Chapter 1
Basic Measure Theory

In this chapter, we introduce the classes of sets that allow for a systematic treatment of events and random observations in the framework of probability theory. Furthermore, we construct measures, in particular probability measures, on such classes of sets. Finally, we define random variables as measurable maps.

1.1 Classes of Sets

In the following, let $\Omega \neq \emptyset$ be a nonempty set and let $\mathcal{A} \subset 2^{\Omega}$ (set of all subsets of Ω) be a class of subsets of Ω. Later, Ω will be interpreted as the space of elementary events and $\mathcal{A}$ will be the system of observable events. In this section, we introduce names for classes of subsets of Ω that are stable under certain set operations and we establish simple relations between such classes.

Definition 1.1 A class of sets $\mathcal{A}$ is called

- $\cap$-closed (closed under intersections) or a π-*system* if $A \cap B \in \mathcal{A}$ whenever $A, B \in \mathcal{A}$,
- σ-$\cap$-closed (closed under countable[1] intersections) if $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$ for any choice of countably many sets $A_1, A_2, \ldots \in \mathcal{A}$,
- $\cup$-closed (closed under unions) if $A \cup B \in \mathcal{A}$ whenever $A, B \in \mathcal{A}$,
- σ-$\cup$-closed (closed under countable unions) if $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ for any choice of countably many sets $A_1, A_2, \ldots \in \mathcal{A}$,
- $\setminus$-closed (closed under differences) if $A \setminus B \in \mathcal{A}$ whenever $A, B \in \mathcal{A}$, and
- closed under complements if $A^c := \Omega \setminus A \in \mathcal{A}$ for any set $A \in \mathcal{A}$.

[1] By "countable" we always mean either finite or countably infinite.

A. Klenke, *Probability Theory*, Universitext, DOI 10.1007/978-1-4471-5361-0_1,

Definition 1.2 (σ-algebra) A class of sets $\mathcal{A} \subset 2^{\Omega}$ is called a σ*-algebra* if it fulfills the following three conditions:

(i) $\Omega \in \mathcal{A}$.
(ii) $\mathcal{A}$ is closed under complements.
(iii) $\mathcal{A}$ is closed under countable unions.

Sometimes a σ-algebra is also named a σ-field. As we will see, we can define probabilities on σ-algebras in a consistent way. Hence these are the natural classes of sets to be considered as *events* in probability theory.

Theorem 1.3 *If $\mathcal{A}$ is closed under complements, then we have the equivalences*

$$\begin{aligned} \mathcal{A} \textit{ is } \cap\textit{-closed} &\iff \mathcal{A} \textit{ is } \cup\textit{-closed}, \\ \mathcal{A} \textit{ is } \sigma\text{-}\cap\textit{-closed} &\iff \mathcal{A} \textit{ is } \sigma\text{-}\cup\textit{-closed}. \end{aligned}$$

Proof The two statements are immediate consequences of de Morgan's rule (reminder: $(\bigcup A_i)^c = \bigcap A_i^c$). For example, let $\mathcal{A}$ be σ-$\cap$-closed and let $A_1, A_2, \ldots \in \mathcal{A}$. Hence

$$\bigcup_{n=1}^{\infty} A_n = \left(\bigcap_{n=1}^{\infty} A_n^c \right)^c \in \mathcal{A}.$$

Thus $\mathcal{A}$ is σ-$\cup$-closed. The other cases can be proved similarly. □

Theorem 1.4 *Assume that $\mathcal{A}$ is $\setminus$-closed. Then the following statements hold*:

(i) *$\mathcal{A}$ is $\cap$-closed.*
(ii) *If in addition $\mathcal{A}$ is σ-$\cup$-closed, then $\mathcal{A}$ is σ-$\cap$-closed.*
(iii) *Any countable (respectively finite) union of sets in $\mathcal{A}$ can be expressed as a countable (respectively finite) disjoint union of sets in $\mathcal{A}$.*

Proof (i) Assume that $A, B \in \mathcal{A}$. Hence also $A \cap B = A \setminus (A \setminus B) \in \mathcal{A}$.

(ii) Assume that $A_1, A_2, \ldots \in \mathcal{A}$. Hence

$$\bigcap_{n=1}^{\infty} A_n = \bigcap_{n=2}^{\infty} (A_1 \cap A_n) = \bigcap_{n=2}^{\infty} A_1 \setminus (A_1 \setminus A_n) = A_1 \setminus \bigcup_{n=2}^{\infty} (A_1 \setminus A_n) \in \mathcal{A}.$$

(iii) Assume that $A_1, A_2, \ldots \in \mathcal{A}$. Hence a representation of $\bigcup_{n=1}^{\infty} A_n$ as a countable disjoint union of sets in $\mathcal{A}$ is

$$\bigcup_{n=1}^{\infty} A_n = A_1 \uplus (A_2 \setminus A_1) \uplus \big((A_3 \setminus A_1) \setminus A_2\big) \uplus \Big(\big((A_4 \setminus A_1) \setminus A_2\big) \setminus A_3\Big) \uplus \ldots.$$

□

Remark 1.5 Sometimes the disjoint union of sets is denoted by the symbol $\biguplus$. Note that this is not a new operation but only stresses the fact that the sets involved are mutually disjoint. ◊

Definition 1.6 A class of sets $\mathcal{A} \subset 2^{\Omega}$ is called an *algebra* if the following three conditions are fulfilled:

(i) $\Omega \in \mathcal{A}$.
(ii) $\mathcal{A}$ is $\setminus$-closed.
(iii) $\mathcal{A}$ is $\cup$-closed.

If $\mathcal{A}$ is an algebra, then obviously $\emptyset = \Omega \setminus \Omega$ is in $\mathcal{A}$. However, in general, this property is weaker than (i) in Definition 1.6.

Theorem 1.7 *A class of sets $\mathcal{A} \subset 2^{\Omega}$ is an algebra if and only if the following three properties hold*:

(i) $\Omega \in \mathcal{A}$.
(ii) $\mathcal{A}$ *is closed under complements*.
(iii) $\mathcal{A}$ *is closed under intersections*.

Proof This is left as an exercise. □

Definition 1.8 A class of sets $\mathcal{A} \subset 2^{\Omega}$ is called a *ring* if the following three conditions hold:

(i) $\emptyset \in \mathcal{A}$.
(ii) $\mathcal{A}$ is $\setminus$-closed.
(iii) $\mathcal{A}$ is $\cup$-closed.

A ring is called a σ-*ring* if it is also σ-$\cup$-closed.

Definition 1.9 A class of sets $\mathcal{A} \subset 2^{\Omega}$ is called a *semiring* if

(i) $\emptyset \in \mathcal{A}$,
(ii) for any two sets $A, B \in \mathcal{A}$ the difference set $B \setminus A$ is a finite union of mutually disjoint sets in $\mathcal{A}$,
(iii) $\mathcal{A}$ is $\cap$-closed.

Definition 1.10 A class of sets $\mathcal{A} \subset 2^{\Omega}$ is called a λ-*system* (or Dynkin's λ-system) if

(i) $\Omega \in \mathcal{A}$,
(ii) for any two sets $A, B \in \mathcal{A}$ with $A \subset B$, the difference set $B \setminus A$ is in $\mathcal{A}$, and
(iii) $\biguplus_{n=1}^{\infty} A_n \in \mathcal{A}$ for any choice of countably many pairwise disjoint sets $A_1, A_2, \ldots \in \mathcal{A}$.

Example 1.11

(i) For any nonempty set Ω, the classes $\mathcal{A} = \{\emptyset, \Omega\}$ and $\mathcal{A} = 2^{\Omega}$ are the trivial examples of algebras, σ-algebras and λ-systems. On the other hand, $\mathcal{A} = \{\emptyset\}$ and $\mathcal{A} = 2^{\Omega}$ are the trivial examples of semirings, rings and σ-rings.
(ii) Let $\Omega = \mathbb{R}$. Then $\mathcal{A} = \{A \subset \mathbb{R} : A \text{ is countable}\}$ is a σ-ring.

(iii) $\mathcal{A} = \{(a, b] : a, b \in \mathbb{R},\ a \le b\}$ is a semiring on $\Omega = \mathbb{R}$ (but is not a ring).
(iv) The class of finite unions of bounded intervals is a ring on $\Omega = \mathbb{R}$ (but is not an algebra).
(v) The class of finite unions of arbitrary (also unbounded) intervals is an algebra on $\Omega = \mathbb{R}$ (but is not a σ-algebra).
(vi) Let E be a finite nonempty set and let $\Omega := E^{\mathbb{N}}$ be the set of all E-valued sequences $\omega = (\omega_n)_{n\in\mathbb{N}}$. For any $\omega_1, \ldots, \omega_n \in E$, let

$$[\omega_1, \ldots, \omega_n] := \left\{\omega' \in \Omega : \omega'_i = \omega_i \text{ for all } i = 1, \ldots, n\right\}$$

be the set of all sequences whose first n values are $\omega_1, \ldots, \omega_n$. Let $\mathcal{A}_0 = \{\emptyset\}$. For $n \in \mathbb{N}$, define

$$\mathcal{A}_n := \left\{[\omega_1, \ldots, \omega_n] : \omega_1, \ldots, \omega_n \in E\right\}. \tag{1.1}$$

Hence $\mathcal{A} := \bigcup_{n=0}^{\infty} \mathcal{A}_n$ is a semiring but is not a ring (if $\#E > 1$).
(vii) Let Ω be an arbitrary nonempty set. Then

$$\mathcal{A} := \left\{A \subset \Omega : A \text{ or } A^c \text{ is finite}\right\}$$

is an algebra. However, if $\#\Omega = \infty$, then $\mathcal{A}$ is not a σ-algebra.
(viii) Let Ω be an arbitrary nonempty set. Then

$$\mathcal{A} := \left\{A \subset \Omega : A \text{ or } A^c \text{ is countable}\right\}$$

is a σ-algebra.
(ix) Every σ-algebra is a λ-system.
(x) Let $\Omega = \{1, 2, 3, 4\}$ and $\mathcal{A} = \{\emptyset, \{1, 2\}, \{1, 4\}, \{2, 3\}, \{3, 4\}, \{1, 2, 3, 4\}\}$. Hence $\mathcal{A}$ is a λ-system but is not an algebra. ◊

Theorem 1.12 (Relations between classes of sets)

(i) *Every σ-algebra also is a λ-system, an algebra and a σ-ring.*
(ii) *Every σ-ring is a ring, and every ring is a semiring.*
(iii) *Every algebra is a ring. An algebra on a finite set Ω is a σ-algebra.*

Proof (i) This is obvious.

(ii) Let $\mathcal{A}$ be a ring. By Theorem 1.4, $\mathcal{A}$ is closed under intersections and is hence a semiring.

(iii) Let $\mathcal{A}$ be an algebra. Then $\emptyset = \Omega \setminus \Omega \in \mathcal{A}$, and hence $\mathcal{A}$ is a ring. If in addition Ω is finite, then $\mathcal{A}$ is finite. Hence any countable union of sets in $\mathcal{A}$ is a finite union of sets. □

Definition 1.13 (liminf and limsup) Let $A_1, A_2, \ldots$ be subsets of Ω. The sets

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m \quad \text{and} \quad \limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m$$

are called *limes inferior* and *limes superior*, respectively, of the sequence $(A_n)_{n\in\mathbb{N}}$.

Remark 1.14

(i) lim inf and lim sup can be rewritten as

$$\liminf_{n\to\infty} A_n = \big\{\omega \in \Omega : \#\{n \in \mathbb{N} : \omega \notin A_n\} < \infty\big\},$$
$$\limsup_{n\to\infty} A_n = \big\{\omega \in \Omega : \#\{n \in \mathbb{N} : \omega \in A_n\} = \infty\big\}.$$

In other words, limes inferior is the event where *eventually all* of the A_n occur. On the other hand, limes superior is the event where *infinitely many* of the A_n occur. In particular, $A_* := \liminf_{n\to\infty} A_n \subset A^* := \limsup_{n\to\infty} A_n$.

(ii) We define the *indicator function* on the set A by

$$\mathbb{1}_A(x) := \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases} \tag{1.2}$$

With this notation,

$$\mathbb{1}_{A_*} = \liminf_{n\to\infty} \mathbb{1}_{A_n} \quad \text{and} \quad \mathbb{1}_{A^*} = \limsup_{n\to\infty} \mathbb{1}_{A_n}.$$

(iii) If $\mathcal{A} \subset 2^\Omega$ is a σ-algebra and if $A_n \in \mathcal{A}$ for every $n \in \mathbb{N}$, then $A_* \in \mathcal{A}$ and $A^* \in \mathcal{A}$. ◊

Proof This is left as an exercise. □

Theorem 1.15 (Intersection of classes of sets) *Let I be an arbitrary index set, and assume that $\mathcal{A}_i$ is a σ-algebra for every $i \in I$. Hence the intersection*

$$\mathcal{A}_I := \{A \subset \Omega : A \in \mathcal{A}_i \text{ for every } i \in I\} = \bigcap_{i\in I} \mathcal{A}_i$$

is a σ-algebra. The analogous statement holds for rings, σ-rings, algebras and λ-systems. However, it fails for semirings.

Proof We give the proof for σ-algebras only. To this end, we check (i)–(iii) of Definition 1.2.

(i) Clearly, $\Omega \in \mathcal{A}_i$ for every $i \in I$, and hence $\Omega \in \mathcal{A}_I$.
(ii) Assume $A \in \mathcal{A}_I$. Hence $A \in \mathcal{A}_i$ for any $i \in I$. Thus also $A^c \in \mathcal{A}_i$ for any $i \in I$. We conclude that $A^c \in \mathcal{A}_I$.
(iii) Assume $A_1, A_2, \ldots \in \mathcal{A}_I$. Hence $A_n \in \mathcal{A}_i$ for every $n \in \mathbb{N}$ and $i \in I$. Thus $A := \bigcup_{n=1}^\infty A_n \in \mathcal{A}_i$ for every $i \in I$. We conclude $A \in \mathcal{A}_I$.

Counterexample for semirings: Let $\Omega = \{1, 2, 3, 4\}$, $\mathcal{A}_1 = \{\emptyset, \Omega, \{1\}, \{2, 3\}, \{4\}\}$ and $\mathcal{A}_2 = \{\emptyset, \Omega, \{1\}, \{2\}, \{3, 4\}\}$. Then $\mathcal{A}_1$ and $\mathcal{A}_2$ are semirings but $\mathcal{A}_1 \cap \mathcal{A}_2 = \{\emptyset, \Omega, \{1\}\}$ is not. □

Theorem 1.16 (Generated σ-algebra) *Let $\mathcal{E} \subset 2^{\Omega}$. Then there exists a smallest σ-algebra $\sigma(\mathcal{E})$ with $\mathcal{E} \subset \sigma(\mathcal{E})$:*

$$\sigma(\mathcal{E}) := \bigcap_{\substack{\mathcal{A} \subset 2^{\Omega} \text{ is a } \sigma\text{-algebra} \\ \mathcal{A} \supset \mathcal{E}}} \mathcal{A}.$$

$\sigma(\mathcal{E})$ is called the σ-algebra generated by *$\mathcal{E}$. $\mathcal{E}$ is called a* generator *of $\sigma(\mathcal{E})$. Similarly, we define $\delta(\mathcal{E})$ as the λ-system generated by $\mathcal{E}$.*

Proof $\mathcal{A} = 2^{\Omega}$ is a σ-algebra with $\mathcal{E} \subset \mathcal{A}$. Hence the intersection is nonempty. By Theorem 1.15, $\sigma(\mathcal{E})$ is a σ-algebra. Clearly, it is the smallest σ-algebra that contains $\mathcal{E}$. For λ-systems the proof is similar. □

Remark 1.17 The following three statements hold:

(i) $\mathcal{E} \subset \sigma(\mathcal{E})$.
(ii) If $\mathcal{E}_1 \subset \mathcal{E}_2$, then $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_2)$.
(iii) $\mathcal{A}$ is a σ-algebra if and only if $\sigma(\mathcal{A}) = \mathcal{A}$.

The same statements hold for λ-systems. Furthermore, $\delta(\mathcal{E}) \subset \sigma(\mathcal{E})$. ◊

Theorem 1.18 ($\cap$-closed λ-system) *Let $\mathcal{D} \subset 2^{\Omega}$ be a λ-system. Then*

$$\mathcal{D} \text{ is a } \pi\text{-system} \quad \Longleftrightarrow \quad \mathcal{D} \text{ is a } \sigma\text{-algebra}.$$

Proof "$\Longleftarrow$" This is obvious.
"$\Longrightarrow$" We check (i)–(iii) of Definition 1.2.

