \overline{K}$ was assumed, $\bigcap_{i\in J} F_i = \emptyset$.

'2.⇒1.' Let $O_i \in \mathcal{O}, i \in I$ be a covering of $\overline{K}$, i.e. $\overline{K} \subseteq \bigcup_{i\in I} O_i$. Set $F_i = O_i^c \cap \overline{K}$, then $F_i^c \in \mathcal{O}$ and $\bigcap_{i\in I} F_i = \overline{K} \cap \left(\bigcup_{i\in I} O_i\right)^c = \emptyset$. So there is $J \subseteq_f I$ with $\bigcap_{i\in J} F_i = \emptyset$. Therefore, $\overline{K}^c \cup \bigcup_{i\in J} O_i = \bigcup_{i\in J} F_i^c = \Omega$, so $\bigcup_{i\in J} O_i \supseteq \overline{K}$. So we found a finite subcovering. In other words, $\overline{K}$ is compact.

'2,⇒3.' Let $\omega_1, \omega_2, \ldots \in K$. We set $F_n = \overline{\{\omega_n, \omega_{n+1}, \ldots\}} \subseteq \overline{K}$. Suppose there is no convergent subsequence of $\omega_1, \omega_2, \ldots$ Then $\bigcap_{n=1}^\infty F_n = \emptyset$. From 2. it then follows that there is a $N \in \mathbb{N}$ with $\emptyset = .\bigcap_{n=1}^N F_n = F_N$. This is a contradiction, since F_N is not empty by construction; therefore there is a convergent subsequence.

'3.⇒1.' if $(\Omega, \mathcal{O})$ is separable. Let Ω' be countable with $\overline{\Omega'} = \Omega$ and $\mathcal{B} := \{B_{1/n}(\omega) : \omega \in \Omega', n \in \mathbb{N}\}$. Then, $\mathcal{B}$ is a countable basis of $\mathcal{O}$. We write $\mathcal{B} = \{B_1, B_2, \ldots\}$.

Suppose $\overline{K}$ is not compact. That is, there is a cover $A_i \in \mathcal{O}, i \in I$ (for some infinite I) with $\overline{K} \subseteq \bigcup_{i\in I} A_i$ and there is no finite subcover. We set for $i \in I$

$$J_i = \{j \in \mathbb{N} : B_j \subseteq A_i\} \subseteq \mathbb{N}$$

and $J := \bigcup_{i\in I} J_i \subseteq \mathbb{N}$. Thus $A_i = \bigcup_{j\in J_i} B_j$, and

$$\overline{K} \subseteq \bigcup_{i\in I} A_i = \bigcup_{i\in I} \bigcup_{j\in J_i} B_j = \bigcup_{j\in J} B_j.$$

This shows that $B_j \in \mathcal{O}, j \in J$ is a countable cover of $\overline{K}$. Since there is no finite subcover for $A_i, i \in I$, there can also be no finite finite subcover for $B_j, j \in J$. (If there would be a finite subcover $B_j, j \in J$, we could take $A_j \supseteq B_j, j \in J$ and find a finite subcover $A_j, j \in J$, contradiction.) We write $J = \{j_1, j_2, ...\}$. For $n \in \mathbb{N}$ we set $\omega_n \in \overline{K} \setminus \bigcup_{i=1}^n B_{j_i}$. (Note that this set is non-empty, otherwise a finite subcover would exist.) By assumption, the sequence $\omega_1, \omega_2, \ldots \in \overline{K}$ has a cluster point $\omega \in \overline{K}$. Since $\overline{K} \subseteq \bigcup_{j \in J} B_j$, there is $k \in J \subseteq \mathbb{N}$ with $\omega \in B_k$. So, on the one hand (since B_k is open) there are infinitely many of the ω_n in B_k, on the other hand, $\omega_i \notin B_k$ for all $i \geq k$ by construction. This is a contradiction, so $\overline{K}$ is compact.

'4.$\Leftarrow$'3. If (Ω, r) is complete: Let $\omega_1, \omega_2, \ldots \in K$. We are going to construct a Cauchy subsequence. This converges since (Ω, r) is complete and K is found to be relatively sequentially compact. In order to construct the subsequence, choose a sequence $\varepsilon_1, \varepsilon_2, \ldots > 0$ with $\varepsilon_n \downarrow 0$. Since K is totally bounded, there are finitely many ε_1-balls covering K. At least one of these balls must contain infinitely many of the ω_n. These have each at most distance $2\varepsilon_1$. Choose ω_{k_1} as one of these infinitely many points. Since this ε_1-ball is covered by finitely many ε_2-balls, there is one of these ε_2-balls, which contains infinitely many of the ω_n. These each have at most distance $2\varepsilon_2$. Choose $\omega_{k_2} \neq \omega_{k_1}$ as one of these infinitely many points. By proceeding further we obtain a sequence $\omega_{k_1}, \omega_{k_2}, \ldots \in K$ such that $r(\omega_{k_n}, \omega_{k_m}) \leq 2\varepsilon_{m \wedge n}$. With other words, as announced, we have found a Cauchy subsequence in K. □

Lemma A.11 (Compact metric spaces are Polish)**.** . *Let* (Ω, r) *be a metric space and* $\mathcal{O}$ *be the topology generated by* r*. If* Ω *is compact, then* $(\Omega, \mathcal{O})$ *is Polish.*

Proof. For the proof, we need to show both, completeness of (Ω, r) and separability of $(\Omega, \mathcal{O})$. For completeness, let $\omega_1, \omega_2, \ldots \in \Omega$ be a Cauchy sequence. Since K is relatively sequentially compact according to Proposition A.9, there is $\omega \in \Omega$ and a subsequence $\omega_{k_1}, \omega_{k_2}, \ldots$ converging to ω. Let $\varepsilon > 0$ and $N \in \mathbb{N}$ be such that $r(\omega_m, \omega_n) < \varepsilon/2$ for $m, n > N$ and $r(\omega_{k_n}, \omega) < \varepsilon/2$ for $k_n > N$. Then for $m > N$ it holds that $r(\omega_m, \omega) \leq r(\omega_m, \omega_{k_n}) + r(\omega_{k_n}, \omega) \leq \varepsilon$. It follows that $\omega_n \xrightarrow{n\to\infty} \omega$. For separability of $(\Omega, \mathcal{O})$, let $\varepsilon_1, \varepsilon_2, \ldots > 0$ with $\varepsilon_n \downarrow 0$. Since K is totally bounded, for all $n \in \mathbb{N}$ there is a k_n and $\omega_{n1}, \ldots, \omega_{nk_n}$ with $K \subseteq \bigcup_{k=1}^{k_n} B_{\varepsilon_n}(\omega_{nk_n})$. Let $\Omega' = \omega_{nk} : n \in \mathbb{N}, k = 1, \ldots, k_n\}$. Then Ω' is countable and for each $\omega \in \Omega$ and each $n \in \mathbb{N}$ there is a $k(\omega, n) \in \{1, \ldots, k_n\}$ with $r(\omega_{k(\omega,n)}, \omega) < \varepsilon_n$. Thus, $(\omega_{k(\omega,n)}, \omega) \xrightarrow{n\to\infty} \omega$. So $\overline{\Omega'} = \Omega$. □