(i) Clearly, $\Omega \in \mathcal{D}$.
(ii) (Closedness under complements) Let $A \in \mathcal{D}$. Since $\Omega \in \mathcal{D}$ and by property (ii) of the λ-system, we get that $A^c = \Omega \setminus A \in \mathcal{D}$.
(iii) (σ-$\cup$-closedness) Let $A, B \in \mathcal{D}$. By assumption, $A \cap B \in \mathcal{D}$, and trivially $A \cap B \subset A$. Thus $A \setminus B = A \setminus (A \cap B) \in \mathcal{D}$. This implies that $\mathcal{D}$ is $\setminus$-closed. Now let $A_1, A_2, \ldots \in \mathcal{D}$. By Theorem 1.4(iii), there exist mutually disjoint sets $B_1, B_2, \ldots \in \mathcal{D}$ with $\bigcup_{n=1}^{\infty} A_n = \biguplus_{n=1}^{\infty} B_n \in \mathcal{D}$. □

Theorem 1.19 (Dynkin's π-λ theorem) *If $\mathcal{E} \subset 2^{\Omega}$ is a π-system, then*

$$\sigma(\mathcal{E}) = \delta(\mathcal{E}).$$

Proof "$\supset$" This follows from Remark 1.17.
"$\subset$" We have to show that $\delta(\mathcal{E})$ is a σ-algebra. By Theorem 1.18, it is enough to show that $\delta(\mathcal{E})$ is a π-system. For any $B \in \delta(\mathcal{E})$ define

$$\mathcal{D}_B := \left\{ A \in \delta(\mathcal{E}) : A \cap B \in \delta(\mathcal{E}) \right\}.$$

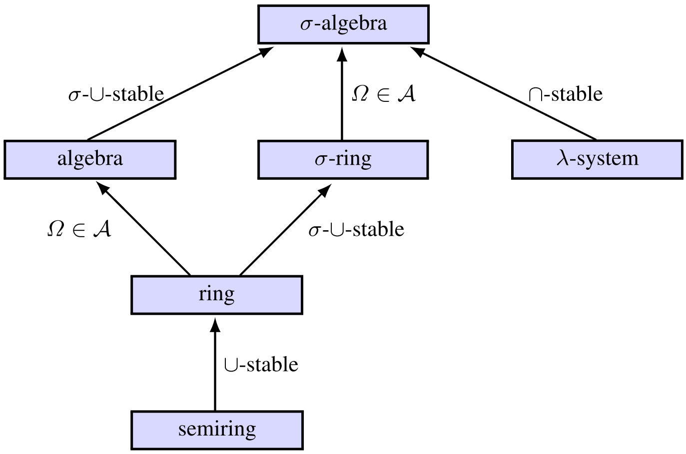

Fig. 1.1 Inclusions between classes of sets $\mathcal{A} \subset 2^{\Omega}$

In order to show that $\delta(\mathcal{E})$ is a π-system, it is enough to show that

$$\delta(\mathcal{E}) \subset \mathcal{D}_B \quad \text{for any } B \in \delta(\mathcal{E}). \tag{1.3}$$

In order to show that $\mathcal{D}_E$ is a λ-system for any $E \in \delta(\mathcal{E})$, we check (i)–(iii) of Definition 1.10:

(i) Clearly, $\Omega \cap E = E \in \delta(\mathcal{E})$; hence $\Omega \in \mathcal{D}_E$.
(ii) For any $A, B \in \mathcal{D}_E$ with $A \subset B$, we have $(B \setminus A) \cap E = (B \cap E) \setminus (A \cap E) \in \delta(\mathcal{E})$.
(iii) Assume that $A_1, A_2, \ldots \in \mathcal{D}_E$ are mutually disjoint. Hence

$$\left(\bigcup_{n=1}^{\infty} A_n\right) \cap E = \biguplus_{n=1}^{\infty} (A_n \cap E) \in \delta(\mathcal{E}).$$

By assumption, $A \cap E \in \mathcal{E}$ if $A, E \in \mathcal{E}$; thus $\mathcal{E} \subset \mathcal{D}_E$ if $E \in \mathcal{E}$. By Remark 1.17(ii), we conclude that $\delta(\mathcal{E}) \subset \mathcal{D}_E$ for any $E \in \mathcal{E}$. Hence we get that $B \cap E \in \delta(\mathcal{E})$ for any $B \in \delta(\mathcal{E})$ and $E \in \mathcal{E}$. This implies that $E \in \mathcal{D}_B$ for any $B \in \delta(\mathcal{E})$. Thus $\mathcal{E} \subset \mathcal{D}_B$ for any $B \in \delta(\mathcal{E})$, and hence (1.3) follows. □

For an illustration of the inclusions between the classes of sets, see Fig. 1.1.

We are particularly interested in σ-algebras that are generated by topologies. The most prominent role is played by the Euclidean space $\mathbb{R}^n$; however, we will also consider the (infinite-dimensional) space $C([0,1])$ of continuous functions $[0,1] \to \mathbb{R}$. On $C([0,1])$ the norm $\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|$ induces a topology. For the convenience of the reader, we recall the definition of a topology.

Definition 1.20 (Topology) Let $\Omega \neq \emptyset$ be an arbitrary set. A class of sets $\tau \subset 2^{\Omega}$ is called a *topology* on Ω if it has the following three properties:

(i) $\emptyset, \Omega \in \tau$.
(ii) $A \cap B \in \tau$ for any $A, B \in \tau$.
(iii) $(\bigcup_{A \in \mathcal{F}} A) \in \tau$ for any $\mathcal{F} \subset \tau$.

The pair (Ω, τ) is called a *topological space*. The sets $A \in \tau$ are called *open*, and the sets $A \subset \Omega$ with $A^c \in \tau$ are called *closed*.

In contrast with σ-algebras, topologies are closed under finite intersections only, but they are also closed under arbitrary unions.

Let d be a metric on Ω, and denote the open ball with radius $r > 0$ centered at $x \in \Omega$ by

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

Then the usual class of open sets is the topology

$$\tau = \left\{\bigcup_{(x,r) \in F} B_r(x) : F \subset \Omega \times (0, \infty)\right\}.$$

Definition 1.21 (Borel σ-algebra) Let (Ω, τ) be a topological space. The σ-algebra

$$\mathcal{B}(\Omega) := \mathcal{B}(\Omega, \tau) := \sigma(\tau)$$

that is generated by the open sets is called the *Borel σ-algebra* on Ω. The elements $A \in \mathcal{B}(\Omega, \tau)$ are called *Borel sets* or *Borel measurable sets*.

Remark 1.22 In many cases, we are interested in $\mathcal{B}(\mathbb{R}^n)$, where $\mathbb{R}^n$ is equipped with the Euclidean distance

$$d(x, y) = \|x - y\|_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}.$$

(i) There are subsets of $\mathbb{R}^n$ that are not Borel sets. These sets are not easy to construct like, for example, *Vitali sets* that can be found in calculus books (see also [37, Theorem 3.4.4]). Here we do not want to stress this point but state that, vaguely speaking, all sets that can be constructed explicitly are Borel sets.
(ii) If $C \subset \mathbb{R}^n$ is a closed set, then $C^c \in \tau$ is in $\mathcal{B}(\mathbb{R}^n)$ and hence C is a Borel set. In particular, $\{x\} \in \mathcal{B}(\mathbb{R}^n)$ for every $x \in \mathbb{R}^n$.
(iii) $\mathcal{B}(\mathbb{R}^n)$ is not a topology. To show this, let $V \subset \mathbb{R}^n$ such that $V \notin \mathcal{B}(\mathbb{R}^n)$. If $\mathcal{B}(\mathbb{R}^n)$ were a topology, then it would be closed under arbitrary unions. As $\{x\} \in \mathcal{B}(\mathbb{R}^n)$ for all $x \in \mathbb{R}^n$, we would get the contradiction $V = \bigcup_{x \in V} \{x\} \in \mathcal{B}(\mathbb{R}^n)$. ◊

In most cases the class of open sets that generates the Borel σ-algebra is too big to work with efficiently. Hence we aim at finding smaller (in particular, countable) classes of sets that generate the Borel σ-algebra and that are more amenable. In some of the examples, the elements of the generating class are simpler sets such as rectangles or compact sets.

We introduce the following notation. We denote by $\mathbb{Q}$ the set of rational numbers and by $\mathbb{Q}^+$ the set of strictly positive rational numbers. For $a, b \in \mathbb{R}^n$, we write

$$a < b \quad \text{if } a_i < b_i \quad \text{for all } i = 1, \dots, n. \tag{1.4}$$

For $a < b$, we define the open *rectangle* as the Cartesian product

$$(a, b) := \bigtimes_{i=1}^{n} (a_i, b_i) := (a_1, b_1) \times (a_2, b_2) \times \ldots \times (a_n, b_n). \tag{1.5}$$

Analogously, we define $[a, b]$, $(a, b]$ and $[a, b)$. Furthermore, we define $(-\infty, b) := \bigtimes_{i=1}^{n}(-\infty, b_i)$, and use an analogous definition for $(-\infty, b]$ and so on. We introduce the following classes of sets:

$$\mathcal{E}_1 := \{A \subset \mathbb{R}^n : A \text{ is open}\}, \qquad \mathcal{E}_2 := \{A \subset \mathbb{R}^n : A \text{ is closed}\},$$

$$\mathcal{E}_3 := \{A \subset \mathbb{R}^n : A \text{ is compact}\}, \qquad \mathcal{E}_4 := \{B_r(x) : x \in \mathbb{Q}^n,\ r \in \mathbb{Q}^+\},$$

$$\mathcal{E}_5 := \{(a, b) : a, b \in \mathbb{Q}^n,\ a < b\}, \qquad \mathcal{E}_6 := \{[a, b) : a, b \in \mathbb{Q}^n,\ a < b\},$$

$$\mathcal{E}_7 := \{(a, b] : a, b \in \mathbb{Q}^n,\ a < b\}, \qquad \mathcal{E}_8 := \{[a, b] : a, b \in \mathbb{Q}^n,\ a < b\},$$

$$\mathcal{E}_9 := \{(-\infty, b) : b \in \mathbb{Q}^n\}, \qquad \mathcal{E}_{10} := \{(-\infty, b] : b \in \mathbb{Q}^n\},$$

$$\mathcal{E}_{11} := \{(a, \infty) : a \in \mathbb{Q}^n\}, \qquad \mathcal{E}_{12} := \{[a, \infty) : a \in \mathbb{Q}^n\}.$$

Theorem 1.23 *The Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$ is generated by any of the classes of sets $\mathcal{E}_1, \dots, \mathcal{E}_{12}$, that is, $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{E}_i)$ for any $i = 1, \dots, 12$.*

Proof We show only some of the identities.

(1) By definition, $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{E}_1)$.

(2) Let $A \in \mathcal{E}_1$. Then $A^c \in \mathcal{E}_2$, and hence $A = (A^c)^c \in \sigma(\mathcal{E}_2)$. It follows that $\mathcal{E}_1 \subset \sigma(\mathcal{E}_2)$. By Remark 1.17, this implies $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_2)$. Similarly, we obtain $\sigma(\mathcal{E}_2) \subset \sigma(\mathcal{E}_1)$ and hence equality.

(3) Any compact set is closed; hence $\sigma(\mathcal{E}_3) \subset \sigma(\mathcal{E}_2)$. Now let $A \in \mathcal{E}_2$. The sets $A_K := A \cap [-K, K]^n$, $K \in \mathbb{N}$, are compact; hence the countable union $A = \bigcup_{K=1}^{\infty} A_K$ is in $\sigma(\mathcal{E}_3)$. It follows that $\mathcal{E}_2 \subset \sigma(\mathcal{E}_3)$ and thus $\sigma(\mathcal{E}_2) = \sigma(\mathcal{E}_3)$.

(4) Clearly, $\mathcal{E}_4 \subset \mathcal{E}_1$; hence $\sigma(\mathcal{E}_4) \subset \sigma(\mathcal{E}_1)$. Now let $A \subset \mathbb{R}^n$ be an open set. For any $x \in A$, define $R(x) = \min(1, \sup\{r > 0 : B_r(x) \subset A\})$. Note that $R(x) > 0$, as A is open. Let $r(x) \in (R(x)/2, R(x)) \cap \mathbb{Q}$. For any $y \in A$ and $x \in (B_{R(y)/3}(y)) \cap \mathbb{Q}^n$, we have $R(x) \geq R(y) - \|x - y\|_2 > \frac{2}{3}R(y)$, and hence $r(x) > \frac{1}{3}R(y)$ and thus $y \in B_{r(x)}(x)$. It follows that $A = \bigcup_{x \in A \cap \mathbb{Q}^n} B_{r(x)}(x)$ is a countable union of sets from $\mathcal{E}_4$ and is hence in $\sigma(\mathcal{E}_4)$. We have shown that $\mathcal{E}_1 \subset \sigma(\mathcal{E}_4)$. By Remark 1.17, this implies $\sigma(\mathcal{E}_1) \subset \sigma(\mathcal{E}_4)$.

(5–12) Exhaustion arguments similar to that in (4) also work for rectangles. If in (4) we take open rectangles instead of open balls $B_r(x)$, we get $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{E}_5)$. For example, we have

$$\bigtimes_{i=1}^{n} [a_i, b_i) = \bigcap_{k=1}^{\infty} \bigtimes_{i=1}^{n} \left(a_i - \frac{1}{k}, b_i\right) \in \sigma(\mathcal{E}_5).$$

The other inclusions $\mathcal{E}_i \subset \sigma(\mathcal{E}_j)$ can be shown similarly. □

Remark 1.24 Any of the classes $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{E}_5, \ldots, \mathcal{E}_{12}$ (but not $\mathcal{E}_4$) is a π-system. Hence, the Borel σ-algebra equals the generated λ-system: $\mathcal{B}(\mathbb{R}^n) = \delta(\mathcal{E}_i)$ for $i = 1, 2, 3, 5, \ldots, 12$. In addition, the classes $\mathcal{E}_4, \ldots, \mathcal{E}_{12}$ are countable. This is a crucial property that will be needed later. ◊

Definition 1.25 (Trace of a class of sets) Let $\mathcal{A} \subset 2^{\Omega}$ be an arbitrary class of subsets of Ω and let $A \in 2^{\Omega} \setminus \{\emptyset\}$. The class

$$\mathcal{A}\big|_A := \{A \cap B : B \in \mathcal{A}\} \subset 2^A \tag{1.6}$$

is called the *trace* of $\mathcal{A}$ on A or the *restriction* of $\mathcal{A}$ to A.

Theorem 1.26 *Let $A \subset \Omega$ be a nonempty set and let $\mathcal{A}$ be a σ-algebra on Ω or any of the classes of Definitions 1.6–1.9. Then $\mathcal{A}\big|_A$ is a class of sets of the same type as $\mathcal{A}$; however, on A instead of Ω. For λ-systems this is not true in general.*

Proof This is left as an exercise. □

Exercise 1.1.1 Let $\mathcal{A}$ be a semiring. Show that any countable (respectively finite) union of sets in $\mathcal{A}$ can be written as a countable (respectively finite) *disjoint* union of sets in $\mathcal{A}$.

Exercise 1.1.2 Give a counterexample that shows that, in general, the union $\mathcal{A} \cup \mathcal{A}'$ of two σ-algebras need not be a σ-algebra.

Exercise 1.1.3 Let (Ω_1, d_1) and (Ω_2, d_2) be metric spaces and let $f : \Omega_1 \to \Omega_2$ be an arbitrary map. Denote by $U_f = \{x \in \Omega_1 : f \text{ is discontinuous at } x\}$ the set of points of discontinuity of f. Show that $U_f \in \mathcal{B}(\Omega_1)$.

Hint: First show that for any $\varepsilon > 0$ and $\delta > 0$ the set

$$U_f^{\delta,\varepsilon} := \big\{x \in \Omega_1 : \text{ there are } y, z \in B_\varepsilon(x) \text{ with } d_2\big(f(y), f(z)\big) > \delta\big\}$$

is open (where $B_\varepsilon(x) = \{y \in \Omega_1 : d_1(x, y) < \varepsilon\}$). Then construct U_f from such $U_f^{\delta,\varepsilon}$.

Exercise 1.1.4 Let Ω be an uncountably infinite set and $\mathcal{A} = \sigma(\{\omega\} : \omega \in \Omega)$. Show that

$$\mathcal{A} = \{A \subset \Omega : A \text{ is countable or } A^c \text{ is countable}\}.$$

Exercise 1.1.5 Let $\mathcal{A}$ be a ring on the set Ω. Show that $\mathcal{A}$ is an Abelian algebraic ring with multiplication "$\cap$" and addition "$\triangle$".

1.2 Set Functions

Definition 1.27 Let $\mathcal{A} \subset 2^\Omega$ and let $\mu : \mathcal{A} \to [0, \infty]$ be a set function. We say that μ is

(i) *monotone* if $\mu(A) \le \mu(B)$ for any two sets $A, B \in \mathcal{A}$ with $A \subset B$,
(ii) *additive* if $\mu(\biguplus_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$ for any choice of finitely many mutually disjoint sets $A_1, \ldots, A_n \in \mathcal{A}$ with $\bigcup_{i=1}^n A_i \in \mathcal{A}$,
(iii) *σ-additive* if $\mu(\biguplus_{i=1}^\infty A_i) = \sum_{i=1}^\infty \mu(A_i)$ for any choice of countably many mutually disjoint sets $A_1, A_2, \ldots \in \mathcal{A}$ with $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$,
(iv) *subadditive* if for any choice of finitely many sets $A, A_1, \ldots, A_n \in \mathcal{A}$ with $A \subset \bigcup_{i=1}^n A_i$, we have $\mu(A) \le \sum_{i=1}^n \mu(A_i)$, and
(v) *σ-subadditive* if for any choice of countably many sets $A, A_1, A_2, \ldots \in \mathcal{A}$ with $A \subset \bigcup_{i=1}^\infty A_i$, we have $\mu(A) \le \sum_{i=1}^\infty \mu(A_i)$.

Definition 1.28 Let $\mathcal{A}$ be a semiring and let $\mu : \mathcal{A} \to [0, \infty]$ be a set function with $\mu(\emptyset) = 0$. μ is called a

- *content* if μ is additive,
- *premeasure* if μ is σ-additive,
- *measure* if μ is a premeasure and $\mathcal{A}$ is a σ-algebra, and
- *probability measure* if μ is a measure and $\mu(\Omega) = 1$.

Definition 1.29 Let $\mathcal{A}$ be a semiring. A content μ on $\mathcal{A}$ is called

(i) *finite* if $\mu(A) < \infty$ for every $A \in \mathcal{A}$ and
(ii) *σ-finite* if there exists a sequence of sets $\Omega_1, \Omega_2, \ldots \in \mathcal{A}$ such that $\Omega = \bigcup_{n=1}^\infty \Omega_n$ and such that $\mu(\Omega_n) < \infty$ for all $n \in \mathbb{N}$.

Example 1.30 (Contents, measures)

(i) Let $\omega \in \Omega$ and $\delta_\omega(A) = \mathbb{1}_A(\omega)$ (see (1.2)). Then δ_ω is a probability measure on any σ-algebra $\mathcal{A} \subset 2^\Omega$. δ_ω is called the *Dirac measure* for the point ω.
(ii) Let Ω be a finite nonempty set. By

$$\mu(A) := \frac{\#A}{\#\Omega} \quad \text{for } A \subset \Omega,$$

we define a probability measure on $\mathcal{A} = 2^\Omega$. This μ is called the *uniform distribution* on Ω. For this distribution, we introduce the symbol $\mathcal{U}_\Omega := \mu$. The resulting triple $(\Omega, \mathcal{A}, \mathcal{U}_\Omega)$ is called a *Laplace space*.

(iii) Let Ω be countably infinite and let

$$\mathcal{A} := \big\{A \subset \Omega : \#A < \infty \text{ or } \#A^c < \infty\big\}.$$

Then $\mathcal{A}$ is an algebra. The set function μ on $\mathcal{A}$ defined by

$$\mu(A) = \begin{cases} 0, & A \text{ is finite}, \\ \infty, & A^c \text{ is finite}, \end{cases}$$

is a content but is not a premeasure. Indeed, $\mu\big(\bigcup_{\omega\in\Omega}\{\omega\}\big) = \mu(\Omega) = \infty$, but $\sum_{\omega\in\Omega} \mu(\{\omega\}) = 0$.

(iv) Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of measures (premeasures, contents) and let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative numbers. Then also $\mu := \sum_{n=1}^{\infty} \alpha_n \mu_n$ is a measure (premeasure, content).

(v) Let Ω be an (at most) countable nonempty set and let $\mathcal{A} = 2^\Omega$. Further, let $(p_\omega)_{\omega\in\Omega}$ be nonnegative numbers. Then $A \mapsto \mu(A) := \sum_{\omega\in A} p_\omega$ defines a σ-finite measure on 2^Ω. We call $p = (p_\omega)_{\omega\in\Omega}$ the *weight function* of μ. The number p_ω is called the weight of μ at point ω.

(vi) If in (v) the sum $\sum_{\omega\in\Omega} p_\omega$ equals one, then μ is a probability measure. In this case, we interpret p_ω as the probability of the elementary event ω. The vector $p = (p_\omega)_{\omega\in\Omega}$ is called a *probability vector*.

(vii) If in (v) $p_\omega = 1$ for every $\omega \in \Omega$, then μ is called *counting measure* on Ω. If Ω is finite, then so is μ.

(viii) Let $\mathcal{A}$ be the ring of finite unions of intervals $(a, b] \subset \mathbb{R}$. For $a_1 < b_1 < a_2 < b_2 < \ldots < b_n$ and $A = \biguplus_{i=1}^{n} (a_i, b_i]$, define

$$\mu(A) = \sum_{i=1}^{n} (b_i - a_i).$$

Then μ is a σ-finite content on $\mathcal{A}$ (even a premeasure) since $\bigcup_{n=1}^{\infty} (-n, n] = \mathbb{R}$ and $\mu((-n, n]) = 2n < \infty$ for all $n \in \mathbb{N}$.

(ix) Let $f : \mathbb{R} \to [0, \infty)$ be continuous. In a similar way to (viii), we define

$$\mu_f(A) = \sum_{i=1}^{n} \int_{a_i}^{b_i} f(x)\, dx.$$

Then μ_f is a σ-finite content on $\mathcal{A}$ (even a premeasure). The function f is called the *density* of μ and plays a role similar to the weight function p in (v). ◊

Lemma 1.31 (Properties of contents) *Let $\mathcal{A}$ be a semiring and let μ be a content on $\mathcal{A}$. Then the following statements hold.*

(i) *If $\mathcal{A}$ is a ring, then $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$ for any two sets $A, B \in \mathcal{A}$.*
(ii) *μ is monotone. If $\mathcal{A}$ is a ring, then $\mu(B) = \mu(A) + \mu(B \setminus A)$ for any two sets $A, B \in \mathcal{A}$ with $A \subset B$.*
(iii) *μ is subadditive. If μ is σ-additive, then μ is also σ-subadditive.*
(iv) *If $\mathcal{A}$ is a ring, then $\sum_{n=1}^{\infty} \mu(A_n) \le \mu(\bigcup_{n=1}^{\infty} A_n)$ for any choice of countably many mutually disjoint sets $A_1, A_2, \ldots \in \mathcal{A}$ with $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.*

Proof (i) Note that $A \cup B = A \uplus (B \setminus A)$ and $B = (A \cap B) \uplus (B \setminus A)$. As μ is additive, we obtain

$$\mu(A \cup B) = \mu(A) + \mu(B \setminus A) \quad \text{and} \quad \mu(B) = \mu(A \cap B) + \mu(B \setminus A).$$

This implies (i).

(ii) Let $A \subset B$. Since $A \cap B = A$, we obtain $\mu(B) = \mu(A \uplus (B \setminus A)) = \mu(A) + \mu(B \setminus A)$ if $B \setminus A \in \mathcal{A}$. In particular, this is true if $\mathcal{A}$ is a ring. If $\mathcal{A}$ is only a semiring, then there exists an $n \in \mathbb{N}$ and mutually disjoint sets $C_1, \ldots, C_n \in \mathcal{A}$ such that $B \setminus A = \biguplus_{i=1}^{n} C_i$. Hence $\mu(B) = \mu(A) + \sum_{i=1}^{n} \mu(C_i) \ge \mu(A)$ and thus μ is monotone.

(iii) Let $n \in \mathbb{N}$ and $A, A_1, \ldots, A_n \in \mathcal{A}$ with $A \subset \bigcup_{i=1}^{n} A_i$. Define $B_1 = A_1$ and

$$B_k = A_k \setminus \bigcup_{i=1}^{k-1} A_i = \bigcap_{i=1}^{k-1} \big(A_k \setminus (A_k \cap A_i)\big) \quad \text{for } k = 2, \ldots, n.$$

By the definition of a semiring, any $A_k \setminus (A_k \cap A_i)$ is a finite disjoint union of sets in $\mathcal{A}$. Hence there exists a $c_k \in \mathbb{N}$ and sets $C_{k,1}, \ldots, C_{k,c_k} \in \mathcal{A}$ such that $\biguplus_{i=1}^{c_k} C_{k,i} = B_k \subset A_k$. Similarly, there exist $d_k \in \mathbb{N}$ and $D_{k,1}, \ldots, D_{k,d_k} \in \mathcal{A}$ such that $A_k \setminus B_k = \biguplus_{i=1}^{d_k} D_{k,i}$. Since μ is additive, we have

$$\mu(A_k) = \sum_{i=1}^{c_k} \mu(C_{k,i}) + \sum_{i=1}^{d_k} \mu(D_{k,i}) \ge \sum_{i=1}^{c_k} \mu(C_{k,i}).$$

Again due to additivity and monotonicity, we get

$$\begin{aligned} \mu(A) &= \mu\left(\biguplus_{k=1}^{n} \biguplus_{i=1}^{c_k} (C_{k,i} \cap A) \right) = \sum_{k=1}^{n} \sum_{i=1}^{c_k} \mu(C_{k,i} \cap A) \\ &\le \sum_{k=1}^{n} \sum_{i=1}^{c_k} \mu(C_{k,i}) \le \sum_{k=1}^{n} \mu(A_k). \end{aligned}$$

Hence μ is subadditive. By a similar argument, σ-subadditivity follows from σ-additivity.

(iv) Let $\mathcal{A}$ be a ring and let $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$. Since μ is additive (and thus monotone), we have by (ii)

$$\sum_{n=1}^{m} \mu(A_n) = \mu\left(\biguplus_{n=1}^{m} A_n\right) \le \mu(A) \quad \text{for any } m \in \mathbb{N}.$$

It follows that $\sum_{n=1}^{\infty} \mu(A_n) \le \mu(A)$. □

Remark 1.32 The inequality in (iv) can be strict (see Example 1.30(iii)). In other words, there are contents that are not premeasures. ◊

Theorem 1.33 (Inclusion–exclusion formula) *Let $\mathcal{A}$ be a ring and let μ be a content on $\mathcal{A}$. Let $n \in \mathbb{N}$ and $A_1, \ldots, A_n \in \mathcal{A}$ such that $\mu(A_1 \cup \ldots \cup A_n) < \infty$. Then the following inclusion and exclusion formulas hold*:

$$\mu(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\{i_1,\ldots,i_k\} \subset \{1,\ldots,n\}} \mu(A_{i_1} \cap \ldots \cap A_{i_k}),$$

$$\mu(A_1 \cap \ldots \cap A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{\{i_1,\ldots,i_k\} \subset \{1,\ldots,n\}} \mu(A_{i_1} \cup \ldots \cup A_{i_k}).$$

Here summation is over all subsets of $\{1, \ldots, n\}$ *with* k *elements.*

Proof This is left as an exercise. *Hint:* Use induction on n. □

The next goal is to characterize σ-subadditivity by a certain continuity property (Theorem 1.36). To this end, we agree on the following conventions.

Definition 1.34 Let $A, A_1, A_2, \ldots$ be sets. We write

- $A_n \uparrow A$ and say that $(A_n)_{n\in\mathbb{N}}$ increases to A if $A_1 \subset A_2 \subset \ldots$ and $\bigcup_{n=1}^{\infty} A_n = A$, and
- $A_n \downarrow A$ and say that $(A_n)_{n\in\mathbb{N}}$ decreases to A if $A_1 \supset A_2 \supset A_3 \supset \ldots$ and $\bigcap_{n=1}^{\infty} A_n = A$.

Definition 1.35 (Continuity of contents) Let μ be a content on the ring $\mathcal{A}$.

(i) μ is called *lower semicontinuous* if $\mu(A_n) \stackrel{n\to\infty}{\longrightarrow} \mu(A)$ for any $A \in \mathcal{A}$ and any sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ with $A_n \uparrow A$.

(ii) μ is called *upper semicontinuous* if $\mu(A_n) \stackrel{n\to\infty}{\longrightarrow} \mu(A)$ for any $A \in \mathcal{A}$ and any sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{A}$ with $\mu(A_n) < \infty$ for some (and then eventually all) $n \in \mathbb{N}$ and $A_n \downarrow A$.

(iii) μ is called $\emptyset$-*continuous* if (ii) holds for $A = \emptyset$.

In the definition of upper semicontinuity, we needed the assumption $\mu(A_n) < \infty$ since otherwise we would not even have $\emptyset$-continuity for an example as simple as the counting measure μ on $(\mathbb{N}, 2^{\mathbb{N}})$. Indeed, $A_n := \{n, n+1, \ldots\} \downarrow \emptyset$ but $\mu(A_n) = \infty$ for all $n \in \mathbb{N}$.

Theorem 1.36 (Continuity and premeasure) *Let μ be a content on the ring $\mathcal{A}$. Consider the following five properties.*

(i) *μ is σ-additive* (*and hence a premeasure*).
(ii) *μ is σ-subadditive.*
(iii) *μ is lower semicontinuous.*
(iv) *μ is $\emptyset$-continuous.*
(v) *μ is upper semicontinuous.*

Then the following implications hold:

$$\text{(i)} \iff \text{(ii)} \iff \text{(iii)} \Longrightarrow \text{(iv)} \iff \text{(v)}.$$

If μ is finite, then we also have (iv) $\Longrightarrow$(iii).

Proof "(i) $\Longrightarrow$(ii)" Let $A, A_1, A_2, \ldots \in \mathcal{A}$ with $A \subset \bigcup_{i=1}^{\infty} A_i$. Define $B_1 = A_1$ and $B_n = A_n \setminus \bigcup_{i=1}^{n-1} A_i \in \mathcal{A}$ for $n = 2, 3, \ldots$. Then $A = \biguplus_{n=1}^{\infty} (A \cap B_n)$. Since μ is monotone and σ-additive, we infer

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A \cap B_n) \le \sum_{n=1}^{\infty} \mu(A_n).$$

Hence μ is σ-subadditive.

"(ii) $\Longrightarrow$(i)" This follows from Lemma 1.31(iv).

"(i) $\Longrightarrow$(iii)" Let μ be a premeasure and $A \in \mathcal{A}$. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{A}$ such that $A_n \uparrow A$ and let $A_0 = \emptyset$. Then

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

"(iii) $\Longrightarrow$(i)" Assume now that (iii) holds. Let $B_1, B_2, \ldots \in \mathcal{A}$ be mutually disjoint, and assume that $B = \biguplus_{n=1}^{\infty} B_n \in \mathcal{A}$. Define $A_n = \bigcup_{i=1}^{n} B_i$ for all $n \in \mathbb{N}$. Then it follows from (iii) that

$$\mu(B) = \lim_{n\to\infty} \mu(A_n) = \sum_{i=1}^{\infty} \mu(B_i).$$

Hence μ is σ-additive and therefore a premeasure.

"(iv) $\Longrightarrow$(v)" Let $A, A_1, A_2, \ldots \in \mathcal{A}$ with $A_n \downarrow A$ and $\mu(A_1) < \infty$. Define $B_n = A_n \setminus A \in \mathcal{A}$ for all $n \in \mathbb{N}$. Then $B_n \downarrow \emptyset$. This implies $\mu(A_n) - \mu(A) = \mu(B_n) \stackrel{n\to\infty}{\longrightarrow} 0$.

"(v) $\Longrightarrow$(iv)" This is evident.

"(iii) $\Longrightarrow$(iv)" Let $A_1, A_2, \ldots \in \mathcal{A}$ with $A_n \downarrow \emptyset$ and $\mu(A_1) < \infty$. Then $A_1 \setminus A_n \in \mathcal{A}$ for any $n \in \mathbb{N}$ and $A_1 \setminus A_n \uparrow A_1$. Hence

$$\mu(A_1) = \lim_{n\to\infty} \mu(A_1 \setminus A_n) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n).$$

Since $\mu(A_1) < \infty$, we have $\lim_{n\to\infty} \mu(A_n) = 0$.

"(iv) $\Longrightarrow$(iii)" (for finite μ) Assume that $\mu(A) < \infty$ for every $A \in \mathcal{A}$ and that μ is $\emptyset$-continuous. Let $A, A_1, A_2, \ldots \in \mathcal{A}$ with $A_n \uparrow A$. Then we have $A \setminus A_n \downarrow \emptyset$ and

$$\mu(A) - \mu(A_n) = \mu(A \setminus A_n) \overset{n\to\infty}{\longrightarrow} 0.$$

Hence (iii) follows. □

Example 1.37 (Compare Example 1.30(iii)) Let Ω be a countable set, and define

$$\mathcal{A} = \big\{A \subset \Omega : \#A < \infty \text{ or } \#A^c < \infty\big\},$$

$$\mu(A) = \begin{cases} 0, & A \text{ is finite,} \\ \infty, & A \text{ is infinite.} \end{cases}$$

Then μ is an $\emptyset$-continuous content but not a premeasure. ◊

Definition 1.38

(i) A pair $(\Omega, \mathcal{A})$ consisting of a nonempty set Ω and a σ-algebra $\mathcal{A} \subset 2^\Omega$ is called a *measurable space*. The sets $A \in \mathcal{A}$ are called *measurable sets*. If Ω is at most countably infinite and if $\mathcal{A} = 2^\Omega$, then the measurable space $(\Omega, 2^\Omega)$ is called *discrete*.

(ii) A triple $(\Omega, \mathcal{A}, \mu)$ is called a *measure space* if $(\Omega, \mathcal{A})$ is a measurable space and if μ is a measure on $\mathcal{A}$.

(iii) If in addition $\mu(\Omega) = 1$, then $(\Omega, \mathcal{A}, \mu)$ is called a *probability space*. In this case, the sets $A \in \mathcal{A}$ are called *events*.

(iv) The set of all finite measures on $(\Omega, \mathcal{A})$ is denoted by $\mathcal{M}_f(\Omega) := \mathcal{M}_f(\Omega, \mathcal{A})$. The subset of probability measures is denoted by $\mathcal{M}_1(\Omega) := \mathcal{M}_1(\Omega, \mathcal{A})$. Finally, the set of σ-finite measures on $(\Omega, \mathcal{A})$ is denoted by $\mathcal{M}_\sigma(\Omega, \mathcal{A})$.

Exercise 1.2.1 Let $\mathcal{A} = \{(a, b] \cap \mathbb{Q} : a, b \in \mathbb{R},\ a \le b\}$. Define $\mu : \mathcal{A} \to [0, \infty)$ by $\mu((a, b] \cap \mathbb{Q}) = b - a$. Show that $\mathcal{A}$ is a semiring and μ is a content on $\mathcal{A}$ that is lower and upper semicontinuous but is not σ-additive.

1.3 The Measure Extension Theorem

In this section, we construct measures μ on σ-algebras. The starting point will be to define the values of μ on a smaller class of sets; that is, on a semiring. Under a mild consistency condition, the resulting set function can be extended to the whole σ-algebra.

Before we develop the complete theory, we begin with two examples.

Example 1.39 (Lebesgue measure) Let $n \in \mathbb{N}$ and let

$$\mathcal{A} = \big\{(a, b] : a, b \in \mathbb{R}^n,\ a < b\big\}$$

be the semiring of half open rectangles $(a, b] \subset \mathbb{R}^n$ (see (1.5)). The n-dimensional volume of such a rectangle is

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

Can we extend the set function μ to a (uniquely determined) measure on the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{A})$? We will see that this is indeed possible. The resulting measure is called Lebesgue measure (or sometimes Lebesgue–Borel measure) λ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. ◊

Example 1.40 (Product measure, Bernoulli measure) We construct a measure for an infinitely often repeated random experiment with finitely many possible outcomes. Let E be the set of possible outcomes. For $e \in E$, let $p_e \geq 0$ be the probability that e occurs. Hence $\sum_{e \in E} p_e = 1$. For a fixed realization of the repeated experiment, let $\omega_1, \omega_2, \ldots \in E$ be the observed outcomes. Hence the space of *all* possible outcomes of the repeated experiment is $\Omega = E^{\mathbb{N}}$. As in Example 1.11(vi), we define the set of all sequences whose first n values are $\omega_1, \ldots, \omega_n$:

$$[\omega_1, \ldots, \omega_n] := \big\{\omega' \in \Omega : \omega'_i = \omega_i \text{ for any } i = 1, \ldots, n\big\}. \tag{1.7}$$

Let $\mathcal{A}_0 = \{\emptyset\}$. For $n \in \mathbb{N}$, define the class of cylinder sets that depend only on the first n coordinates

$$\mathcal{A}_n := \big\{[\omega_1, \ldots, \omega_n] : \omega_1, \ldots, \omega_n \in E\big\}, \tag{1.8}$$

and let $\mathcal{A} := \bigcup_{n=0}^{\infty} \mathcal{A}_n$.

We interpret $[\omega_1, \ldots, \omega_n]$ as the event where the outcome of the first experiment is ω_1, the outcome of the second experiment is ω_2 and finally the outcome of the nth experiment is ω_n. The outcomes of the other experiments do not play a role for the occurrence of this event. As the individual experiments ought to be independent, we should have for any choice $\omega_1, \ldots, \omega_n \in E$ that the probability of the event $[\omega_1, \ldots, \omega_n]$ is the product of the probabilities of the individual events; that is,

$$\mu\big([\omega_1, \ldots, \omega_n]\big) = \prod_{i=1}^{n} p_{\omega_i}.$$

This formula defines a content μ on the semiring $\mathcal{A}$, and our aim is to extend μ in a unique way to a probability measure on the σ-algebra $\sigma(\mathcal{A})$ that is generated by $\mathcal{A}$.

Before we do so, we make the following definition. Define the (ultra-)metric d on Ω by

$$d\big(\omega, \omega'\big) = \begin{cases} 2^{-\inf\{n \in \mathbb{N}: \omega_n \neq \omega'_n\}}, & \omega \neq \omega', \\ 0, & \omega = \omega'. \end{cases} \tag{1.9}$$

Hence (Ω, d) is a compact metric space. Clearly,

$$[\omega_1, \ldots, \omega_n] = B_{2^{-n}}(\omega) = \big\{\omega' \in \Omega : d\big(\omega, \omega'\big) < 2^{-n}\big\}.$$

The complement of $[\omega_1, \ldots, \omega_n]$ is an open set, as it is the union of $(\#E)^n - 1$ open balls

$$[\omega_1, \ldots, \omega_n]^c = \bigcup_{(\omega'_1, \ldots, \omega'_n) \neq (\omega_1, \ldots, \omega_n)} \big[\omega'_1, \ldots, \omega'_n\big].$$

Since Ω is compact, the closed subset $[\omega_1, \ldots, \omega_n]$ is compact. As in Theorem 1.23, it can be shown that $\sigma(\mathcal{A}) = \mathcal{B}(\Omega, d)$.

Exercise: Prove the statements made above. ◊

The main result of this chapter is Carathéodory's measure extension theorem.

Theorem 1.41 (Carathéodory) *Let $\mathcal{A} \subset 2^\Omega$ be a ring and let μ be a σ-finite premeasure on $\mathcal{A}$. There exists a unique measure $\widetilde{\mu}$ on $\sigma(\mathcal{A})$ such that $\widetilde{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$. Furthermore, $\tilde{\mu}$ is σ-finite.*

We prepare for the proof of this theorem with a couple of lemmas. In fact, we will show a slightly stronger statement in Theorem 1.53.

Lemma 1.42 (Uniqueness by an $\cap$-closed generator) *Let $(\Omega, \mathcal{A}, \mu)$ be a σ-finite measure space and let $\mathcal{E} \subset \mathcal{A}$ be a π-system that generates $\mathcal{A}$. Assume that there exist sets $\Omega_1, \Omega_2, \ldots \in \mathcal{E}$ such that $\bigcup_{n=1}^\infty \Omega_n = \Omega$ and $\mu(\Omega_n) < \infty$ for all $n \in \mathbb{N}$. Then μ is uniquely determined by the values $\mu(E)$, $E \in \mathcal{E}$.*

If μ is a probability measure, the existence of the sequence $(\Omega_n)_{n\in\mathbb{N}}$ is not needed.

Proof Let ν be a (possibly different) σ-finite measure on $(\Omega, \mathcal{A})$ such that

$$\mu(E) = \nu(E) \quad \text{for every } E \in \mathcal{E}.$$

Let $E \in \mathcal{E}$ with $\mu(E) < \infty$. Consider the class of sets

$$\mathcal{D}_E = \big\{A \in \mathcal{A} : \mu(A \cap E) = \nu(A \cap E)\big\}.$$

In order to show that $\mathcal{D}_E$ is a λ-system, we check the properties of Definition 1.10:

(i) Clearly, $\Omega \in \mathcal{D}_E$.

(ii) Let $A, B \in \mathcal{D}_E$ with $A \supset B$. Then

$$\mu\big((A \setminus B) \cap E\big) = \mu(A \cap E) - \mu(B \cap E)$$
$$= \nu(A \cap E) - \nu(B \cap E) = \nu\big((A \setminus B) \cap E\big).$$

Hence $A \setminus B \in \mathcal{D}_E$.

(iii) Let $A_1, A_2, \ldots \in \mathcal{D}_E$ be mutually disjoint and $A = \bigcup_{n=1}^{\infty} A_n$. Then

$$\mu(A \cap E) = \sum_{n=1}^{\infty} \mu(A_n \cap E) = \sum_{n=1}^{\infty} \nu(A_n \cap E) = \nu(A \cap E).$$

Hence $A \in \mathcal{D}_E$.

Clearly, $\mathcal{E} \subset \mathcal{D}_E$; hence $\delta(\mathcal{E}) \subset \mathcal{D}_E$. Since $\mathcal{E}$ is a π-system, Theorem 1.19 yields

$$\mathcal{A} \supset \mathcal{D}_E \supset \delta(\mathcal{E}) = \sigma(\mathcal{E}) = \mathcal{A}.$$

Hence $\mathcal{D}_E = \mathcal{A}$.

This implies $\mu(A \cap E) = \nu(A \cap E)$ for any $A \in \mathcal{A}$ and $E \in \mathcal{E}$ with $\mu(E) < \infty$. Now let $\Omega_1, \Omega_2, \ldots \in \mathcal{E}$ be a sequence such that $\bigcup_{n=1}^{\infty} \Omega_n = \Omega$ and $\mu(\Omega_n) < \infty$ for all $n \in \mathbb{N}$. Let $E_n := \bigcup_{i=1}^{n} \Omega_i$, $n \in \mathbb{N}$, and $E_0 = \emptyset$. Hence $E_n = \biguplus_{i=1}^{n} (E_{i-1}^c \cap \Omega_i)$. For any $A \in \mathcal{A}$ and $n \in \mathbb{N}$, we thus get

$$\mu(A \cap E_n) = \sum_{i=1}^{n} \mu\big((A \cap E_{i-1}^c) \cap \Omega_i\big) = \sum_{i=1}^{n} \nu\big((A \cap E_{i-1}^c) \cap \Omega_i\big) = \nu(A \cap E_n).$$

Since $E_n \uparrow \Omega$ and since μ and ν are lower semicontinuous, we infer

$$\mu(A) = \lim_{n\to\infty} \mu(A \cap E_n) = \lim_{n\to\infty} \nu(A \cap E_n) = \nu(A).$$

The additional statement is trivial as $\tilde{\mathcal{E}} := \mathcal{E} \cup \{\Omega\}$ is a π-system that generates $\mathcal{A}$, and the value $\mu(\Omega) = 1$ is given. Hence one can choose the constant sequence $E_n = \Omega$, $n \in \mathbb{N}$. However, note that it is not enough to assume that μ is finite. In this case, in general, the total mass $\mu(\Omega)$ is not uniquely determined by the values $\mu(E)$, $E \in \mathcal{E}$; see Example 1.45(ii). □

Example 1.43 Let $\Omega = \mathbb{Z}$ and $\mathcal{E} = \{E_n : n \in \mathbb{Z}\}$ where $E_n = (-\infty, n] \cap \mathbb{Z}$. Then $\mathcal{E}$ is a π-system and $\sigma(\mathcal{E}) = 2^{\Omega}$. Hence a finite measure μ on $(\Omega, 2^{\Omega})$ is uniquely determined by the values $\mu(E_n)$, $n \in \mathbb{Z}$.

However, a σ-finite measure on $\mathbb{Z}$ is not uniquely determined by the values on $\mathcal{E}$: Let μ be the counting measure on $\mathbb{Z}$ and let $\nu = 2\mu$. Hence $\mu(E) = \infty = \nu(E)$ for all $E \in \mathcal{E}$. In order to distinguish μ and ν one needs a generator that contains sets of finite measure (of μ). Do the sets $\tilde{F}_n = [-n, n] \cap \mathbb{Z}$, $n \in \mathbb{N}$ do the trick? Indeed, for any σ-finite measure μ, we have $\mu(\tilde{F}_n) < \infty$ for all $n \in \mathbb{N}$. However, the sets $\tilde{F}_n$ do not generate 2^{Ω} (but which σ-algebra?). We get things to work out better if we modify the definition: $F_n = [-n/2, (n+1)/2] \cap \mathbb{Z}$. Now $\sigma(\{F_n,\ n \in \mathbb{N}\}) = 2^\Omega$, and hence $\mathcal{E} = \{F_n,\ n \in \mathbb{N}\}$ is a π-system that generates 2^Ω and such that $\mu(F_n) < \infty$ for all $n \in \mathbb{N}$. The conditions of the theorem are fulfilled as $F_n \uparrow \Omega$. ◊

Example 1.44 (Distribution function) A probability measure μ on the space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ is uniquely determined by the values $\mu((-\infty, b])$ (where $(-\infty, b] = \times_{i=1}^n (-\infty, b_i]$, $b \in \mathbb{R}^n$). In fact, these sets form a π-system that generates $\mathcal{B}(\mathbb{R}^n)$ (see Theorem 1.23). In particular, a probability measure μ on $\mathbb{R}$ is uniquely determined by its *distribution function* $F : \mathbb{R} \to [0, 1]$, $x \mapsto \mu((-\infty, x])$. ◊

Example 1.45

(i) Let $\Omega = \{1, 2, 3, 4\}$ and $\mathcal{E} = \{\{1, 2\}, \{2, 3\}\}$. Clearly, $\sigma(\mathcal{E}) = 2^\Omega$ but $\mathcal{E}$ is not a π-system. In fact, here a probability measure μ is not uniquely determined by the values, say $\mu(\{1, 2\}) = \mu(\{2, 3\}) = \frac{1}{2}$. We give just two different possibilities: $\mu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_3$ and $\mu' = \frac{1}{2}\delta_2 + \frac{1}{2}\delta_4$.
(ii) Let $\Omega = \{1, 2\}$ and $\mathcal{E} = \{\{1\}\}$. Then $\mathcal{E}$ is a π-system that generates 2^Ω. Hence a probability measure μ is uniquely determined by the value $\mu(\{1\})$. However, a *finite* measure is not determined by its value on $\{1\}$, as $\mu = 0$ and $\nu = \delta_2$ are different finite measures that agree on $\mathcal{E}$. ◊

Definition 1.46 (Outer measure) A set function $\mu^* : 2^\Omega \to [0, \infty]$ is called an *outer measure* if

(i) $\mu^*(\emptyset) = 0$, and
(ii) μ^* is monotone,
(iii) μ^* is σ-subadditive.

Lemma 1.47 *Let $\mathcal{A} \subset 2^\Omega$ be an arbitrary class of sets with $\emptyset \in \mathcal{A}$ and let μ be a nonnegative set function on $\mathcal{A}$ with $\mu(\emptyset) = 0$. For $A \subset \Omega$, define the set of countable coverings $\mathcal{F}$ with sets $F \in \mathcal{A}$:*

$$\mathcal{U}(A) = \left\{ \mathcal{F} \subset \mathcal{A} : \mathcal{F} \text{ is at most countable and } A \subset \bigcup_{F \in \mathcal{F}} F \right\}.$$

Define

$$\mu^*(A) := \inf\left\{ \sum_{F \in \mathcal{F}} \mu(F) : \mathcal{F} \in \mathcal{U}(A) \right\},$$

where $\inf \emptyset = \infty$. *Then μ^* is an outer measure. If in addition μ is σ-subadditive, then $\mu^*(A) = \mu(A)$ for all $A \in \mathcal{A}$.*

Proof We check properties (i)–(iii) of an outer measure.

(i) Since $\emptyset \in \mathcal{A}$, we have $\{\emptyset\} \in \mathcal{U}(\emptyset)$; hence $\mu^*(\emptyset) = 0$.
(ii) If $A \subset B$, then $\mathcal{U}(A) \supset \mathcal{U}(B)$; hence $\mu^*(A) \le \mu^*(B)$.

(iii) Let $A_n \subset \Omega$ for any $n \in \mathbb{N}$ and let $A \subset \bigcup_{n=1}^{\infty} A_n$. We show that $\mu^*(A) \le \sum_{n=1}^{\infty} \mu^*(A_n)$. Without loss of generality, assume $\mu^*(A_n) < \infty$ and hence $\mathcal{U}(A_n) \neq \emptyset$ for all $n \in \mathbb{N}$. Fix $\varepsilon > 0$. For every $n \in \mathbb{N}$, choose a covering $\mathcal{F}_n \in \mathcal{U}(A_n)$ such that

$$\sum_{F \in \mathcal{F}_n} \mu(F) \le \mu^*(A_n) + \varepsilon 2^{-n}.$$

Then $\mathcal{F} := \bigcup_{n=1}^{\infty} \mathcal{F}_n \in \mathcal{U}(A)$ and

$$\mu^*(A) \le \sum_{F \in \mathcal{F}} \mu(F) \le \sum_{n=1}^{\infty} \sum_{F \in \mathcal{F}_n} \mu(F) \le \sum_{n=1}^{\infty} \mu^*(A_n) + \varepsilon.$$

Let $A \in \mathcal{A}$. Since $\{A\} \in \mathcal{U}(A)$, we have $\mu^*(A) \le \mu(A)$. If μ is σ-subadditive, then for any $\mathcal{F} \in \mathcal{U}(A)$, we have $\sum_{F \in \mathcal{F}} \mu(F) \ge \mu(A)$; hence $\mu^*(A) \ge \mu(A)$. □

Definition 1.48 (μ^*-measurable sets) Let μ^* be an outer measure. A set $A \in 2^{\Omega}$ is called *μ^*-measurable* if

$$\mu^*(A \cap E) + \mu^*\big(A^c \cap E\big) = \mu^*(E) \quad \text{for any } E \in 2^{\Omega}. \tag{1.10}$$

We write $\mathcal{M}(\mu^*) = \{A \in 2^{\Omega} : A \text{ is } \mu^*\text{-measurable}\}$.

Lemma 1.49 *$A \in \mathcal{M}(\mu^*)$ if and only if*

$$\mu^*(A \cap E) + \mu^*\big(A^c \cap E\big) \le \mu^*(E) \quad \textit{for any } E \in 2^{\Omega}.$$

Proof As μ^* is subadditive, the other inequality is trivial. □

Lemma 1.50 *$\mathcal{M}(\mu^*)$ is an algebra.*

Proof We check properties (i)–(iii) of an algebra from Theorem 1.7.

(i) $\Omega \in \mathcal{M}(\mu^*)$ is evident.
(ii) (Closedness under complements) By definition, $A \in \mathcal{M}(\mu^*) \iff A^c \in \mathcal{M}(\mu^*)$.
(iii) (π-system) Let $A, B \in \mathcal{M}(\mu^*)$ and $E \in 2^{\Omega}$. Then

$$\begin{aligned}
&\mu^*\big((A \cap B) \cap E\big) + \mu^*\big((A \cap B)^c \cap E\big)\\
&\quad = \mu^*(A \cap B \cap E) + \mu^*\big(\big(A^c \cap B \cap E\big) \cup \big(A^c \cap B^c \cap E\big) \cup \big(A \cap B^c \cap E\big)\big)\\
&\quad \le \mu^*(A \cap B \cap E) + \mu^*\big(A^c \cap B \cap E\big)\\
&\qquad + \mu^*\big(A^c \cap B^c \cap E\big) + \mu^*\big(A \cap B^c \cap E\big)\\
&\quad = \mu^*(B \cap E) + \mu^*\big(B^c \cap E\big)\\
&\quad = \mu^*(E).
\end{aligned}$$

Here we used $A \in \mathcal{M}(\mu^*)$ in the last but one equality and $B \in \mathcal{M}(\mu^*)$ in the last equality. □

Lemma 1.51 *An outer measure μ^* is σ-additive on $\mathcal{M}(\mu^*)$.*

Proof Let $A, B \in \mathcal{M}(\mu^*)$ with $A \cap B = \emptyset$. Then

$$\mu^*(A \cup B) = \mu^*\big(A \cap (A \cup B)\big) + \mu^*\big(A^c \cap (A \cup B)\big) = \mu^*(A) + \mu^*(B).$$

Inductively, we get (finite) additivity. By definition, μ^* is σ-subadditive; hence we conclude by Theorem 1.36 that μ^* is also σ-additive. □

Lemma 1.52 *If μ^* is an outer measure, then $\mathcal{M}(\mu^*)$ is a σ-algebra. In particular, μ^* is a measure on $\mathcal{M}(\mu^*)$.*

Proof By Lemma 1.50, $\mathcal{M}(\mu^*)$ is an algebra and hence a π-system. By Theorem 1.18, it is sufficient to show that $\mathcal{M}(\mu^*)$ is a λ-system.

Hence, let $A_1, A_2, \ldots \in \mathcal{M}(\mu^*)$ be mutually disjoint, and define $A := \biguplus_{n=1}^{\infty} A_n$. We have to show $A \in \mathcal{M}(\mu^*)$; that is,

$$\mu^*(A \cap E) + \mu^*\big(A^c \cap E\big) \le \mu^*(E) \quad \text{for any } E \in 2^{\Omega}. \tag{1.11}$$

Let $B_n = \bigcup_{i=1}^{n} A_i$ for all $n \in \mathbb{N}$. For all $n \in \mathbb{N}$, we have

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

Inductively, we get $\mu^*(E \cap B_n) = \sum_{i=1}^{n} \mu^*(E \cap A_i)$. The monotonicity of μ^* now implies that

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap B_n) + \mu^*\big(E \cap B_n^c\big)\\ &\ge \mu^*(E \cap B_n) + \mu^*\big(E \cap A^c\big)\\ &= \sum_{i=1}^{n} \mu^*(E \cap A_i) + \mu^*\big(E \cap A^c\big).\end{aligned}$$

Letting $n \to \infty$ and using the σ-subadditivity of μ^*, we conclude

$$\mu^*(E) \ge \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*\big(E \cap A^c\big) \ge \mu^*(E \cap A) + \mu^*\big(E \cap A^c\big).$$

Hence (1.11) holds and the proof is complete. □

We come to an extension theorem for measures that makes slightly weaker assumptions than Carathéodory's theorem (Theorem 1.41).

Theorem 1.53 (Extension theorem for measures) *Let $\mathcal{A}$ be a semiring and let $\mu : \mathcal{A} \to [0, \infty]$ be an additive, σ-subadditive and σ-finite set function with $\mu(\emptyset) = 0$.*
Then there is a unique σ-finite measure $\widetilde{\mu} : \sigma(\mathcal{A}) \to [0, \infty]$ such that $\widetilde{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$.

Proof As $\mathcal{A}$ is a π-system, uniqueness follows by Lemma 1.42.

In order to establish the existence of $\widetilde{\mu}$, we define as in Lemma 1.47

$$\mu^*(A) := \inf\left\{ \sum_{F \in \mathcal{F}} \mu(F) : \mathcal{F} \in \mathcal{U}(A) \right\} \quad \text{for any } A \in 2^{\Omega}.$$

By Lemma 1.47, μ^* is an outer measure and $\mu^*(A) = \mu(A)$ for any $A \in \mathcal{A}$. We have to show that $\mathcal{M}(\mu^*) \supset \sigma(\mathcal{A})$. Since $\mathcal{M}(\mu^*)$ is a σ-algebra (Lemma 1.52), it is enough to show $\mathcal{A} \subset \mathcal{M}(\mu^*)$.

To this end, let $A \in \mathcal{A}$ and $E \in 2^{\Omega}$ with $\mu^*(E) < \infty$. Fix $\varepsilon > 0$. Then there is a sequence $E_1, E_2, \ldots \in \mathcal{A}$ such that

$$E \subset \bigcup_{n=1}^{\infty} E_n \quad \text{and} \quad \sum_{n=1}^{\infty} \mu(E_n) \le \mu^*(E) + \varepsilon.$$

Define $B_n := E_n \cap A \in \mathcal{A}$. Since $\mathcal{A}$ is a semiring, for every $n \in \mathbb{N}$ there is an $m_n \in \mathbb{N}$ and sets $C_n^1, \ldots, C_n^{m_n} \in \mathcal{A}$ such that $E_n \setminus A = E_n \setminus B_n = \biguplus_{k=1}^{m_n} C_n^k$. Hence

$$E \cap A \subset \bigcup_{n=1}^{\infty} B_n, \qquad E \cap A^c \subset \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{m_n} C_n^k \quad \text{and} \quad E_n = B_n \uplus \biguplus_{k=1}^{m_n} C_n^k.$$

By the definition of the outer measure and since μ is assumed to be (finitely) additive, we get

$$\begin{aligned}
\mu^*(E \cap A) + \mu^*\big(E \cap A^c\big) &\le \sum_{n=1}^{\infty} \mu(B_n) + \sum_{n=1}^{\infty} \sum_{k=1}^{m_n} \mu\big(C_n^k\big) \\
&= \sum_{n=1}^{\infty} \left(\mu(B_n) + \sum_{k=1}^{m_n} \mu\big(C_n^k\big) \right) \\
&= \sum_{n=1}^{\infty} \mu(E_n) \\
&\le \mu^*(E) + \varepsilon.
\end{aligned}$$

Hence $\mu^*(E \cap A) + \mu^*(E \cap A^c) \le \mu^*(E)$ and thus $A \in \mathcal{M}(\mu^*)$, which implies $\mathcal{A} \subset \mathcal{M}(\mu^*)$. Now define $\widetilde{\mu} : \sigma(\mathcal{A}) \to [0, \infty]$, $A \mapsto \mu^*(A)$. By Lemma 1.51, $\widetilde{\mu}$ is a measure and $\widetilde{\mu}$ is σ-finite since μ is σ-finite. □

Example 1.54 (Lebesgue measure, continuation of Example 1.39) We aim at extending the volume $\mu((a,b]) = \prod_{i=1}^n (b_i - a_i)$ that was defined on the class of rectangles $\mathcal{A} = \{(a,b] : a, b \in \mathbb{R}^n,\ a < b\}$ to the Borel σ-algebra $\mathcal{B}(\mathbb{R}^n)$. In order to check the assumptions of Theorem 1.53, we only have to check that μ is σ-subadditive. To this end, let $(a,b],\ (a(1),b(1)],\ (a(2),b(2)],\ldots \in \mathcal{A}$ with

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

We show that

$$\mu\big((a,b]\big) \le \sum_{k=1}^\infty \mu\big(\big(a(k),b(k)\big]\big). \tag{1.12}$$

For this purpose we use a compactness argument to reduce (1.12) to finite additivity. Fix $\varepsilon > 0$. For any $k \in \mathbb{N}$, choose $b_\varepsilon(k) > b(k)$ such that

$$\mu\big(\big(a(k),b_\varepsilon(k)\big]\big) \le \mu\big(\big(a(k),b(k)\big]\big) + \varepsilon 2^{-k-1}.$$

Further choose $a_\varepsilon \in (a,b)$ such that $\mu((a_\varepsilon,b]) \ge \mu((a,b]) - \frac{\varepsilon}{2}$. Now $[a_\varepsilon,b]$ is compact and

$$\bigcup_{k=1}^\infty \big(a(k),b_\varepsilon(k)\big) \supset \bigcup_{k=1}^\infty \big(a(k),b(k)\big] \supset (a,b] \supset [a_\varepsilon,b],$$

whence there exists a K_0 such that $\bigcup_{k=1}^{K_0}(a(k),b_\varepsilon(k)) \supset (a_\varepsilon,b]$. As μ is (finitely) subadditive (see Lemma 1.31(iii)), we obtain

$$\begin{aligned}\mu\big((a,b]\big) &\le \frac{\varepsilon}{2} + \mu\big((a_\varepsilon,b]\big) \le \frac{\varepsilon}{2} + \sum_{k=1}^{K_0} \mu\big(\big(a(k),b_\varepsilon(k)\big]\big)\\ &\le \frac{\varepsilon}{2} + \sum_{k=1}^{K_0}\big(\varepsilon 2^{-k-1} + \mu\big(\big(a(k),b(k)\big]\big)\big) \le \varepsilon + \sum_{k=1}^\infty \mu\big(\big(a(k),b(k)\big]\big).\end{aligned}$$

Letting $\varepsilon \downarrow 0$ yields (1.12); hence μ is σ-subadditive. ◊

Combining the last example with Theorem 1.53, we have shown the following theorem.

Theorem 1.55 (Lebesgue measure) *There exists a uniquely determined measure λ^n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ with the property that*

$$\lambda^n\big((a,b]\big) = \prod_{i=1}^n (b_i - a_i) \quad \textit{for all } a, b \in \mathbb{R}^n \textit{ with } a < b.$$

λ^n is called the Lebesgue measure *on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ or* Lebesgue–Borel measure.

Example 1.56 (Lebesgue–Stieltjes measure) Let $\Omega = \mathbb{R}$ and $\mathcal{A} = \{(a,b] : a,b \in \mathbb{R},\ a \le b\}$. $\mathcal{A}$ is a semiring and $\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$. Furthermore, let $F : \mathbb{R} \to \mathbb{R}$ be monotone increasing and right continuous. We define a set function

$$\tilde{\mu}_F : \mathcal{A} \to [0,\infty), \quad (a,b] \mapsto F(b) - F(a).$$

Clearly, $\tilde{\mu}_F(\emptyset) = 0$ and $\tilde{\mu}_F$ is additive.

Let $(a,b], (a(1),b(1)], (a(2),b(2)], \ldots \in \mathcal{A}$ such that $(a,b] \subset \bigcup_{n=1}^{\infty}(a(n), b(n)]$. Fix $\varepsilon > 0$ and choose $a_\varepsilon \in (a,b)$ such that $F(a_\varepsilon) - F(a) < \varepsilon/2$. This is possible, as F is right continuous. For any $k \in \mathbb{N}$, choose $b_\varepsilon(k) > b(k)$ such that

$$F\big(b_\varepsilon(k)\big) - F\big(b(k)\big) < \varepsilon 2^{-k-1}.$$

As in Example 1.54, it can be shown that $\tilde{\mu}_F((a,b]) \le \varepsilon + \sum_{k=1}^{\infty} \tilde{\mu}_F((a(k),b(k)])$. This implies that $\tilde{\mu}_F$ is σ-subadditive. By Theorem 1.53, we can extend $\tilde{\mu}_F$ uniquely to a σ-finite measure μ_F on $\mathcal{B}(\mathbb{R})$. ◊

Definition 1.57 (Lebesgue–Stieltjes measure) The measure μ_F on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by

$$\mu_F\big((a,b]\big) = F(b) - F(a) \quad \text{for all } a,b \in \mathbb{R} \text{ with } a < b$$

is called the *Lebesgue–Stieltjes measure* with distribution function F.

Example 1.58 Important special cases for the Lebesgue–Stieltjes measure are the following:

(i) If $F(x) = x$, then $\mu_F = \lambda^1$ is the Lebesgue measure on $\mathbb{R}$.

(ii) Let $f : \mathbb{R} \to [0,\infty)$ be continuous and let $F(x) = \int_0^x f(t)\,dt$ for all $x \in \mathbb{R}$. Then μ_F is the extension of the premeasure with *density* f that was defined in Example 1.30(ix).

(iii) Let $x_1, x_2, \ldots \in \mathbb{R}$ and $\alpha_n \ge 0$ for all $n \in \mathbb{N}$ such that $\sum_{n=1}^{\infty} \alpha_n < \infty$. Then $F = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{[x_n,\infty)}$ is the distribution function of the finite measure $\mu_F = \sum_{n=1}^{\infty} \alpha_n \delta_{x_n}$.

(iv) Let $x_1, x_2, \ldots \in \mathbb{R}$ such that $\mu = \sum_{n=1}^{\infty} \delta_{x_n}$ is a σ-finite measure. Then μ is a Lebesgue–Stieltjes measure if and only if the sequence $(x_n)_{n\in\mathbb{N}}$ does not have a limit point. Indeed, if $(x_n)_{n\in\mathbb{N}}$ does not have a limit point, then by the Bolzano–Weierstraß theorem, $\#\{n \in \mathbb{N} : x_n \in [-K,K]\} < \infty$ for every $K > 0$. If we let $F(x) = \#\{n \in \mathbb{N} : x_n \in [0,x]\}$ for $x \ge 0$ and $F(x) = -\#\{n \in \mathbb{N} : x_n \in [x,0)\}$, then $\mu = \mu_F$. On the other hand, if μ is a Lebesgue–Stieltjes measure, this is $\mu = \mu_F$ for some F, then $\#\{n \in \mathbb{N} : x_n \in (-K,K]\} = F(K) - F(-K) < \infty$ for all $K > 0$; hence $(x_n)_{n\in\mathbb{N}}$ does not have a limit point.

(v) If $\lim_{x\to\infty}(F(x) - F(-x)) = 1$, then μ_F is a probability measure. ◊

We will now have a closer look at the case where μ_F is a probability measure.

Definition 1.59 (Distribution function) A right continuous monotone increasing function $F : \mathbb{R} \to [0, 1]$ with $F(-\infty) := \lim_{x\to-\infty} F(x) = 0$ and $F(\infty) := \lim_{x\to\infty} F(x) = 1$ is called a (proper) *probability distribution function* (p.d.f.). If we only have $F(\infty) \le 1$ instead of $F(\infty) = 1$, then F is called a (possibly) defective p.d.f. If μ is a (sub-)probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then $F_\mu : x \mapsto \mu((-\infty, x])$ is called the distribution function of μ.

Clearly, F_μ is right continuous and $F(-\infty) = 0$, since μ is upper semicontinuous and finite (Theorem 1.36). Since μ is lower semicontinuous, we have $F(\infty) = \mu(\mathbb{R})$; hence F_μ is indeed a (possibly defective) distribution function if μ is a (sub-)probability measure.

The argument of Example 1.56 yields the following theorem.

Theorem 1.60 *The map $\mu \mapsto F_\mu$ is a bijection from the set of probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to the set of probability distribution functions, respectively from the set of sub-probability measures to the set of defective distribution functions.*

We have established that every finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a Lebesgue–Stieltjes measure for some function F. For σ-finite measures, the corresponding statement does not hold in this generality as we saw in Example 1.58(iv).

We come now to a theorem that combines Theorem 1.55 with the idea of Lebesgue–Stieltjes measures. Later we will see that the following theorem is valid in greater generality. In particular, the assumption that the factors are of Lebesgue–Stieltjes type can be dropped.

Theorem 1.61 (Finite products of measures) *Let $n \in \mathbb{N}$ and let $\mu_1, \ldots, \mu_n$ be finite measures or, more generally, Lebesgue–Stieltjes measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then there exists a unique σ-finite measure μ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ such that*

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

We call $\mu =: \bigotimes_{i=1}^n \mu_i$ the product measure *of the measures $\mu_1, \ldots, \mu_n$.*

Proof The proof is the same as for Theorem 1.55. One has to check that the intervals $(a, b_\varepsilon]$ and so on can be chosen such that $\mu((a, b_\varepsilon]) < \mu((a, b]) + \varepsilon$. Here we employ the right continuity of the increasing function F_i that belongs to μ_i. The details are left as an exercise. □

Remark 1.62 Later we will see in Theorem 14.14 that the statement holds even for arbitrary σ-finite measures $\mu_1, \ldots, \mu_n$ on arbitrary (even different) measurable spaces. One can even construct infinite products if all factors are probability spaces (Theorem 14.36). ◊

Example 1.63 (Infinite product measure, continuation of Example 1.40) Let E be a finite set and let $\Omega = E^{\mathbb{N}}$ be the space of E-valued sequences. Further, let $(p_e)_{e\in E}$ be a probability vector. Define a content μ on $\mathcal{A} = \{[\omega_1,\ldots,\omega_n] : \omega_1,\ldots,\omega_n \in E,\ n\in\mathbb{N}\}$ by

$$\mu\big([\omega_1,\ldots,\omega_n]\big) = \prod_{i=1}^{n} p_{\omega_i}.$$

We aim at extending μ to a measure on $\sigma(\mathcal{A})$. In order to check the assumptions of Theorem 1.53, we have to show that μ is σ-subadditive. As in the preceding example, we use a compactness argument.

Let $A, A_1, A_2, \ldots \in \mathcal{A}$ and $A \subset \bigcup_{n=1}^{\infty} A_n$. We are done if we can show that there exists an $N \in \mathbb{N}$ such that

$$A \subset \bigcup_{n=1}^{N} A_n. \tag{1.13}$$

Indeed, due to the (finite) subadditivity of μ (see Lemma 1.31(iii)), this implies $\mu(A) \le \sum_{n=1}^{N} \mu(A_n) \le \sum_{n=1}^{\infty} \mu(A_n)$; hence μ is σ-subadditive.

We now give two different proofs for (1.13).

1st Proof. The metric d from (1.9) induces the product topology on Ω; hence, as remarked in Example 1.40, (Ω, d) is a compact metric space. Every $A \in \mathcal{A}$ is closed and thus compact. Since every A_n is also open, A can be covered by finitely many A_n; hence (1.13) holds.

2nd Proof. We now show by *elementary* means the validity of (1.13). The procedure imitates the proof that Ω is compact. Let $B_n := A \setminus \bigcup_{i=1}^{n} A_i$. We assume $B_n \neq \emptyset$ for all $n \in \mathbb{N}$ in order to get a contradiction. By Dirichlet's pigeonhole principle (recall that E is finite), we can choose $\omega_1 \in E$ such that $[\omega_1] \cap B_n \neq \emptyset$ for infinitely many $n \in \mathbb{N}$. Since $B_1 \supset B_2 \supset \ldots$, we obtain

$$[\omega_1] \cap B_n \neq \emptyset \quad \text{for all } n \in \mathbb{N}.$$

Successively choose $\omega_2, \omega_3, \ldots \in E$ in such a way that

$$[\omega_1,\ldots,\omega_k] \cap B_n \neq \emptyset \quad \text{for all } k, n \in \mathbb{N}.$$

B_n is a disjoint union of certain sets $C_{n,1},\ldots,C_{n,m_n} \in \mathcal{A}$. Hence, for every $n \in \mathbb{N}$ there is an $i_n \in \{1,\ldots,m_n\}$ such that $[\omega_1,\ldots,\omega_k] \cap C_{n,i_n} \neq \emptyset$ for infinitely many $k \in \mathbb{N}$. Since $[\omega_1] \supset [\omega_1,\omega_2] \supset \ldots$, we obtain

$$[\omega_1,\ldots,\omega_k] \cap C_{n,i_n} \neq \emptyset \quad \text{for all } k, n \in \mathbb{N}.$$

For fixed $n \in \mathbb{N}$ and large k, we have $[\omega_1,\ldots,\omega_k] \subset C_{n,i_n}$. Hence $\omega = (\omega_1,\omega_2,\ldots) \in C_{n,i_n} \subset B_n$. This implies $\bigcap_{n=1}^{\infty} B_n \neq \emptyset$, contradicting the assumption. ◇

Combining the last example with Theorem 1.53, we have shown the following theorem.

Theorem 1.64 (Product measure, Bernoulli measure) *Let E be a finite nonempty set and $\Omega = E^{\mathbb{N}}$. Let $(p_e)_{e\in E}$ be a probability vector. Then there exists a unique probability measure μ on $\sigma(\mathcal{A}) = \mathcal{B}(\Omega)$ such that*

$$\mu\big([\omega_1,\ldots,\omega_n]\big) = \prod_{i=1}^{n} p_{\omega_i} \quad \text{for all } \omega_1,\ldots,\omega_n \in E \text{ and } n\in\mathbb{N}.$$

μ is called the product measure *or* Bernoulli measure *on Ω with weights $(p_e)_{e\in E}$. We write $\left(\sum_{e\in E} p_e\delta_e\right)^{\otimes\mathbb{N}} := \mu$. The σ-algebra $(2^E)^{\otimes\mathbb{N}} := \sigma(\mathcal{A})$ is called the product σ-algebra on Ω.*

We will study product measures in a systematic way in Chapter 14.

The measure extension theorem yields an abstract statement of existence and uniqueness for measures on $\sigma(\mathcal{A})$ that were first defined on a semiring $\mathcal{A}$ only. The following theorem, however, shows that the measure of a set from $\sigma(\mathcal{A})$ can be well approximated by finite and countable operations with sets from $\mathcal{A}$.

Denote by

$$A \triangle B := (A\setminus B)\cup(B\setminus A) \quad \text{for } A, B\subset\Omega \tag{1.14}$$

the *symmetric difference* of the two sets A and B.

Theorem 1.65 (Approximation theorem for measures) *Let $\mathcal{A}\subset 2^{\Omega}$ be a semiring and let μ be a measure on $\sigma(\mathcal{A})$ that is σ-finite on $\mathcal{A}$.*

(i) *For any $A\in\sigma(\mathcal{A})$ and $\varepsilon>0$, there exist mutually disjoint sets $A_1, A_2,\ldots\in\mathcal{A}$ such that $A\subset\bigcup_{n=1}^{\infty} A_n$ and $\mu(\bigcup_{n=1}^{\infty} A_n\setminus A)<\varepsilon$.*
(ii) *For any $A\in\sigma(\mathcal{A})$ with $\mu(A)<\infty$ and any $\varepsilon>0$, there exists an $n\in\mathbb{N}$ and mutually disjoint sets $A_1,\ldots,A_n\in\mathcal{A}$ such that $\mu(A\triangle\bigcup_{k=1}^{n} A_k)<\varepsilon$.*
(iii) *For any $A\in\mathcal{M}(\mu^*)$, there are sets $A_-, A_+\in\sigma(\mathcal{A})$ with $A_-\subset A\subset A_+$ and $\mu(A_+\setminus A_-)=0$.*

Remark 1.66 (iii) implies that (i) and (ii) also hold for $A\in\mathcal{M}(\mu^*)$ (with μ^* instead of μ). If $\mathcal{A}$ is an algebra, then in (ii) for any $A\in\sigma(\mathcal{A})$, we even have $\inf_{B\in\mathcal{A}}\mu(A\triangle B)=0$. ◊

Proof (ii) As μ and the outer measure μ^* coincide on $\sigma(\mathcal{A})$ and since $\mu(A)$ is finite, by the very definition of μ^* (see Lemma 1.47) there exists a covering $B_1, B_2,\ldots\in\mathcal{A}$ of A such that

$$\mu(A)\geq\sum_{i=1}^{\infty}\mu(B_i)-\varepsilon/2.$$

Let $n \in \mathbb{N}$ with $\sum_{i=n+1}^{\infty} \mu(B_i) < \frac{\varepsilon}{2}$ (such an n exists since $\mu(A) < \infty$). For any three sets C, D, E, we have

$$C \triangle D = (D \setminus C) \cup (C \setminus D) \subset (D \setminus C) \cup \big(C \setminus (D \cup E)\big) \cup E \subset \big(C \triangle (D \cup E)\big) \cup E.$$

Choosing $C = A$, $D = \bigcup_{i=1}^{n} B_i$ and $E = \bigcup_{i=n+1}^{\infty} B_i$, this yields

$$\begin{aligned} \mu\left(A \triangle \bigcup_{i=1}^{n} B_i\right) &\le \mu\left(A \triangle \bigcup_{i=1}^{\infty} B_i\right) + \mu\left(\bigcup_{i=n+1}^{\infty} B_i\right) \\ &\le \mu\left(\bigcup_{i=1}^{\infty} B_i\right) - \mu(A) + \frac{\varepsilon}{2} \le \varepsilon. \end{aligned}$$

As $\mathcal{A}$ is a semiring, there exist a $k \in \mathbb{N}$ and $A_1, \ldots, A_k \in \mathcal{A}$ such that

$$\bigcup_{i=1}^{n} B_i = B_1 \uplus \biguplus_{i=2}^{n} \bigcap_{j=1}^{i-1} (B_i \setminus B_j) =: \biguplus_{i=1}^{k} A_i.$$

(i) Let $A \in \sigma(\mathcal{A})$ and $E_n \uparrow \Omega$, $E_n \in \sigma(\mathcal{A})$ with $\mu(E_n) < \infty$ for any $n \in \mathbb{N}$. For every $n \in \mathbb{N}$, choose a covering $(B_{n,m})_{m \in \mathbb{N}}$ of $A \cap E_n$ with

$$\mu(A \cap E_n) \ge \sum_{m=1}^{\infty} \mu(B_{n,m}) - 2^{-n}\varepsilon.$$

(This is possible due to the definition of the outer measure μ^*, which coincides with μ on $\mathcal{A}$.) Let $\bigcup_{m,n=1}^{\infty} B_{n,m} = \biguplus_{n=1}^{\infty} A_n$ for certain $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ (Exercise 1.1.1). Then

$$\begin{aligned} \mu\left(\biguplus_{n=1}^{\infty} A_n \setminus A\right) &= \mu\left(\bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} B_{n,m} \setminus A\right) \\ &\le \mu\left(\bigcup_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \big(B_{n,m} \setminus (A \cap E_n)\big)\right) \\ &\le \sum_{n=1}^{\infty} \left(\left(\sum_{m=1}^{\infty} \mu(B_{n,m})\right) - \mu(A \cap E_n)\right) \le \varepsilon. \end{aligned}$$

(iii) Let $A \in \mathcal{M}(\mu^*)$ and $(E_n)_{n \in \mathbb{N}}$ as above. For any $m, n \in \mathbb{N}$, choose $A_{n,m} \in \sigma(\mathcal{A})$ such that $A_{n,m} \supset A \cap E_n$ and $\mu^*(A_{n,m}) \le \mu^*(A \cap E_n) + \frac{2^{-n}}{m}$.

Define $A_m := \bigcup_{n=1}^{\infty} A_{n,m} \in \sigma(\mathcal{A})$. Then $A_m \supset A$ and $\mu^*(A_m \setminus A) \le \frac{1}{m}$. Define $A_+ := \bigcap_{m=1}^{\infty} A_m$. Then $\sigma(\mathcal{A}) \ni A_+ \supset A$ and $\mu^*(A_+ \setminus A) = 0$. Similarly, choose $(A_-)^c \in \sigma(\mathcal{A})$ with $(A_-)^c \supset A^c$ and $\mu^*((A_-)^c \setminus A^c) = 0$. Then $A_+ \supset A \supset A_-$ and $\mu(A_+ \setminus A_-) = \mu^*(A_+ \setminus A_-) = \mu^*(A_+ \setminus A) + \mu^*(A \setminus A_-) = 0$. □

Remark 1.67 (Regularity of measures) (Compare with Theorem 13.6.) Let λ^n be the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Let $\mathcal{A}$ be the semiring of rectangles of the form $(a,b] \subset \mathbb{R}^n$; hence $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{A})$ by Theorem 1.23. By the approximation theorem, for any $A \in \mathcal{B}(\mathbb{R}^n)$ and $\varepsilon > 0$, there exist countably many $A_1, A_2, \ldots \in \mathcal{A}$ with $A \subset \bigcup_{i=1}^{\infty} A_i$ and

$$\lambda^n\left(\bigcup_{i=1}^{\infty} A_i \setminus A\right) < \varepsilon/2.$$

For any A_i, there exists an *open* rectangle $B_i \supset A_i$ with $\lambda^n(B_i \setminus A_i) < \varepsilon 2^{-i-1}$ (upper semicontinuity of λ^n). Hence $U = \bigcup_{i=1}^{\infty} B_i$ is an open set $U \supset A$ with

$$\lambda^n(U \setminus A) < \varepsilon.$$

This property of λ^n is called *outer regularity*.

If $\lambda^n(A)$ is finite, then for any $\varepsilon > 0$ there exists a compact $K \subset A$ such that

$$\lambda^n(A \setminus K) < \varepsilon.$$

This property of λ^n is called *inner regularity*. Indeed, let $N > 0$ be such that $\lambda^n(A) - \lambda^n(A \cap [-N,N]^n) < \varepsilon/2$. Choose an open set $U \supset (A \cap [-N,N]^n)^c$ such that $\lambda^n(U \setminus (A \cap [-N,N]^n)^c) < \varepsilon/2$, and let $K := [-N,N]^n \setminus U \subset A$. ◊

Definition 1.68 (Null set) Let $(\Omega, \mathcal{A}, \mu)$ be a measure space.

(i) A set $A \in \mathcal{A}$ is called a *μ-null set*, or briefly a null set, if $\mu(A) = 0$. By $\mathcal{N}_\mu$ we denote the class of all subsets of μ-null sets.

(ii) Let $E(\omega)$ be a property that a point $\omega \in \Omega$ can have or not have. We say that E holds *μ-almost everywhere* (a.e.) or for *almost all* (a.a.) ω if there exists a null set N such that $E(\omega)$ holds for every $\omega \in \Omega \setminus N$. If $A \in \mathcal{A}$ and if there exists a null set N such that $E(\omega)$ holds for every $\omega \in A \setminus N$, then we say that E holds almost everywhere on A.

If $\mu = P$ is a probability measure, then we say that E holds *P-almost surely* (a.s.), respectively almost surely on A.

(iii) Let $A, B \in \mathcal{A}$ be such that $\mu(A \triangle B) = 0$. Then we write $A = B \pmod{\mu}$.

Definition 1.69 A measure space $(\Omega, \mathcal{A}, \mu)$ is called *complete* if $\mathcal{N}_\mu \subset \mathcal{A}$.

Remark 1.70 (Completion of a measure space) Let $(\Omega, \mathcal{A}, \mu)$ be a σ-finite measure space. There exists a unique smallest σ-algebra $\mathcal{A}^* \supset \mathcal{A}$ and an extension μ^* of μ to $\mathcal{A}^*$ such that $(\Omega, \mathcal{A}^*, \mu^*)$ is complete. $(\Omega, \mathcal{A}^*, \mu^*)$ is called the *completion* of $(\Omega, \mathcal{A}, \mu)$. With the notation of Theorem 1.53, this completion is

$$\big(\Omega, \mathcal{M}(\mu^*), \mu^*|_{\mathcal{M}(\mu^*)}\big).$$

Furthermore,

$$\mathcal{M}(\mu^*) = \sigma(\mathcal{A} \cup \mathcal{N}_\mu) = \{A \cup N : A \in \mathcal{A},\ N \in \mathcal{N}_\mu\}$$

and $\mu^*(A \cup N) = \mu(A)$ for any $A \in \mathcal{A}$ and $N \in \mathcal{N}_\mu$.

In the following, we will not need these statements. Hence, instead of giving a proof, we refer to the textbooks on measure theory (e.g., [37]). ◊

Example 1.71 Let λ be the Lebesgue measure (more accurately, the Lebesgue–Borel measure) on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Then λ can be extended uniquely to a measure λ^* on

$$\mathcal{B}^*(\mathbb{R}^n) = \sigma\big(\mathcal{B}(\mathbb{R}^n) \cup \mathcal{N}\big),$$

where $\mathcal{N}$ is the class of subsets of Lebesgue–Borel null sets. $\mathcal{B}^*(\mathbb{R}^n)$ is called the σ-algebra of Lebesgue measurable sets. For the sake of distinction, we sometimes call λ the *Lebesgue–Borel measure* and λ^* the *Lebesgue measure*. However, in practice, this distinction will not be needed in this book. ◊

Example 1.72 Let $\mu = \delta_\omega$ be the Dirac measure for the point $\omega \in \Omega$ on some measurable space $(\Omega, \mathcal{A})$. If $\{\omega\} \in \mathcal{A}$, then the completion is $\mathcal{A}^* = 2^\Omega$, $\mu^* = \delta_\omega$. In the extreme case of a trivial σ-algebra $\mathcal{A} = \{\emptyset, \Omega\}$, however, the empty set is the only null set, $\mathcal{N}_\mu = \{\emptyset\}$; hence $\mathcal{A}^* = \{\emptyset, \Omega\}$, $\mu^* = \delta_\omega$. Note that, on the trivial σ-algebra, Dirac measures for different points $\omega \in \Omega$ cannot be distinguished. ◊

Definition 1.73 Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and $\Omega' \in \mathcal{A}$. On the trace σ-algebra $\mathcal{A}\big|_{\Omega'}$, we define a measure by

$$\mu\big|_{\Omega'}(A) := \mu(A) \quad \text{for } A \in \mathcal{A} \text{ with } A \subset \Omega'.$$

This measure is called the *restriction* of μ to Ω'.

Example 1.74 The restriction of the Lebesgue–Borel measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ to $[0, 1]$ is a probability measure on $([0, 1], \mathcal{B}(\mathbb{R})\big|_{[0,1]})$. More generally, for a measurable $A \in \mathcal{B}(\mathbb{R})$, we call the restriction $\lambda\big|_A$ the *Lebesgue measure* on A. Often this measure will be denoted by the same symbol λ when there is no danger of ambiguity.

Later we will see (Corollary 1.84) that $\mathcal{B}(\mathbb{R})\big|_A = \mathcal{B}(A)$, where $\mathcal{B}(A)$ is the Borel σ-algebra on A that is generated by the (relatively) open subsets of A. ◊

Example 1.75 (Uniform distribution) Let $A \in \mathcal{B}(\mathbb{R}^n)$ be a measurable set with n-dimensional Lebesgue measure $\lambda^n(A) \in (0, \infty)$. Then we can define a probability measure on $\mathcal{B}(\mathbb{R}^n)\big|_A$ by

$$\mu(B) := \frac{\lambda^n(B)}{\lambda^n(A)} \quad \text{for } B \in \mathcal{B}\big(\mathbb{R}^n\big) \text{ with } B \subset A.$$

This measure μ is called the *uniform distribution* on A and will be denoted by $\mathcal{U}_A := \mu$. ◊

Exercise 1.3.1 Show the following generalization of Example 1.58(iv): A measure $\sum_{n=1}^{\infty} \alpha_n \delta_{x_n}$ is a Lebesgue–Stieltjes measure for a suitable function F if and only if $\sum_{n:|x_n|\le K} \alpha_n < \infty$ for all $K > 0$.

Exercise 1.3.2 Let Ω be an uncountably infinite set and let $\omega_0 \in \Omega$ be an arbitrary element. Let $\mathcal{A} = \sigma(\{\omega\} : \omega \in \Omega \setminus \{\omega_0\})$.

(i) Give a characterization of $\mathcal{A}$ as in Exercise 1.1.4 (p. 11).
(ii) Show that $(\Omega, \mathcal{A}, \delta_{\omega_0})$ is complete.

Exercise 1.3.3 Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence of finite measures on the measurable space $(\Omega, \mathcal{A})$. Assume that for any $A \in \mathcal{A}$ there exists the limit $\mu(A) := \lim_{n\to\infty} \mu_n(A)$.

Show that μ is a measure on $(\Omega, \mathcal{A})$.

Hint: In particular, one has to show that μ is $\emptyset$-continuous.

1.4 Measurable Maps

A major task of mathematics is to study homomorphisms between objects; that is, structure-preserving maps. For topological spaces, these are the continuous maps, and for measurable spaces, these are the measurable maps.

In the rest of this chapter, we let $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ be measurable spaces.

Definition 1.76 (Measurable maps)

(i) A map $X : \Omega \to \Omega'$ is called $\mathcal{A}$–$\mathcal{A}'$-measurable (or, briefly, measurable) if $X^{-1}(\mathcal{A}') := \{X^{-1}(A') : A' \in \mathcal{A}'\} \subset \mathcal{A}$; that is, if

$$X^{-1}\big(A'\big) \in \mathcal{A} \quad \text{for any } A' \in \mathcal{A}'.$$

If X is measurable, we write $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$.

(ii) If $\Omega' = \mathbb{R}$ and $\mathcal{A}' = \mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$, then $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is called an $\mathcal{A}$-measurable real map.

Example 1.77

(i) The identity map $\mathrm{id} : \Omega \to \Omega$ is $\mathcal{A}$–$\mathcal{A}$-measurable.
(ii) If $\mathcal{A} = 2^{\Omega}$ or $\mathcal{A}' = \{\emptyset, \Omega'\}$, then any map $X : \Omega \to \Omega'$ is $\mathcal{A}$–$\mathcal{A}'$-measurable.
(iii) Let $A \subset \Omega$. The indicator function $\mathbb{1}_A : \Omega \to \{0, 1\}$ is $\mathcal{A}$–$2^{\{0,1\}}$-measurable if and only if $A \in \mathcal{A}$. ◊

Theorem 1.78 (Generated σ-algebra) *Let $(\Omega', \mathcal{A}')$ be a measurable space and let Ω be a nonempty set. Let $X : \Omega \to \Omega'$ be a map. The preimage*

$$X^{-1}\big(\mathcal{A}'\big) := \big\{X^{-1}\big(A'\big) : A' \in \mathcal{A}'\big\} \tag{1.15}$$

is the smallest σ-algebra with respect to which X is measurable. We say that $\sigma(X) := X^{-1}(\mathcal{A}')$ is the σ-algebra on Ω that is generated *by X.*

Proof This is left as an exercise. □

We now consider σ-algebras that are generated by more than one map.

Definition 1.79 (Generated σ-algebra) Let Ω be a nonempty set. Let I be an arbitrary index set. For any $i \in I$, let $(\Omega_i, \mathcal{A}_i)$ be a measurable space and let $X_i : \Omega \to \Omega_i$ be an arbitrary map. Then

$$\sigma(X_i,\, i \in I) := \sigma\left(\bigcup_{i \in I} \sigma(X_i)\right) = \sigma\left(\bigcup_{i \in I} X_i^{-1}(\mathcal{A}_i)\right)$$

is called the σ-algebra on Ω that is *generated* by $(X_i,\, i \in I)$. This is the smallest σ-algebra with respect to which all X_i are measurable.

As with continuous maps, the composition of measurable maps is again measurable.

Theorem 1.80 (Composition of maps) *Let $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$ and $(\Omega'', \mathcal{A}'')$ be measurable spaces and let $X : \Omega \to \Omega'$ and $X' : \Omega' \to \Omega''$ be measurable maps. Then the map $Y := X' \circ X : \Omega \to \Omega''$, $\omega \mapsto X'(X(\omega))$ is $\mathcal{A}$–$\mathcal{A}''$-measurable.*

Proof Obvious, since $Y^{-1}(\mathcal{A}'') = X^{-1}((X')^{-1}(\mathcal{A}'')) \subset X^{-1}(\mathcal{A}') \subset \mathcal{A}$. □

In practice, it is often not possible to check if a map X is measurable by checking if all preimages $X^{-1}(A')$, $A' \in \mathcal{A}'$ are measurable. Most σ-algebras $\mathcal{A}'$ are simply too large. Thus it comes in very handy that it is sufficient to check measurability on a generator of $\mathcal{A}'$ by the following theorem.

Theorem 1.81 (Measurability on a generator) *Let $\mathcal{E}' \subset \mathcal{A}'$ be a class of $\mathcal{A}'$-measurable sets. Then $\sigma(X^{-1}(\mathcal{E}')) = X^{-1}(\sigma(\mathcal{E}'))$ and hence*

$$X \text{ is } \mathcal{A}\text{–}\sigma(\mathcal{E}')\text{-measurable} \iff X^{-1}(E') \in \mathcal{A} \quad \text{for all } E' \in \mathcal{E}'.$$

If in particular $\sigma(\mathcal{E}') = \mathcal{A}'$, then

$$X \text{ is } \mathcal{A}\text{–}\mathcal{A}'\text{-measurable} \iff X^{-1}(\mathcal{E}') \subset \mathcal{A}.$$

Proof Clearly, $X^{-1}(\mathcal{E}') \subset X^{-1}(\sigma(\mathcal{E}')) = \sigma(X^{-1}(\sigma(\mathcal{E}')))$. Hence also

$$\sigma\big(X^{-1}(\mathcal{E}')\big) \subset X^{-1}\big(\sigma(\mathcal{E}')\big).$$

For the other inclusion, consider the class of sets

$$\mathcal{A}'_0 := \{A' \in \sigma(\mathcal{E}') : X^{-1}(A') \in \sigma(X^{-1}(\mathcal{E}'))\}.$$

We first show that $\mathcal{A}'_0$ is a σ-algebra by checking (i)–(iii) of Definition 1.2:

(i) Clearly, $\Omega' \in \mathcal{A}'_0$.

(ii) (Stability under complements) If $A' \in \mathcal{A}'_0$, then

$$X^{-1}((A')^c) = (X^{-1}(A'))^c \in \sigma(X^{-1}(\mathcal{E}'));$$

hence $(A')^c \in \mathcal{A}'_0$.

(iii) (σ-$\cup$-stability) Let $A'_1, A'_2, \ldots \in \mathcal{A}'_0$. Then

$$X^{-1}\left(\bigcup_{n=1}^{\infty} A'_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A'_n) \in \sigma(X^{-1}(\mathcal{E}'));$$

hence $\bigcup_{n=1}^{\infty} A'_n \in \mathcal{A}'_0$.

Now $\mathcal{A}'_0 = \sigma(\mathcal{E}')$ since $\mathcal{E}' \subset \mathcal{A}'_0$. Hence $X^{-1}(A') \in \sigma(X^{-1}(\mathcal{E}'))$ for any $A' \in \sigma(\mathcal{E}')$ and thus $X^{-1}(\sigma(\mathcal{E}')) \subset \sigma(X^{-1}(\mathcal{E}'))$. □

Corollary 1.82 (Measurability of composed maps) *Let I be a nonempty index set and let $(\Omega, \mathcal{A})$, $(\Omega', \mathcal{A}')$ and $(\Omega_i, \mathcal{A}_i)$ be measurable spaces for any $i \in I$. Further, let $(X_i : i \in I)$ be a family of measurable maps $X_i : \Omega' \to \Omega_i$ with $\mathcal{A}' = \sigma(X_i : i \in I)$. Then the following holds*: *A map $Y : \Omega \to \Omega'$ is $\mathcal{A}$–$\mathcal{A}'$-measurable if and only if $X_i \circ Y$ is $\mathcal{A}$–$\mathcal{A}_i$-measurable for all $i \in I$.*

Proof If Y is measurable, then by Theorem 1.80 every $X_i \circ Y$ is measurable. Now assume that all of the composed maps $X_i \circ Y$ are $\mathcal{A}$–$\mathcal{A}_i$-measurable. By assumption, the set $\mathcal{E}' := \{X_i^{-1}(A'') : A'' \in \mathcal{A}_i,\ i \in I\}$ is a generator of $\mathcal{A}'$. Since all $X_i \circ Y$ are measurable, we have $Y^{-1}(A') \in \mathcal{A}$ for any $A' \in \mathcal{E}'$. Hence Theorem 1.81 yields that Y is measurable. □

Recall the definition of the trace of a class of sets from Definition 1.25.

Corollary 1.83 (Trace of a generated σ-algebra) *Let $\mathcal{E} \subset 2^{\Omega}$ and assume that $A \subset \Omega$ is nonempty. Then $\sigma(\mathcal{E}\big|_A) = \sigma(\mathcal{E})\big|_A$.*

Proof Let $X : A \hookrightarrow \Omega$, $\omega \mapsto \omega$ be the canonical inclusion; hence $X^{-1}(B) = A \cap B$ for all $B \subset \Omega$. By Theorem 1.81, we have

$$\begin{aligned}
\sigma(\mathcal{E}\big|_A) &= \sigma(\{E \cap A : E \in \mathcal{E}\}) \\
&= \sigma(\{X^{-1}(E) : E \in \mathcal{E}\}) = \sigma(X^{-1}(\mathcal{E})) \\
&= X^{-1}(\sigma(\mathcal{E})) = \{A \cap B : B \in \sigma(\mathcal{E})\} = \sigma(\mathcal{E})\big|_A.
\end{aligned}$$

□

Recall that, for any subset $A \subset \Omega$ of a topological space (Ω, τ), the class $\tau\big|_A$ is the topology of relatively open sets (in A). We denote by $\mathcal{B}(\Omega, \tau) = \sigma(\tau)$ the Borel σ-algebra on (Ω, τ).

Corollary 1.84 (Trace of the Borel σ-algebra) *Let (Ω, τ) be a topological space and let $A \subset \Omega$ be a nonempty subset of Ω. Then*

$$\mathcal{B}(\Omega, \tau)\big|_A = \mathcal{B}(A, \tau\big|_A).$$

Example 1.85

(i) Let Ω' be countable. Then $X : \Omega \to \Omega'$ is $\mathcal{A}$–$2^{\Omega'}$-measurable if and only if $X^{-1}(\{\omega'\}) \in \mathcal{A}$ for all $\omega' \in \Omega'$. If Ω' is uncountably infinite, this is wrong in general. (For example, consider $\Omega = \Omega' = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$, and $X(\omega) = \omega$ for all $\omega \in \Omega$. Clearly, $X^{-1}(\{\omega\}) = \{\omega\} \in \mathcal{B}(\mathbb{R})$. If, on the other hand, $A \subset \mathbb{R}$ is not in $\mathcal{B}(\mathbb{R})$, then $A \in 2^{\mathbb{R}}$, but $X^{-1}(A) \notin \mathcal{B}(\mathbb{R})$.)

(ii) For $x \in \mathbb{R}$, we agree on the following notation for rounding:

$$\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \le x\} \quad \text{and} \quad \lceil x \rceil := \min\{k \in \mathbb{Z} : k \ge x\}. \tag{1.16}$$

The maps $\mathbb{R} \to \mathbb{Z}$, $x \mapsto \lfloor x \rfloor$ and $x \mapsto \lceil x \rceil$ are $\mathcal{B}(\mathbb{R})$–$2^{\mathbb{Z}}$-measurable since for all $k \in \mathbb{Z}$ the preimages $\{x \in \mathbb{R} : \lfloor x \rfloor = k\} = [k, k+1)$ and $\{x \in \mathbb{R} : \lceil x \rceil = k\} = (k-1, k]$ are in $\mathcal{B}(\mathbb{R})$. By the composition theorem (Theorem 1.80), for any measurable map $f : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ the maps $\lfloor f \rfloor$ and $\lceil f \rceil$ are also $\mathcal{A}$–$2^{\mathbb{Z}}$-measurable.

(iii) A map $X : \Omega \to \mathbb{R}^d$ is $\mathcal{A}$–$\mathcal{B}(\mathbb{R}^d)$-measurable if and only if

$$X^{-1}\big((-\infty, a]\big) \in \mathcal{A} \quad \text{for any } a \in \mathbb{R}^d.$$

In fact $\sigma((-\infty, a], a \in \mathbb{R}^d) = \mathcal{B}(\mathbb{R}^d)$ by Theorem 1.23. The analogous statement holds for any of the classes $\mathcal{E}_1, \dots, \mathcal{E}_{12}$ from Theorem 1.23. ◊

Example 1.86 Let $d(x, y) = \|x - y\|_2$ be the usual Euclidean distance on $\mathbb{R}^n$ and let $\mathcal{B}(\mathbb{R}^n, d) = \mathcal{B}(\mathbb{R}^n)$ be the Borel σ-algebra with respect to the topology generated by d. For any subset A of $\mathbb{R}^n$, we have $\mathcal{B}(A, d) = \mathcal{B}(\mathbb{R}^n, d)\big|_A$. ◊

We want to extend the real line by the points $-\infty$ and $+\infty$. Thus we define

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

From a topological point of view, $\overline{\mathbb{R}}$ will be considered as the so-called two point compactification by considering $\overline{\mathbb{R}}$ as topologically isomorphic to $[-1, 1]$ via the map

$$\varphi : [-1, 1] \to \overline{\mathbb{R}}, \quad x \mapsto \begin{cases} \tan(\pi x/2), & x \in (-1, 1), \\ -\infty, & x = -1, \\ \infty, & x = +1. \end{cases}$$

In fact, $\bar{d}(x, y) = |\varphi^{-1}(x) - \varphi^{-1}(y)|$ for $x, y \in \overline{\mathbb{R}}$ defines a metric on $\overline{\mathbb{R}}$ such that φ and φ^{-1} are continuous. Hence φ is a topological isomorphism. We denote by $\bar{\tau}$ the corresponding topology induced on $\overline{\mathbb{R}}$ and by τ the usual topology on $\mathbb{R}$.

Corollary 1.87 *With the above notation,* $\bar{\tau}\big|_{\mathbb{R}} = \tau$ *and hence* $\mathcal{B}(\overline{\mathbb{R}})\big|_{\mathbb{R}} = \mathcal{B}(\mathbb{R})$.
In particular, if $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *is measurable, then in a canonical way* X *is also an* $\overline{\mathbb{R}}$*-valued measurable map.*

Thus $\overline{\mathbb{R}}$ is really an extension of the real line, and the inclusion $\mathbb{R} \hookrightarrow \overline{\mathbb{R}}$ is measurable.

Theorem 1.88 (Measurability of continuous maps) *Let* (Ω, τ) *and* (Ω', τ') *be topological spaces and let* $f : \Omega \to \Omega'$ *be a continuous map. Then* f *is* $\mathcal{B}(\Omega)$–$\mathcal{B}(\Omega')$*-measurable.*

Proof As $\mathcal{B}(\Omega') = \sigma(\tau')$ and by Theorem 1.81, it is sufficient to show that $f^{-1}(A') \in \sigma(\tau)$ for all $A' \in \tau'$. However, since f is continuous, we even have $f^{-1}(A') \in \tau$ for all $A' \in \tau'$. □

For $x, y \in \overline{\mathbb{R}}$, we agree on the following notation.

$$
\begin{aligned}
x \vee y &= \max(x, y) && \text{(maximum)},\\
x \wedge y &= \min(x, y) && \text{(minimum)},\\
x^+ &= \max(x, 0) && \text{(positive part)},\\
x^- &= \max(-x, 0) && \text{(negative part)},\\
|x| &= \max(x, -x) = x^- + x^+ && \text{(modulus)},\\
\operatorname{sign}(x) &= \mathbb{1}_{\{x>0\}} - \mathbb{1}_{\{x<0\}} && \text{(sign function)}.
\end{aligned}
$$

Analogously, for measurable real maps we write, for example, $X^+ = \max(X, 0)$. The maps $x \mapsto x^+$, $x \mapsto x^-$ and $x \mapsto |x|$ are continuous (and hence measurable by the preceding theorem). Clearly, the map $x \mapsto \operatorname{sign}(x)$ also is measurable. Using Corollary 1.82, we thus get the following corollary.

Corollary 1.89 *If* X *is a real or* $\overline{\mathbb{R}}$*-valued measurable map, then the maps* X^-, X^+, $|X|$ *and* $\operatorname{sign}(X)$ *also are measurable.*

Theorem 1.90 (Coordinate maps are measurable) *Let* $(\Omega, \mathcal{A})$ *be a measurable space and let* $f_1, \ldots, f_n : \Omega \to \mathbb{R}$ *be maps. Define* $f := (f_1, \ldots, f_n) : \Omega \to \mathbb{R}^n$. *Then*

$$f \text{ is } \mathcal{A}\text{–}\mathcal{B}\big(\mathbb{R}^n\big)\text{-measurable} \quad \iff \quad \text{each } f_i \text{ is } \mathcal{A}\text{–}\mathcal{B}(\mathbb{R})\text{-measurable}.$$

The analogous statement holds for $f_i : \Omega \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$.

Proof For $b \in \mathbb{R}^n$, we have $f^{-1}((-\infty, b)) = \bigcap_{i=1}^n f_i^{-1}((-\infty, b_i))$. If each f_i is measurable, then $f^{-1}((-\infty, b)) \in \mathcal{A}$. However, the rectangles $(-\infty, b)$, $b \in \mathbb{R}^n$, generate $\mathcal{B}(\mathbb{R}^n)$, and hence f is measurable. Now assume that f is measurable. For $i = 1, \ldots, n$, let $\pi_i : \mathbb{R}^n \to \mathbb{R}$, $x \mapsto x_i$ be the projection on the ith coordinate. Clearly, π_i is continuous and thus $\mathcal{B}(\mathbb{R}^n)$–$\mathcal{B}(\mathbb{R})$-measurable. Hence $f_i = \pi_i \circ f$ is measurable by Theorem 1.80. □

In the following theorem, we agree that $\frac{x}{0} := 0$ for all $x \in \mathbb{R}$.

Theorem 1.91 *Let $(\Omega, \mathcal{A})$ be a measurable space. Let $h : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $f, g : (\Omega, \mathcal{A}) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ be measurable maps. Then also the maps $f + g$, $f - g$, $f \cdot h$ and f/h are measurable.*

Proof The map $\pi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $(x, \alpha) \mapsto \alpha \cdot x$ is continuous and thus measurable. By Theorem 1.90, $(f, h) : \Omega \to \mathbb{R}^n \times \mathbb{R}$ is measurable. Hence also the composed map $f \cdot h = \pi \circ (f, h)$ is measurable. Similarly, we obtain the measurability of $f + g$ and $f - g$.

In order to show measurability of f/h, we define the map $H : \mathbb{R} \to \mathbb{R}$, $x \mapsto 1/x$. Note that by our convention $H(0) = 0$. Hence $f/h = f \cdot H \circ h$. Thus it is enough to show that H is measurable. Clearly, $H\big|_{\mathbb{R}\setminus\{0\}}$ is continuous. For any open set $U \subset \mathbb{R}$, $U \setminus \{0\}$ is also open and hence $H^{-1}(U \setminus \{0\}) \in \mathcal{B}(\mathbb{R})$. Furthermore, $H^{-1}(\{0\}) = \{0\}$. Concluding, we get $H^{-1}(U) = H^{-1}(U \setminus \{0\}) \cup (U \cap \{0\}) \in \mathcal{B}(\mathbb{R})$. □

Theorem 1.92 *Let $X_1, X_2, \ldots$ be measurable maps $(\Omega, \mathcal{A}) \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$. Then the following maps are also measurable:*

$$\inf_{n\in\mathbb{N}} X_n, \qquad \sup_{n\in\mathbb{N}} X_n, \qquad \liminf_{n\to\infty} X_n, \qquad \limsup_{n\to\infty} X_n.$$

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

$$\Big(\inf_{n\in\mathbb{N}} X_n\Big)^{-1}\big([-\infty, a)\big) = \bigcup_{n=1}^{\infty} X_n^{-1}\big([-\infty, a)\big) \in \mathcal{A}.$$

By Theorem 1.81, this implies that $\inf_{n\in\mathbb{N}} X_n$ is measurable. The proof for $\sup_{n\in\mathbb{N}} X_n$ is similar.

For any $n \in \mathbb{N}$, we define $Y_n := \inf_{m\geq n} X_m$. Note that Y_n is measurable and hence $\liminf_{n\to\infty} X_n := \sup_{n\in\mathbb{N}} Y_n$ also is measurable. The proof for the limes superior is similar. □

We come to an important example of measurable maps $(\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the so-called simple functions.

Definition 1.93 (Simple function) Let $(\Omega, \mathcal{A})$ be a measurable space. A map $f : \Omega \to \mathbb{R}$ is called a *simple function* if there is an $n \in \mathbb{N}$ and mutually disjoint measurable sets $A_1, \ldots, A_n \in \mathcal{A}$, as well as numbers $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, such that

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

Remark 1.94 A measurable map that assumes only finitely many values is a simple function. (Exercise: Show this!) ◊

Definition 1.95 Assume that $f, f_1, f_2, \ldots$ are maps $\Omega \to \overline{\mathbb{R}}$ such that

$$f_1(\omega) \le f_2(\omega) \le \ldots \quad \text{and} \quad \lim_{n\to\infty} f_n(\omega) = f(\omega) \quad \text{for any } \omega \in \Omega.$$

Then we write $f_n \uparrow f$ and say that $(f_n)_{n\in\mathbb{N}}$ increases (pointwise) to f. Analogously, we write $f_n \downarrow f$ if $(-f_n) \uparrow (-f)$.

Theorem 1.96 *Let $(\Omega, \mathcal{A})$ be a measurable space and let $f : \Omega \to [0, \infty]$ be measurable. Then the following statements hold.*

(i) *There exists a sequence $(f_n)_{n\in\mathbb{N}}$ of nonnegative simple functions such that $f_n \uparrow f$.*
(ii) *There are measurable sets $A_1, A_2, \ldots \in \mathcal{A}$ and numbers $\alpha_1, \alpha_2, \ldots \ge 0$ such that $f = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{A_n}$.*

Proof (i) For $n \in \mathbb{N}_0$, define $f_n = (2^{-n} \lfloor 2^n f \rfloor) \wedge n$. Then f_n is measurable (by Theorem 1.92 and Example 1.85(ii)) and assumes at most $n2^n + 1$ different values. Hence it is a simple function. Clearly, $f_n \uparrow f$.

(ii) Let f_n be as above. Let $B_{n,i} := \{\omega : f_n(\omega) - f_{n-1}(\omega) = i2^{-n}\}$ and $\beta_{n,i} = i2^{-n}$ for $n \in \mathbb{N}$ and $i = 1, \ldots, 2^n$. Hence $f_n - f_{n-1} = \sum_{i=1}^{2^n} \beta_{n,i} \mathbb{1}_{B_{n,i}}$. By changing the numeration $(n, i) \mapsto m$, we get $(\alpha_m)_{m\in\mathbb{N}}$ and $(A_m)_{m\in\mathbb{N}}$ such that

$$f = f_0 + \sum_{n=1}^{\infty} (f_n - f_{n-1}) = \sum_{m=1}^{\infty} \alpha_m \mathbb{1}_{A_m}.$$

□

As a corollary to this statement on the structure of $[0, \infty]$-valued measurable maps, we show the following factorization lemma.

Corollary 1.97 (Factorization lemma) *Let $(\Omega', \mathcal{A}')$ be a measurable space and let Ω be a nonempty set. Let $f : \Omega \to \Omega'$ be a map. A map $g : \Omega \to \overline{\mathbb{R}}$ is $\sigma(f)$–$\mathcal{B}(\overline{\mathbb{R}})$-measurable if and only if there is a measurable map $\varphi : (\Omega', \mathcal{A}') \to (\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ such that $g = \varphi \circ f$.*

Proof "$\Longleftarrow$" If φ is measurable and $g = \varphi \circ f$, then g is measurable by Theorem 1.80.

"$\Longrightarrow$" Now assume that g is $\sigma(f)$–$\mathcal{B}(\overline{\mathbb{R}})$-measurable. First consider the case $g \geq 0$. Then there exist measurable sets $A_1, A_2 \ldots \in \sigma(f)$ as well as numbers $\alpha_1, \alpha_2, \ldots, \in [0, \infty)$ such that $g = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{A_n}$. By the definition of $\sigma(f)$, for any $n \in \mathbb{N}$ there is a set $B_n \in \mathcal{A}'$ such that $f^{-1}(B_n) = A_n$; that is, such that $\mathbb{1}_{A_n} = \mathbb{1}_{B_n} \circ f$. Define $\varphi : \Omega' \to \overline{\mathbb{R}}$ by

$$\varphi = \sum_{n=1}^{\infty} \alpha_n \mathbb{1}_{B_n}.$$

Clearly, φ is $\mathcal{A}'$–$\mathcal{B}(\overline{\mathbb{R}})$-measurable and $g = \varphi \circ f$.

Now drop the assumption that g is nonnegative. Then there exist measurable maps φ^- and φ^+ such that $g^- = \varphi^- \circ f$ and $g^+ = \varphi^+ \circ f$. Hence $\varphi := \varphi^+ - \varphi^-$ does the trick. □

A measurable map transports a measure from one space to another.

Definition 1.98 (Image measure) Let $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$ be measurable spaces and let μ be a measure on $(\Omega, \mathcal{A})$. Further, let $X : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ be measurable. The *image measure* of μ under the map X is the measure $\mu \circ X^{-1}$ on $(\Omega', \mathcal{A}')$ that is defined by

$$\mu \circ X^{-1} : \mathcal{A}' \to [0, \infty], \quad A' \mapsto \mu\big(X^{-1}(A')\big).$$

Example 1.99 Let μ be a measure on $\mathbb{Z}^2$ and let $X : \mathbb{Z}^2 \to \mathbb{Z}$, $(x, y) \mapsto x + y$. Then

$$\mu \circ X^{-1}\big(\{x\}\big) = \sum_{y \in \mathbb{Z}} \mu\big(\{(x - y, y)\}\big).$$

◊

Example 1.100 Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a linear bijection and let λ be the Lebesgue measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Then $\lambda \circ L^{-1} = |\det(L)|^{-1}\lambda$. This is clear since for any $a, b \in \mathbb{R}^n$ with $a < b$, the parallelepiped $L^{-1}((a, b])$ has volume $|\det(L^{-1})| \times \prod_{i=1}^{n}(b_i - a_i)$. ◊

As a generalization of the last example, we state without proof the transformation formula for measures with continuous densities under differentiable maps. The proof can be found in textbooks on calculus.

Theorem 1.101 (Transformation formula in $\mathbb{R}^n$) *Let μ be a measure on $\mathbb{R}^n$ that has a continuous* (*or piecewise continuous*) *density $f : \mathbb{R}^n \to [0, \infty)$. That is,*

$$\mu\big((-\infty, x]\big) = \int_{-\infty}^{x_1} dt_1 \ldots \int_{-\infty}^{x_n} dt_n\, f(t_1, \ldots, t_n) \quad \textit{for all } x \in \mathbb{R}^n.$$

Let $A \subset \mathbb{R}^n$ be an open or a closed subset of $\mathbb{R}^n$ with $\mu(\mathbb{R}^n \setminus A) = 0$. Further, let $B \subset \mathbb{R}^n$ be open or closed. Finally, assume that $\varphi : A \to B$ is a continuously *differentiable bijection with derivative* φ'. *Then the image measure* $\mu \circ \varphi^{-1}$ *has the density*

$$f_\varphi(x) = \begin{cases} \frac{f(\varphi^{-1}(x))}{|\det(\varphi'(\varphi^{-1}(x)))|}, & \text{if } x \in B, \\ 0, & \text{if } x \in \mathbb{R}^n \setminus B. \end{cases}$$

Exercise 1.4.1 Let $f : \mathbb{R} \to \mathbb{R}$, $x \mapsto |x|$. Show that a Borel measurable map $g : \mathbb{R} \to \mathbb{R}$ is $\sigma(f) = f^{-1}(\mathcal{B}(\mathbb{R}))$-measurable if and only if g is even.

Exercise 1.4.2 Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $f : \Omega \to \mathbb{R}$ be measurable. Assume that $g : \Omega \to \mathbb{R}$ fulfills $g = f$ μ-almost everywhere. Show that g need not be measurable.

Exercise 1.4.3 Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable with derivative f'. Show that f' is $\mathcal{B}(\mathbb{R})$–$\mathcal{B}(\mathbb{R})$-measurable.

Exercise 1.4.4 (Compare Examples 1.40 and 1.63.) Let $\Omega = \{0, 1\}^{\mathbb{N}}$ and let $\mathcal{A} = (2^{\{0,1\}})^{\otimes \mathbb{N}}$ be the σ-algebra generated by the cylinder sets

$$\big\{[\omega_1, \ldots, \omega_n] : n \in \mathbb{N}, \omega_1, \ldots, \omega_n \in \{0, 1\}\big\}.$$

Further, let $\mu = (\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)^{\otimes \mathbb{N}}$ be the Bernoulli measure on Ω with equal weights on 0 and 1. For all $n \in \mathbb{N}$, let $X_n : \Omega \to \{0, 1\}$, $\omega \mapsto \omega_n$ be the nth coordinate map. Finally, let

$$U(\omega) = \sum_{n=1}^{\infty} X_n(\omega) 2^{-n} \quad \text{for } \omega \in \Omega.$$

(i) Show that $\mathcal{A} = \sigma(X_n : n \in \mathbb{N})$.
(ii) Show that U is $\mathcal{A}$–$\mathcal{B}([0, 1])$-measurable.
(iii) Determine the image measure $\mu \circ U^{-1}$ on $([0, 1], \mathcal{B}([0, 1]))$.
(iv) Determine an $\Omega_0 \in \mathcal{A}$ such that $\tilde{U} := U\big|_{\Omega_0}$ is bijective.
(v) Show that $\tilde{U}^{-1}$ is $\mathcal{B}([0, 1])$–$\mathcal{A}\big|_{\Omega_0}$-measurable.
(vi) Give an interpretation of the map $X_n \circ \tilde{U}^{-1}$.

Exercise 1.4.5 (Lusin's theorem) Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel measurable map. Show that for any $\varepsilon > 0$, there exists a closed set $C \subset \mathbb{R}$ with $\lambda(\mathbb{R} \setminus C) < \varepsilon$ such that the restriction $f\big|_C$ of f to C is continuous. (Note: Clearly, this does not mean that f would be continuous in every point $x \in C$.)

Hint: Use the inner regularity of Lebesgue measure λ (Remark 1.67) to show the assertion first for indicator functions. Next construct a sequence of maps that approximates f uniformly on a suitable set C.

1.5 Random Variables

The fundamental idea of modern probability theory is to model one or more random experiments as a probability space $(\Omega, \mathcal{A}, \mathbf{P})$. The sets $A \in \mathcal{A}$ are called *events*. In most cases, the events of Ω are not observed directly. Rather, the observations are aspects of the single experiments that are coded as measurable maps from Ω to a space of possible observations. In short, every random observation (the technical term is *random variable*) is a measurable map. The probabilities of the possible random observations will be described in terms of the distribution of the corresponding random variable, which is the image measure of $\mathbf{P}$ under X. Hence we have to develop a calculus to determine the distributions of, for example, sums of random variables.

Definition 1.102 (Random variables) Let $(\Omega', \mathcal{A}')$ be a measurable space and let $X : \Omega \to \Omega'$ be measurable.

(i) X is called a *random variable* with values in $(\Omega', \mathcal{A}')$. If $(\Omega', \mathcal{A}') = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then X is called a real random variable or simply a random variable.
(ii) For $A' \in \mathcal{A}'$, we denote $\{X \in A'\} := X^{-1}(A')$ and $\mathbf{P}[X \in A'] := \mathbf{P}[X^{-1}(A')]$. In particular, we let $\{X \geq 0\} := X^{-1}([0, \infty))$ and define $\{X \leq b\}$ similarly and so on.

Definition 1.103 (Distributions) Let X be a random variable.

(i) The probability measure $\mathbf{P}_X := \mathbf{P} \circ X^{-1}$ is called the *distribution* of X.
(ii) For a real random variable X, the map $F_X : x \mapsto \mathbf{P}[X \leq x]$ is called the *distribution function of* X (or, more accurately, of $\mathbf{P}_X$). We write $X \sim \mu$ if $\mu = \mathbf{P}_X$ and say that X has distribution μ.
(iii) A family $(X_i)_{i \in I}$ of random variables is called *identically distributed* if $\mathbf{P}_{X_i} = \mathbf{P}_{X_j}$ for all $i, j \in I$. We write $X \stackrel{\mathcal{D}}{=} Y$ if $\mathbf{P}_X = \mathbf{P}_Y$ ($\mathcal{D}$ for *distribution*).

Theorem 1.104 *For any distribution function F, there exists a real random variable X with $F_X = F$.*

Proof We explicitly construct a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and a random variable $X : \Omega \to \mathbb{R}$ such that $F_X = F$.

The simplest choice would be $(\Omega, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $X : \mathbb{R} \to \mathbb{R}$ the identity map and $\mathbf{P}$ the Lebesgue–Stieltjes measure with distribution function F (see Example 1.56).

A more instructive approach is based on first constructing, independently of F, a sort of standard probability space on which we define a random variable with uniform distribution on $(0, 1)$. In a second step, this random variable will be transformed by applying the inverse map F^{-1}: Let $\Omega := (0, 1)$, $\mathcal{A} := \mathcal{B}(\mathbb{R})\big|_{\Omega}$ and let $\mathbf{P}$

be the Lebesgue measure on $(\Omega, \mathcal{A})$ (see Example 1.74). Define the left continuous inverse of F:

$$F^{-1}(t) := \inf\bigl\{x \in \mathbb{R} : F(x) \geq t\bigr\} \quad \text{for } t \in (0,1).$$

Then

$$F^{-1}(t) \leq x \quad \Longleftrightarrow \quad t \leq F(x).$$

In particular, $\{t : F^{-1}(t) \leq x\} = (0, F(x)] \cap (0,1)$; hence $F^{-1} : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is measurable and

$$\mathbf{P}\bigl[\bigl\{t : F^{-1}(t) \leq x\bigr\}\bigr] = F(x).$$

Concluding, $X := F^{-1}$ is the random variable that we wanted to construct. □

Example 1.105 We present some prominent distributions of real random variables X. These are some of the most important distributions in probability theory, and we will come back to these examples in many places.

(i) Let $p \in [0,1]$ and $\mathbf{P}[X = 1] = p$, $\mathbf{P}[X = 0] = 1 - p$. Then $\mathbf{P}_X =: \mathrm{Ber}_p$ is called the *Bernoulli distribution* with parameter p; formally

$$\mathrm{Ber}_p = (1-p)\delta_0 + p\delta_1.$$

Its distribution function is

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1-p, & x \in [0,1), \\ 1, & x \geq 1. \end{cases}$$

The distribution $\mathbf{P}_Y$ of $Y := 2X - 1$ is sometimes called the *Rademacher distribution* with parameter p; formally $\mathrm{Rad}_p = (1-p)\delta_{-1} + p\delta_{+1}$. In particular, $\mathrm{Rad}_{1/2}$ is called *the* Rademacher distribution.

(ii) Let $p \in [0,1]$ and $n \in \mathbb{N}$, and let $X : \Omega \to \{0, \ldots, n\}$ be such that

$$\mathbf{P}[X = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Then $\mathbf{P}_X =: b_{n,p}$ is called the *binomial distribution* with parameters n and p; formally

$$b_{n,p} = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \delta_k.$$

(iii) Let $p \in (0,1]$ and $X : \Omega \to \mathbb{N}_0$ with

$$\mathbf{P}[X = n] = p(1-p)^n \quad \text{for any } n \in \mathbb{N}_0.$$

Then $\gamma_p := b^-_{1,p} := \mathbf{P}_X$ is called the *geometric distribution*[2] with parameter p; formally

$$\gamma_p = \sum_{n=0}^{\infty} p(1-p)^n \delta_n.$$

Its distribution function is $F(x) = 1 - (1-p)^{\lfloor x+1 \rfloor \vee 0}$ for $x \in \mathbb{R}$.

We can interpret $X + 1$ as the waiting time for the first success in a series of independent random experiments, any of which yields a success with probability p. Indeed, let $\Omega = \{0, 1\}^{\mathbb{N}}$ and let $\mathbf{P}$ be the product measure $((1-p)\delta_0 + p\delta_1)^{\otimes \mathbb{N}}$ (Theorem 1.64), as well as $\mathcal{A} = \sigma([\omega_1, \ldots, \omega_n] : \omega_1, \ldots, \omega_n \in \{0, 1\},\ n \in \mathbb{N})$. Define

$$X(\omega) := \inf\{n \in \mathbb{N} : \omega_n = 1\} - 1,$$

where $\inf \emptyset = \infty$. Clearly, any map

$$X_n : \Omega \to \mathbb{R}, \quad \omega \mapsto \begin{cases} n-1, & \omega_n = 1, \\ \infty, & \omega_n = 0, \end{cases}$$

is $\mathcal{A}$–$\mathcal{B}(\overline{\mathbb{R}})$-measurable. Thus also $X = \inf_{n \in \mathbb{N}} X_n$ is $\mathcal{A}$–$\mathcal{B}(\overline{\mathbb{R}})$-measurable and is hence a random variable. Let $\omega^0 := (0, 0, \ldots) \in \Omega$. Then $\mathbf{P}[X \geq n] = \mathbf{P}[[\omega_1^0, \ldots, \omega_n^0]] = (1-p)^n$. Hence

$$\mathbf{P}[X = n] = \mathbf{P}[X \geq n] - \mathbf{P}[X \geq n+1] = (1-p)^n - (1-p)^{n+1} = p(1-p)^n.$$

(iv) Let $r > 0$ (note that r need not be an integer) and let $p \in (0, 1]$. We denote by

$$b^-_{r,p} := \sum_{k=0}^{\infty} \binom{-r}{k} (-1)^k p^r (1-p)^k \delta_k \tag{1.17}$$

the *negative binomial distribution* or *Pascal distribution* with parameters r and p. (Here $\binom{x}{k} = \frac{x(x-1)\ldots(x-k+1)}{k!}$ for $x \in \mathbb{R}$ and $k \in \mathbb{N}$ is the generalized binomial coefficient.) If $r \in \mathbb{N}$, then one can show as in the preceding example that $b^-_{r,p}$ is the distribution of the waiting time for the rth success in a series of random experiments. We come back to this in Example 3.4(iv).

(v) Let $\lambda \in [0, \infty)$ and let $X : \Omega \to \mathbb{N}_0$ be such that

$$\mathbf{P}[X = n] = e^{-\lambda} \frac{\lambda^n}{n!} \quad \text{for any } n \in \mathbb{N}_0.$$

Then $\mathbf{P}_X =: \mathrm{Poi}_\lambda$ is called the *Poisson distribution* with parameter λ.

[2]Warning: For some authors, the geometric distribution is shifted by one to the right; that is, it is a distribution on $\mathbb{N}$.

(vi) Consider an urn with $B \in \mathbb{N}$ black balls and $W \in \mathbb{N}$ white balls. Draw $n \in \mathbb{N}$ balls from the urn without replacement. A little bit of combinatorics shows that the probability of drawing exactly $b \in \{0, \ldots, n\}$ black balls is given by the *hypergeometric distribution* with parameters $B, W, n \in \mathbb{N}$:

$$\mathrm{Hyp}_{B,W;n}\big(\{b\}\big) = \frac{\binom{B}{b}\binom{W}{n-b}}{\binom{B+W}{n}} \quad \text{for } b \in \{0, \ldots, n\}. \tag{1.18}$$

This generalizes easily to the situation of k colors and B_i balls of color $i = 1, \ldots, k$. As above, we get that the probability of drawing out of n balls exactly b_i balls of color i for each $i = 1, \ldots, k$ (with the restriction $b_1 + \ldots + b_k = n$ and $b_i \le B_i$ for all i) is given by the *generalized hypergeometric distribution*

$$\mathrm{Hyp}_{B_1,\ldots,B_k;n}\big(\{(b_1, \ldots, b_k)\}\big) = \frac{\binom{B_1}{b_1}\cdots\binom{B_k}{b_k}}{\binom{B_1+\ldots+B_k}{n}}. \tag{1.19}$$

(vii) Let $\mu \in \mathbb{R}$, $\sigma^2 > 0$ and let X be a real random variable with

$$\mathbf{P}[X \le x] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt \quad \text{for } x \in \mathbb{R}.$$

Then $\mathbf{P}_X =: \mathcal{N}_{\mu,\sigma^2}$ is called the Gaussian *normal distribution* with parameters μ and σ^2. In particular, $\mathcal{N}_{0,1}$ is called the standard normal distribution.

(viii) Let $\theta > 0$ and let X be a nonnegative random variable such that

$$\mathbf{P}[X \le x] = \mathbf{P}\big[X \in [0, x]\big] = \int_0^x \theta\, e^{-\theta t}\, dt \quad \text{for } x \ge 0.$$

Then $\mathbf{P}_X$ is called the *exponential distribution* with parameter θ (in shorthand, $\exp_\theta$).

(ix) Let $\mu \in \mathbb{R}^d$ and let Σ be a positive definite symmetric $d \times d$ matrix. Let X be an $\mathbb{R}^d$-valued random variable such that

$$\mathbf{P}[X \le x] = \det(2\pi\,\Sigma)^{-1/2} \int_{(-\infty,x]} \exp\left(-\frac{1}{2}\big\langle t-\mu,\, \Sigma^{-1}(t-\mu)\big\rangle\right) \lambda^d(dt)$$

for $x \in \mathbb{R}^d$ (where $\langle\,\cdot\,,\,\cdot\,\rangle$ denotes the inner product in $\mathbb{R}^d$). Then $\mathbf{P}_X =: \mathcal{N}_{\mu,\Sigma}$ is the d-dimensional normal distribution with parameters μ and Σ. ◊

Definition 1.106 If the distribution function $F : \mathbb{R}^n \to [0, 1]$ is of the form

$$F(x) = \int_{-\infty}^{x_1} dt_1 \ldots \int_{-\infty}^{x_n} dt_n\, f(t_1, \ldots, t_n) \quad \text{for } x = (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

for some integrable function $f : \mathbb{R}^n \to [0, \infty)$, then f is called the *density* of the distribution.

Example 1.107

(i) Let $\theta, r > 0$ and let $\Gamma_{\theta,r}$ be the distribution on $[0, \infty)$ with density

$$x \mapsto \frac{\theta^r}{\Gamma(r)} x^{r-1} e^{-\theta x}.$$

(Here Γ denotes the gamma function.) Then $\Gamma_{\theta,r}$ is called the *Gamma distribution* with scale parameter θ and shape parameter r.

(ii) Let $r, s > 0$ and let $\beta_{r,s}$ be the distribution on $[0, 1]$ with density

$$x \mapsto \frac{\Gamma(r+s)}{\Gamma(r)\Gamma(s)} x^{r-1}(1-x)^{s-1}.$$

Then $\beta_{r,s}$ is called the *Beta distribution* with parameters r and s.

(iii) Let $a > 0$ and let Cau_a be the distribution on $\mathbb{R}$ with density

$$x \mapsto \frac{1}{a\pi} \frac{1}{1 + (x/a)^2}.$$

Then Cau_a is called the *Cauchy distribution* with parameter a. ◊

Exercise 1.5.1 Use the identity $\binom{-n}{k}(-1)^k = \binom{n+k-1}{k}$ to deduce (1.17) by combinatorial means from its interpretation as a waiting time.

Exercise 1.5.2 Give an example of two normally distributed random variables X and Y such that (X, Y) is not (two-dimensional) normally distributed.

Exercise 1.5.3 Use the transformation formula (Theorem 1.101) to show the following statements.

(i) Let $X \sim \mathcal{N}_{\mu,\sigma^2}$ and let $a \in \mathbb{R} \setminus \{0\}$ and $b \in \mathbb{R}$. Then $(aX + b) \sim \mathcal{N}_{a\mu+b,a^2\sigma^2}$.
(ii) Let $X \sim \exp_\theta$ and $a > 0$. Then $aX \sim \exp_{\theta/a}$.

Exercise 1.5.4 Show that $F : \mathbb{R}^2 \to [0, 1]$ is the distribution function of a (uniquely determined) probability measure μ on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ if and only if

(i) F is monotone increasing and right-continuous
(ii) $F(-x) \to 0$ and $F(x) \to 1$ as $x \to \infty$,
(iii) $F((y_1, y_2)) - F((y_1, x_2)) - F((x_1, y_2)) + F((x_1, x_2)) \geq 0$ for all $x_1 \leq y_1$ and $x_2 \leq y_2$.

Exercise 1.5.5

(i) Let F and G be distribution functions on $\mathbb{R}$. Use Exercise 1.5.4 to show that $(x, y) \mapsto F(x) \wedge G(y)$ is a distribution function on $\mathbb{R}^2$.
(ii) Give an example of two distribution functions F and G on $\mathbb{R}^2$ such that $(x, y) \mapsto F(x) \wedge G(y)$ is *not* a distribution function on $\mathbb{R}^4$.

Hint: First use the inclusion-exclusion formula (Theorem 1.33) to derive a criterion similar to that in Exercise 1.5.4(iii).